On Identifying Folk Song Melodies Employing Recurring Motifs

Peter van Kranenburg,[*1] Anja Volk,[#2] Frans Wiering[#3]
[*]*Meertens Institute, Amsterdam, Netherlands*
[#]*Utrecht University, Department of Information and Computing Sciences*
[1]peter.van.kranenburg@meertens.knaw.nl, [2,3]{A.Volk,F.Wiering}@uu.nl

ABSTRACT

The recurrence of characteristic motifs plays an important role in the identification of a folk song melody as member of a tune family. Based on a unique data set with expert annotations of motif occurrences in a collection of Dutch folk song melodies, we define 15 abstract motif classes. Taking a computational approach, we evaluate to what extent these 15 motif classes contribute to automatic identification of folk songs. We define various similarity measures for melodies represented as sequences of motif occurrences. In a retrieval experiment, alignment measures appear the most successful. The results are additionally improved by taking into account the phrase position of motif occurrences. These insights motivate future research to improve automatic motif detection and retrieval performance, and to determine similarity between melodies on the basis of motifs.

I. INTRODUCTION

The study of melodic variation among folk song tunes is of interest for various fields of research. In Ethnomusicology, the variation of melodies in oral circulation is a natural topic of interest. During the 20th century, many field-recordings have been made of Western folk songs. One of the most prominent examples is the work done by Béla Bartók (see e.g., Bartók, 1997). To categorize the resulting collections of melodies, many classification systems have been developed (see e.g., Bohlman, 1988).

From the perspective of Music Cognition the study of folk song melodies is also relevant. Since folk songs are part of oral culture, the transmission of a melody from singer to singer heavily depends on the features of the underlying human processes of memorization and reproduction. Rather than an unchanged note-to-note reproduction of a melody, minor or major parts of the melody change when people sing the songs back from memory. Therefore, in the course of oral transmission, a melody can change considerably. Yet, despite the changes, the melody usually retains its 'identity', namely that is is recognizable as the 'same' song. This raises the question what parts or aspects of the melody are stable in oral transmission and what parts or aspects can change without obscuring the identity of the melody. The research that we present in the current paper relates to this question.

Finally, in Music Information Retrieval, the availability of large collections of digitized monophonic folk song melodies provides data that can be used to evaluate melodic similarity measures or complete retrieval systems.

A central concept in the study of folk song melodies is *tune family*. This concept was introduced in the 1950s by Samuel Bayard to denote a group of melody instances that supposedly 'descend' from one single tune through the process of oral transmission. Later on, an extension of the concept was proposed by James Cowdery (1984). He considered other types of melodic relatedness than having a common 'ancestor' for establishing tune family membership. One of these is melodies being composed from the same 'pool of motifs'. Such a pool of motifs consist of concrete melodic material that is available to the folk musician for constructing a melody.

In a previous project, we investigated melodic similarity among a collection of Dutch folk song melodies (Volk & Van Kranenburg, rint). One of our findings was that recurring melodic motifs are the most important basis for musicological specialists to classify melodies into tune families. Occurrences of such motifs serve the recognition of a melody as member of a tune family. This finding suggests that these melodic motifs are elements of melodies that are relatively stable within the process of oral transmission, and presumably also in human melodic memory.

Since the recurrence of motifs seems very important, our current aim is to use motifs instead of individual notes to compare melodies with each other using a computational approach. We employ 15 explicitly defined motif classes based on a unique data set with expert annotations of motif classes and occurrences in a collection of Dutch folk-song melodies. We represent folk song melodies as sequences of motifs, and we investigate to what extent these sequences provide information to identify a melody as a member of a certain tune family. It is not our aim to improve results of existing note-based melodic similarity measures.

In this paper we test the following hypotheses:

1. Folk song melodies can be identified as member of a tune family solely based on occurrences of motifs.
2. The time order of the occurrences of the motifs is important for the identification of the melodies.
3. The position in the phrase where the motifs occur is important for the identification of the melodies.

