CROSSING PHRASE BOUNDARIES IN MUSIC

Zuzana Cenkerová
Slovak Academy of Sciences
zuzana.cenkerova@savba.sk

Martín Hartmann
University of Jyväskylä
martin.hartmann@jyu.fi

Petri Toiviainen
University of Jyväskylä
petri.toiviainen@jyu.fi

ABSTRACT

This paper presents a new model for segmenting symbolic music data into phrases. It is based on the idea that melodic phrases tend to consist of notes, which increase rather than decrease in length towards the phrase end. Previous research implies that the timing of note events might be a stronger predictor of both theoretical and perceived segmentation than pitch information. Our approach therefore relies only on temporal information about note onsets. Phrase boundaries are predicted at those points in a melody where the difference between subsequent note-to-note intervals reaches minimal values. On its own, the proposed model is parameter-free, does not require adjustments to fit a particular dataset, and is not biased towards metrical music. We have tested the model on a set of 6226 songs and compared it with existing rule-based segmentation algorithms that had been previously identified as good performers: LBDM and Grouper. Next, we investigated two additional predictors: meter and the presence of pauses. Finally, we integrated all approaches into a meta-classifier, which yielded a significantly better performance than each of the individual models.

1. INTRODUCTION

Melodic segmentation refers to the subdivision of melodies into smaller meaningful groups. When hearing a piece of music, listeners - whether musically trained or not - will perceive stronger or weaker points of closure in the melody. Experts in music theory are able to identify these points by analyzing the score. Researchers in Music Information Retrieval (MIR) have attempted to automatically identify and reliably predict points in melody indicated as group borders by either listeners or music theorists. The quality of any proposed model is therefore established as the degree of agreement with human ratings. In MIR, segmentation has been mostly investigated at phrase level, rather than at lower levels in the structural hierarchy, such as motifs. What complicates the task is the fact that the task does not have a single "correct" solution. The perception of boundaries between phrases is subjective, especially where the border is weak or ambiguous. Musicians, as well as musically untrained listeners, have been shown to display differences (between groups and within groups) when indicating phrase boundaries (e.g. Deliège, 1987, Peretz, 1989, Thom et al., 2002; Bozkurt et al., 2014, Hartmann et al., 2017). In this sense, no model can predict human judgments infallibly.

The formal definition of the criteria used by listeners for melodic segmentation laid out by Lerdahl and Jackendoff (1983) in *A Generative Theory of Tonal Music* (GTTM) has been frequently used as a point of departure for the MIR segmentation task. GTTM derives from the rules defined by Gestalt psychology, such as proximity and similarity. The Grouping Preference Rules of Lerdahl and Jackendoff postulate that listeners cluster tones into groups on the basis of a set of rules, including temporal proximity (slur/rest - GPR 2a; inter-onset-interval - GPR 2b); degree of change in register (GPR 3a), dynamics (GPR 3b), articulation (GPR 3c) or length (GPR 3d); symmetry (GPR 5), and motivic similarity (GPR 6). A boundary is perceived in places where the temporal proximity, or the change in the individual properties mentioned above, is greater than that of the neighbouring transitions. The last two rules describe listeners' preference for group shapes that are symmetrical (GPR 5), and the tendency to place parallel shapes into parallel groups (GPR 6). While the authors of GTTM separated grouping from meter and treated them as independent entities, they emphasized that grouping and meter interact, and structures are perceived most clearly where they are in mutual accordance.

