Comparing computational approaches to Rhythmic and Melodic Similarity in Folksong Research

Anja Volk, Jörg Garbers, Peter van Kranenburg,
Frans Wiering, Louis Grijp, Remco C. Veltkamp
Utrecht University, Meertens Institute
anja.volk@meertens.knaw.nl

May 8, 2007

Abstract

In this paper we compare computational approaches to rhythmic and melodic similarity in order to find relevant features characterizing similarity in a large collection of Dutch folksongs. Similarity rankings based on Transportation Distances are compared to an approach of rhythmic similarity based on Inner Metric Analysis proposed in this paper. The comparison between the two models demonstrates the important impact of rhythmic organization on melodic similarity.

1 Introduction

Computational approaches to melodic similarity contribute to the study of melodies in the different areas of music cognition, ethnomusicology and music information retrieval. In this paper we study rhythmic similarity in the context of melodic similarity as a first step within the interdisciplinary enterprise of the WITCHCRAFT project (Utrecht University and Meertens Institute). The project makes use of and contributes to methods in these three areas in order to develop a content based retrieval system for a large collection of Dutch folksongs. The retrieval system will give access to the collection *Onder de groene linde* (see [5]) hosted by the Meertens Institute to both the general public and musical scholars. For the latter it is of special interest to be able to classify, identify and trace melodic variants with the help of the retrieval system to be designed.

The similarity between different variants of a folksong melody is based on a variety of musical dimensions, such as rhythm, contour or cadence notes. According to cognitive studies metric and rhythmic structures play a central role in the perception of melodic similarity.

In this paper we focus on rhythmic similarity by comparing similarity rankings based on *Inner Metric Analysis* (IMA) to *Transportation Distances* (see [6] and [1]). Transportation Distances have been successfully applied to the measurement of melodic similarity of, for instance, RISM incipits or karaoke pieces (see [6]). By excluding the pitch factor we apply the Transportation Distances in this paper to rhythm only in order to study the impact of rhythm on melodic similarity. Inner Metric Analysis has been successfully applied to the study of metric structures of musical pieces in the context of music analysis in [3], music cognition in [7] and classification in [2]. We therefore propose in this paper an approach to the measurement of rhythmic-metric similarity based on IMA and compare the results to those of the Transportation Distances.

2 Two computational approaches to rhythmic similarity

2.1 Transportation distances

Transportation distances consider melodies as weighted point sets. A similarity (or distance) measure between two melodies is defined on the basis of weight flows between these point sets that have to be minimized. We compare two instances of these distances, namely the Earth Mover's Distance (EMD) and the Proportional Transportation Distance (PTD). Both distance measures and their application to melodies are described in detail in [6]. In the application to melodies, every note is a point in the Euclidean space with the two coordinates *pitch* and *onset time*, the *duration* of the note determines its weight. In this article we apply these distances to rhythms instead of melodies, hence the coordinates of the points are determined by the onset time only.

Figure 1 gives an example for two short rhythms with the minimal flow of weights according to the EMD. The arrows indicate which amount of weight from the first rhythm is transported to which note in the second rhythm. For instance, the first rhythm is being described with the point set r1 = $\{(0.0, 1.0), (1.0, 0.5), (1.5, 0.25), (1.75, 0.25), (2.0, 0.5), (2.5, 0.5), (3.0, 1.0)\}$. The first coordinate of each point is the onset time of the note, the second is its weight (which equals its duration).

The difference between EMD and PTD becomes evident in the similarity comparison between pieces of different total weight, which is in this case the total length. The EMD realizes partial matching, hence ignores the extra notes in the longer piece. Within the PTD approach the total weight of each piece is normalized in order to prevent partial matching, hence the existence of extra notes that can not be matched in the second melody is effectively penalized.

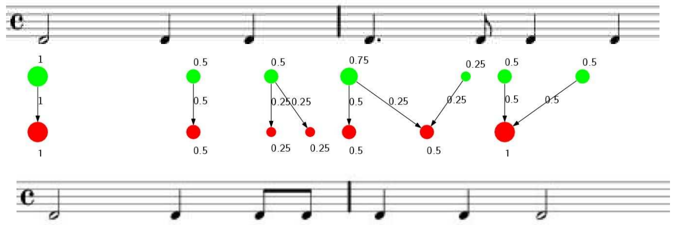

Figure 1: Example of minimal weight flow between two rhythms. The point size depicts the weight.

