

# Melodic Segmentation: Structure, Cognition, Algorithms

Zuzana Cenkerová / [zuzana.cenkerova@savba.sk](mailto:zuzana.cenkerova@savba.sk)

Institute of Musicology, Slovak Academy of Sciences, Bratislava, SK

## Abstract

Segmentation of melodies into smaller units (phrases, themes, motifs, etc.) is an important process in both music analysis and music cognition. Also, segmentation is a necessary pre-processing step for various tasks in music information retrieval. Several algorithms for automatic segmentation have been proposed, based on different music-theoretical backgrounds and computing approaches. Rule-based models operate on a given set of logical conditions. Learning-based models, originating in linguistics, compute segmentation criteria on the basis of statistical parameters of a training corpus and/or of the given composition. The aim of this preliminary study is to propose and describe a new segmentation algorithm that is rule-based, parsimonious, and unambiguous.

## Keywords

segmentation, melody, phrase, motif, music information retrieval

## Introduction

The term *melodic segmentation*, also called *grouping*, refers to the subdivision of melodies into musically meaningful smaller units<sup>1</sup>. In musical terms, we might talk about motifs, figures, themes, periodes, phrases, etc. In this sense, melodic segmentation – the task of breaking down the melody into smaller sections on multiple hierarchical levels, identifying the ones that contribute most to the composition’s identity, and explaining how the composer uses them to create a musical structure, has traditionally been performed by expert musicologists in the field of music theory and analysis. However, segmentation as a cognitive process is also routinely performed by the minds of musically untrained listeners. The importance of segmentation mechanisms for the processing and memorizing of musical structures has, more recently, been recognized and researched in the field of music cognition.

The aim of this paper is to introduce the task of automatic melodic segmentation, which falls into the area of music information retrieval (MIR). MIR is a small but growing interdisciplinary field within musicology, which emerged following a formalization trend in music theory, as well as advances in technology, and draws inspiration from developments in linguistics. After providing a theoretical background, we will discuss existing algorithms for automatic music segmentation. Finally, a new segmentation algorithm will be outlined. This being a preliminary study, the presentation is limited to our theoretical motivations, main hypotheses and future possibilities for the algorithm’s use.

## Segmentation in music theory and analysis

In music theory and analysis, a piece of music may be segmented in different ways, depending on the theoretical approach, and the hierarchical level chosen (e.g., phrases can be further divided into subphrases, motifs, etc.). However, there is no single, unambiguous definition of terms like motif, theme, period, or phrase. Some segment types are usually related to specific musical forms or styles (e.g. *countersubject* in the fugue, *riff* in popular music styles). But even with segments used relatively consistently across styles, the interpretation of borders between successive segments is subjective. The same is true for segment nomenclature; in individual cases, the dividing line between, say, motif and phrase may be thin.

A *motif* is commonly understood as the shortest subdivision of a theme or phrase that still maintains its identity as an idea.<sup>2</sup> However, the importance of its rhythmic, melodic and harmonic constituents is a matter of some debate. Riemann, for example,

---

1 PEARCE, Marcus – MÜLLENSIEFEN, Daniel – WIGGINS, Geraint. Melodic grouping in music information retrieval: New methods and applications. In: *Advances in music information retrieval*. Springer, Berlin Heidelberg, 2010, pp. 364–388.

2 DRABKIN, William. Motif. *Grove Music Online*. Oxford Music Online. Oxford University Press. Web. 20. 12. 2016. <<http://www.oxfordmusiconline.com/subscriber/article/grove/music/19221>>.

emphasizes the aspect of meter and rhythm,<sup>3</sup> while Schenker emphasizes the intervallic aspect, and sees rhythm and contour as secondary.<sup>4</sup> Nevertheless, most authors agree that an important characteristic of a motif is that it occurs repeatedly within a piece of music; self-reference in terms of structural repetition and similarity is a factor that plays a crucial role in making music intelligible.<sup>5</sup>

Most automatic segmentation models operate on the level of *phrases*. As a theoretical unit, the phrase generally falls between motif and period. The span of a musical phrase is as contestable as that of its linguistic counterpart: in common practice, it frequently spans 4 measures, but may be shorter or longer.<sup>6</sup> *Phrasing*, a related term, refers to the manner in which the performer expressively interprets both individual phrases and their combination in the piece, and is indicative of the performer's individual style.<sup>7</sup>