The validity of GPRs was largely supported in listening experiments; the effects of long inter-onset-intervals (IOI) and rests (GPR 2b and 2a) were typically found stronger than those of pitch changes (e. g. Deliège, 1987; Frankland and Cohen, 2004; Peretz, 1989). Of course, the durational proportions of events in music scores are not the same as in live music. Yet, the use of long notes at phrase ends is frequently further accented in performance practice, in that performers lengthen the last note of the phrase, and insert a micropause (Friberg et al. 1998); although Cambouropoulos (2001) suggests that it is not always the last note that is lengthened, but, in some cases, the penultimate note, resulting in the delay of the phrase-final note onset. Pauses are salient dividers between consecutive phrases, as they often precede phrase starts (Temperley, 2001). In vocal music at least, singers need a pause to breathe in, and this is usually reflected in the music score. Bruderer's (2008) results show that in Western popular music, IOIs, pauses, and timbre change contribute most to the perception of boundaries. Repetition and motivic similarity also provide the listener with cues about phrase boundaries. While the identification of pattern similarity is a non-trivial task, there are indications that metrical context plays a strong role in similarity perception: repetitions often begin at phrase starts (Temperley, 2001), and patterns sharing the same meter are more often rated as similar (Gabrielsson, 1973). Ahlbäck (2007) shows that listeners might not recognize even exact repetitions if the repeated fragments start at different points in the metrical hierarchy.

We introduce a model based on the assumption that inter-onset-intervals (IOI) within music phrases tend to get longer as the phrase progresses, and longest towards the phrase end. This pattern is best seen on simple recitative songs, such as Gregorian chant. Our aim is to create a parsimonious, general, parameter-free model which does not require additional changes to reflect the specifics of the music data. Previous research suggests that temporal information provides more salient grouping cues than pitch information; our model is therefore based solely on the timing of note events.

We are interested in comparing the predictive power of this approach with that offered by other kinds of temporal information, such as offset-to-onset intervals (rests) and metrical symmetry, and with existing rule-based, Gestalt-derived models.

2. COMPUTATIONAL SEGMENTATION

Several algorithms have been proposed to deal with segmentation of both symbolic and audio music data. Existing models for automatic segmentation of melodies in the symbolic form are largely limited to monophonic data and can be roughly divided into two groups. Rule-based models derive from expert knowledge and intuitions of music theorists, and are usually based on the principles of Gestalt psychology. The second group of models is driven by computational rather than musical knowledge, using supervised and unsupervised machine learning. In the first case, algorithms are trained on a portion of the data (or another music corpus), and the collected information serves to make predictions on the remaining data. With unsupervised learning, phrases are predicted based on statistical regularities of note-to-note relationships (for a recent review, see Rodríguez Lopez, 2016).

The model we propose belongs to the Gestalt tradition. Tenney and Polansky (1980) are commonly credited as authors of one of the earliest Gestalt-based models, but the ones that we will review here in more detail are LBDM and Grouper, as they have been consistently reported as best-performing when tested on larger datasets.

2.1 LBDM

The Local Boundary Detection Model (LBDM) introduced by Cambouropoulos (2001) operates on local changes in pitch, IOIs, and rests. The Change Rule places a phrase border between any two consecutive intervals that are not identical with respect to these three parameters. The border strength is proportional to the degree of change between the two intervals. The Proximity Rule assigns a higher border strength on the larger interval out of any two consecutive, non-identical intervals. The default settings for the relative weights of pitch, IOI, and rest intervals are set to .25, .5, and .25, respectively. LBDM results in a profile of border strengths, and uses thresholding for separating borders from non-borders. Studies that have tested LBDM on larger amounts of data include Thom et al. (2002), whose dataset comprised over 2600 songs from the Essen collection. The authors performed an optimization for their dataset by trying out different combinations from a predefined grid of parameter values, and with the best-performing setting obtained in this way reported a mean F-score of .50. Pearce et al. (2010) used a smaller Essen subset (1705 songs); their implementation using the default weights and a threshold of .05 yielded a mean F of .63.

