MUSICAL FEATURE AND NOVELTY CURVE CHARACTERIZATIONS AS PREDICTORS OF SEGMENTATION ACCURACY

Martin Hartmann
Finnish Centre for
Interdisciplinary Music Research
University of Jyväskylä, Finland
martin.hartmann@jyu.fi

Olivier Lartillot
Department of Musicology
University of Oslo, Norway
olartillot@gmail.com

Petri Toiviainen
Finnish Centre for
Interdisciplinary Music Research
University of Jyväskylä, Finland
petri.toiviainen@jyu.fi

ABSTRACT

Novelty detection is a well-established method for analyzing the structure of music based on acoustic descriptors. Work on novelty-based segmentation prediction has mainly concentrated on enhancement of features and similarity matrices, novelty kernel computation and peak detection. Less attention, however, has been paid to characteristics of musical features and novelty curves, and their contribution to segmentation accuracy. This is particularly important as it can help unearth acoustic cues prompting perceptual segmentation and find new determinants of segmentation model performance. This study focused on spectral, rhythmic and harmonic prediction of perceptual segmentation density, which was obtained for six musical examples from 18 musician listeners via an annotation task. The proposed approach involved comparisons between perceptual segment density and novelty curves; in particular, we investigated possible predictors of segmentation accuracy based on musical features and novelty curves. For pitch and rhythm, we found positive correlates between segmentation accuracy and both local variability of musical features and mean distance between subsequent local maxima of novelty curves. According to the results, segmentation accuracy increases for stimuli with milder local changes and fewer novelty peaks. Implications regarding prediction of listeners' segmentation are discussed in the light of theoretical postulates of perceptual organization.

1. INTRODUCTION

Musical segments are a representation of the perceived structure of music, and hence carry its multifaceted, interwoven, and hierarchically organized nature. Transitional points or regions between segments, which are called *perceptual segment boundaries*, can emerge from temporal changes in one or more musical attributes, or from more complex configurations involving e.g. repetition and cadences. Exhibited acoustic change can be measured to yield an estimate of *novelty* with respect to previous and upcoming musical events, with the aim to delineate perceptual segment boundaries: for each instant, its degree of acoustic novelty is expected to predict the likelihood of indicating a boundary. Segmentation accuracies for different musical features, such as timbre and harmony descriptors, can be obtained via Novelty detection (Foote, 2000) approaches. The success of predicting musical segment boundaries from acoustic estimates of novelty depends on the level of structural complexity of a musical piece. For instance, music that unambiguously evokes few sharp segment boundaries and clear continuity within segments to listeners should exhibit high accuracy, whereas pieces that prompt many boundaries and a rather heterogeneous profile might be more challenging for prediction. Music is however multidimensional, which implies that accuracy for a given stimulus could depend on the musical feature or features under study.

This work focuses on the assessment of possible predictors of segmentation accuracy that could be obtained directly from musical features or from novelty points derived from these features. We explored musical features of different types, because listeners rarely focus on a single dimension of music. To illustrate possible applications of the proposed approach, one could imagine a segmentation system that could extract candidate features from a particular musical piece (Peiszer, Lidy, & Rauber, 2008), and discard any features that would not seem to be informative of musical changes in order to focus only on those changes that would be deemed relevant for a listener. This would result in more efficient prediction, as the system would not require to compute subsequent structure analysis steps for irrelevant features. Our endeavor is motivated by the direct impact of music segmentation on other areas of computational music analysis, including music summarization, chorus detection, and music transcription, and its relevance for the study of human perception, as it can deepen our understanding on how listeners parse temporally unfolding processes.

MIR (Music Information Retrieval) studies on segmentation and structure analysis typically require perceptual data for algorithm evaluation. Often, data collection involves the indication of segment boundaries from one or few listeners for a large amount of musical examples; this results in a set of time points for each piece. In contrast, approaches on music segmentation within the field of music perception and cognition commonly involve the collection of perceptual boundaries from multiple participants in listening experiments; the collected data is often aggregated across listeners for its analysis. Kernel Density Estimation (*KDE*,

Silverman, 1986) has been used in recent segmentation studies (e.g. Bruderer, 2008; Hartmann, Lartillot, & Toiviainen, 2016b) to obtain a precise aggregation across boundary data. This approach consists of obtaining a probability density estimate of the boundary data using a Gaussian function; in a KDE curve, temporal regions that prompted boundary indication from many listeners are represented as peaks of perceptual boundary density. This continuous representation of perceptual segment boundary probability has been used to compare different stimuli, groups of participants, and segmentation tasks (Bruderer, 2008; Hartmann et al., 2016b).

Prediction of perceptual boundary density in the audio domain often involves the computation of novelty curves (Foote, 2000), which roughly describe the extent to which a temporal context is characterized by two continuous segments separated by a discontinuity with respect to a given musical feature. Recent studies have compared novelty curves with perceptual boundary density curves (Hartmann, Lartillot, & Toiviainen, 2016a) and compared peaks derived from both curves (Mendoza Garay, 2014), showing that novelty detection can predict segmentation probabilities derived from numerous participants.

