

# EXPLORING PHRASE FORM STRUCTURES. PART I: EUROPEAN FOLK SONGS.

Klaus Frieler

Liszt School of Music Weimar

{klaus.frieler}@hfm-weimar.de

## ABSTRACT

In this explorative study, we investigate the phrase form structures of 7821 folk songs taken from the EsAC database. The main purpose was to test whether phrase form structure is a useful feature for computational folk song analysis, e. g. classification tasks. To this end, we examined the self-similarity of phrase sequences in folk songs with regard to semitone intervals and duration classes as well as in combination. Phrase form structure can be characterised by coherence values measuring the amount of repetition. Duration-based form structures show a coherence about twice as high than interval-based form structures, i. e., rhythmic models are much more likely to be repeated than interval sequences. The most common interval-based form is ABCD, whereas the the most frequent duration-based form model is AAAA. Folk songs show a strong preference for even-numbered form lengths, with 4, 6, and 8 phrases being the most common. In comparing sub-collections, significant differences between folk songs of different origin were found, particularly, between Irish and Polish tunes on one hand, and Central European songs on the other hand. Interestingly, the sub-collection of German children songs showed the most diverse set of different form structures, whereas the *Kolberg* collection of Polish song was the most homogeneous set. All in all, this shows that phrase form structure might indeed be a useful starting point to derive features for other computational applications.

## 1. INTRODUCTION

Exact and varied repetitions of musical units constitute musical form, which can be viewed as self-similarity within a piece of music. Since musical units are often hierarchically structured, musical form can be found on different hierarchical levels as well. One reasonable level to consider is that of the musical phrase, which can roughly defined as coherent Gestalt-like units. Hence, the investigation of phrase form structure (PFS) (also called phrase relationships, c.f. Sagrillo, 1999) of melodies is a good starting point for examing musical structure. This is the topic of the current paper, more specifically, the phrasal structures in a certain set of Central, Eastern, and Western Europe folk songs as contained in the well-known EsAC database (Schaffrath, 1995; Dahlig, 2000). The EsAC database is unique compared to most other folk song collections, since it provides phrase information. The annotation was mostly done by the transcribers, either of the original sources or the coders of the source, and are hence not fully objective. Furthermore, experiments in segment perception (e.g. Pearce et al., 2008; Spevak et al., 2002) showed that perceived phrase boundaries are subject to individual variation—a fact already well-known to folk song researchers. However, in the case of the EsAC database one can safely assume that the transcribers were highly trained experts

who chose the most likely and reasonable segmentation, particularly in the presence of lyrics (see Sagrillo, 1999 for a discussion, also Spevak et al., 2002).

The next higher level of form of parts such as verse, refrain, bridges etc, is beyond scope of the present paper, since the EsAC creators only coded one verse and one refrain for a songs if more than one were present. Unfortunately, they did not indicate it anywhere how many verses there were and which part might be considered a verse or a refrain. Hence, it is hardly possible to discern larger form structures (e. g. such as proposed in Lomax & Grauer, 1967) from the current data.

The next lower level, that of motifs would also be a good candidate, but the distinction between phrases and motifs is often hard to draw. Sub-phrases are most likely to be perceived if some smaller unit, i. e., a motif, is repeated within the same phrase or in another phrase of the song. Hence, the existence of motifs is already tightly connected to similarity, and thus a topic for pattern mining. In contrast, phrases boundaries are often determined by longer (breathing) rests and cognitive load (Temperley, 2001), as well as – particularly in the case of folk songs – by semantical and grammatical language units in the lyrics.

Sagrillo (1999) carried out a very similar investigation of phrase form structure for the *Luxemburg* collection, using the old EsAC analysis software ANA, which does not easily run anymore on modern computers. The software provided pitch-based and rhythmic form strings in the same manner as used here, but it is not known on which algorithm(s) the automated detection was based. Moreover, Sagrillo was considering variations and sequentiation of phrases, which is not done in this study, but nevertheless some of his results could be roughly reproduced as a subset of our results.