Some cognition-oriented music theories, notably Lerdahl and Jackendoff's *Generative Theory of Tonal Music*<sup>8</sup> and Narmour's *Implication/Realization Theory*<sup>9</sup>, have defined perceptual grouping rules, which in turn have become the basis of some segmentation algorithms. *Generative Theory of Tonal Music* (GTTM) is a highly formalized music theory that constitutes a "formal description of the musical intuitions of a listener experienced in a musical idiom".<sup>10</sup> The authors have defined preference rules for grouping and meter; according to these, borders between two melodic events are predicted by pauses, long inter-onset intervals, and sudden changes of register, dynamics, or articulation. Unlike GTTM, which operates statically on an entire music piece, the *Implication/Realization Model* emphasizes the dynamic processes as music unfolds in time<sup>11</sup>. It defines rules for so-called "implicative intervals", which are open and give rise to expectation, and "realized intervals", which are closed and represent the termination of an ongoing melodic structure. It is worth noting that both of these theoretical models are limited to tonal music.

3 DRABKIN, Ref. 2.

4 SCHENKER, Heinrich. *Harmony*. Chicago: University of Chicago Press, 1980.

5 CAMBOUROPOULOS, Emilios. *Music Perception: An Interdisciplinary Journal*, 2006, 23(3), pp. 249–268.

6 BELLINGHAM, Jane. *Phrase*. *The Oxford Companion to Music*. Ed. Alison Latham. Oxford Music Online. Oxford University Press. Web. 20. 12. 2016. <<http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e5148>>.

7 BELLINGHAM, Ref. 6.

8 LERDAHL, Fred – JACKENDOFF, Ray. *A Generative Grammar of Tonal Music*. Cambridge, Mass.: The MIT Press, 1983.

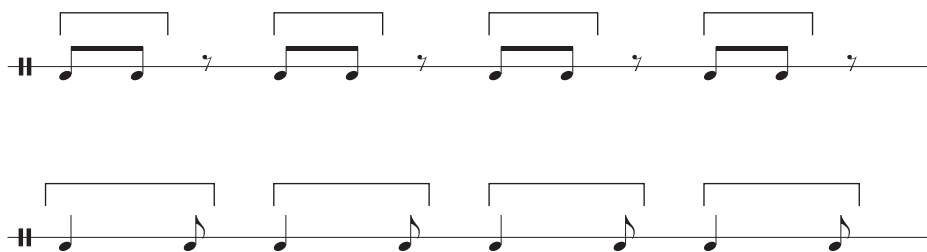
9 NARMOUR, Eugene. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization model*. Chicago: University of Chicago Press, 1990; NARMOUR, Eugene. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press, 1992.

10 LERDAHL – JACKENDOFF, Ref. 8, p. 1.

11 PEARCE – MÜLLENSIEFEN – WIGGINS, Ref. 1.

## Segmentation in cognition

Melodic segmentation as a cognitive task has implications for many music-processing mechanisms, including music perception, attention, memory, and performance. The way in which single sound events are perceptually organized into larger groups is commonly believed to be ruled by laws defined by Gestalt psychology, such as proximity and similarity<sup>12</sup> (Fig. 1).



**Fig. 1** Perceptual grouping of notes based on temporal proximity (above) and similarity (below).

Experimental studies of grouping perception show high degrees of agreement on the perception of melodic segment borders between musically trained and untrained listeners. Deliège<sup>13</sup> asked musicians and non-musicians to identify group borders in music from Bach to Stravinsky. The borders identified by her subjects proved convergent with GTTM predictions. Frankland and Cohen<sup>14</sup> obtained similar results with nursery rhymes and tonal melodies from the classical repertoire.

The effect that segmentation has on memory has been investigated in several melody recognition studies, where listeners were presented with a short excerpt and asked whether it appeared in the stimulus. Excerpts corresponding with phrases were recognized more often than excerpts straddling two phrases.<sup>15</sup>

In music performance, expressive changes of articulation, tempo, and dynamics depend on metric and grouping structure of the piece. Performers typically decrease tempo and dynamics at the end of phrases<sup>16</sup>. Patterns of rubato reflect the hierarchical levels

12 WERTHEIMER, Max: Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 1923, 4(1), pp. 301–350.

13 DELIÈGE, Irène. Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception: An Interdisciplinary Journal*, 1987, 4(4), pp. 325–359.

14 FRANKLAND, Bradley – COHEN, Annabel. Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception: An Interdisciplinary Journal*, 2004, 21(4), pp. 499–543.