A preliminary step in novelty detection and other segmentation frameworks consists of the extraction of a musical feature, which will determine the type of musical contrast to be detected. Timbre, tonality, and to some extent rhythm (Jensen, 2007) have been considered to be important features for structural analysis. Relatively high prediction of musical structure has been found for two musical features: MFCCs (Mel-Frequency Cepstral Coefficients) for timbre description (Foote, 2000) and Chromagram or similar features (Serrà, Muller, Grosche, & Arcos, 2014) for description of pitch changes; also combined approaches have been proposed (Eronen, 2007). The segmentation accuracy achieved by novelty curves seems to highly depend on the musical feature that is used, and on the choice of temporal parameters for feature extraction (Peeters, 2004). In addition, different musical pieces might require different musical features for optimal prediction (Peiszer et al., 2008); as pointed out by McFee and Ellis (2014), structure in pop and rock is frequently determined by harmonic change, whereas jazz is often sectioned based on instrumentation. No single feature can optimally predict all musical examples; certain features are more appropriate than others depending on particular aspects of musical pieces (Smith & Chew, 2013).

This mechanism is however not well understood at present: it is unclear what characteristics of musical features contribute to the segmentation accuracy for a given musical piece. Addressing this issue would be important since it would enable the possibility to select optimal musical features for further novelty detection, avoiding the computation of novelty curves that would not yield satisfactory results; it would also help to develop better alternatives to the novelty detection approach, with the aim of reducing computational costs.

To analyze the impact of different factors associated to novelty curves upon segmentation and compare different segmentation algorithms, a number of performance measures have been proposed, such as precision, recall, and F-measure (Lukashevich, 2008); also correlation between time series has been applied for this purpose (Hartmann et al., 2016a). One of the factors that has been shown to highly contribute to the segmentation accuracy is the width of the novelty kernel (e.g. Hartmann et al., 2016a), which roughly refers to the temporal context with respect to which novelty is estimated for each time point.

However, to the best of our knowledge, no study has systematically investigated what specific aspects of novelty curves contribute to their accuracy for a given stimulus. It would be relevant to investigate what characteristics of novelty curves relate to their accuracy, as this could allow to predict the relative suitability of a novelty curve for a given stimulus without the need of direct comparison against ground truth, and to bypass computation of novelty curves that would be assumed not to deliver satisfactory performance with regard to a particular stimulus. From the viewpoint of music perception, it would be useful to better understand the extent to which musical characteristics perceived by listeners are directly apparent from novelty curves, and to gain more knowledge on the types of musical changes that prompt both boundary perception and high novelty scores.

Recently, Hartmann et al. (2016a) studied segmentation accuracy achieved for concatenated musical pieces using different novelty curves. It was found that optimal prediction of perceptual segment boundary density involves the use of large kernel widths; the study also highlights the role of rhythmic and pitch-based features on segmentation prediction. This study is a follow-up to the paper by Hartmann et al. (2016a), as it focused further on prediction of perceptual segmentation density via novelty detection, and examined the same musical pieces; in particular, we investigated one of the perceptual segmentation sets (an *annotation* segmentation task performed by musician listeners) studied by Hartmann et al. (2016a), and explored perceptual segmentation density and novelty curves for individual musical stimuli. The aim of the present study was to understand whether or not local variability of musical features and distance between novelty peaks are related with the accuracy of segmentation models. The following research questions guided our investigation:

1. What specific aspects of musical stimuli that account for segmentation accuracy can be directly described from musical features?

2. What stimulus-specific attributes of novelty curves determine optimal segmentation accuracy?

As regards the first research question, we expected to find an inverse relationship, dependent on musical stimulus, between magnitude of local feature variation and accuracy obtained via novelty detection. For instance, musical stimuli displaying unfrequent local tonal contrast would yield optimal segmentation accuracy via tonal novelty curves. The rationale behind this hypothesis is that if there is not much local change in a feature, then the local changes that occur in that feature should be more salient. One of the most basic

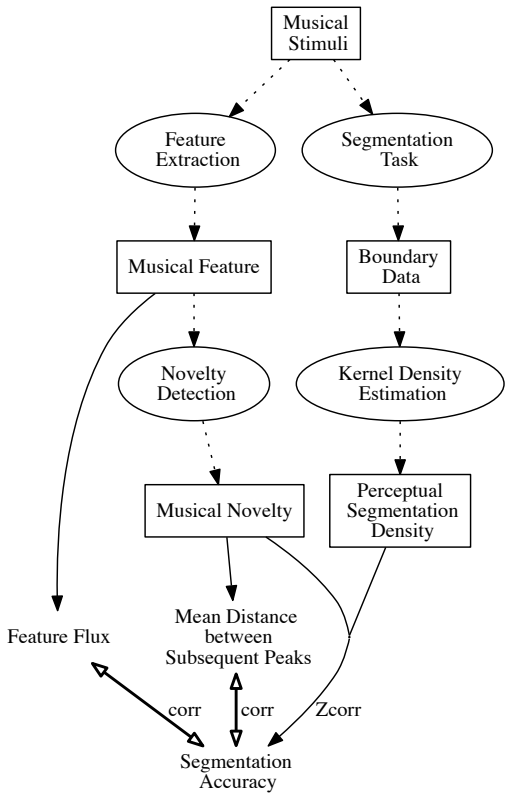

Figure 1. General design of the study.