2.2 Grouper

The Grouper Program proposed by Temperley (2001) relies only on temporal information (note-to-note intervals and meter) and analyses the score as a whole, performing all possible analyses and selecting the favourites using three criteria. The Gap Score is the sum of IOI and OOI (offset-to-onset interval) of two consecutive notes; phrases are assigned a bonus based on the Gap Score between the notes at the border. Secondly, phrases receive a logarithmic penalty for deviating from an optimal phrase length, set by default to 8. As a third step, Grouper penalizes phrases that are not metrically "in phase" with the preceding phrase, meaning that they do not start on the same beat on the highest and second-highest metrical levels. For example, a phrase starting on the first beat should be preferably followed by a phrase that also starts on the first beat. The list of beats is provided by the Meter Program, which calculates a metrical structure for a given melody using a division of timepoints into small units called pips. The experimental test of Grouper computed by Thom et al. (2002) gave a mean F-measure of .62; Pearce et al. (2010) obtained an average F-score of .66 with their data.

3. NEW MODEL: IOI DIFFERENCES

In this model, boundaries are selected by calculating differences between successive IOIs (ΔIOI), or, in other words, second-order differences between note onset times. OOIs are not considered; any note followed by a pause is treated as if it was lengthened by the duration of the pause. Boundary candidates are chosen only from negative IOI differences, that is, from transitions where a longer IOI is followed by a shorter IOI.

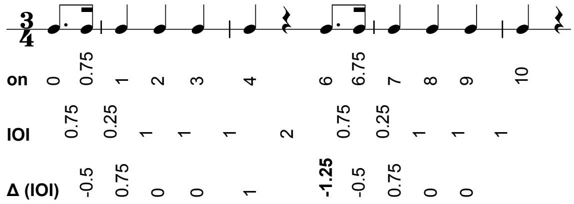

Figure 1. Rhythmic representation of "Happy Birthday": on = note onset (in beats), IOI = inter-onset-intervals (in beats), ΔIOI = difference in successive inter-onset-intervals (in beats).

In the example in Figure 1, there are three occurrences of a negative ΔIOI: we could theoretically place a boundary after the first dotted note, after the pause, and after the second dotted note.It is obvious that due to rhythmic variability within phrases, creating boundaries at all negative ΔIOIs would lead to oversegmentation. In this way, we would perhaps detect many true boundaries, but at the same time, too many of the identified boundaries would be false. Because we expect the largest negative IOIs to occur at phrase ends, we considered only the lowest value of ΔIOI in each song. In the example above, a boundary would be placed before the second dotted note where ΔIOI = -1.25, splitting the music sample into two equal groups. Any other ΔIOI equal to -1.25 occurring earlier or later in the same melody would also be chosen as a boundary.

4. METHOD

4.1 Dataset

The Essen Folksong Collection (Schaffrath, 1995) is a set of 6,236 mostly Germanic folksongs in symbolic format, with phrases annotated by music experts. It comprises simple diatonic melodies with a clear metrical structure. Because large collections of symbolic music data with annotated phrase borders are not readily available, the Essen corpus has been widely used in testing segmentation models. The obvious disadvantage of using the same dataset repeatedly is that findings cannot be generalized to other music styles and traditions. Its advantage, however, is that it presents a convenient basis against which different models can be tested and their performances compared.

4.2 Models and Implementation

For comparison of performance, we chose models based on Gestalt-like rules coming from musical knowledge. LBDM, Grouper, and a simplified version of GPR 2a (Pause). To account for symmetry, we added two models based on metrical structure: Temperley's Meter, and Meter Finder. To make outputs comparable, all models were set up to return a binary vector of ones (boundaries) and zeros (non-boundaries). We assumed there was always a boundary before the first note onset of every song. As a final step, we combined all models to make a compound model.

4.2.1 Previous Models

1. LBDM was implemented in the version offered by MIDI Toolbox (Eerola and Toiviainen, 2004). Instead of using default settings, LBDM was optimized for the Essen dataset, and computed with weights w1(pitch) = .10, w2(IOI) = .23, w3(rest) = .67, and a threshold of .20. Rather than choosing predefined values for these parameters, we used a Genetic Algorithm search heuristic (single-objective optimization with bound constraints) to find optimal parameter values. An objective function was used to calculate the F-measure of each song for a given set of weight and threshold values, and to finally obtain the negative of the mean F-measure across songs. The algorithm converged to an optimal solution, that is, a set of parameter values that would minimize the negative of the mean F-measure across songs. It is worth noting that the optimal combination of parameters assigned the lowest weight to the pitch rule, compared to the other two rules.