II. DATA

The data set we work with is the collection *Onder de groene linde*, which contains around 7,000 audio recordings of Dutch folk songs, hosted at the Meertens Institute, Amsterdam (Grijp, 2008). Around 2,500 of these recordings were encoded for computational processing. We also use folk songs from written sources. Altogether, for the current study, we use a collection of 4,830

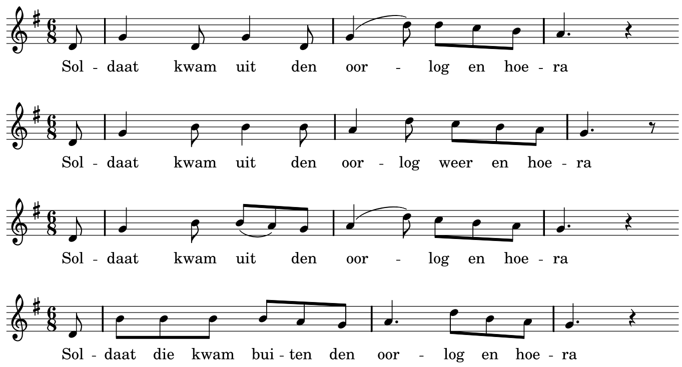

Figure 1. Incipits of four members of the tune family *Soldaat kwam uit de oorlog.*

songs. The pitches are represented according to the base-40 pitch encoding (Hewlett, 1992).

One of the tasks of the collection specialists of the institute is to classify the melodies into tune families. The collection only contains 'end points' in the process of oral transmission. The full history of a melody, comprising the ancestral variants, is lost. Therefore, in practice, the assignment of a recorded tune to a tune family is done based on similarity relations between the melodies. As an example of the degree of variation that can be found among tune family members, Figure 1 shows four melodies from the tune family *Soldaat kwam uit de oorlog*.

In a previous study, our aim was to understand how these assignments are established (Volk & Van Kranenburg, rint). We asked the collection specialists to annotate similarity relations among melodies in a subset of 360 melodies in 26 tune families: the *Annotated Corpus*. One of our findings was that recurrence of characteristic motifs plays a major role in the classification of the melodies, even more so than similarity of melodic contour or rhythm. Therefore, we asked the collection specialists to annotate occurrences of such characteristic motifs in the set of 360 melodies. Since occurrences of a motif can show some variation, we introduce the terms *motif class* and *motif occurrence*. A motif class is a common label for occurrences of the 'same' motif. In total c. 1400 motif occurrences in c. 100 motif classes were annotated, which is a rich source of musical data.

A. Definition of Motif Classes

Since we have the melodies available as sequences of notes, we need to convert the melodies into sequences of motifs. Due to the bottom-up procedure that was taken during the annotation phase, motif classes were only defined within tune families (Volk et al., 2008). Therefore, the annotated motifs are not general enough for a computational classification experiment. Our approach is to define a limited set of motif classes at a level that is more abstract than that of the annotated motif classes, and to implement a detection rule for each of them. After this step, occurrences of these abstract motif classes can be detected in all melodies.

A manual inventory of the entire set of annotated motifs showed that at least 15 *types* of motif classes recur in more than one tune family. Together, these 15 types of motif classes account for at least 85 of the 100 motif classes in the Annotated Corpus. The remaining annotated motif classes do not show features that are shared with other classes. They are really specific for a certain tune family.

Although the 15 types of motif classes are more general than the annotated motif classes, we will also use the term motif class for them in the remainder of this paper. For all our investigations, we use these 15 motif classes rather than the, unformalized, annotated motif classes. For each motif class, we define a detection rule. We aim at simple detection rules, which implies that motif occurrences that show relatively complex differences with respect to the defined motif classes could be left undetected. As a consequence, the results in this paper show to what extent our specific set of 15 motif classes is useful for melody retrieval. The classes are the following:

Rhythmic Syncope A note is considered a rhythmic syncope if during the note a metrical moment occurs with stronger metrical weight than the weight at the onset of the note, given the notated meter.[1]

Repercussion A sequence of three or more notes with the same pitch.

Long Notes A sequence of three or more notes that are relatively long. Relatively long means that the durations are greater than the average duration of all notes in the song, and that the occurrence rates of the durations are lower than the occurrence rates of both the most frequent and the second most frequent durations in the entire song.