15 DOWLING, William. Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics*, 1973, 14(1), pp. 37–40.

16 PALMER, Caroline. Music performance. *Annual Review of Psychology*, 1997, 48(1), pp. 115–138.

of phrases; the more important the segment, the greater the phrase-final lengthening.<sup>17</sup> The beginnings and endings of phrases appear to be more salient than the middle parts. In Sloboda's experiment<sup>18</sup>, performers were less likely to notice an intentional error in the score if it was placed mid-phrase, and play the harmonically correct tone instead.

## Automatic segmentation models

Melodic segmentation performed on symbolic representations of music (scores) is part of many music information and retrieval tasks. Large music databases, such as Répertoire International des Sources Musicales (RISM), often use the first phrase of a composition as an identifying label. As a pre-processing step, melodic segmentation enables quantitative corpus analyses (including melodic feature computation that can serve as a basis for style comparison), studies of similarity and repetition, as well as studies of expressive timing, which may in turn be used in the design of quasi-human performance algorithms.

Several segmentation algorithms have been proposed, using different computational approaches. This section offers a brief review based on Pearce, Müllensiefen, and Wiggins<sup>19</sup>; see their study and the original studies referenced here for more details. The efficiency of the models is usually assessed by comparing the model's predictions to judgments of expert musicologists (or a „ground truth“ derived from these judgments; as previously discussed, different approaches to music analysis may lead to different interpretations of segment borders). The model fit is calculated in terms of *precision* and *recall*. Precision refers to the percentage (or fraction) of relevant segment borders found by the algorithm out of all segment borders identified by the model. Recall is the fraction of relevant segment borders found by the model out of all relevant segment borders as judged by human analysts.

*Rule-based models*, as their name suggests, operate on pre-defined sets of segmentation rules. An early model by Boroda<sup>20</sup> is an adaptation of a linguistic segmentation algorithm based on non-descending rhythmic values. The set of four rules used by Boroda relies exclusively on differences in duration between the given tone, its first and second predecessor, and its successor. Although the author offers examples of the resulting segmentations, to our knowledge the model has not been tested on a larger dataset. Frankland and Cohen<sup>21</sup> based their model on GTTM grouping preference rules. The

17 REPP, Bruno H. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *The Journal of the Acoustical Society of America*, 1992, 92(5), pp. 2546–2568.

18 SLOBODA, John. The effect of item position on the likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology*, 1976, 30(4), pp. 228–237.

19 PEARCE – MÜLLENSIEFEN – WIGGINS, Ref. 1.

20 BORODA, Moisej. Zur Bestimmung einer phrasenähnlichen melodischen Informationseinheit in der Musik. In: Orlov, J.K., Boroda, M.G., Nadarejšvili, I.Š. (Eds.): *Sprache, Text, Kunst. Quantitative Analysen* (pp. 222–230). Bochum: Brockmeyer, 1982.

21 FRANKLAND – COHEN, Ref. 14.

strength of a boundary between two notes is therefore calculated as a function of rest length (if there is a rest between the two notes), distance between neighbouring notes, register change, and length change. The *Local Boundary Detection Model* (LBDM) by Cambouropoulos<sup>22</sup> consists of two rules. The *change* rule is used to set border strength as a function of change, or dissimilarity, between consecutive entities (Cambouropoulos uses pitch, inter-onset interval and rest). The *proximity* rule states that if two consecutive intervals between entities are different, the boundary placed on the larger interval will be stronger. The model requires some manual calibration; depending on the conditions used, Cambouropoulos reports recall values between 63–74%, and precision at 55%. Temperley's *Grouper* algorithm<sup>23</sup> operates on three rules. The *gap rule* refers to the preference of placing boundaries at large inter-onset and offset-to-onset intervals (after long tones and rests). The *phrase length* rule is a preference for phrases consisting of a certain number of notes; Temperley, working with music from the Essen Folk Song Collection, used phrases of about 10 notes. According to the *metrical parallelism* rule, successive segments are preferred to start at the same beat within a bar (e. g. both on the first beat). Grouper outperformed LBDM with recall scores of 76% and a precision of 74%.