## 2. METHOD

The analysis was carried out using the version of the EsAC database included in the MeloSpySuite<sup>1</sup> (Frieler et al., 2013), which contains currently 7821 folk songs in 13 collections. To extract the form information, similarity values between each phrase of a song were calculated using edit distance (Levenshtein, 1965) on a interval-based or duration-class-based representation of the melody. Intervals are signed semitone differences, and durations classes range from “very short” to “very long” with the beat level

<sup>1</sup> Available at <http://jazzomat.hfm-weimar.de>

as the mid point. Form strings (such as AAAA, ABCD etc.) were extracted from the resulting self-similarity matrices using fixed numerical thresholds. Two phrases  $p_i, p_j$  were deemed identical if

$$\sigma_I(p_i, p_j) \geq 0.6$$

for intervals and

$$\sigma_D(p_i, p_j) \geq 0.7$$

for duration classes, where  $\sigma(p_i, p_j)$  is the normed edit similarity taking values in  $[0, 1]$  (Müllensiefen & Frieler, 2004a). Edit distances are generally the most successful algorithms in modeling perceived similarities and were applied in folk song research earlier (e.g., Müllensiefen & Frieler, 2004b; van Kranenburg et al., 2013). This particular thresholds were chosen after some informal testing while trying to replicate manual phrase form analysis as carried out by the author and as found in Sagrillo (1999). As any fixed thresholds, these are of course not perfect, and could surely be optimised further with more strict procedures. However, it turned out that these values are high enough to capture “true” similarities, since any value higher than  $\approx 0.3$  is already very unlikely to occur by chance (Müllensiefen & Frieler, 2004b), and low enough to still find a reasonable amount of PFS variation. Moreover, due to the smaller event space of duration classes spurious similarities are more likely to occur, the rhythm form threshold is slightly higher.

The analysis was carried out using the `melfeature` commandline tool from the `MeloSpySuite` (Frieler et al., 2013). Actually, the software allows to set two independent thresholds for labelling (nearly) exact repetitions and varied repetition, which were set equal here to disregard varied repetition to simplify further analysis. The resulting form strings were imported into R (R Core Team, 2013) for further analysis.

### 3. RESULTS

In total, 1008 different interval form strings (IF) and 1385 duration-based forms (DF) were found, which gave rise to 990 distinct combined form classes (CF). Combining was done by enumerating each unique pair of IF and DF symbols of a song with a new form symbol. The most common forms are ABCD for IF (17.7%) as well as CF (19.4%), and AAAA for DF (5.8%) (c.f. Sagrillo, 1999). The most common DF going along with the IF ABCD is AAAA (14.1%).

Form lengths range from 1 to 20 elements, with a median of 5 elements and a 75%-quartile of 7 elements. Very long forms are relatively rare, only 276 (3.5%) tunes have more than 10 phrases. The majority of songs consists of 4 phrases, followed closely by songs with 6 and 8 phrases. These results are very similar to Sagrillo (1999). Generally, a clear preference for an even number of form parts can be found; there are nearly twice as many forms with an even length than with an odd length. The number of different parts in a form is often smaller than the total form length. IFs and CFs have a median of 4 and DFs a median

of 3 different parts per form. One can define the *coherence* (or *redundancy*) of a form as the amount of contained repetition, i.e. the number of unique elements divided by the total length of the form (subtracted from 1 for better interpretation). A coherence of 0 means that no form part is repeated, whereas a coherence of 1 can only be reached in the limit of a single, infinitely repeated part. We found a median coherence for IFs of 0.2, for CFs of 0.1667, and for DFs of 0.5. Thus, DFs contain about more than twice as much repetitions than IFs. This means, that songs often exhibit roughly constant rhythmic models over more varying tonal content, often in consecutive phrases, since the probability for an IF part to be repeated is only 5%, but 35% for a DF part. Finally, we defined the *coherence ratio* as the ratio of IF and DF coherence. We found a median coherence ratio of 0.67, which means that in the average the coherence of IFs is about two-thirds of the DF coherence. Only in 277 songs (3.5%), the IF coherence was larger than the DF coherence.

