# Symbolic Segmentation: A Corpus-Based Analysis of Melodic Phrases

Marcelo Rodríguez-López and Anja Volk

Department of Information and Computing Sciences,
Utrecht University, Utrecht, NL.
{m.e.rodriguezlopez,a.volk}@uu.nl

**Abstract.** Gestalt-based segmentation models constitute the current state of the art in automatic segmentation of melodies. These models commonly assume that segment boundary perception is mainly triggered by local discontinuities, i.e. by abrupt changes in pitch and/or duration of neighbouring note events. This paper presents a statistical study of melodic phrase segments to test this assumption. The study focuses on the analysis of pitch and duration in the neighbourhood of annotated phrase boundaries. Our analysis reveals stability in duration information (specifically inter-onset-intervals), and variability in pitch information (specifically chromatic intervals). We conclude that pitch discontinuities, while important for segmentation, can not be addressed as a local, idiom independent phenomenon.

## 1 Introduction

Music segmentation is a basic problem in research fields concerned with automated music description and processing. Segmenting musical input is concerned with modelling the formation of temporal units holding musical content. Historically, the tasks of music segmentation that have received more attention within music content research fall into three groups: (*a*) the segmentation of musical audio into notes, as part of transcription systems [1], (*b*) the segmentation of symbolic encodings of music into phrases [2, 3], and (*c*) the segmentation of audio/symbolic music files into sections [4]. In this paper we focus on the study on units of the second kind, i.e. those resembling the musicological concept of the *phrase*. Given that phrase-level segmentation deals mainly with monophonic music, this area is commonly referred to as *melodic segmentation*. In this paper we address the problem of melodic segmentation from the perspective of Music Information Retrieval (MIR), as part of the MUSIVA project [5]. In the project we pay special attention to the role that segments play in assessing the variation and similarity between melodies.

In MIR the task of melodic segmentation consist in identifying cognitively plausible segment boundaries. Melodic segmentation has commonly been modelled from three perspectives. The first assumes that boundary perception is mostly related to self-organizing principles in the brain, so that the process is mainly idiom independent and thus the information needed can be found in the immediate neighbourhood of the boundary [6, 7]. The second assumes that the main cue for boundary detection is related to melodic self-similarity [8], supporting the view of boundary perception as idiom independent, yet requiring

access to information over the large sections of the melody. The last perspective defends the view of exposure [9, 10], which assumes that idiom-related factors, such as tonal/melodic/form-level prototypical patterns are important cues for segmentation.

Recent comparative studies [2, 3, 11, 12] report results of only modest success (F-scores [13] peaking at $0.60 - 0.66$ when evaluated in large melodic corpora). Two models that have consistently ranked higher in these studies are LBDM [6] and Grouper [7]. Both models are based on heuristics inspired by Gestalt principles, i.e. idiom independent rules based on local information. Gestalt-based models commonly search for discontinuities in pitch and rhythmic parametrisations of melodies. In respect to rhythm this usually corresponds to relatively long durations or extended silences, and in respect to pitch to relatively large intervals. In this paper we report on a statistical analysis of local pitch and duration information surrounding annotated segment boundaries. We aim to provide empirical evidence that supports/refutes the assumptions of locality and universality of discontinuity made by Gestalt-based segmentation models.

The remainder of this paper is organized as follows. In §2 we summarize work in corpus analysis of melodies and melodic phrase structure. In §3 we present our statistical analysis of phrase boundary neighbourhoods and discuss our results. Finally, in §4 we draw conclusions and outline future work.

## 2 Previous Work in Corpus Analysis of Melodic Phrases

In general, statistical analyses of melodic corpora with annotated phrases have focused on characterising within-phrase regularities, such as prototypical contours [14] and pitch-interval-size-shrinking [15]. Studies have also assessed the effect of within-phrase information in other musical processes, such as pulse induction [16] and melodic similarity [17].

Conversely, studies focusing on the immediate vicinity of annotated boundaries or considering successive phrases are scarcer. In [18] Brown et al. examined if the conditions of closure proposed by Narmour [19] could be observed in phrase and score endings of the Essen Collection (EFSC). They found strong evidence for the occurrence of 2 (out of 6) conditions at phrase/score ends when compared to the total population of notes, namely for durational expansion and tonal resolution. Similarly, in [20] Bod analyses the EFSC looking for phrase joints that challenge the assumptions of Gestalt-based segmentation models. Bod observed a class of phrase joints that he labelled "jump-phrases", which referred to phrases that contained a pitch interval jump at the beginning/end of a phrase (or in both) rather than at the joint. Bod reports that more than 32% of the subset of the EFSC used in the study (1000 songs) contained at least one jump-phrase.