2.2 Inner Metric Analysis

Inner Metric Analysis (see [4], [3]) describes the *inner* metric structure of a piece of music generated by the actual notes *inside* the bars as opposed to the *outer* metric structure associated with a given abstract grid such as the bar lines. The model assigns a metric weight to each note of the piece.

Figure 2 gives an example for the song *OGL 19914*[1] from the collection [5] belonging to the melody group *Deze morgen*. The notes of the first phrase are shown in the top example of figure 4. For each note a line depicts the metric weight such that the higher the line, the higher the corresponding weight. The background gives the bar lines for orientation. The metric weight profile corresponds to the typical accent hierarchy of a 6/8 bar.

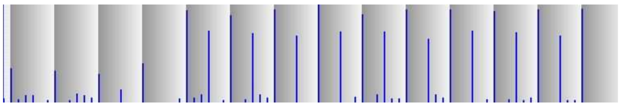

Figure 2: Metric weight of *OGL 19914*, melody group *Deze morgen* in 6/8

2.2.1 The metric weights

The details of the model have been described in [3] or [2]. The general idea is to search for all pulses (chains of equally spaced events) of a given piece and then to assign a *metric weight* to each note. The pulses are chains of equally spaced onsets of the notes of the piece called *local meters*. Let On denote the set of all onsets of notes in a given piece. We consider every subset $m \subset On$ of equally spaced onsets as a local meter if it contains at least three onsets and is not a subset of any other subset of equally spaced onsets. Let k denote

---

[1] OGL as the abbreviation of *Onder der groene linde*

the number of onsets a local meter consists of minus 1. Hence k counts the number of repetition of the period (distance between consecutive onsets of the local meter) within the local meter. The metric weight of an onset is then calculated as the weighted sum of the length k of all local meters m_k that coincide at this onset.

Let $M(\ell)$ be the set of all local meters of the piece of length at least ℓ. The general metric weight of an onset, $o \in On$, is as follows:

$$W_{\ell,p}(o) = \sum_{\{m \in M(\ell): o \in m_k\}} k^p.$$

In all examples of this paper we have set the parameter $\ell = 2$, hence we consider all local meters that exist in the piece. In order to obtain stable layers in the metric weights of the folksongs we have chosen $p = 3$.

2.2.2 Defining similarity based on Inner Metric Analysis

Metric weights of short fragments of musical pieces have been used in [2] to classify dance rhythms of the same meter and tempo using the correlation coefficient. In this article we want to modify this approach to measure the rhythmic-metric similarity between two complete melodies. The similarity measurement is hence carried out on the analytical information given by the metric weights.

Since the metric weight is defined only on note onsets, we define in a first step for each of the two pieces the metric weight of all silence events as zero and hence obtain the *metric grid weight.* The silence events are inserted along the finest grid of the piece which is determined by the shortest existing interval between two consecutive onsets of the piece. Thus we obtain a weight for all events e of the piece along the finest onset grid.

We want to compare the consecutive weights within cells of equal total duration (for instance 4 quarter notes in length) of the two pieces. Therefore in cases where the finest onset grids of the two pieces differ, we adapt the grids of the pieces to a common finer grid by adding events e with the weight zero along the finer grid.

In the second step, the metric grid weight is split into consecutive segments that cover an area of equal duration in the piece. These segments contain the weights to be compared with the correlation coefficient, we therefore call them *correlation windows.* The first correlation window of each piece starts with the first full bar, hence the weights of an upbeat are disregarded. For all examples of this article we have set the size of the correlation window to one bar. Figure 3 shows an example for two metric grid weights with the first 3 correlation windows. For the computation of the similarity measure both grid weights are completely covered with correlations

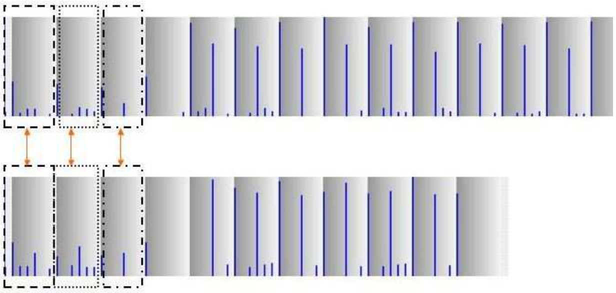

Figure 3: The first three correlation windows of two metric weights to be compared