Broken Chord A sequence of three or more notes of which the pitches form a triad in any inversion.

Leaps A succession of three or more melodic leaps greater than a second.

Big Leap A melodic leap of a perfect fifth or greater.

Tritone Leap A melodic leap of an augmented fourth or diminished fifth.

V-I Upbeat A V-I Upbeat occurs at the beginning of a phrase if the first bar has an upbeat of which the pitch is either a fourth below or a fifth above the first note of the first full bar.

Leap before Final The final note of the melody is preceded by a leap.

Ends on Third The final note of the melody is the (major) third in the scale. To determine this, a histogram is constructed according to the occurrence rates of each of the 40 pitch classes in the entire melody. If the final note is the third of the scale,

[1] For the computation of the metric weight we use the python toolbox *music21* (http://mit.edu/music21).

the tone a minor second above it is the fourth of the scale. The corresponding bin in the histogram is strongly expected not to be empty. The same applies to the tone a major second below the final note of the melody, which would be the second of the scale if the final note is the third. The bins of both control tones are expected to be empty in the case that the final note is the tonic. Thus, the final tone is considered to be the major third if the bins of both control tones are not empty.

Unaccented Final This occurs if the final note of a phrase has metric weight < 0.5, which is the weight of a secondary accent in music21.

Ascending Line Succession of three or more ascending intervals.

Descending Line Succession of three or more descending intervals.

Peak The highest note in a phrase, if it is the only occurrence of that pitch in the phrase.

Trough The lowest note in a phrase, if it is the only occurrence of that pitch in the phrase.

Occurrences of these motif classes are not allowed to span phrase boundaries, which are given in the data. As an example, Figure 2 shows both detected and annotated motifs for three melodies from the tune family *Soldaat kwam uit de oorlog*. The number of detected motif is larger than the number of annotated motifs, which is only one for these melodies. It was not the aim of the annotators to be exhaustive. They annotated the most salient motifs. In most cases in Figure 2, the annotated motif corresponds with a combination of Peak and Descending Line in the detected motifs.

Although the construction of this set of 15 motif classes is performed in a bottom-up, data-driven, way, the resulting set does show some internal structure. There is one rhythmical motif class: Rhythmic Syncope. Among the remaining 14 pitch-based motif classes, there are four classes that reflect special properties of sequences of three or more notes: Repercussion, Long Notes, Broken Chord, and Leaps. Three of the motif classes are special kinds of leaps: Big Leap, Tritone Leap, and V-I Upbeat. Three motif classes are related to the final note of the melody: Leap before Final, Ends on Third, and Unaccented Final. Finally, there are four motif classes that are related to contour: Ascending Line, Descending Line, Peak, and Trough.

Together, these 15 motif classes are a of major importance for the identity of folk song melodies by the collection specialists of the Meertens Institute. Moreover, from a cognitive point of view, these motif classes reflect kinds of melodic structures that serve the recognition of melodies.

III. METHOD

A. Similarity Measures for Motif Sequences

To compute the similarity of two melodies that both are represented as a sequence of motif occurrences, we compare two approaches: a similarity measure in which the time order is not taken into account, and an alignment approach in which the time order is preserved. The comparison of the results of these two approaches serves to test hypothesis 2. Furthermore, we will include the position of the motif occurrences within the phrase in the alignment approach, which addresses hypothesis 3. As a reference, we compare the results of the motif-based alignment with a previous alignment method that operates at note-level.

1) Bag-of-Motifs. At first, we take a bag-of-motifs approach. Each melody is represented by a 15-dimensional vector. Each element captures the relative importance of a motif class according to the *tf-idf* weighting scheme, in which the term frequency (tf) of a motif class in the melody is weighted by the inverse document frequency (idf) of that motif class, where tf indicates the relative importance of the motif in a single melody and idf the relative importance of the motif in the corpus as a whole (cf. Manning et al. (2008, Section 6.2)). As similarity measure we take the cosine similarity between two vectors (cf. Manning et al. (2008, Section 6.3)).