## 3 Corpus Analysis of Annotated Phrases

**Aims and Contribution:** In this paper we take as our working hypothesis that pitch and duration discontinuities (defined here as "jumps" in respect to a local context) can be considered relatively strong, universal cues for segment boundary perception. We aim to test this hypothesis based on empirical evidence

provided by corpus statistics. To this end, we carry out a statistical analysis of pitch and duration *interval sizes* measured at phrase joints, and compare these interval sizes to interval sizes found in a local context around the boundary. The characterization of interval sizes resulting from this analysis contributes both to the deepening of the understanding of boundary perception, and consequently to the development of more robust models of melodic segmentation.
**Scope:** The analysis presented here considers only pitch and duration information, and focuses on vocal folk melodies sampled from the EFSC. Our choice of parametric representation of notes is taken so that it agrees with the information used by most Gestalt-based segmentation models. Likewise, our choice of melodic corpus is motivated by its widespread use in the testing of segmentation models, as it can be considered the de facto benchmark.

EFSC melodies comprise of mainly vocal music. Hence, the influence of instrumental tessitura in the distributional properties of interval sizes cannot be directly attested. Moreover, EFSC melodies are limited to a single genre, and despite the collection's large size, the cultural traditions included are not equally-well represented. Thus, if the entire corpus is used, a characterization of interval size in respect to stylistic traits is also difficult to measure. For these reasons, we focus on the two most distinct and well represented cultural traditions within the corpus, namely German and Chinese. In this study we assume that these two traditions are distinct enough so that the evidence of regularities (or lack thereof) can be used as an indicator (albeit modest) to assess the universality and locality of discontinuity-related boundary cues.

### 3.1 About the Essen Folk Song Collection (EFSC)

The Essen Folk Song Collection contains over 20,000 songs, of which 6,251 are currently publicly available. The EFSC data was compiled and encoded from notated sources by a team of ethnomusicologists and folklorists lead by Helmuth Schaffrath. The songs are available both in EsAC and **kern formats, and include information on pitch, duration, meter, barlines, rests, and phrase markings. Text accompanying the songs is not available in the encodings.

### 3.2 Parametric Representations Considered

The most common melodic parametrisation used in segmentation models consists of pitch and duration intervals. In this paper we measure pitch intervals (`PI`s) in semitones, and measure duration using inter-onset-intervals (`IOI`s) in seconds.

For the analysis of pitch and duration intervals, first `PI`s and `IOI`s of different melodies have to be made comparable. `PI`s are a relative, transposition-invariant measure, so no further processing is needed. `IOI`s, on the other hand, are an absolute measure. Thus, to achieve invariance to tempo, we simply normalized the durations of each `IOI` collected for a local context by the total duration of the context. We abbreviate normalized `IOI`s as `NIOI`s.

### 3.3 Procedure

The subsets of the EFSC used for our analysis were obtained processing the **kern encodings with a combination of Python and Matlab scripting. We used

the entire '`Deutschl`' and '`China`' sets for our experiments. Following [15], we filtered out songs which contained rests, and also excluded songs with just one phrase. The filtering resulted in 1,268 German folk songs, ranging in length from 13 to 184 notes. In total, this accounted for 5289 phrase-pairs, with phrases ranging from 2 to 21 notes. Similarly, we obtained 1416 Chinese songs, ranging from 17 to 270 notes. From this subset 4624 phrase-pairs were extracted, with phrases ranging from 2 to 28 notes.

We denote the extracted sequential phrase-pairs as $ph_{a,b}$ (with $a, b$ denoting the left- and right- most phrases of the phrase pair). We collected statistics for intervals occurring within a local context established around the boundary of $ph_{a,b}$. The local context for analysis spans an interval of $[-2, 2]$, i.e. two notes to the left and two notes to the right of the boundary bisecting $ph_{a,b}$[1]. Here we consider a joint as the last note belonging to $ph_a$ and the first note belonging to $ph_b$, and denote the interval computed at the joint as $j(ph_{a,b})$. Similarly, we denote the contextual intervals surrounding the joint as $c(ph_a)$ and $c(ph_b)$, when referring to them individually, and $c(ph_{a,b})$ when referring to both.

### 3.4 Results

From our working hypothesis we can derive three main assumptions used by Gestalt models: `1.` discontinuities in pitch and/or duration are strong cues for segmentation, `2.` discontinuities can be modelled as a local phenomenon, `3.` discontinuity cues are universal. In the following we describe three experiments, using simple descriptive statistics and statistical hypothesis testing, to investigate the validity of these assumptions.

