CONTINUOUS WAVELET-LIKE TRANSFORM BASED FEATURES FOR CONTENT-BASED MUSIC SIMILARITY AND RETRIEVAL

Aliaksandr Paradzinets
Laboratoire LIRIS CNRS
UMR 5205
Ecole Centrale de Lyon
(University of Lyon)
aliaksandr.paradzinets@ec-lyon.fr

Liming Chen
Laboratoire LIRIS CNRS
UMR 5205
Ecole Centrale de Lyon
(University of Lyon)
liming.chen@ec-lyon.fr

ABSTRACT

For the MIREX Audio Music Similarity contest we propose several high-level musical similarity features that can be used in automatic music navigation, classification and recommendation. The features we propose use Continuous Wavelet-like Transform as a basic time-frequency analysis of a musical signal. Rhythmic similarity measurement characteristic is presented as a novel 2D beat histogram. This extended abstract presents also high-level musical similarity features based on note detection algorithm. Evaluation of proposed similarity metrics was done by a listening test in comparison with “random” similarity as well as by playlist composition where reinterpreted songs were searched.

1. CONTINUOUS WAVELET TRANSFORM VS. FFT

In our works we use a continuous wavelet-like transform as a basic time-frequency transform presented in [1].

The Fast Fourier Transform and the Short-Time Fourier Transform have been the traditional techniques in signal analysis for detecting pitches. However, the frequency and time resolution is linear and constant across the frequency scale while the frequency scale of notes as well as human perception of a sound is logarithmic (low frequencies have higher frequency resolution and lower temporal resolution while high frequencies have high temporal and low frequency resolutions [2]).

2. ACCOUSTIC SIMILARITY FEATURES

2.1. 2D beat histogram for rhythmic similarity

The idea of building a beat histogram is not novel [3]. Simple 1D beat histogram can be used in genre classification, tempo induction as well as music similarity search. In our work we propose a modified histogram – a two-dimensional one. Unlike 1D histogram this 2D histogram is free from beat detection threshold issue.

The beat/onset detection algorithm being described in this paper is based on Continuous Wavelet-like Transform as all other algorithms in our work. Here the signal processed by CWT is treated a grayscale image. Thus, we apply image treatment operators like Sobel. In the resulting spectrogram image distinct vertical lines are likely to represent beats or onsets.

Further, the enhanced spectrogram $W^{*}(t, scale)$ is processed by calculating a beat curve in the following way. A small 5-sample window together with preceding large 100-sample window is moved across the enhanced spectrogram. The value of the beat curve in each time moment is the number of points in the small window with values higher than a threshold which is obtained from the average value of points in the large window. Numerous beat curves may be computed separately by dividing the spectrum into bands. For the general question of beat detection the only one beat curve is used.

The probable beats are situated in beat curve’s peaks. However, the definition of final beat threshold for the beat curve is problematic. Adaptive and none-adaptive algorithms for peak detection may be unstable. Many weak beats can be missed while some false beats can be detected.

Recall that our aim is the use of the rhythmical information for music similarity estimation. One of rhythmical information representation is the beat histogram. A classical one-dimensional beat histogram provides some knowledge only about the different beat periods while the distribution of beats in the meaning of their strength is not clear. At the same time beat detection algorithm and its parameters affect the form of the histogram. In order to avoid the dependency from the beat detection algorithm parameters we propose a 2D form of beat histogram, which is built with a beats period on the X axis and with amplitude (strength) of a beat on the Y axis (Figure 1). The information about beat strength in the proposed histogram is implicit since the histogram is computed upon the threshold used in beat detection. It is hence possible to avoid the disadvantage of recording conditions dependency (e.g. volume) and peak detection method. The range of threshold variation is taken from 1 to the found maximum-1. Thus, the beat strength is taken relatively and the volume dependency is avoided.

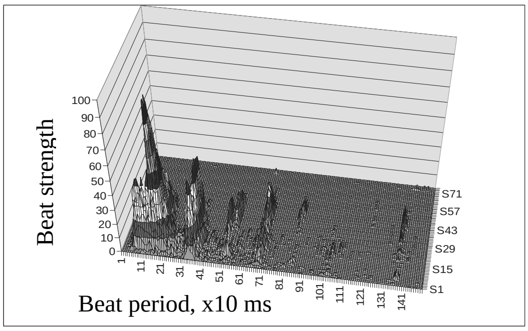


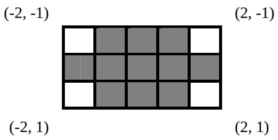
Figure 1. A 2-D beat histogram.