2) Alignment-Based Similarity Measures. Since there is a linear ordering in the motif occurrences, sequence alignment is an appropriate choice to compare two melodies. The Needleman-Wunsch sequence alignment algorithm finds the optimal alignment of two sequences of symbols given the two sequences, a similarity measure for symbols, and a gap penalty (Needleman & Wunsch, 1970). The similarity measure for symbols indicates to what extent we want the two symbols to be aligned. By providing an appropriate definition for this similarity measure, the alignment algorithm can be employed for a specific problem. In our case, the symbols represent occurrences of motif classes and the melody is a sequence of occurrences of motif classes. Since more than one motif class can occur at the same place, the symbols are actually sets of motif classes. We denote the first melody with $\mathbf{x} : x_1, \ldots, x_i, \ldots, x_n$, and the second melody with $\mathbf{y} : y_1, \ldots, y_j, \ldots, y_m$, in which symbols x_i and y_j are the sets of motif classes that respective occur at position i in sequence $\mathbf{x}$ and position j in sequence $\mathbf{y}$. We define two variants of the similarity measure for symbols. Firstly, $S(x_i, y_j)$, which is the number of matching motifs minus the number of mismatching motifs between the sets x_i and y_j, and secondly, $S_{prh}(x_i, y_j)$, which is $S(x_i, y_j)$ weighted by the difference between the positions of x_i and y_i in their respective phrases. The position of a symbol in the phrase, $phr(x)$, is expressed as a real valued number between 0 and 1, where 0 corresponds with the onset time of the first note and 1 corresponds with the onset time of the last note of the phrase. For the phrase position of a motif occurrence the phrase position of the note at which the motif occurrence was detected is taken. The absolute value of the difference between $phr(x_i)$ and $phr(y_j)$ results in a number between 0 and 1 that can be used as weighting factor. The less the positions correspond, the lower the resulting value of S_{phr}. We apply the weighting for positive values of $S(x_i, y_j)$ only, which corrects the similarity score for pairs

Figure 2. Three melodies from the tune family *Soldaat kwam uit de oorlog*. Both the manually annotated motifs (left) and the detected motif classes (right) are shown. All labels are above the staves.

of symbols with a net match of motifs, but does not affect the, already low, score for symbols with a net mismatch. The definitions are as follows:

$$S(x_i,y_j) = |x_i \cap y_j| - |(x_i \cup y_j) \setminus (x_i \cap y_j)|$$

$$S_{phr}(x_i,y_j) = \begin{cases} (1-|phr(x_i)-phr(y_j)|)S(x_i,y_j) & \text{if } S(x_i,y_j) > 0 \\ S(x_i,y_j) & \text{if } S(x_i,y_j) \leq 0 \end{cases}$$

These similarity measures are used in the alignment algorithm by filling a n by m dynamic programming matrix D recursively according to:

$$D(i,j) = \max \begin{cases} D(i-1,j-1)+S(x_i,y_j) \\ D(i-1,j)-\gamma \\ D(i,j-1)-\gamma \end{cases}$$

in which γ is the gap penalty, $D(i,0) = 0$ for $0 < i \leq n$, and $D(0,j) = 0$ for $0 < j \leq m$. For the gap penalty we take $\gamma = 0.5$, by which we make the insertion of a gap 'cheaper' than a net mismatch of one motif. After computing the matrix, $D(n,m)$ contains the score of the alignment, which is the sum of the similarity values of the symbols that are aligned with each other and the gap penalties for the inserted gaps. We take this score as a measure of similarity of the two melodies: the better the two melodies can be aligned, the more similar they are. By comparing the differences between employing $S(x_i,y_j)$ and $S_{phr}(x_i,y_j)$, we test the hypothesis whether including the position of the motifs in the phrase improves recognition and retrieval of melodies.

