SUBMISSION TO MIREX 2013 SYMBOLIC MELODIC SIMILARITY

Carles Roig, Lorenzo J. Tardón, Ana M. Barbancho, Isabel Barbancho
Dept. Ingeniera de Comunicaciones, E.T.S. Ingeniera de Telecomunicacin, Universidad de Málaga
carles@ic.uma.es, lorenzo@ic.uma.es, abp@ic.uma.es, ibp@ic.uma.es

ABSTRACT

In the present submission to MIREX 2013, we provide an algorithm to handle the symbolic melody similarity task. The method goal is to find the ten most similar melodies, according to one used as a query, among a database. Both query and melodies in the database are monophonic MIDI files. The novelty of the proposed algorithm is the tracking and evaluating of musical features scheme. Taking into account the developments of previous existing papers, it is observed that the comparison of the rhythmic patterns and the melodic curves are crucial parameters to determine the grade of similarity between songs. So, different levels of similarity between these features have been determined. The selected approach is based on a scoring system related to the number of common elements in both the query and the candidate excerpts. A weighting scheme has also been developed to consider the different importance of the features selected.

1. INTRODUCTION

Our melody similarity evaluation approach is based on an feature-wise comparison. The music elements selected for our analysis are rhythm and pitch. They were chosen because tracking them is simple and the information that can be obtained from them is very useful for the characterization of music [2]. From each music element, the features to be evaluated will be the following:

Rhythm	Downbeat onset. Passing notes onset.
Pitch	Pitch direction. Interval distances. Transposition

Table 1. Features to be evaluated.

This work has been funded by the Ministerio de Economía y Competitividad of the Spanish Governmet under Project No. TIN2010-21089-C03-02 and Project No. IPT-2011-0885-430000, by the Junta de Andalucía under Project No. P11-TIC-7154 and by the Ministerio de Educación, Cultura y Deporte through the 'Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i 2008-2011, prorrogado por Acuerdo de Consejo de Ministros de 7 de octubre de 2011'. This work has been done in the setting Campus de Excelencia Internacional Andalucía Tech, Universidad de Málaga.

However, a strict comparison between two streams of MIDI files identifying the presence or not of a certain representative element can obtain an inaccurate result since music is built making use of elements like modulations and ornaments in order to create variations of the same melody. According to [5], human listeners identify such melodies as similar considering that the transformed is a variation of the first one. However, such process of generation of a new version the same melody changes the number of elements, their durations and pitch, so the result of a direct binary comparison approach would return a bad score where a human listener would give a high punctuation.

In order to taking into account this issue, we have included some analysis stages that will be able to recognize the function of each element within the melody. This is a simplified model of the processing done by a human listener to understand the music and detect similarities between several melodies [3].

In the following sections, these features used to the evaluation of similarity will be explained in detail.

2. RHYTHM

Paying attention to the rhythm, it must be noted that downbeat detection is crucial to obtain the schematic rhythm pattern. According to [5], the rhythmic sequence of the original melody and the variation, omitting the passing notes and ornaments, are similar and almost equivalent. Taking into account this fact, the first analysis step will consist on the identification of the downbeats and the extraction of the rhythmic schema of both the query and the candidate melodies.

Since each measure signature has a different downbeat structure depending of the subdivision (consider binary division: 2/4, 3/4, 4/4, versus ternary division: 6/8, 9/8,), it has to be estimated. The method used in the proposed algorithm is based on the analysis of the time position of the elements in the melody. If the majority of the notes are located in time instants that are multiple of the triple of the beat duration, then the subdivision will be considered to be ternary. Otherwise, the subdivision will be considered to be binary.

After the measure has been estimated, it is easy to identify the downbeats. In the binary case, the notes whose onset time is a multiple of $2t$ (where t is the beat duration) will be considered downbeats. In the ternary case, the downbeats will be identified if the onset time is a multiple of $3t$. Thus, a previous step has to be performed, estimate the beat duration, t.

The algorithm used is described in [6] and is based on the analysis of the histogram of the inter-onset interval plus a post-processing stage for fine adjustment.

Using the beat duration, the schematic rhythm pattern is obtained and a comparison stage with this feature can be performed. The Downbeat Score (DownS) is calculated by the following expression:

$$DownS = n_{dc}/N_d \tag{1}$$

where $DownS$ represents the Downbeat Score, it is a number between 0 and 1, where 0 means completely different and 1 the same rhythm. n_{dc} is the number of downbeat coincidences, and N_d is the total number of downbeats in the original melody (query).

Once the downbeat comparison is done, and in order to define a penalty score related to the presence of extra passing notes, a new measure is obtained: Passing Score (PassS). This score penalizes the similarity measure according to the presence of extra elements in the melody evaluated. The $PassS$ is obtained from:

$$PassS = (Pass_q - (N_c - n_{dc}))/N_q \tag{2}$$

where PassS is the Passing Score. The lower the value of PassS, the lower the number of extra notes, so the similarity should be higher. $Pass_q$ is the number of passing notes in the query, N_c is the number of elements in the candidate, n_{dc} is the number of coincidences concerning downbeats between candidate and query, and N_q is the number elements in the query.

Note that different scores do not have the same importance. In Section 4, the weighting that will be applied to each score will be presented.

3. PITCH

In our proposal, the pitch rating is closely related to the melodic contour [1]. According to [3], since we are children we are able to memorize a melody just by focusing on the melodic contour. This is the key feature used in our proposal to obtain the pitch similarity measure between to melodies. The pitch contour is a continuous curve with the information of the pitch evolution in the time domain. Since the symbolic representation is discrete, the curve is created by joining the 2D position of each element of the MIDI file, assuming that the onset times correspond to the x component and the pitch number is the y coordinate.