**Experiment 1** First we sought to assess the level of strength of pitch and duration discontinuities, by testing the intuition that a higher strength should correlate with a strong presence of discontinuities at phrase joints. To this end we simply computed the proportion of phrase-pairs having a local discontinuity in respect to all phrase-pairs of each (German and Chinese) subset. We also computed the number of discontinuities that can be observed as a proportion of the phrase-pairs of each song. Table 3.4 lists the observed proportions.

In Table 3.4 we can observe a fairly weak presence of pitch discontinuities, both when compared to the total amount of phrase-pairs of each subset and to individual songs. Duration discontinuities, on the other hand, have a higher presence, especially for the Chinese subset ($\approx 73\%$ of all phrase-pairs, $\approx 58\%$ of songs have duration discontinuities in all boundaries of each song, and only $\approx 9\%$ of songs have no presence of duration discontinuities in the boundaries of each song).

We can also observe a general tendency of pitch and duration discontinuity to have higher presence in the Chinese subset (the relative increase in respect to the German subset varying between $\approx 12 - 18\%$ for pitch and $\approx 13 - 34\%$ for duration).

[1] We take a context size of $[-2, 2]$ as it commonly constitutes the upper limit for context sizes in comparative studies of melodic segmentation models [2, 3, 11, 12] (beyond this value the performance of Gestalt based models either drops or does not seems to result in significant improvements).

**Table 1.** Proportion of discontinuities observed in German and Chinese subsets. On the upper section a peak $p$ refers to the case: $j(ph_{a,b}) > c(ph_{a,b})$ for every $ph_{a,b}$ *in each subset*. On the lower section a peak $p$ refers to the case: $j(ph_{a,b}) > c(ph_{a,b})$ for every $ph_{a,b}$ *in a song*, and $np$ refers to the case: $j(ph_{a,b}) < c(ph_{a,b})$ for every $ph_{a,b}$ *in a song*.

| pitch | duration | Germany | China |
|---|---|---|---|
| $p$ | - | 32.41 % | 43.73% |
| - | $p$ | 49.57 % | 72.56% |
| $p$ | - | 04.51 % | 22.53% |
| - | $p$ | 24.13 % | 57.70% |
| $np$ | - | 31.49 % | 28.53% |
| - | $np$ | 23.26 % | 09.04% |

**Experiment 2** In our second experiment, we once more assessed the level of strength of pitch and duration discontinuities, this time by testing the intuition that higher strength should correlate with low overlap between the interval distributions of $j(ph_{a,b})$ and $c(ph_{a,b})$. To this end we give a visual depiction of the quartile spread of pitch and duration intervals in Fig. 1.

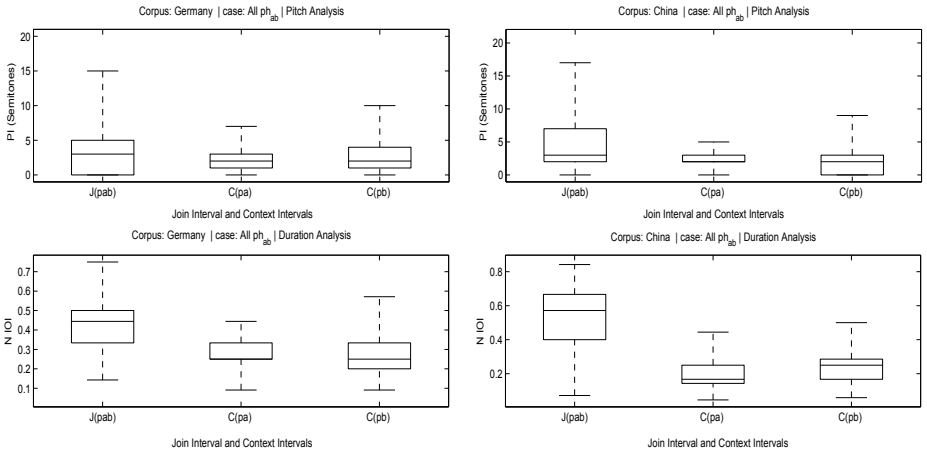


**Fig. 1.** Box plots of pitch and duration intervals of German and Chinese folk songs.

Fig. 1 shows that pitch intervals surrounding and at joints present overlap in both subsets. The overlap is visibly higher for the German subset, where not only box sizes and medians are close, but also the fact that $j(ph_{a,b})$ and $c(ph_b)$ present inverse skewness patterns suggests that phrase joints tend to have intervals of comparable or smaller size than their surrounding context. On the contrary, for the case of duration we can observe a clear dominance of larger intervals at the joint, which results in marginally overlapping inter-quartile ranges (depicted

by the boxes) for German songs, and completely non-overlapping inter-quartile ranges for Chinese songs.