B. Retrieval Evaluation

To evaluate the similarity measures, we take each of the 360 melodies of the Annotated Corpus as query. For each of the 360 queries a ranked list is constructed containing all remaining 4829 melodies ordered according to the computed similarity values. The relevant items are all melodies that belong to the same tune family as the query. For evaluation, we use the mean average precision (MAP) and the recognition rate. The MAP value reflects how well *all* relevant items are retrieved (see Manning et al. (2008, Section 8.4)). The *recognition rate* is the fraction of queries that has a relevant item at the top-position of the ranked list. This measure indicates how well the distance measure is able to serve the recognition of a melody, since for recognizing a melody as member of a tune family, only one relevant result can be enough. As a related measure, we also compute the fraction of queries that have at least one relevant item among the first 10 positions, the *10nn-recognition rate*.

IV. RESULTS

Table 1 summarizes the values of the evaluation measures for the various approaches. Figure 3 shows the relation between the average precision and average recall with the rank as parameter: for each configuration, we take for each rank the average of both the precision and the recall values over all 360 ranked lists, resulting in a pair of average precision and average recall for each rank $r \in [1,2,...,4830]$. In the diagram, for each configuration, the pairs of average precision and average recall are connected.

Similarity Measure	MAP	RR	10nn-RR
Bag-of-Motifs	0.10	0.26	0.55
Motif Alignment	0.29	0.68	0.85
Motif Alignment + Phrase Pos.	0.34	0.75	0.90
Note Alignment	0.68	0.92	0.98

Table 1. Retrieval Evaluation Measures for the various similarity measures. MAP is the mean average precision, RR is the recognition rate, 10nn-RR is the fraction of queries that has at least one relevant item among the first 10 items on the ranked list.

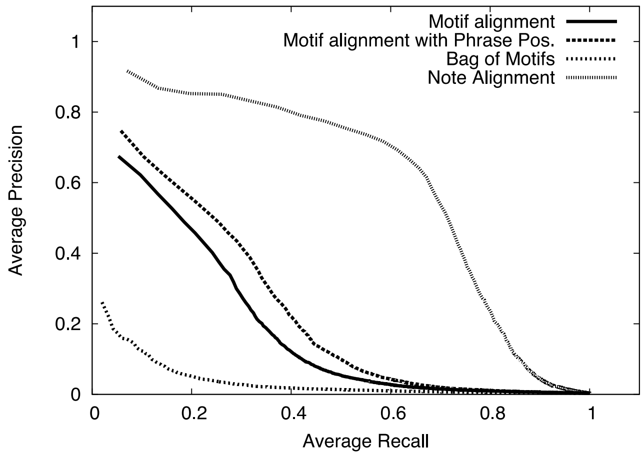

Figure 3. Diagrams of Average Precision vs. Average Recall for the various similarity measures.

We test whether the differences between the similarity measures are significant by performing a non-parametric Friedman test, with significance level $\alpha = 0.05$, using for each similarity measure the average precisions of the 360 individual queries. With a post hoc Tukey HSD test, we determine which of the measures differ significantly. The Friedman test shows that there are significant differences, $\chi^2(3,N=360) = 861, p < 0.0001$. The Tukey HSD test points out that all differences are significant.

In general, given the small size of the tune families (typically between 10 and 20 members), the results for the retrieval of tune family members from the corpus of 4,830 melodies based on characteristic motifs support hypothesis 1. The basic alignment approach only using the sequences of motif occurrences achieves a recognition rate of 67.78%. However, the low MAP of 0.287 shows that many melodies from other tune families are at relatively high positions, and that some of the tunes that should be found are only listed in the lower ranks of the result lists.

The results for the alignment approach show a large and significant improvement compared to those of the bag-of-motifs approach. This implies that the time order of the motifs is important for the recognition of melodies, which confirms hypothesis 2.

The motif alignment approach can be improved through taking into account the phrase position of motifs: it obviously is relevant whether for instance a broken chord motif begins, or ends a phrase. Such an influence could be observed indeed. The recognition rate improves to 75%, and the MAP to 0.337, respectively. This supports hypothesis 3. Still, the tune recognition based on motifs is outperformed by a comparable approach we took in a previous study, in which we aligned sequences of notes rather than sequences of motifs (Van Kranenburg, 2010).