2. Grouper was computed using original code by Temperley (2001). While its author does not recommend it, it is possible to run Grouper without metrical information provided by the Meter Program. This variant uses just the first two of its three rules, that is, it bases the analysis only on IOIs, OOIs, and a preferred phrase length. We considered it instructive to include Grouper in both versions - with (**2a**) and without (**2b**) the beat list output from Meter - to see how much the metrical information adds to its performance.

3. Meter is a program which generates a beat list to be used by Grouper (Temperley, 2001). We decided to use it on its own to obtain a segmentation based on metrical regularity itself. Meter returns a hierarchy of beats on five different metrical levels, with 0 as the lowest and 4 as the highest level. Most commonly, the downbeat comes with the first note onset. Some songs, however, start with an anacrusis, and the downbeat comes with a later note onset. For the purposes of segmentation, we are interested in a regular distribution of strong beats on the highest metrical level, starting with the first note onset. The reason for this is that if a phrase starts on a particular beat, it is likely that the next phrase will start on the same-level beat, as seen in the example in Figure 1. Our implementation therefore shifts the beat list given by Meter cyclically, so that the first strong (level 4) beat occurs at the first onset, and phrase boundaries are then placed at every note with a level 4 onset.

4. Pause. To test for the importance of pauses in segmenting melodies, we have included a simple all-or-nothing rest rule reminiscent of GPR 2a. With this rule, a boundary is placed after any pause, irrespective of its length.

4.2.2 New Approaches

5. ΔIOI is the implementation of our segmentation model as described earlier (see 3 above).

6. Meter Finder is an algorithm proposed by Toiviainen and Eerola (2005). The authors used autocorrelation to classify melodies into "double meter" or "triple meter" categories, and reported a correct classification rate of over 90% for the Essen corpus. This straightforward classification approach lends itself well for comparison with Meter. Based on the Meter Finder categorization, a simple metrical grid has been constructed for every piece. Melodies in double meter had a boundary placed every 8 beats from the first note onset; melodies in triple meter, every 6 beats. If a predicted boundary did not coincide with a note onset, it was left out.

7. ΔIOI OR Meter Finder OR Pause. Based on preliminary results we decided to incorporate a model that uses evidence from three aspects of note events timing: IOIs, rests, and metrical information. This rule generates a phrase border at any timepoint where either of the three models, 5, 6, or 4, assumes a phrase break. Each of these models has a relatively low recall rate on its own; taking a disjunction of their prediction sets was expected to enhance recall.