Gestalt principles, the law of *Prägnanz*, relates to this rationale, because it states that, under given conditions, perceptual organization will be as "good" as possible, i.e. percepts tend to have the simplest, most stable and most regular organization that fits the sensory pattern (Koffka, 1935).

Regarding the second question, we expected that stimuli yielding higher accuracy for a given feature would exhibit a relatively large temporal distance between novelty peaks for that feature. Although this relationship could depend on other factors such as tempo and duration, we believed that the absolute time span between peaks would serve as a rough predictor of accuracy for the reasons above mentioned regarding the law of Prägnanz. To give an example, optimal accuracy was expected for stimuli characterized by long and uniform segments with respect to instrumentation that would be delimited by important timbral changes, whereas music characterized by more frequent timbral change and gradual transitions would yield lower accuracy for a timbre-based novelty measure.

2. METHOD

Figure 1 illustrates the research design utilized in our investigation. The upper part of the figure concerns computational prediction of segmentation via novelty detection and a perceptual modelling of collected segmentation data via perceptual segmentation density estimation. These two topics were more thoroughly covered in Hartmann et al. (2016b) and Hartmann et al. (2016a), respectively. The bottom part of the figure, specifically the solid line connections, refer to investigation of correlates of perceptual boundary density prediction, which is the main focus of this study.

2.1 Segmentation Task

To obtain perceptual segment boundary density of the stimuli, we collected boundary data from participants via a listening task that involved an offline annotation. This non real-time segmentation task, called *annotation task*, is described with more detail in Hartmann et al. (2016b). The reason for analyzing prediction of non real-time segmentation was to reduce the number of intervening factors: compared to a real-time segmentation, boundary placements obtained via offline segmentation are probably better aligned in time with respect to musical changes; also, inter-subject agreement has been found to be higher for offline than for online tasks (Hartmann et al., 2016b).

2.1.1 Subjects

18 musicians (11 males, 7 females) with a mean age of 27.61 years (SD = 4.45) and an average musical training of 14.39 years (SD = 7.49) took part of this experiment. They played classical (12 participants) and non-classical musical styles (6 participants) such as rock, pop, folk, funk, and jazz.

2.1.2 Stimuli

We selected 6 instrumental music pieces; two of them lasted 2 minutes and the other four were trimmed down to this length for a total experiment duration of around one hour. The pieces (see Hartmann et al., 2016a) comprise a variety of styles and considerably differ from one another in terms of musical form; further, they emphasize aspects of musical change of varying nature and complexity.

2.1.3 Apparatus

An interface in Sonic Visualizer (Cannam, Landone, & Sandler, 2010) was prepared to obtain segmentation boundary indications from participants. Stimuli were presented in randomized order; for each stimulus, the interface showed its waveform as a visual-spatial cue over which boundaries would be positioned (subjects were asked to focus solely on the music). Participants used headphones to play back the music at a comfortable listening level, and both keyboard and mouse were required to complete the segmentation task.

2.1.4 Procedure

Written instructions were given to participants; these included a presentation of the interface tools and a task description, which consisted of the following steps:

1. Listen to the whole musical example.

2. Indicate significant instants of change while listening to the music by pressing the Enter key of the computer.

3. Freely play back from different parts of the musical example and make the segmentation more precise by adjusting the position of boundaries; also removal of any boundaries indicated by mistake is allowed.

4. Rate the perceived strength of each boundary (ratings of boundary strength were collected for another study). Start over from the first step for the next musical example.

2.2 Perceptual Segment Boundary Density

We obtained a perceptual boundary density estimate across the segmentation data collected from musician participants; this estimate would be further compared against novelty curves to assess their accuracy. The perceptual boundary data of all participants was used to obtain a curve of perceptual segmentation density using a KDE bandwidth of 1.5 s; values around this bandwidth were found optimal for comparison between perceptual segmentation densities (Hartmann et al., 2016b). From each tail of the perceptual density curves, 6.4 s were trimmed for more accurate comparisons with novelty curves (see below).

