THE SIMILE ALGORITHM FOR MELODIC SIMILARITY

Klaus Frieler
University of Hamburg
Hopfenstrasse 22
20359 Hamburg, Germany
kgf@omniversum.de

Daniel Müllensiefen
University of Hamburg
Christoph-Probst-Weg 8
20251 Hamburg, Germany
dmuellensiefen@freenet.de

ABSTRACT

This abstract sketches the approach and the employed means for constructing measures of melodic similarity in our software toolbox SIMILE. For different tasks related to melodic similarity different hybrid algorithms proved to be optimal. The hybrid algorithms consist of several single algorithms mostly known in the literature that are taken together in a linear combination. One special linear combination is chosen to serve for the given task in this contest. The single measures contained in this hybrid algorithm are explained briefly and references to more detailed publications are given.

Keywords: Melodic similarity, edit distance, n-grams, harmonic weighting.

1 BASIC APPROACH

The approach of the SIMILE toolbox for melodic analysis and similarity measurement is to combine algorithms that cover different aspects of a specific task and build hybrid algorithms from them.

2 BACKGROUND: EMPIRICAL RESEARCH ON MELODIC SIMILARITY

For melodic similarity analysis we implemented about 50 algorithms that can be used to determine the similarity between to given melodies. Most of the methods are standard algorithms from the literature (e.g. edit distance on pitch values, n-grams on intervals, correlation of interpolated pitch contour values), some of them were created by our own (similarity of implicit tonalities, edit distance on categorised rhythm values). A detailed overview over the implemented algorithms is given in Müllensiefen & Frieler (2004).

For several specific tasks related to melodic similarity we gathered rating data by human experts in experiments. This data was used as ground truth to evaluate and optimise the implemented similarity measures for that given task. Among those tasks were a) judging the similarity of melodic variants to one reference melody, b) judging the similarity of melodic variants and completely different melodies (non-variants) to one reference melody, and c) judging the similarity of variants and non-variants of a short reference phrase. The empirical results for tasks a) and b) are published in detail in Müllensiefen & Frieler (2004).

We used linear regression to explain the human judgements by taking the different similarity measures as predictors. For different tasks we obtained different optimal combinations of similarity measures to predict the human data, suggesting that human experts base their judgement on different aspects of the melodies according to the task demands. Three different optimal linear combinations of similarity algorithms can again be found in Müllensiefen & Frieler (2004). Another combination of different similarity measures that resulted to be optimal for short phrases as in the incipit material for this contest, is described below.

3 THE HYBIRD ALGORITHM FOR SHORT MELODIC PHRASES

For the task in the present contest, which consists in estimating the similarity between one short reference phrase and variants and non-variants of this phrase, we chose an hybrid measure that resulted from the data anylsis of an experiment on similarity of melodic phrases that has not been published yet. We call this hybrid measure **optiP**. It represents a linear combination of six different algorithms that differ in the musical dimension considered, the abstraction method for this dimension, and the similarity algorithm. The abstraction and similarity methods we used are explained briefly in the two following paragraphs.

3.1 Melodic dimensions and abstraction methods

The abstraction methods we implemented so far for representing data in different musical dimensions of single line melodies are:

- **Pitch:** MIDI quantisation, leap/step-quantisation, Parsons Code, see Müllensiefen & Frieler (2004).
- **Rhythm:** Categorisation to five duration classes, representation as 'gaussified' values, see Müllensiefen & Frieler (2004), Frieler (2004).
- **Contour:** Different methods for smoothing coarse directional movements, Fourier transform, see e.g. Steinbeck (1982).
- **Implied tonality:** Categorisation according to harmonic content as based on Krumhansl's tonality vector, e.g. Krumhansl (1990).
- **Accent structure:** Combinations of Gestalt-like accent rules from psychological literature, e.g. Jones (1987).

3.2 Similarity algorithms

Data from any abstraction method can be combined with most of the following similarity algorithms that we employ for comparision. But in the actual implementation of the similarity measures not all combinations of abstraction methods and similaruity algorithms made sense. The implemented general similarity algorithms are:

- **Edit Distance**: e.g. Mongeau & Sankoff, (1990)
- **n-grams**: e.g. Downie (1999)
- **Geometric distance**: Steinbeck (1982); O'Maidin (1998)
- **Correlation coefficient**: e.g. Steinbeck, (1982)

3.3 Optimised similarity algorithm for short melodic phrases

$$\begin{aligned} optP = {} & 0.31 \cdot joint412 + 1.37 \cdot rawEd \cdot connEd \\ & + 0.643 \cdot rawEd \cdot harmCoEA \\ & - 1.55 \cdot connEd \cdot bGrCoorF \\ & + 0.65 \cdot bGrCoorF \cdot rhytFuzz \\ & - 0.39 \cdot harmCoEA \cdot rhytFuzz - 0.133 \end{aligned} \quad (1)$$