The measure of rhythmic distance can be defined in numerous ways. In our experiments we have find out the following equation which takes into account slight variation of rhythm of musical pieces being compared.

$$Dist_{H1,H2} = \sum_{x=1,y=1}^{N,M} \frac{1}{2}\left(\min_R\left(\left|H1_{x,y} - H2_{(x,y)+R}\right|\right) + \min_R\left(\left|H1_{(x,y)+R} - H2_{x,y}\right|\right)\right)$$

where

H1, H2 – beat histograms to compare (M×N)
R – an area of the following form (to allow slight variations)



2.2. Transcription-derived similarity features

This paragraph covers aspects of higher level musical similarity metrics. Algorithms described in the paragraph are based on automated transcription (multiple F0 estimation) of polyphonic music with the use of Continuous Wavelet-like Transform described in [1].

The transcription algorithm issues for each window a list of detected f0`s together with relative amplitudes of their partials. This information is then used for building several kinds of statistical characteristics (histograms).

The simplest way to calculate a similarity distance is to calculate a distance between note histograms. Note histogram (profile) is computed across the whole musical title or its part and serves for estimation of musical similarity by tonality as well as tonality (musical key) itself. Tonality in music is a definition of note set used in a piece which is characterized by tonic or key note and mode (e.g. minor, major). Each tonality has its own distribution of notes involved in a play and it can be obtained from the note histogram [4]. To compare two musical titles in the meaning of tonal similarity we calculate a similarity of two note profiles. These profiles must be either aligned by the detected tonality's key note (e.g. by Re for D-dur or D-mol) or a maximal similarity across all possible combinations of tonalities must be searched.

Another musical similarity metric we propose in the current work is a similarity based on note successions histogram. Here probabilities of 3-note chains are collected and their histogram is then used as a "fingerprint" of musical title. A musical basis of such similarity metric is that if same passages are frequent in two musical compositions, it gives a chance that these two compositions have similarities in melody or harmony.

The procedure is note successions histogram calculation is following. First, note extraction over the whole piece is carried out with a step of 320 samples (20ms). Then detected notes are grouped in local note histograms in order to find a dominant note in each grouping window. The size of the grouping window may vary from 100ms to 1 sec. Finally, all loudest notes are extracted from local histograms and their chains are collected in the note successions histogram. The resulting histogram is 3-dimensional histogram where each axe is a note of 3-note chain found in the musical piece being analyzed.

The third characteristic we extract from a musical piece is a timbre histogram. In general, "voiced" instruments differ from each other also by their timbre – profile of their partials. In our work we collect all detected notes with relative amplitude of their harmonics. Further, relative amplitudes of harmonics are reduced to 3-4 bits and attached together in order to form a number. Histogram of these numbers is then computed. Comparing of such histograms gives one more possibility of a similarity measurement.

2.3. Combining of similarity types

While pure similarity metrics could be interesting for exact matching of musical pieces by certain criteria, a combination of similarities have in goal building of "general" similarities like human listener could do (e.g. finding a piece with the same rhythm and key type could issue two slow sad melodies which are judged similar by a human listener).

In our work we have studied two variants of combining. A liner combining - is simple weighted sum of distances and another version of liner combining is a weighted sum of ratings. In this case for every kind of similarity being combined its rating or position in a sorted list of similar titles is obtained. Final distance is computed as a weighted sum of ratings.

3. EXPERIMENTS

Our main experiments have in aim an estimation of musical similarity accuracy. They consist of two evaluation parts – listening test and playlist relevance evaluation.

3.1. Listening test

Preliminary experiments with musical similarity search were carried out. A database of approximately 1000 musical composition of different artists, genres and rhythms has been processed. The system retrieved by different combinations of similarity metrics the 5 most similar songs from the database for a given example.