*Learning-based models*, such as Bod's *Data Oriented Parsing*<sup>24</sup> (based on supervised learning), or Brent's *Transition Probabilities*<sup>25</sup> and *Pointwise Mutual Information*<sup>26</sup> (based on unsupervised learning), originate in models used in computational linguistics and rely on statistical information (probabilities calculated as the frequency of occurrence of single events or combinations of successive events in the training set). The *Information Dynamics of Music* (IDyOM) model proposed by Pearce, Müllensiefen, and Wiggins<sup>27</sup> also takes an unsupervised approach. IDyOM follows the logic of Narmour's theory of implicative and realized intervals<sup>28</sup>. Implicative intervals are followed by continuations that are predictable, because the implication is for the melody to continue in a certain way. However, what happens after a realized interval is difficult to predict, because a realized interval terminates an ongoing structure and does not give us information about the upcoming structure. The quantification of unexpectedness in statistics is the measure of information content: low information content indicates low predictability, and vice versa. Therefore, IDyOM is based on the assumption that boundaries are perceived before events of low information content in terms of pitch, inter-onset and offset-to-onset interval. The length of the preceding event chain, upon which the prediction is calcu-

22 CAMBOUROPOULOS, Emiliós. The local boundary detection model (LBDM) and its application in the study of expressive timing. In: *Proceedings of the International Computer Music Conference* (pp. 17–22). San Francisco: ICMA, 2001.

23 TEMPERLEY, David. *The Cognition of Basic Musical Structures*. Cambridge, Mass., The MIT Press, 2001.

24 BOD, Rens. Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, 2001, 30(3), pp. 27–36.

25 BRENT, Michael. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 1999a, 34(1–3), pp. 71–105.

26 BRENT, Michael. Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, 1999b, 3(8), pp. 294–301.

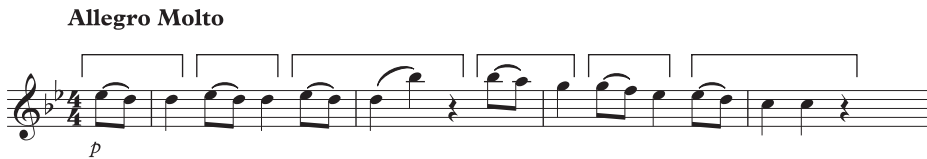
27 PEARCE – MÜLLENSIEFEN – WIGGINS, Ref. 1.

28 NARMOUR, Ref. 9.

lated, is optimized for each point in the melody. Compared to other models, IDyOM gained better precision scores than Grouper or LBDM, but lower recall scores. Surprisingly, for the particular corpus tested (Germanic folk song melodies from the Essen Folk Song Collection), the best precision scores were achieved using a single criterion out of GTTM – pause length – suggesting pauses as the strongest segmenting criterion for the given dataset. While none of the models achieved the performance reported by Bod<sup>29</sup> for his supervised learning model, Data Oriented Parsing was not compared to the other models on the same dataset, so the scores are not directly comparable.

## New model proposal

The segmenting model we propose is based on a simple principle of non-descending note durations, whereby the duration of pauses is added to the duration of the note preceding the pause (see example in Fig. 2). The idea is inspired by the algorithm designed by Boroda<sup>30</sup>. A preliminary analysis suggests that using only the first out of Boroda's four rules might generate more accurate results than the four-rule combination.



**Fig. 2** An example of a melody segmented by the criterion of non-descending note durations (W. A. Mozart, Symphony No. 40 in G minor, KV. 550, Mov. 1)

The fact that pauses lengthen the duration of preceding notes is likely to result in placing borders after pauses; in this respect we hypothesize that our model will identify many of the same borders as the GTTM pause length rule (see previous chapter). However, as illustrated by the example in Figure 2, the proposed model will probably return shorter segments, corresponding to motifs rather than phrases. One of the main questions therefore is to what degree music is rhythmically organized in this way, that is, if motifs tend to consist of ascending, rather than descending, durations. The next question is if average segment size can be used as a style-differentiating criterion. For example, if we apply our segmentation model on music that is largely isochronous, i.e. consisting mostly of notes of the same duration, the segments obtained will be much longer than in the example shown in Figure 2. Measuring the average motif/phrase length would open space to further research questions, not only in music, but also for music-language comparisons. For example, Patel and Daniele<sup>31</sup> compared rhythmic variability in speech

29 BOD, Ref. 24.

30 BORODA, Ref. 20.

31 PATEL, Aniruddh – DANIELE, Joseph. An empirical comparison of rhythm in language and music. *Cognition*, 2003, 87(1), pp. B35-B45.