2.3 Feature Extraction and Novelty Detection

We computed novelty curves from 5 musical features describing timbre (Subband Flux), rhythm (Fluctuation Patterns), pitch class (Chromagram) and tonality (Key Strength, Tonal Centroid) using MIRtoolbox 1.6.1 (Lartillot & Toiviainen, 2007); see Hartmann et al. (2016a) for a description of the features used for novelty detection. For each feature, a self-similarity matrix was obtained by computing the cosine distance between all possible pairs of feature frames. Novelty for each time point was computed via convolution between each self-similarity matrix and a Gaussian checkerboard kernel (Foote, 2000) with half width of 11 s; large kernel sizes have been previously used to overcome high levels of detail in novelty curves (e.g. Hartmann et al., 2016a).

As done with the perceptual density curves, we truncated the novelty curves by trimming 6.4 s from each extreme to avoid edge effects. We chose the smallest value that would eliminate, for all stimuli, any novelty spikes caused by the contrast between music and silence in the beginning and end of tracks; trimming the extremes of the novelty curves also increased the number of novelty points that derive from a full checkerboard kernel. Once the novelty curves were computed, we also trimmed 6.4 s from the extremes of each dimension of the musical features in order to obtain the predictors of accuracy described below.

2.4 Characterization of Musical Features and Novelty Curves

Subsequently, we computed two characterizations in order to estimate feature local discontinuity and temporal distance between music structure changes. From each musical feature matrix F we calculated mean *Feature Flux*, an estimate of the amount of local variation; Feature Flux is the Euclidean distance between successive feature frames. First, for each time series F_d, where d corresponds to a feature dimension, the squared difference between successive time points is obtained. Next, a flux time series v is obtained as the squared root of the sum across dimensions:

$$v_t = \sqrt{\sum_{d=1}^{N}(F_d(t) - F_d(t-1))^2}$$

Finally, mean Feature Flux is obtained by averaging the flux time series v across time points:

$$\textit{Feature Flux} = \frac{1}{K}\sum_{t=1}^{K} v_t$$

From each novelty curve, we obtained *Mean Distance Between Subsequent Peaks* (MDSP), which describes the peak-to-peak duration (in seconds) of novelty curves. To compute this estimate, we first obtained from the novelty curve a vector of novelty peak locations v, where v_i corresponds to the ith peak, and N corresponds to the number of peaks; MDSP was calculated as follows:

$$\textit{MDSP} = \frac{1}{N}\sum_{i=1}^{N}(v_i - v_{i-1})$$

2.5 Segmentation Accuracy and its Correlates

We compared perceptual segmentation density and novelty curves to obtain segmentation accuracy. To this end, we performed correlations between novelty and perceptual segmentation density for each stimulus and musical feature.

We focused on the possible relationship between accuracy and the aforementioned characterizations of musical features and novelty peaks. Hence, for each feature we correlated across stimuli the accuracy with Feature Flux and with MDSP. In order to perform these correlations, the accuracies of each feature required to follow an approximately normal distribution, so we subsequently transformed accuracies via Fisher's z transformation of r. The normalization of Z_r involved the calculation of effective degrees of freedom to correct for temporal autocorrelation (Pyper & Peterman, 1998).

3. RESULTS

3.1 Prediction of Perceptual Segmentation Density

Figure 2 shows the correlation between novelty curves and perceptual boundary density for each stimulus. The prediction accuracies were found to vary depending on stimulus; for instance, *Smetana* yielded very high correlations, whereas these were rather low for *Couperin*. Also, for any given stimulus, accuracy differed according to the feature and feature type used for novelty detection; for instance, *Smetana* yielded higher accuracy for pitch-based features than for rhythmic and spectral-based features. Interestingly, no single novelty feature successfully predicted perceptual boundary density for all stimuli.

At this point, it might be relevant to illustrate the data analyzed in this study and at the same time explore two

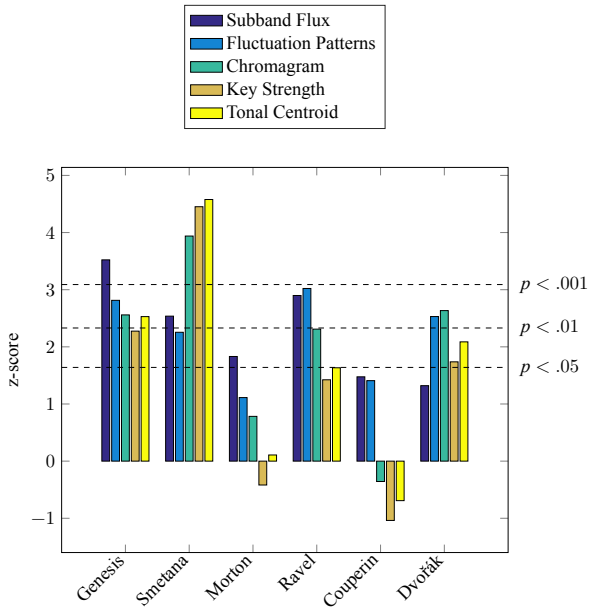

Figure 2. Correlation between novelty curves and perceptual boundary density.