On a global level, many differences with regard to PFS can be found between the collections. All ANOVAs using form length and the four coherence variables as dependent variables became highly significant ( $p < 0.0001$ ) due to the large dataset, but effect sizes were rather low (all  $R_{adj}^2 < 0.05$ , except for form length with  $R_{adj}^2 = 0.21$ ). Generally, the Central European songs, mostly from Germany, Lorraine, and Luxembourg, which make up approx. 90% of the database, were more similar to each other than to the 62 Irish songs and the 676 songs from the two Polish sub-collections. Interestingly, the most diverse spectrum of forms as measured by the entropy of the form distribution can be found in the *Kinderlieder* (German children songs), which has a nearly flat distribution. The most common IF for *Kinderlieder* is ABCD with a relative frequency of only 8%. In contrast, the *Kolberg* collection from Poland has the most homogeneous distribution. Here the most common IF is AB used in over more than half of the songs (54%). Furthermore, the *Irland* (Ireland), *Warmia* (North-East Poland), and *Kolberg* collection are the only collections in which the most frequent IF type is *not* ABCD, but AABA, AABC, and AB respectively. On the other hand, there are only two collections, *Lothringen* (Lorraine) and *Kolberg* (Poland), in which AAAA is *not* among the first two most frequent DF form types.

### 4. CONCLUSION

In this explorative study, we found interesting statistical details of phrase form types in a certain set of European folksongs. Collections from different cultural origin exhibit clear differences with respect to form types and derived features. Hence, we suggest that descriptors of phrasal form structure might be useful features for manual and automated classification and other types of folk song research.

## 5. REFERENCES

- Dahlig, E. (2000). An integrated system of encoding, analysing and processing of one-part melodies. In und Joachim Stange-Elbe, B. E. (Ed.), *Musik im virtuellen Raum*, volume 3 of *Musik und Neue Technologie*, (pp. 427–433)., Osnabrück.
- Frieler, K., Abeßer, J., Zaddach, W.-G., & Pfeleiderer, M. (2013). Introducing the Jazzomat Project and the Melo(S)py Library. In van Kranenburg, P., C. Anagnostopoulou, C., & Volk, A. (Eds.), *Proceedings of the Third International Workshop on Folk Music Analysis, Meertens Institute and Utrecht University Department of Information and Computing Sciences*, (pp. 76–78).
- Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Doklady Akademii Nauk SSSR*, 163(4), 845–848. Englische Übersetzung in: *Soviet Physics Doklady*, 10(8) S. 707-710, 1966.
- Lomax, A. & Grauer, V. (1967). *Cantometrics Coding Book*.
- Müllensiefen, D. & Frieler, K. (2004a). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, 13, 147–176.
- Müllensiefen, D. & Frieler, K. (2004b). Optimizing measures of melodic similarity for the exploration of a large folk-song database. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR2004)*, *Universitat Pompeu Fabra*.
- Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2008). A Comparison of Statistical and Rule-Based Models of Melodic Segmentation. In *Proceedings of the Ninth International Conference on Music Information Retrieval*, (pp. 89–94)., Philadelphia, USA. Drexel University.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Sagrillo, D. (1999). *Melodiegestalten im luxemburgischen Volkslied: Zur Anwendung computergestützter Verfahren bei der Klassifikation von Volksliedabschnitten*. Bonn: Holo.
- Schaffrath, H. (1995). The Essen Folksong Collection in Kern Format. In Huron, D. (Ed.), *Computer Database*, Menlo Park, CA.
- Spevak, C., Thom, B., & Hoethker, K. (2002). Issues in Melodic Segmentation. In *Proceedings of the Second International Conference on Music and Artificial Intelligence, Edinburgh, Scotland, 2002*, Springer's LNCS/LNAI series.
- Temperley, D. (2001). *Cognition of Basic Musical Structures*. Cambridge: MIT Press.
- van Kranenburg, P., Volk, A., & Wiering, F. (2013). A Comparison between Global and Local Features for Computational Classification of Folk Song Melodies. *Journal of New Music Research*, 42, 1–18.