V. CONCLUSIONS AND FUTURE WORK

In this paper we set out to evaluate the extent to which occurrences of melodic motifs contribute to the establishing of tune family membership of folk song melodies. The 15 pre-defined melodic motif classes we employ are based on annotations of melodic motifs that are considered of importance for the identity of melodies.

Our approach to retrieve folk song melodies solely based on occurrences of these pre-defined melodic motif classes, proves to be successful. From a collection as large as 4,830 melodies, many members of the same tune family could be retrieved. Our best performing configuration achieved a recognition rate of 0.75 and a MAP of 0.34. 90% of the queries have a relevant item among the first 10 ranks. This means that the approach is suited for identification of melodies, for which only one or a few relevant results are sufficient. Considering the size of the collection, this shows that occurrences of melodic motifs provide much information on the identity of a song as member of a tune family. Nevertheless, there are many false positives and false negatives. Therefore, the current method is not suited to retrieve all tune family members given a query.

As was proposed, the results show that melody identification based on motifs works better if the order of motifs is considered, than when a bag-of-motifs approach is used as a similarity measure. This confirms the findings of Zbikowski (2002) that time-order is of crucial importance in music.

The position of a motif in a phrase proved to be an important additional cue for the recognition of melodies, which agrees with the result of Margulis (2012) that literal repetitions of patterns are not detected by listeners if they do not occur at corresponding phrase positions.

Despite the promising results, the note-alignment still performs better than the motif-alignment. This shows that the melodies contain more information about their identity than currently is captured by the 15 motif classes. To better understand the differences between the note-level alignment and the motif-level alignment one of our next steps will be to analyze the actual result lists for both methods.

The annotated motifs show differences per tune family. Therefore, to evaluate to what extent different tune families are characterized by different motifs, we plan to perform a feature selection procedure in order to obtain the optimal set of motif classes per tune family.

Moreover, it would be interesting to test the method described in this paper on different corpora, e.g., the Essen folksong collection. This would indicate to what extent our sparse set of motifs can also contribute to the detection of melodic similarity between tunes from different cultural backgrounds.

On the long term, we will work on better models of musical motifs and on more robust automatic detection. This will be, in our opinion, a significant step towards understanding the process of oral transmission, and to predict what kind of variations can occur. Eventually, our aim is to automatically derive an appropriate set of motif classes from a corpus of melodies.

One of the grand challenges of computational musicology is to model and understand the changes that occur to melodies in oral transmission. Eventually, we expect our research to lead to new insights into the musical parameters that influence the perception and recall of musical motifs, and their stability or instability in oral transmission.

ACKNOWLEDGMENTS

This work has been carried out within the Tunes and Tales project, which is funded by the Royal Netherlands Academy of Arts and Sciences through the Computational Humanities program. Anja Volk is supported by the Netherlands Organization for Scientific Research, NWO-VIDI grant 276-35-001.

REFERENCES

Bartók, B. (1997). *Bela Bartók studies in ethnomusicology*. Lincoln: University of Nebraska Press.

Bayard, S. (1950). Prolegomena to a study of the principal melodic families of british-american folk song. *Journal of American Folklore*, 63(247), 1–44.

Bohlman, P. (1988). *The Study of Folk Music in the Modern World*. Bloomington: Indiana University Press.

Cowdery, J. (1984). A fresh look at the concept of tune family. *Ethnomusicology*, 28(3), 495–504.

Grijp, L. P. (2008). Introduction. In L. P. Grijp & I. van Beersum (Eds.), *Under the Green Linden — 163 Dutch Ballads from the Oral Tradition* (pp. 18–27). Amsterdam: Meertens Institute + Music & Words.

Hewlett, W. B. (1992). A base-40 number-line representation of musical pitch. *Musikometrika*, 4, 1–14.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Margulis, E. (2012). Musical repetition detection across multiple exposures. *Music Perception*, In Print.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3), 443–453.

Van Kranenburg, P. (2010). *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*. PhD thesis, Utrecht University, Utrecht.

Volk, A. & Van Kranenburg, P. (in print). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*.

Zbikowski, L. (2002). *Conceptualizing Music*. Oxford University Press.