Since the data we are dealing with is a symbolic discrete representation, the pitch comparison performed element to element is by no means exact, specially if the notes are not placed in the same time instant. Thus, the use of a continuous representation of the pitch fixes the problem of dealing with different rhythmic patterns. Firstly, a continuous representation enables the sample to sample (time to time) comparison. Secondly, the instantaneous data of both curves provides a real pitch representation independent to different rhythm structures.

Concerning the pitch features, three different scores are defined: Pitch Direction Score (DirS), Interval Score (InterS) and Transposition Score (TransS).

3.1 Pitch Direction Score

The first score to compute will be based on the evaluation, beat by beat, of the direction of the pitch contour. This feature is related to the relative pitch. Recall that the relative pitch capacity is far more common among humans listeners [4] than the absolute pitch capacity.

On the other hand, note that the comparison performed beat by beat provides accurate results since this choice to perform the comparison overcomes the problem that can arise when the songs compared do not follow the same rhythm.

The pitch direction score is a coarse evaluation of the pitch curve similarity. It omits the distance, in pitch, between an element and the following one, but it checks the direction of the pitch change.

This score is obtained using the following expression:

$$DirS = \frac{1}{N_b}\sum_{i=1}^{N_b} DSim_i \tag{3}$$

where $DirS$ is the Pitch Direction Score, 0 corresponds to opposite pitch contour and 1 means that both contours are equivalent. N_b is the number of beats of the comparison (this value, in our system, is defined as the smallest number of beats among the number of beats in the query and the candidate melodies). Finally, $DSim_i$ is the similarity value of the i-th beat, given by:

$$DSim_i = \begin{cases} sign(PC_q[i]) = sign(PC_c[i]) & 1 \\ sign(PC_q[i]) \neq sign(PC_c[i]) & 0 \end{cases} \tag{4}$$

where $PC_q[i]$ and $PC_c[i]$ are the i-th pitch contour value of the query and the candidate melodies, respectively.

3.2 Interval Score

Unlikely the Pitch Direction Score, the Interval Score takes into account the pitch distance between two consecutive notes.

In this case the score is obtained by computing the area between both pitch contours: the one of the query melody and the one of the candidate melody:

$$InterS = \frac{1}{MaxErr}\sum_{i=1}^{N_b} |PC_q[i] - PC_c[i]| \tag{5}$$

where $InterS$ is the Interval Score, the lower the score, the more similar the melodies should be. If the pitch contour is very similar, the area between both curves will be small. $MaxErr$ is value determined by the pitch contour of the query melody to normalize the area. This value is obtained considering the worst case in which the opposite pitch contour curve would appear ($PC_c[i] = -PC_q[i]$).

3.3 Transposition Score

Finally, cases can be found in which the pitch contour is the same and the rhythm is also the same in the melodies compared but the original melody may have been transposed to another tone. A human listener can identify this variation, thus, it will be immediate to notice that a transformation has performed. Even being very similar, both melodies will be considered different (because of the pitch transposition), but with a high similarity rating. Consequently, we consider the addition of a small penalty score to take into account the case in which a transposition has been done.

Since the melody transposition can be observed as a constant difference between the pitch contours, the Transposition Score will be highly related to the variance of the difference. If the difference is constant (var(x) $\simeq$ 0), it will be assumed that a transposition has been done, so a small penalty will be added. Otherwise, the similarity will decrease since they will be considered different curves, and a severe penalization will be added since the Interval Score will decrease.

The Transposition Score is obtained according to:

$$TransS = 1 - \frac{1}{12}\sqrt{var(|PC_q[i] - PC_c[i]|)} \quad (6)$$

4. WEIGHTING

A priori, the rhythm seem to be the key feature to determine the similarity between two melodies. However, different rhythm patterns with the same melodic contour lead a human listener to perceive a notable degree of similarity between the melodies. So, the scores of the rhythm and pitch are considered to have the same weight.

Then, within each category the weights attempt to account for their importance for the human similarity perception. Note that, the weights are heuristically defined but based on music theory concepts [2]. This weights are presented in Table 2.

50%	Rhythm	80%	Downbeat
		20%	Passing
50%	Pitch	70%	Direction
		20%	Interval
		10%	Transpose

Table 2. Weight considered for the different features selected for melody similarity evaluation.

5. REFERENCES

[1] W. Appel: "Harvard Dictionary of Music," 2nd Edition, *The Belknap Press of Harvard University*, 2000.

[2] B. Benward: "Music: In Theory and Practice," 7th Edition, Vol. 1, *McGraw-HIll Companies*, 2003.

[3] W. J. Downling, B. Tillmann, and F. A. Dan: "Memory and the Experience of Hearing Music," *Music Perception*, 19/2, 249-276, 2002.

[4] L. Humphries: "Learning to Sight-Sing: The Mental Mechanism of Aural Imagery," *Minneapolis: Thinking Applied*, No. 1, 2008.

[5] S. McAdams, and D. Matzkin: "Similarity, Invariance, and Musical Variation," *The Biological Foundations of Music*, 62-74, 2001.

[6] C. Roig, I. Barbancho, E. Molina, L. J. Tardón, and A. M. Barbancho: "Rumbator: a Flamenco Rumba Cover Version Generator Based on Audio Processing at Note-Leve," *in Proceedings of DAFX'13*, 2013.