musical stimuli that were found to be contrasting with respect to their prediction. Figure 3 visually compares representations for two stimuli that yielded optimal and worst accuracies for Tonal Centroid, *Smetana* and *Couperin*, respectively. The upper graphs show the Tonal Centroid time series for each stimulus; dimensions 1-2, 3-4 and 5-6 correspond to pairs of numerical coordinates for the circle of fifths, minor thirds, and major thirds, respectively. The middle graphs present novelty curves for Tonal Centroid, which result from the representations in the upper graphs. The lower graph shows perceptual segmentation density based on listeners' segmentation. Some clear differences can be seen between the profiles for *Smetana* and *Couperin*. The three most contrasting peaks of perceptual density for *Smetana* correspond to changes of key: at 40 s the music changes from an F key to a fugue in G min, at 66 s the F melody is reprised in C♯, and at 104 s begins a similar fugue transposed to F min. These changes are quite clear in the Tonal Centroid representation, yielding a relatively adequate novelty prediction, although the rather gradual transition at around 66 s does not yield a stark novelty peak.

In contrast, the Tonal Centroid time series of *Couperin* does not allow for a straightforward acoustic interpretation of the perceptual segmentation density profile. According to the peaks of perceptual density, listeners seem to base their indications primarily on the ending of cadences; these are characterized by the use of heavy ornamentation (e.g. mordents) and chords, which are rather salient because the piece is almost exclusively two-voiced. Due to these cues, listeners seem to place more segment boundaries on endings of melodies than on beginnings, for instance at 23 s. In comparison to this, Tonal Centroid and its corresponding novelty curve would describe slightly delayed musical changes, as harmonic transitions become apparent only after enough development of subsequent melodic material. Another reason behind the highly inaccurate prediction is that listeners placed boundaries for changes of register, rests and durational changes, which might partly explain the higher accuracies obtained for spectral and rhythmic features.

3.2 Finding Predictors of Segmentation Accuracy

Musical Feature	Feature Flux	MDSP
Subband Flux	.54	−.03
Fluctuation Patterns	−.30	.11
Chromagram	−.68	.27
Key Strength	**−.74**	.47
Tonal Centroid	−.66	**.50**

Table 1. Correlation between z-transformed accuracy and characterizations of musical features (Feature Flux) and novelty curves (MDSP).

Subsequently, we analyzed characterizations of extracted features and of novelty curves derived therefrom, looking for correlates of accuracy. We focused on Feature Flux, a global estimate of the extracted features, and on MDSP, which was obtained from novelty curves, to find whether or not these would be indicative of novelty curve accuracy. Table 1 shows the correlation between segmentation accuracies and the obtained characterizations of musical features and novelty curves. We found a strong negative correlation between accuracy and Feature Flux for pitch-based features; as regards accuracy and MDSP, we obtained moderate to strong positive correlations for tonal features (*strong* and *moderate* mean $|r| > .5$ and $.3 < |r| < .5$ respectively, following Cohen, 1988). Although these results did not reach statistical significance at $p < .05$, some interpretations can still be made. According to the results, accuracy increases for stimuli with fewer local change in pitch content and less peaks in pitch-based novelty curves. A similar pattern of results was found for rhythm; we obtained for Fluctuation Patterns a moderate negative correlation between Feature Flux and accuracy, and a weak positive correlation between MDSP and accuracy. Timbre seemed to yield an opposite trend, at least for Feature Flux; Subband Flux exhibited a strong positive correlation between Feature Flux and accuracy, and no or very weak correlation between MDSP and accuracy. This suggests that high accuracy is associated with more local variability of spectral fluctuation.

4. DISCUSSION

Understanding which specific aspects of musical pieces influence novelty-based segmentation prediction is a crucial but challenging issue. One possible way to address this problem is to focus on the particulars of this approach and tackle the question of what characteristics of musical features and their respective novelty curves predict segmentation accuracy for different musical pieces. This study tries to fill the gap in this respect, and aims to open a discussion on the possibility of predicting accuracy directly from