**Experiment 3** In this experiment we examined if variation on the statistical behaviour of interval sizes could be attributed to non-local factors, and also investigated discontinuity strength from a third perspective. To test a possible influence of non-local factors, we examined the influence of phrase-size by grouping all intervals according to the sizes of phrase-pairs $ph_{a,b}$. We investigated discontinuity strength by testing statistical significance of central tendencies, i.e. a t-test comparison of the average size of all $j(ph_{a,b})$ per size-group, and the average size of all $c(ph_{a,b})$ per size-group. Our results are presented in Fig. 2.

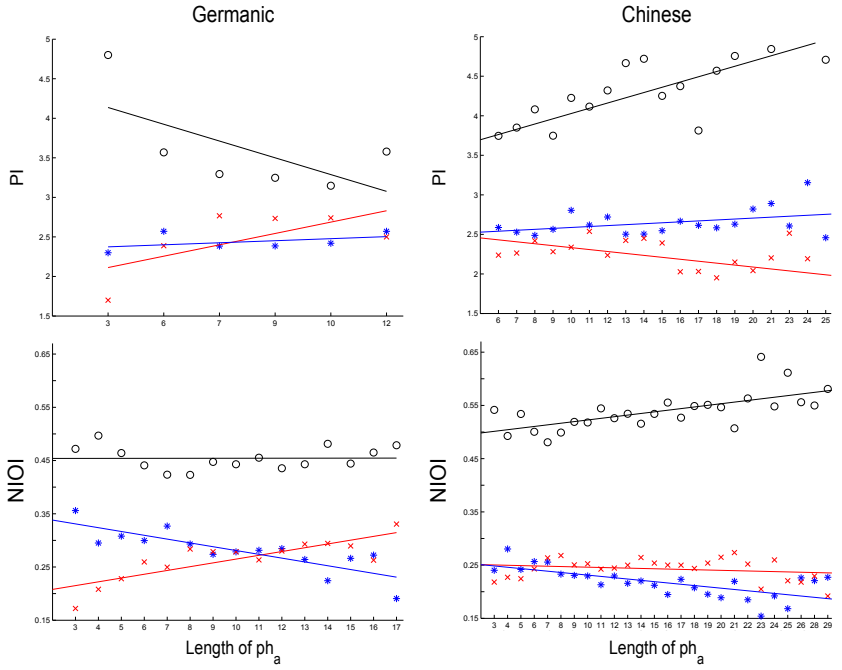


**Fig. 2.** Averages of $j(ph_{a,b})$ (circles), $c(ph_a)$ (crosses) and $c(ph_b)$ (asterisks), when grouped according to $ph_{a,b}$ size. Solid lines indicate corresponding regression slopes. We carried out a pairwise t-test between means at joints $j(ph_{a,b})$ and means at $c(ph_{a,b})$. The x-axis shows only size groups with statistically significant differences (at the 5% level). Total # of $ph_{a,b}$ size groups are: Chinese=26, German=19.

In the x-axis of Fig. 2 we present only size groups for which the differences between joints and contextual intervals are statistically significant[2]. In respect to

[2] To avoid a bias of subset size and phrase-group size on the statistical significance testing, we created equal size groups of both subsets and of all phrase-size-groups using random sampling.

pitch intervals, we can observe a higher number phrase-size-groups of statistically significant differences of Chinese intervals when compared to German intervals (19 out of 26 in the first case, and only 6 out of 19 in the second). This once again points to a higher degree of strength of pitch discontinuity in the Chinese subset. In pitch intervals we can also observe a clear influence of phrase length in the size of the interval: a progressive increase in the separation between means of $j(ph_{a,b})$ and $c(ph_{a,b})$ for Chinese songs, and, on the contrary, a decrease between means in German songs. The last observation suggests once again than in German songs local pitch discontinuities are relatively weak predictors of phrase boundaries. In respect to duration, the high number of statistically significant differences and the steady (German) and divergent (Chinese) tendencies shown by the regression lines support the findings of our previous experiments. That is, local leaps in duration make for a relatively strong predictor of phrase boundaries. This observation agrees with comparative studies [2, 3, 11, 21]. We can not, however, generalize this finding to all music, as our study only investigated two musical traditions, and may contain biases introduced by the boundary annotation process used in the EFSC.