windows. Let $w_{i,\, i=1,\ldots,n}$ denote the consecutive correlation windows of the first piece and $v_{j,\, j=1,\ldots,m}$ those of the second piece. Let $c_{k,\, k=1,\ldots,min(n,m)}$ denote the correlation coefficient between the grid weights that are covered by the windows w_k and v_k. Then we define the similarity $IMA_{c,s}$ that is defined on the subsets of the two pieces from the beginning until the end of the shorter piece as the mean of all correlation coefficients:

$$IMA_{c,s} = \frac{1}{min(n,m)} \sum_{k=1}^{min(n,m)} c_k$$

The partial similarity $IMA_{c,s}$ disregards all extra notes at the end of the longer piece. However, in many contexts it might be important to add a penalty for these extra notes that have no counterpart in the shorter piece. Therefore we define the correlation coefficient value between the additional correlation windows of the longer piece to the empty correlation windows of the shorter piece as zero $c_{k,\, k=min(n,m)+1,\ldots max(n,m)} = 0$. Hence we obtain the similarity measure $IMA_{c,e}$ taking into consideration the entire pieces:

$$IMA_{c,e} = \frac{1}{max(n,m)} \sum_{k=1}^{max(n,m)} c_k$$

We will use the latter measure for the application to rhythmic similarity ranking in section 3.

3 Evaluation of the rhythmic similarity approaches

In this section we compare the similarity measurements based on IMA, EMD and PTD in a first and simple approach to rhythmic similarity of melodies. The application of these measurements is simple in so far as it does not contain a segmentation procedure and the search for similar segments that are shifted in time. Since the pieces contain musically meaningful segments (phrases) we applied IMA, EMD and PTD to both single phrases and complete pieces.

The evaluation of the similarity measurements is based on melody groups of the collection *Onder de groene linde* (OGL) of related songs. The melodies belonging to one group are being considered as musically similar. The current test corpus of digitized melodies contains 141 songs which are segmented into 567 phrases in total. One melody (or melody phrase) of such a group is selected as the query and the similarity measure to all other melodies (or melody phrases) in the test corpus is calculated and ordered (the ordered list starts with the most similar melody).

A good similarity measurement should therefore list the members of the group the query belongs to among the top hits of the list, if the members are more similar to each other than to members of other melody groups. The melody groups in our test corpus that have been constructed by musicologists fulfill this condition to a certain extent. However, sometimes a very similar song was assigned to a different group based on other than musical reasons - for instance, because of the text. A typical comparison of those ranking lists include the number of melodies that should have been found within the top hits of the list (because they are group members), but get a very low rank ("false negatives") and the number of melodies that end up high in the ranking list but do not belong to the same melody group as the query ("false positives"). Since our melody groups have not been tested to always contain the most similar melodies, we will in our comparison not only count the false positives but check whether they are nevertheless musically similar. In addition we calculate the mean for all ranks of melodies belonging to the group of the query as a measurement of the performance of the similarity approaches.

In the following section 3.1 we discuss one example in detail using the melody group *Deze morgen*, in section 3.2 we summarize very briefly further results of the comparison.

3.1 A detailed comparison on the melody group Deze morgen

The melody group called *Deze morgen* contains 12 melodies which are very similar to each other. However, two songs have one phrase less than the others. First we want to compare the results of the ranking lists for a single phrase, in the second step we will use the entire piece.

As the query for the single phrase we used the first phrase of *OGL 19914* (the top melody from figure 4). For the evaluation of the ranking list we focus on the ranks that have been assigned to the other first phrases of melodies in this group, since they are all rhythmically very similar to the query.

The ranking list according to IMA contains among the first 19 elements
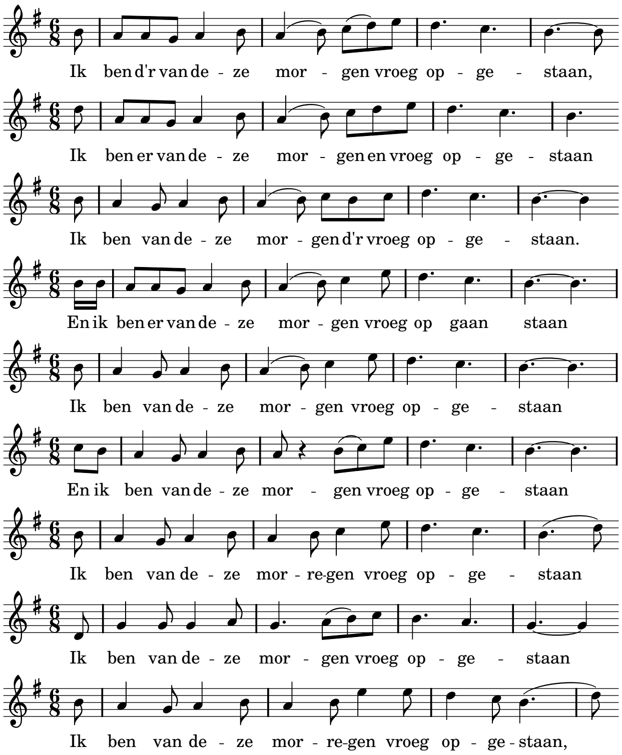
Figure 4: Excerpt from the top hits from the list according to IMA (melodies of same rhythm excluded, the listed rhythms cover the first 19 matches).