8. Compound Model is based on a different approach from all the other models, in that it uses a machine learning component. We performed logistic regression to make a meta-classifier with all the models from the previous analysis to see if their combination would result in improved results. In addition to the binary outputs generated by models 1, 2a, 2b, 3, 4, 5, and 6, we also included probabilistic model versions where possible. *LBDM(prob)* is the profile of boundary strengths returned by LBDM before thresholding. The weights used are identical to the binary version. *Meter(prob)* is a modification of the original output of the Meter Program. As explained earlier, Meter returns a profile of beats with assigned metrical hierarchy values (0 to 4). We took the exponential of these values and shifted them cyclically in the same way as in the binary version, so that the first note onset of any song had the highest hierarchical level (4). *Meter Finder(prob)* returns a metrical hierarchy profile on three levels. For songs in double meter, the boundary on every 16th beat is the strongest, with a weight of 1. A weaker boundary (w = .66) coincides with every eighth beat, the weakest (w = .33) with every fourth beat. Melodies in triple meter get a similar hierarchy profile, with decreasing weights placed on every 12th, 6th, and 3rd beat. Boundaries are only considered (and weighted) if they coincide with a note onset. Δ *IOI(prob)* calculates all ΔIOI values of the piece and replaces the positive ΔIOI values with zeros. Our model selection approach consisted of an exhaustive search for the best fit out of each possible combination of the 14 models (7 binary, 4 probabilistic, 3 interactions) based on its log likelihood and then, using the Akaike and Bayesian information criterion (AIC and BIC) estimators, penalizing its log likelihood according to the number of predictor variables of the model. The lowest AIC was obtained combining all models except the binary version of ΔIOI. The combination with the lowest BIC excluded the binary versions of LBDM and ΔIOI. We report the results of the latter model with the lowest BIC, as it uses less elements. Possible overfitting was investigated by computing 10 times 10-fold cross-validation on the compound model. The mean F-measure across 10 folds and 10 iterations was very similar to the one reported (mean F = .76; min = .75, max = .78). Note that this cross-validation analysis, as well as the AIC/BIC calculations, was computed for concatenated data from all songs.

5. RESULTS

5.1 Preliminary phrase analysis

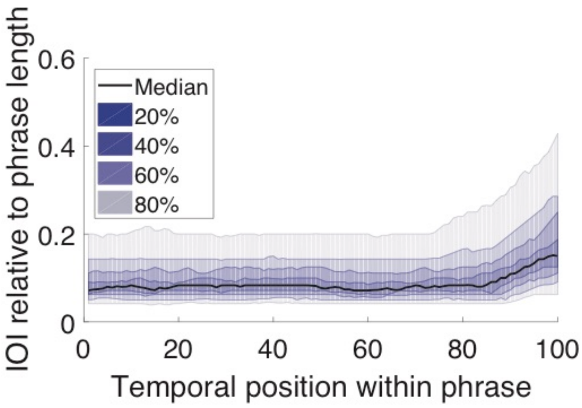

Figure 2. IOI relative to phrase length for all phrases of the dataset, interpolated by 100. The dark line denotes the median (percentile 50). The various degrees of shading (darker near the centre of the range, fainter towards the edge of the range) depict the dispersion of the middle 20%, 40%, 60%, and 80% of the observed values.

As a first step, we performed an analysis of all phrases of the Essen collection to see if the assumption that notes tend to lengthen towards the phrase endings has any merit. The number of note-onset events in the annotated phrases range from 2 to 49, with a median of 8, suggesting that the songs have sometimes been segmented on different hierarchical levels. To estimate changes in IOI for different temporal positions within phrases, we first computed IOI vectors for all individual phrases. Each vector included the last IOI in the phrase: to do this, we added an extra note to the end of each phrase. This added note corresponded either to the first note onset of the following phrase or, if it was the final phrase of the song, to a phantom note at the end of the song. All IOI vectors obtained in this way were divided by the number of note onsets within the phrase, excluding the added note onset at the end of the phrase. Subsequently, to compare the IOI vectors across all phrases from all songs, each vector of within-phrase IOIs was interpolated to a length of 100 points. As a result, all vectors were scaled to have the same length, regardless of their absolute phrase length (Figure 2). The IOIs displayed an ascending tendency at the ends of phrases.

5.2 Model comparison

The results are summarized in Table 1, with the highest values obtained for each of the three measures marked in bold. Paired two-sample sign tests between the F-scores obtained for each song of the 7 binary models (LBDM, Grouper with Meter, Grouper without Meter, Meter, Pause, ΔIOI, and Meter Finder) show that out of the 21 possible model pairs, there are three cases where the differences between models do not reach significance at an alpha of .001: LBDM - Meter Finder, Meter - Pause, and Grouper - Meter Finder. The Compound performed significantly better than any other model, based on the sign test results.