Where

- **joint412**: Accent measure consisting of four Gestalt-rules; the accents from all four rules are summed for every note and the resulting sequences of number symbols are compared with the edit distance
- **rawEd**: Edit distance on sequence of raw (transposed) pitch values
- **connEd**: Contour measure: All pitch values between melodic turning points are interpolated and the resulting number sequences are compared with the edit distance
- **harmCoEA**: Harmonic measure: The main tonality of the melody is calculated according to Krumhansl's algorithm and the two tonalities are compared with the edit distance; for short phrases due to transposition the measure only separates major from minor tonalities
- **bGrCoorF**: Bi-grams of intervall categories (=three-note sequences) are counted for both melodies and are compared using the coordinate matching measure.
- **rhytFuzz:** Edit distance of sequences of categorised rhythm values

4 CONTEST RESULTS

The goal of the MIREX 2005 symbolic melodic similarity contest was to retrieve the most similar from 585 RISM A/II collection incipits for 11 different incipits as queries. The result to each query was a list ranked for similarity to the original. The ranked lists were compared with human ratings for the same queries and the same collection. Our algorithm reached only the seventh rank among the seven participants. The absolute measures of comparision between our results and the ratings of the human subjects are quite low as well as can be seen from the following table:

Average Dynamic Recall	51.81%
Normalised Recall at Group Boundaries	45.10%
Average precision (non-interpolated)	33.93%
Precision at N documents	33.71%
Runtime	54.593 s

Evaluation measures for the optP algorithm from the symbolic melodic similarity contest

The rather poor contest results of our algorithm demands detailed explanations which we have not investigated so far, but some points should and can already be made.

1. We did not optimise our algorithm to the training data for this contest, because we wanted a independent cross-check of our hybrid approach with a different data set. Unfortunately, it turned out that the optimisation to our data set did not very well fit to the data used in the contest. This is the usual danger of statistical data fitting and looking at the complexity of our algorithm it can be suspected that it suffered from over-fitting. It would be interesting to run all other algorithms of the contest over our data.
2. As explained in Müllensiefen & Frieler (2004), our data was collected from music experts rating the similarity of pop music phrases and corresponding variants. This is quite a different task than the one used in the contest, and our optimisation aimed at the explicit modelling of absolute expert ratings and not rank lists. Moreover, we excluded more than half of the subject from the data analysis because they did not meet our rigid criteria of reliability and validity as could be seen from their ratings. So the populations from which of our subjects and from which the subjects for the ground truth data was drawn, might differ.
3. Our experimental material (current pop melodies) and the RISM repertoire (incipits from melodic themes from classical music until around 1800) differ substantially. The results for our algorithm may therefore indicate that our initial assumption about general cognitive strategies for judging melodic phrases from (early) classical music and modern pop music is not adequate.
4. Despite the results we still believe that our hybrid approach is valuable, because we always found its advantage over more simple approaches with our experimental data. The lesson that can be drawn from this contest is, that it still needs more effort and more data to achieve true general supremacy.
5. Finally, the overall results of all algorithms in fulfilling the task are not totally satisfactory which shows that symbolic melodic similarity is a rather complex issue and still far from being regarded as solved.

We like to thank the organisators and contributors to this contest for the great work they have done and which we believe will help to improve the scientific advance in this field!

REFERENCES

[1] Downie, J.S. (1999). *Evaluating a simple approach to music information retrieval. Conceiving melodic n-grams as text*. PhD Thesis, University of Western Ontario.

[2] Frieler, K. (2004). Beat and Meter Extrraction Using gaussified onsets. ISMIR Proceedings Barcelona, 2004

[3] Jones, M.R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, 41, p. 621-634.

[4] Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

[5] Mongeau, M., and Sankoff, D. (1990). "Comparison of musical sequences". *Computers and the Humanities* 24, 161-175.

[6] Müllensiefen, D., and Frieler, K. (2004). "Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgements”. *Computing in Musicology* 13, 147-176.

[7] O`Maidin, D. (1998). A Geometrical Algorithm for Melodic Difference in Melodic Similarity. *Melodic Similarity: Concepts, Procedures, and Applications. Computing in Musicology* 11. Ed. Walter B. Hewlett & Eleanor Selfridge-Field. Cambridge: MIT Press.

[8] Steinbeck, W. (1982). *Struktur und Ähnlichkeit. Methoden automatisierter Melodieanalyse*. Kieler Schriften zur Musikwissenschaft XXV. Kassel: Bärenreiter.

[9] Uitdenbogerd, A. L (2002). *Music Information Retrieval Technology*. PhD thesis. RMIT University Melbourne Victoria, Australia