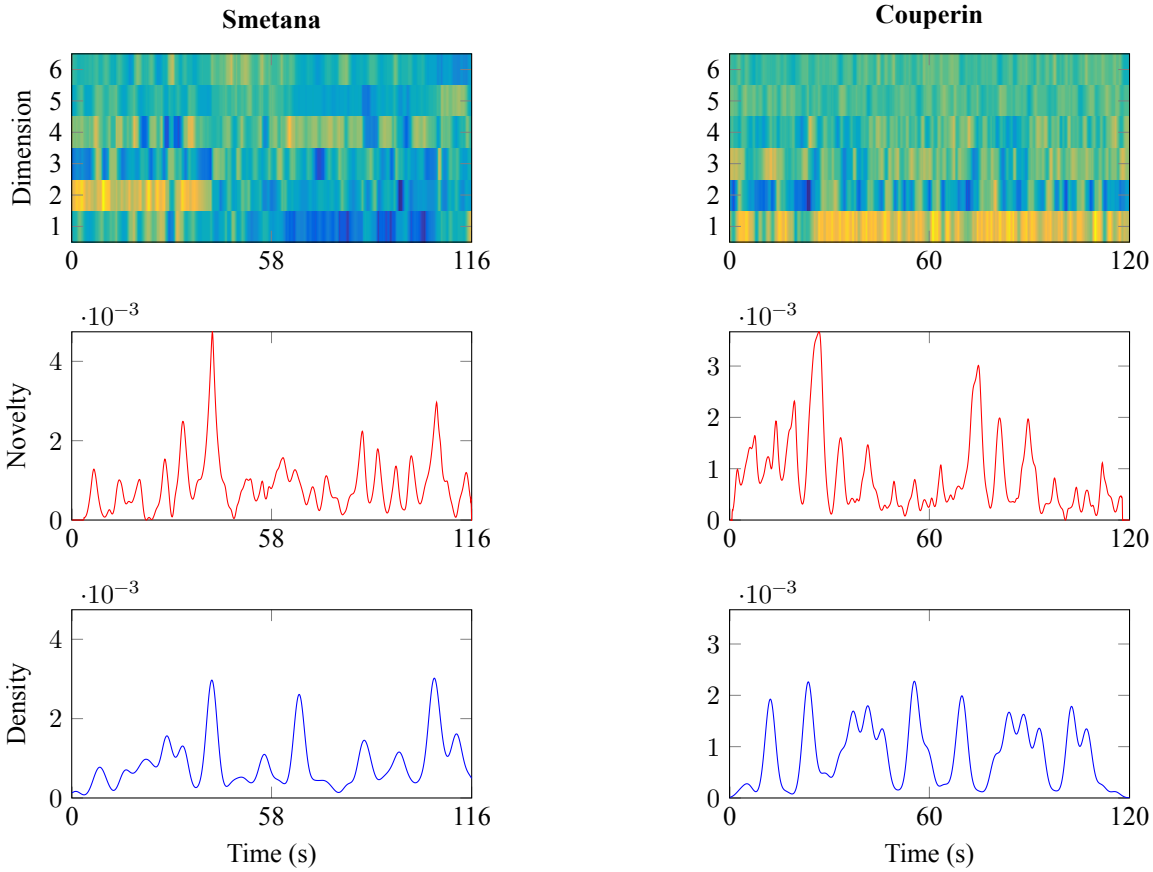

Figure 3. Tonal Centroid, novelty of Tonal Centroid and perceptual boundary density for stimuli *Smetana* and *Couperin*.

musical feature characteristics; we believe that automatic segmentation systems could optimize their parameters for stimuli of different idioms and that such an approach would lead to an increase in accuracy.

Regarding the validity of the approach, we also highlight that our method for studying segmentation involves amassing annotations from multiple listeners for the same stimulus; this possibility has been previously explored only by few studies on segmentation prediction (Mendoza Garay, 2014; Hartmann et al., 2016a). In contrast, MIR studies on music segmentation are often based on data coming from one or few annotators; the computation of a KDE is usually not needed since the set of annotated segment boundaries is directly compared against peaks picked from a novelty curve. In this sense, analyses of perceptual segmentation based on data that is probably more representative of the musician population should be useful for better understanding both perception and prediction of musical structure.

This section examines our research questions in light of the proposed analysis, and assesses the extent to which the stated hypotheses could be supported. Finally, we conclude this article with possible directions for future research.

4.1 Segmentation Accuracy

The first step to address the research questions was to investigate the accuracy to predict perceived boundaries yielded by different musical features for different musical examples. As shown in Figure 2, accuracy seems to highly vary according to musical piece, motivating further analyses on novelty detection that focus on each piece separately. It is also apparent that no single algorithm is robust for prediction of all examples, suggesting the importance of incorporating combinations of multiple features (Hartmann et al., 2016a; Müller, Chew, & Bello, 2016), multiple time scales (Kaiser & Peeters, 2013; Gaudefroy, Papadopoulos, & Kowalski, 2015), and other aspects of segmentation (e.g. repetition principles, see Gaudefroy et al., 2015; Paulus & Klapuri, 2006) into novelty approaches. Overall, however, the results seem to support the idea that performance in structural analysis heavily may depend more on musical stimuli than on the algorithm used or the choice of parameters (Peiszer et al., 2008), which could serve to justify subsequent steps of our analysis.

4.2 Feature-based Prediction of Novelty Detection Performance

Our first hypothesis was that accuracy obtained via novelty detection would increase for stimuli with low local variation of musical features. In support with this hypothesis, we overall found Feature Flux to be a good predictor of correlation between perceptual segmentation density and novelty (Table 1), and a same pattern of results for pitch-based and rhythmic features. This suggests, for these features, that increased local feature continuity may be indicative of higher accuracy of novelty curves.