(French and English) and instrumental compositions of French and English composers. They found spoken French to be more isochronous than spoken English. Accordingly, rhythmic variability of French compositions was shown to be lower than that of English compositions. It would be interesting to make an analogous comparison for other languages and compositions. Another issue to be explored is the quantitative relationship between long and short music segments (e.g. motifs and phrases). According to the Menzerath-Altmann law in linguistics<sup>32</sup>, the length of a clause is inversely proportional to the length of its constituents: the longer a sentence, the shorter its clauses; the longer a word, the shorter its syllables. If an analogous relationship existed in music, longer phrases would consist of shorter motifs. To our knowledge, this issue has not yet been explored. The value of the proposed algorithm, and its relevance in terms of precision and recall, will only become clearer after its testing on musical data.

## Bibliography

- ALTMANN, Gabriel – SCHWIBBE, Michael. *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Hildesheim/Zürich/New York: Olms, 1989.
- BELLINGHAM, Jane. Phrase. *The Oxford Companion to Music*. Ed. Alison Latham. Oxford Music Online. Oxford University Press. Web. 20. 12. 2016. <<http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e5148>>.
- BOD, Rens. Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, 2001, 30(3), pp. 27–36.
- BORODA, Moisej. Zur Bestimmung einer phrasenähnlichen melodischen Informationseinheit in der Musik. In: Orlov, J.K., Boroda, M.G., Nadarejšvili, I.Š. (Eds.): *Sprache, Text, Kunst. Quantitative Analysen* (pp. 222–230). Bochum: Brockmeyer, 1982.
- BRENT, Michael. An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning*, 1999a, 34(1–3), pp. 71–105.
- BRENT, Michael. Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences*, 1999b, 3(8), pp. 294–301.
- CAMBOUROPOULOS, Emiliós. Musical parallelism and melodic segmentation: A computational approach. *Music Perception: An Interdisciplinary Journal*, 2006, 23(3), pp. 249–268.
- CAMBOUROPOULOS, Emiliós. The local boundary detection model (LBDM) and its application in the study of expressive timing. In: *Proceedings of the International Computer Music Conference* (pp. 17–22). San Francisco: ICMA, 2001.
- DELIÈGE, Irène. Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception: An Interdisciplinary Journal*, 1987, 4(4), pp. 325–359.
- DOWLING, William. Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics*, 1973, 14(1), pp. 37–40.
- DRABKIN, William. Motif. *Grove Music Online*. Oxford Music Online. Oxford University Press. Web. 20. 12. 2016. <<http://www.oxfordmusiconline.com/subscriber/article/grove/music/19221>>.

<sup>32</sup> ALTMANN, Gabriel – SCHWIBBE, Michael. *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Hildesheim/Zürich/New York: Olms, 1989.



- FRANKLAND, Bradley – COHEN, Annabel. Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception: An Interdisciplinary Journal*, 2004, 21(4), pp. 499–543.
- LERDAHL, Fred – JACKENDOFF, Ray. *A Generative Grammar of Tonal Music*. Cambridge, Mass.: The MIT Press, 1983.
- NARMOUR, Eugene. *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization model*. Chicago: University of Chicago Press, 1990; NARMOUR, Eugene. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press, 1992.
- PALMER, Caroline. Music performance. *Annual Review of Psychology*, 1997, 48(1), pp. 115–138.
- PATEL, Aniruddh – DANIELE, Joseph. An empirical comparison of rhythm in language and music. *Cognition*, 2003, 87(1), pp. B35-B45.
- PEARCE, Marcus – MÜLLENSIEFEN, Daniel – WIGGINS, Geraint. Melodic grouping in music information retrieval: New methods and applications. In: *Advances in music information retrieval*. Springer, Berlin Heidelberg, 2010, pp. 364–388.
- REPP, Bruno H. Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *The Journal of the Acoustical Society of America*, 1992, 92(5), pp. 2546–2568.
- SCHENKER, Heinrich. *Harmony*. Chicago: University of Chicago Press, 1980.
- SLOBODA, John. The effect of item position on the likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology*, 1976, 30(4), pp. 228–237.
- TEMPERLEY, David. *The Cognition of Basic Musical Structures*. Cambridge, Mass., The MIT Press, 2001.
- WERTHEIMER, Max: Untersuchungen zur Lehre von der Gestalt II. *Psychologische Forschung*, 1923, 4(1), pp. 301–350.