Researches from the laboratory (not necessarily working with music) were taken as listeners. They were proposed to rate random queries from the database with scores from 0 (not similar) to 5 (very similar) according to shown similarity type. Neither songs' titles nor artist names were provided to listeners. Also with a probability of 50% listeners were provided by random and not similar music pieces without being notified of this fact in order to avoid prejudgements.

In our experiments we have used 4 pure similarity metrics: rhythmic, tonality, timbre and melodic; and 4 mixtures where *comb1* was a combination of tonality and rhythm metrics, *comb2* – timbre and rhythm, comb3 – tonality + melody + rhythm, comb4 –timbre + melody + rhythm. For the mentioned mixtures both liner and rating combinations were applied.

Evaluation results obtained in our experiments are presented in the Table 1. Here for each similarity type there is mean and median value of totality of votes. The column "corresponding random" shows the mean and median of listeners' votes for those cases when listeners were proposed random songs as similar. Since listeners were not notified about this fact, they still had to evaluate how similar were the proposed songs. These data are used as background un-truth. All found multiple interpretation of songs were not filtered out and considered as 5 – very similar.

Table 1. Listening test results (mean / median).

Similarity type	Linear combination or single	Rating combination	Corresponding random
rhythmic	2.92 / 2	n/a	0.40 / 0
tonality	3.16 / 3		2.41 / 3
timbre	2.16 / 2		0.81 / 0
melodic	2.23 / 2		1.60 / 2
comb1	3.55 / 4	2.06 / 3	0.94 / 1
comb2	2.78 / 3	3.75 / 4	0.97 / 0
comb3	3.85 / 5	1.80 / 1	0.75 / 0
comb4	2.49 / 3	2.26 / 3	1.01 / 0

3.2. Playlist relevance evaluation

Finally we proceed on analysis of relevance of top 5 songs in playlists generated for seed songs. We considered two types of relevance: number of songs from the same genre and number of songs from the same artists. For the database we took ISMIR2004 genre classification database based on *Magnatune* collection. The database contained totally 729 titles of 128 artists in 6 genres. The obtained results are as following (Table 2).

Table 2. Average number songs in the same genre or from the same artist.

Similarity type	Same genre	Same artist
Comb2_lin	3.58	0.99
Comb2_rt	3.48	0.89
Comb3_lin	3.07	0.86

The next picture (Figure 2) depicts distribution histograms of number of songs in the same genre and from the same artist for the best combination which in this case is comb2_lin.

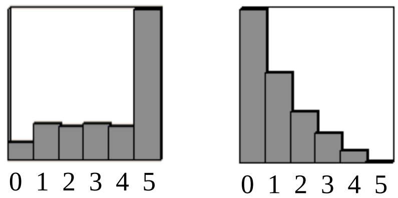


Figure 2. Histogram of number of songs in the same genre in TOP-5 (left), and histogram of number of songs from the same artist in (TOP-5) (right).

4. CONCLUSION

In this paper we described CWT-based approaches of automated music analysis. Rhythmic and several musical similarity metrics have been proposed and evaluated. Significant results were observed in listening test on rhythmic and timbral similarity measurements as well as their combinations while pure tonality and melodic similarities where not evident (especially for listeners without musical education).

While evaluation of metrics by listening test proves semblances between human and algorithmic similarities of music, reinterpreted songs search and playlist relevance analysis only validate proposed musical similarity metrics.

A promising direction can be an improvement of algorithms using listener's feedback as for example a use of NNs instead of linear combiners.

5. MIREX SUBMISSION

The version of the algorithm submitted to MIREX 2007 was supposed to be comb3 linear combination of distances. However, due to a disappointing mistake in the mixing statement only rhythmic distance was taken into account making the algorithm to use only rhythmic similarity.

REFERENCES

[1] Paradzinets A., Harb H., Chen L., Use of Continuous Wavelet-like Transform in Automated Music Transcription. *In: EUSIPCO'06*, Florence, Italy

[2] Tzanetakis G., Essl G., Cook P., *Audio Analysis using the Discrete Wavelet Transform*, Proc. WSES Int. Conf. Acoustics and Music: Theory 2001 and Applications (AMTA 2001) Skiathos, Greece

[3] Tzanetakis G., Essl G., Cook P., (2001), Automatic Musical Genre Classification of Audio Signals, *ISMIR2001*

[4] Ching-Hua Chuan, Elain Chew., Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm, *ICME2005*