The second hypothesis of this study was that accuracy

would increase for stimuli with more distant novelty peaks. Indeed, mean distance between subsequent peaks was found to be somewhat indicative of the accuracy of novelty curves with respect to perceptual segmentation density. This suggests, particularly for tonal features, that longer novelty peak-to-peak duration is associated with higher correlation between detected novelty and perceptual segment density.

Focusing on pitch-based and rhythmic features, the results indicate that high local variability is associated with low accuracy. A possible interpretation is that music characterized by few local changes in pitch or rhythm often involves few, highly contrasting pitch-based or rhythmic structural boundaries. For instance, rhythmically stable melodies are clearly separated by highly discernible rests and long notes in *Dvořák*. This could be interpreted in the light of theoretical approaches to musical expectation (Narmour, 1992), according to which similarity between successive events generates the expectation of another similar event. If eventually this expectation is not satisfied, a sense of closure may be perceived, prompting the indication of a segment boundary; this may explain why, for example, accurate tonal-based segmentation predictions exhibit low variability at a local level and often involve few, perceptually stark boundaries that delimit homogeneous groupings of events.

Related to our previous result, we found that novelty curve characteristics can be used as predictors of accuracy: larger distance between subsequent novelty peaks was found to result in higher novelty accuracy (Table 1), especially for tonal features. As aforementioned, this relationship relates to the properties of "good" organizations proposed by Gestalt theorists (Koffka, 1935). In this sense, music characterized by novelty peaks that are clearly isolated should yield higher accuracy as they would relate to perceptually salient musical changes.

We should highlight that the features yielding highest correlations with accuracy were tonal. It is possible that interpretations derived via perceptual organization rules and expectation violation are better applicable to the case of prediction via tonal features because these features focus unambiguously on changes in perceived tonal context (and not on e.g. loudness changes). In contrast, other descriptions used are somewhat more vague: i) Subband Flux discontinuities encompass changes of instrumentation, register, voicing, articulation, and loudness; ii) Fluctuation Pattern changes could be attributed to rhythmic patterns, tempo, articulation, and use of repetition; iii) changes in Chromagram are manifested in pitch steps, pitch jumps, and use of chords. In this respect, tonal features consider a single dimension of musical change, whereas other features analyzed in this study may yield more intricate descriptions.

Accuracy seemed to increase for stimuli with little local change and more distance between peaks (Table 1), however Subband Flux seemed to yield an opposite trend. In this regard, high local variability of changes in instrumentation, register, loudness, etc., seems to be associated with higher accuracy. It could be the case that musical pieces with high local spectral change and multiple novelty peaks are also characterized by few structural sections of long duration, and yield a relatively straightforward prediction; for instance, *Genesis* contains multiple short sounds and effects, yet its sections are clearly delimited by important instrumentation changes, which probably had a positive effect on accuracy. Again, stylistic information might help to disentangle these and other problems regarding segmentation accuracy.

4.3 General Discussion

One of the main aims of this study was to find out methods to select musical features that would be efficient in segmentation prediction for a given stimulus; to this end, we investigated the relationship between accuracy and characterizations of musical features and novelty curves. According to the results, for most features there is an inverse relationship between local variability and accuracy, and a direct relationship between mean distance between subsequent novelty peaks and accuracy. This suggests that stimuli whose features are characterized by low variation between successive time points, and whose novelty curves have few peaks, are likely to yield higher segmentation prediction accuracy. A possible reason that explains these results is that music with infrequent musical change often yields perceptually salient boundaries; according to the Gestalt rules of perceptual organization, similar events that are proximal in time are grouped together, creating a strong sense of closure whenever a dissimilar event is perceived. Following this interpretation, if a given musical dimension changes frequently, an increase of listeners' attention towards other dimensions evoking strong closure may occur during segmentation.

4.4 Considerations for Further Research

Since this study focused on the analysis of segmentations of the same musical pieces from multiple listeners, the number of segmented stimuli does not suffice to draw solid conclusions about the correlates of segmentation accuracy; this should be considered a major methodological caveat. More musical stimuli are clearly needed to assess the generalization ability of our results; future studies should in this respect increase the number of musical stimuli used for perceptual segmentation tasks while maintaining a satisfactory participant sample size. As regards the sample of participants used, this study only focused on musician listeners, mainly following MIR studies, which recruit expert annotators for the preparation of musical structure data sets. Further work should concentrate on annotation segmentation from nonmusicians in order to understand accuracy of novelty curves with regards to the majority of the population.

Another issue to consider is that the novelty detection approach is designed to yield maximum scores for high continuity within segments and high discontinuity at segmentation points, so in this sense it is not surprising that music exhibiting clear discontinuity between large sequences of homogeneity for a given feature will yield higher segmentation accuracy for that feature. In this regard, our results should be further tested using other approaches; for instance probabilistic methods (e.g. Pauwels, Kaiser, &

Peeters, 2013; Pearce & Wiggins, 2006) would be suitable as they offer alternative assumptions regarding location of actual boundaries.