Model	Precision	Recall	F
Grouper with Meter	.77	.73	.74
Grouper without Meter	.68	.66	.66
LBDM	.81	.60	.65
Meter	.59	.70	.61
Pause (GPR 2a)	**.98**	.48	.60
Compound (BIC)	.92	.68	**.75**
ΔIOI OR Meter Finder OR Pause	.64	**.81**	.68
Meter Finder	.70	.64	.64
ΔIOI	.79	.54	.58

Table 1. Mean Precision, Recall, and F-scores of existing models (top) and new approaches (bottom). Both groups are sorted in order of their F-scores.

6. DISCUSSION

There are several points that we would like to highlight in the discussion.

1. The approach based on minimal IOI differences does not on its own reach the performance of LBDM and Grouper. Its lower predictive power is partially compensated by its simplicity and versatility. The finding that the optimal LBDM setting assigns a minimal weight to the Pitch rule supports the importance of timing over pitch in melodic segmentation. Also, LBDM was outperformed by Grouper, which does not use pitch information.

2. It is interesting to compare our results for LBDM and Grouper with other studies. LBDM with optimized parameters found by the Genetic Algorithm generated better results than those reported by Pearce et al. (2010) on their set of 1705 songs, and the improvement was even more pronounced compared to the results computed by Thom et al. (2002) on a set of over 2600 songs. Grouper, implemented in the original version as designed by its author, obtained an F of .74 on our dataset, which is considerably higher than the .62 and .66 reported in the abovementioned studies.

3. As noted before e.g. by Pearce et al. (2010), pauses constitute highly reliable indicators of phrase borders in the Essen corpus. The Pause rule obtained the highest precision score, implying that almost all rests in the corpus occur between rather than within phrases. As discussed earlier, pauses present an important cognitive divider for music listeners, but there is some debate as to their relevance in symbolic data. It has been argued that with folk tunes, pauses marked in music scores are sometimes only a convention used by transcribers to visually mark segment borders or "breathing points" which the singer may and may not use in their actual performance (Rodríguez Lopez, 2016). This question requires a deeper investigation of the congruence of folk song performances and their transcriptions. One possibility is to use the method employed by Bruderer (2008) with Western popular songs, a style in which scores are also created by transcribing performances. Bruderer manually time-alligned MIDI data with recordings; this approach leads to a more realistic time representation of both pauses and note onsets.

4. For the songs of the Essen collection, metrical symmetry was a relevant predictor of their phrase layout. Meter Finder outperformed Meter, which is remarkable considering that it predicts phrase boundaries at regular intervals starting with the first onset, without any further information about how the melody progresses. At the same time, Meter Finder presumably misses a portion of true boundaries in cases where the prediction does not fall on a note onset. Modifying the model to include imperfectly aligned note onsets resulted in improved recall, but lower precision, with similar F-scores.

5. The model integrating three temporal predictors - minimum IOI differences, pauses, and meter - performed better than all the other individual models, except Grouper with Meter, and yielded the highest recall score. This is an encouraging result, warranting further investigation. Instead of a union (OR) operation, the three temporal criteria could perhaps be combined in a different manner, leading to further performance enhancement.

6. The Compound ensemble classifier generated the best mean F-score, and performed significantly better than each of the other models. Compared to the second-best Grouper with Meter, its slightly lower recall is balanced out by a considerable improvement in precision.

7. Our main focus was on state-of-the-art techniques for rule-based boundary detection; we also tried to address the problem of dimensionality by finding an optimal model based upon AIC/BIC values. However, the selected model is still relatively high-dimensional and thus difficult to interpret. In the future, we will utilize feature selection and dimensionality reduction techniques to assess the relative contribution of predictor variables to the model and reduce the risk of multicollinearity.