11 members of the group and misses among the top 20 hits only one phrase at rank 29 (see figure 5). The mean rank for all first phrases of the group is 8.75. All false positives with a better rank than 29 are musically very similar to the query (for instance, most of them are second phrases from melodies of the same group). Figure 4 lists the best hits from the ranking list according to IMA with the exclusion of melodies that duplicate the rhythmic structure of melodies that are displayed. Hence the displayed 9 melodies stem from the best 19 matches.
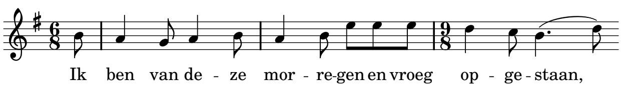
Figure 5: IMA assigns *OGL 37511* rank 29

PTD ranks 10 group members among the first 22 matches and the ranks of all phrases of the group have a mean of 18.6. Thus it misses the first phrase of *OGL 37511* in figure 5 which is placed on rank 68 and a very similar phrase to the latter one is placed on rank 73. The first phrase of *OGL 37511* was ranked lower than all other members of the group (rank 29) by IMA as well, indicating that this rhythm is somewhat less similar to the query. However, the low rank of 68 according to PTD is very drastic. For instance, figure 6 gives three examples of phrases that are assigned a higher similarity to the query according to PTD. These are rhythmically less similar to the query than the missed phrase from figure 5.

One of the reasons for the low rank of *OGL 37511* are two long notes near the end of the query (the notes of the syllables "op" and "staan" in the top melody of figure 4). Both of them do not have a counterpart in *OGL 37511* and therefore their weight is distributed over 5 different notes each that are located much earlier in the piece. In contrast to this, the first phrase shown in figure 6 contains many notes that are located in roughly the same area as the long end notes of the query. Also here the weight is distributed among 4 to 6 notes, but this weight has to be transported only locally and not to notes far apart as in *OGL 37511*. This results in a much higher similarity ranking.
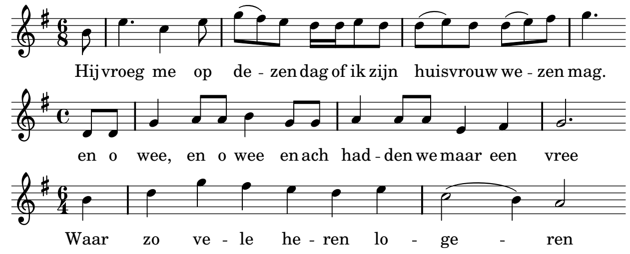
Figure 6: PTD assings rank 11 (top melody), rank 34 (middle melody) and rank 50 (bottom melody)

The false positives within the first 24 matches of the PTD list are all rhythmically similar except rank 11 which is shown in figure 6. On the other hand the ranking list according to IMA contains up to 29 similar elements in the beginning, hence the last elements are being missed by PTD.

EMD ranks 10 group members among the first 24 matches and misses two phrases at the ranks 58 and 79. The ranks of all phrases have a mean of 19.8. The false positive on rank 3 (see figure 7) demonstrates the partial matching of the EMD: since the 5 notes of that very short phrase can be matched with the first 5 notes in the query, this phrase gets a rather high similarity measurement. Among the false positives within the first 24 matches are in sum 4 examples of such shorter melodies that are rhythmically not very similar to the query.