Finally, as an outcome of this study it can be stated that listeners may tend to focus on musical dimensions that do not change often. This interpretation is plausible and highlights the importance of e.g. tonal and tempo stability, as well as the role of repetition and motivic similarity in musical pieces. Future work should test this possibility by conducting listening studies in which listeners would describe what is the most salient dimension for different time points in the music; further, as suggested by Müller et al. (2016), automatic detection of these acoustic description cues should also be a relevant task regarding structural segmentation.

5. REFERENCES

Bruderer, M. (2008). *Perception and modeling of segment boundaries in popular music* (Doctoral dissertation, JF Schouten School for User-System Interaction Research, Technische Universiteit Eindhoven, Eindhoven).

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: an open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the ACM Multimedia International Conference* (pp. 1467–1468). Firenze, Italy.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd). New Jersey: Lawrence Erlbaum.

Eronen, A. (2007). Chorus detection with combined use of MFCC and chroma features and image processing filters. In *Proceedings of the 10th International Conference on Digital Audio Effects* (pp. 229–236). Citeseer. Bordeaux.

Foote, J. T. (2000). Automatic audio segmentation using a measure of audio novelty. In *IEEE International Conference on Multimedia and Expo* (Vol. 1, pp. 452–455). IEEE. New York.

Gaudefroy, C., Papadopoulos, H., & Kowalski, M. (2015). A multi-dimensional meter-adaptive method for automatic segmentation of music. In *13th International Workshop on Content-Based Multimedia Indexing (CBMI)* (pp. 1–6). IEEE.

Hartmann, M., Lartillot, O., & Toiviainen, P. (2016a). Interaction features for prediction of perceptual segmentation: effects of musicianship and experimental task. *Journal of New Music Research*.

Hartmann, M., Lartillot, O., & Toiviainen, P. (2016b). Multi-scale modelling of segmentation: effect of musical training and experimental task. *Music Perception*, *34*(2), 192–217.

Jensen, K. (2007). Multiple scale music segmentation using rhythm, timbre, and harmony. *EURASIP Journal on Applied Signal Processing*, *2007*(1), 159–159.

Kaiser, F. & Peeters, G. (2013). Multiple hypotheses at multiple scales for audio novelty computation within music. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. Vancouver.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace and Company.

Lartillot, O. & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects* (pp. 237–244). Bordeaux.

Lukashevich, H. M. (2008). Towards quantitative measures of evaluating song segmentation. In *Proceedings of the 9th International Conference on Music Information Retrieval* (pp. 375–380).

McFee, B. & Ellis, D. P. (2014). Learning to segment songs with ordinal linear discriminant analysis. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 5197–5201).

Mendoza Garay, J. (2014). *Self-report measurement of segmentation, mimesis and perceived emotions in acousmatic electroacoustic music* (Master's thesis, University of Jyväskylä).

Müller, M., Chew, E., & Bello, J. P. (2016). Computational Music Structure Analysis (Dagstuhl Seminar 16092). *Dagstuhl Reports*, *6*(2), 147–190. doi:http://dx.doi.org/10.4230/DagRep.6.2.147

Narmour, E. (1992). *The analysis and cognition of melodic complexity: the implication-realization model*. University of Chicago Press.

Paulus, J. & Klapuri, A. (2006). Music structure analysis by finding repeated parts. In *Proceedings of the 1st ACM workshop on audio and music computing multimedia* (pp. 59–68).

Pauwels, J., Kaiser, F., & Peeters, G. (2013). Combining harmony-based and novelty-based approaches for structural segmentation. In *Ismir* (pp. 601–606).

Pearce, M. T. & Wiggins, G. (2006). The information dynamics of melodic boundary detection. In *Proceedings of the ninth international conference on music perception and cognition* (pp. 860–865).

Peeters, G. (2004). Deriving musical structures from signal analysis for music audio summary generation: "sequence" and "state" approach. *Computer Music Modeling and Retrieval*, 169–185.

Peiszer, E., Lidy, T., & Rauber, A. (2008). Automatic audio segmentation: segment boundary and structure detection in popular music. In *Proceedings of the 2nd International Workshop on Learning the Semantics of Audio Signals (LSAS)*. Paris, France.

Pyper, B. J. & Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(9), 2127–2140.

Serrà, J., Muller, M., Grosche, P., & Arcos, J. (2014). Unsupervised music structure annotation by time series structure features and segment similarity. *IEEE Transactions on Multimedia*, *16*(5), 1229–1240.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Boca Raton: CRC press.

Smith, J. B. & Chew, E. (2013). A meta-analysis of the mirex structure segmentation task. In *Proc. of the 14th International Society for Music Information Retrieval Conference* (Vol. 16, pp. 45–47). Curitiba.