## 4 Conclusion

In this paper we have presented a statistical 'corpus-based' study of melodic phrases. Our analyses reveal that IOI jumps remain stable across cultural origin and phrase-sizes, and that, under the same conditions, PI jumps show substantial variability. We conclude that pitch discontinuity, while important for segmentation, cannot be accurately modelled assuming locality and universality.

Our findings suggest that, to improve upon the state of the art of automatic segmentation, computational models making use of pitch information can not longer be "blind" to the type of music they have to process. That is, the best way of assessing how a pitch-related local cue should be treated would require previous analysis of the input melody, to look for the presence of stylistic trademarks. We suggest that this can be accomplished by developing graphical models that allow a previous classification of the input melodies, and an inferential engine that decides when and how to use the information available for boundary cue detection. Producing annotated corpora is then an urgent need if we are to study, train, and evaluate models that are aware of the music they are processing.

In future work we will study how the stability of pitch- and rhythm- related boundary cues vary in respect to factors such as musical style and instrumental tessitura. We will also attempt to characterise the performance of segmentation algorithms according to corpora of different characteristics (tradition, and again, style and instrumentation). To this end, we are currently working on the production of Jazz and Rock melodic corpora with phrase annotations.

**Acknowledgments.** We thank F. Wiering and the anonymous reviewers for the useful comments on earlier drafts of this document. M.E. Rodríguez-López and A. Volk are supported by the Netherlands Organization for Scientific Research, NWO-VIDI grant 276-35-001.

## References


1. Benetos, E., Dixon, S.: Polyphonic music transcription using note onset and offset detection. In: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, IEEE (2011) 37–40
2. Pearce, M., Müllensiefen, D., Wiggins, G.: Melodic grouping in music information retrieval: New methods and applications. Advances in music information retrieval (2010) 364–388
3. Pearce, M., Müllensiefen, D., Wiggins, G.: The role of expectation and probabilistic learning in auditory boundary perception: a model comparison. Perception **39**(10) (2010) 1365
4. Paulus, J., Müller, M., Klapuri, A.: State of the art report: Audio-based music structure analysis. In: Proceedings of the 11th international society for music information retrieval conference. (2010) 625–36
5. Volk, A., de Haas, W.B., van Kranenburg, P.: Towards modelling variation in music as foundation for similarity. In: Proc. ICMPC. (2012) 1085–1094
6. Cambouropoulos, E.: The local boundary detection model (lbdm) and its application in the study of expressive timing. In: Proceedings of the International Computer Music Conference (ICMC01). (2001) 232–235
7. Temperley, D.: The cognition of basic musical structures. MIT press (2004)
8. Cambouropoulos, E.: Musical parallelism and melodic segmentation. Music Perception **23**(3) (2006) 249–268
9. Pearce, M.T., Wiggins, G.A.: The information dynamics of melodic boundary detection. In: Proceedings of the Ninth International Conference on Music Perception and Cognition. (2006) 860–865
10. Abdallah, S., Plumbley, M.: Information dynamics: patterns of expectation and surprise in the perception of music. Connection Science **21**(2-3) (2009) 89–117
11. Wiering, F., de Nooijer, J., Volk, A., Tabachneck-Schijf, H.: Cognition-based segmentation for music information retrieval systems. Journal of New Music Research **38**(2) (2009) 139–154
12. Thom, B., Spevak, C., Höthker, K.: Melodic segmentation: Evaluating the performance of algorithms and musical experts. In: Proceedings of the International Computer Music Conference (ICMC). (2002) 65–72
13. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. Volume 999. MIT Press (1999)
14. Huron, D.: The melodic arch in western folksongs. Computing in Musicology **10** (1996) 3–23
15. Shanahan, D., Huron, D.: Interval size and phrase position: A comparison between german and chinese folksongs. (2011)
16. Nettheim, N.: The pulse in german folksong: A statistical investigation. Musikometrika **20**(1) (1997) 94–106
17. Eerola, T., Bregman, M.: Melodic and contextual similarity of folk song phrases. Musicae Scientiae **11**(1 suppl) (2007) 211–233
18. Brown, A.R., Gifford, T., Davidson, R.: Tracking levels of closure in melodies. In: Proc. Of the 12th International Conference on Music Perception and Cognition. (2012) 149–152
19. Narmur, E.: The Analysis and Cognition of Basic Melodic Structures: The Implication–realisation Model. University of Chicago Press (1990)
20. Bod, R.: Probabilistic grammars for music. In: Belgian-Dutch Conference on Artificial Intelligence (BNAIC). (2001)
21. Lartillot, O., Mungan, E.: A more informative segmentation model, empirically compared with state of the art on traditional turkish music. In: Proceedings of the Third International Workshop on Folk Music Analysis (FMA2013). (2013) 63