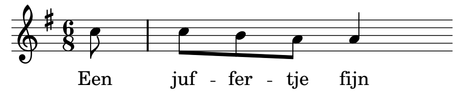

Figure 7: Melodic phrase at rank 3 according to EMD

The low rank 79 for the first phrase of *OGL 25904* (see figure 8) has its main reason in the existence of many shorter melody phrases in the test corpus. These phrases match a part of the query with a lower total weight than the weight of this phrase leading to a higher similarity value.
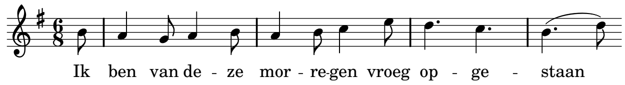
Figure 8: Melodic phrase at rank 79 according to EMD

In summary all three methods miss only very few group members by assigning them a low rank within the list. While the list ordered according to IMA contains 29 rhythmically similar melodies at the top and covers in this range also all first phrases from the group *Deze morgen*, both PTD and EMD miss 2 phrases.

For the comparison of the entire melodies using *OGL 19914* as the query we will give only a very short overview. The ranking list according to IMA contains within the first 11 hits 1 false positive and misses 2 melodies that have one phrase less then all other group members (rank 27, 57). The ranking list according to PTD contains within the first 17 hits 7 false positives at the beginning of the list. The two melodies being missed at the beginning (rank 45 and 107) are again the songs that have one phrase less. The ranking list according to EMD contains within the first 17 hits 9 false positives, three melodies have a much lower rank, two of them are the same being missed by IMA and PTD. In summary, the mean rank for all melodies belonging to the group is 10.75 for IMA, 20.3 for PTD and 20.8 for EMD.

Hence the melody group of *Deze morgen* is an example that demonstrates how much the rhythmic structure alone determines similarity. The comparison of the three models reveals the best results for the IMA, while the EMD has the most false positives due to its partial matching.

3.2 Summary of further results

A problem in the application of the PTD occurred in the comparison of last phrases. For instance, in the ranking list for the last phrases of the melody group called *Heer Halewijn A* both IMA and EMD find 9 similar phrases among the first 10 respective 11 hits. In contrast to this, PTD ranks only 2 of them at the top of the list, the others are ranked lower than rank 31. The reason for this is a difference in the duration of the last note (due to different transcription strategies of the recorded melodies). If the last note in all the examples is adjusted to the same duration, PTD lists 8 of the phrases among the top 11 melodies. Similar effects are observed in the ranking lists for the entire melodies. In most of the examples IMA gains the best result. However, if the query is rhythmically only very little differentiated (such as a quasi continuous eighth note chain), then the results of both Transportation

Distances are more convincing. In general rhythmic similarity seems to be an important component of the similarity of melodies in the current test corpus of melodies from *Onder de groene linde.*

4 Conclusion

The aim of the comparison of the computational approaches to rhythmic similarity in this paper is a first test in how far the different methods are suited for finding rhythmically similar melodies. For the application of the PTD a solution concerning the length of the last note of a phrase has to be found. For the application of the EMD it might be necessary to filter out hits that are much shorter then the query if one is not interested in partial matching. The use of the metric weights obtained by IMA as the weights in the Transportational Distances instead of the duration could be a promising merge of the two models. The application of the Transportational Distances to pitches only while ignoring the rhythm information and a comparison to the results obtained in this paper is a further step towards the investigation of the importance of rhythmic similarity in the context of melodic similarity.

References

[1] Bosma, M., Veltkamp, R. C., Wiering, F. "Muugle: A framework for the comparison of Music Information Retrieval methods". In *Proceedings of the ICMPC 2006*, pp. 1297-1303.

[2] Chew, E., Volk, A., and Lee, C. "Dance Music Classification Using Inner Metric Analysis". In *Proceedings of the 9th INFORMS Computer Society Conference.* Kluwer, 2005, pp. 355-370.

[3] Fleischer (Volk), A. *Die analytische Interpretation. Schritte zur Erschließung eines Forschungsfeldes am Beispiel der Metrik.* dissertation.de - Verlag im Internet Gmbh, Berlin, 2003.

[4] Mazzola, G. *The Topos of Music.* Birkhäuser, Basel, 2002.

[5] *Onder de groene linde. Verhalende liederen uit de mondelinge overlevering.* Uitgeverij Uniepers, Amsterdam, 1987-1991

[6] Typke, R. *Music Retrieval Based on Melodic Similarity.* PhD thesis, Utrecht University, 2007.

[7] Volk, A. "The Empirical Evaluation of a Mathematical Model for Inner Metric Analysis". In *Proceedings of the 5th Triennal ESCOM Conference.* Hanover, 2003.