MIREX 2011 - AMS TASK: MELODY VARIATION PATTERNS FOR AUDIO MUSIC SIMILARITY

Simone Sammartino, Lorenzo J. Tardón, Isabel Barbancho, Cristina de la Bandera
Dept. Ingeniería de Comunicaciones, E.T.S. Ingeniería de Telecomunicación,
Universidad de Málaga, Campus Universitario de Teatinos s/n, 29071, Málaga, Spain
ssammartino@ic.uma.es

ABSTRACT

The dynamics of the melodic line is a crucial feature for the temporal variability of the tonal content of a song, and it determines its characteristic mood pattern. A method based on the measure of the variation of the pitch over the song time is proposed as a tool to characterize the music similarity. The *pitch-variation* vs. *time* matrices of each of the analyzed song are compared to achieve a measure of the closeness of the dynamic melodic content of the excerpts. The algorithm is proposed to be submitted to the Audio Music Similarity task of MIREX 2011, in occasion of the 12th ISMIR Conference.

1. INTRODUCTION

In MIR community, many different approaches for automatic music recommendation are based on the retrieval of content-based descriptors that are able to estimate the audio similarity and, somehow, simulate the performance of the human brain with regard to the evaluation of music similarity. The melody extraction is one of the most discussed and explored field in MIR community [4].

Lemström and Pienimäki [2] talk about a dualistic approach to the melody extraction, based either on a linear unidimensional representation of the main melodic stream of the song or on a more complex bi-dimensional geometrical structure of the melodic 'objects'.

The process of extracting the melodic line of a music excerpt is fundamentally based on the estimation of a series of pitches following the rhythmic structure of the song [1], both on its standard representation [5] and on its symbolized version (MIDI) [3]. The algorithm proposed here is based on the extraction of the melodic line of the song and the calculation of the variations of the pitches, over the time.

2. THE MELODIC SIGNATURE

As mentioned above, the procedure proposed is fundamentally based on the extraction of the monophonic melodic line of the excerpt.

The song is fragmented, the pitch of each fragment is extracted and the temporal variability of these leading pitches is tracked. In particular, the variation of the pitch is determined in relation to its temporal stability. Any change in the pitch is tagged with the frequency difference of the change and the interval at which it has occurred. In this way, a paired information on *how much the sound changed* and *how long it lasted* is tracked over the whole song time. In Figure 1, a brief flow chart of the algorithm is detailed.

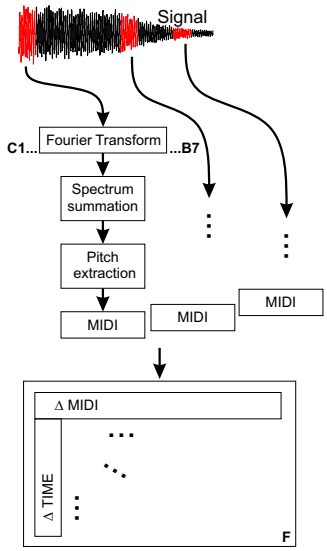

Figure 1. A brief representation of the work flow of the algorithm.

After the signal fractioning and the calculus of the Fourier transform, for each fragment, the pitch extraction process is carried out by summing up the amplitudes of the frequency bins rounding the central frequencies of the 84 notes (half pitch difference above and half pitch difference below each

note), from C1 to B7. From this simplified spectrum, the highest amplitude peak is extracted as the pitch of that fragment. The frequency of the pitch is translated to its corresponding MIDI number.

At this point, the variation of the pitch over the series of fragments is measured as the difference in MIDI number of the pitch, versus the interval at which it occurred (the interval range is limited to 2 seconds). The interval is defined as the number of fragments passed from the first pitch detection to its first change. For each pair of midi-variation/interval, a table of occurrences is assembled, by measuring how many times a certain combination of the two measures occurred.

A ΔMIDI vs. Δtime matrix of occurrences is built, defined as the melody signature of the song analyzed.

In Figure 2, an example of such kind of matrix is shown.

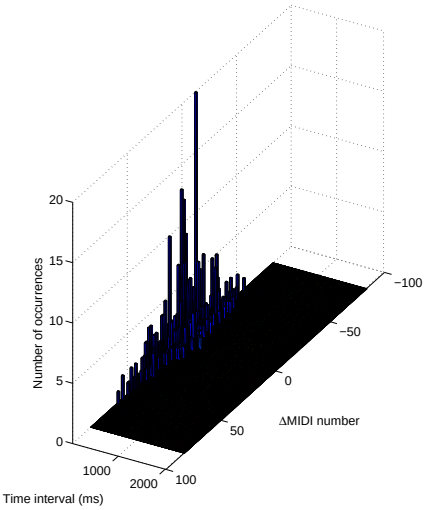

Figure 2. A tridimensional representation of the melody signature matrix.

The interpretation of the meaning of that matrix is quite trivial. The wider the midi variation (the higher the dispersion along the Y axis) the higher the tonal spread of the music excerpt, and vice versa. Conversely, the wider the time variation, the more stable the melodic content of the song.

3. THE CALCULATION OF THE DISTANCE

The degree of similarity between a pair of songs is quantified by the degree of closeness of the shape of the surfaces, in both directions, both as the spread of the tonal content of the song, and as the fastness of its melodic variability. This issue can be conveniently resumed by the measure of the cross-correlation of the pair of matrices.

The bi-dimensional cross correlation matrix is calculated for each pair of signature matrices and its maximum value is extracted. After normalizing this value from 0 to 1, its one's complement is taken into account as the distance between the two analyzed matrices.

Both a full dense matrix and a sparse matrix of the 100 most similar elements for each song are calculated for the submission to the MIREX 2011.

4. REFERENCES

[1] C Isikhan and G Ozcan. A survey of melody extraction techniques for music information retrieval. In *Proceedings of the CIM08 conference on Interdisciplinary Musicology*, pages 1–8, 2008.

[2] Kjell Lemström and Anna Pienimäki. On comparing edit distance and geometric frameworks in content-based retrieval of symbolically encoded polyphonic music. *Musicae Scientiae*, 11(1 suppl):135–152, 2007.

[3] G Ozcan, C Isikhan, and A Alpkock. Melody extraction on midi music files. pages 414–422. Ieee, 2005.

[4] J Pedro, Ponce De León, José Iñesta, and Rizo David. *Mining digital music score collections: melody extraction and genre recognition*. Number November. INTECH, 2008.

[5] Graham E. Poliner and Daniel P. W. Ellis. A classification approach to melody transcription. In *Proceedings of the ISMIR'05*, pages 161–166, 2005.