7. CONCLUSIONS

We conclude that basing melodic segmentation on temporal information alone is a valid approach, and that different elements of timing - the preference for negative IOIs, the presence of rests, and metrical symmetry, converge to an optimal task solution. At this time, our findings are limited to European ethnomusicological material. Studies investigating non-Western music, such as Bozkurt et al. (2014), suggest that models performing well on European folk music do not necessarily give good segmentations of melodies coming from different traditions. The need for corroboration on more diverse music samples applies to computational models of melodic segmentation in general.

Acknowledgments
This work was supported by the Research Grant Agency of the Slovak Ministry of Education (VEGA 1/0637/15) and the Academy of Finland (projects 272250 and 274037).

8. REFERENCES

[1] S. Ahlbäck, "Melodic similarity as a determinant of melody structure", in Musicae Scientiae 11/1, 2007, pp. 235-280.

[2] M. Bruderer, "Perception and Modelling of Segment Boundaries in Popular Music", PhD. Thesis, Technische Universiteit Eindhoven, 2008.

[3] B. Bozkurt, M. K. Karaosmanoğlu, B. Karaçalı, and E. Ünal, "Usul and Makam driven automatic melodic segmentation for Turkish music", in Journal of New Music Research, 43/4, 2014, pp. 375-389.

[4] E. Cambouropoulos, "The Local Boundary Detection Model (LBDM) and its Application in the Study of Expressive Timing," in Proc. Int. Computer Music Conference (ICMA 2001), San Francisco, 2001, pp.17-22.

[5] E. Cambouropoulos, "Musical parallelism and melodic segmentation: A computational approach," in Music Perception, 23/3, 2006, pp. 249-268.

[6] T. Eerola and P. Toiviainen, "The MIDI Toolbox: MATLAB tools for music research", 2004, available at www.jyu.fi/musica/miditoolbox.

[7] I. Deliège, "Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules", in Music Perception, 44/2, 1987, pp. 325-360.

[8] B. Frankland and A. Cohen, "Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music," in Music Perception, 21/4, 2004, pp. 499-543.

[9] A. Friberg et al., "Musical punctuation on the microlevel: Automatic identification and performance of small melodic units", in Journal of New Music Research, 27/3, 1998, pp. 271-292.

[10] A. Gabrielsson, "Similarity ratings and dimension analyses of auditory rhythm patterns II.", in Scandinavian Journal of Psychology 14/1, 1973, pp. 161-176.

[11] M. Hartmann, O. Lartillot, and P. Toiviainen, "Interaction features for prediction of perceptual segmentation: Effects of musicianship and experimental task", in Journal of New Music Research, 46/2, 2017, pp. 156-174.

[12] F. Lerdahl and R. Jackendoff, A Generative Theory of Tonal Music. MIT Press, 1983.

[13] M. Pearce, D. Müllensiefen, and G. Wiggins, "Melodic grouping in music information retrieval: New methods and applications," in Advances in MIR, pp. 364-388. Springer, 2010.

[14] I. Peretz, "Clustering in music: An appraisal of task factors", in International Journal of Psychology, 24, 1989, pp. 157-178.

[15] M. Rodríguez Lopez, "Automatic Melody Segmentation" (PhD. thesis). Utrecht University, 2016.

[16] H. Schaffrath, "The Essen Folksong Collection" (computer database), in D. Huron (Ed.). Menlo Park, CA: CCARH, 1995.

[17] D. Temperley, The Cognition of Basic Musical Structures. MIT Press, 2001. Code available at http://www.link.cs.cmu.edu/melisma/

[18] J. Tenney and L. Polansky, "Temporal Gestalt perception in music," in Contemporary Music Review, 24/2, 1980, pp. 205-241.

[19] B. Thom, Ch. Spevak, and K. Hötker, "Melodic segmentation: evaluating the performance of algorithms and musical experts", in Proc. Int. Computer Music Conference (ICMA 2002), San Francisco, 2002, pp. 65-72.

[20] P. Toiviainen and T. Eerola, "Autocorrelation in meter induction: The role of accent structure," in J. Acoust. Soc. Am., 119/2, 2005, pp. 1164–1170.