Effects of Musicianship and Experimental Task on Perceptual Segmentation

Martin Hartmann,[*1] Olivier Lartillot,[#2] Petri Toiviainen[*3]

[*]*Department of Music, University of Jyväskylä, Finland*

[#]*Department of Architecture, Design and Media Technology, Aalborg University, Denmark*

[1]martin.hartmann@jyu.fi, [2]ol@create.aau.dk, [3]petri.toiviainen@jyu.fi

ABSTRACT

The perceptual structure of music is a fundamental issue in music psychology that can be systematically addressed via computational models. This study estimated the contribution of spectral, rhythmic and tonal descriptors for prediction of perceptual segmentation across stimuli. In a real-time task, 18 musicians and 18 non-musicians indicated perceived instants of significant change for six ongoing musical stimuli. In a second task, 18 musicians parsed the same stimuli using audio editing software to provide non-real-time segmentation annotations. We built computational models based on a non-linear fuzzy integration of basic and interaction descriptors of local musical novelty. We found that musicianship of listeners and segmentation task had an effect on model prediction rate, dimensionality and components. Changes in tonality and rhythm, as well as simultaneous change of these aspects were important to predict segmentation by listeners. Our results suggest that musicians pay attention to more features than non-musicians, including more high-level structure interactions. Prediction of non-real-time annotations involved more features, particularly interactions thereof, suggesting high context dependency. The role of interactions on perception of musical change has an impact on the study of neural, kinetic and speech stream processing.

Topic area: Musical structure, Cognitive modeling of music
Keywords: segmentation density, musical training, segmentation task, audio-based computational modeling

I. BACKGROUND

While listening to music, we spontaneously parse musical structure based on our perception of significant changes and repetitions. This dual process of grouping and segmenting music involves high-level cognitive functions such as memory, attention, and decision-making. Since music listening is a temporally unfolding process, real-time indications of musical boundaries are of great interest for music perception. However, the real-time perception of a succession of events may not guarantee a complete understanding of an underlying structure. Moreover, experience and musicianship in particular might guide our attention towards different characteristics of the musical stream. On top of that, the hierarchical grouping structure of music affords multiple levels for segmentation, such as notes, beats, motifs, phrases, melodies and sectional forms. In this study, we mainly investigate phrase-level musical boundaries, which are understood in this article as instants of significant change in the music. We aimed to systematically investigate the role of timbre, rhythm, and tonality on segmentation by musicians and non-musicians in different tasks. To this end, we proposed a method for polyphonic audio-based computational modeling of perceptual segmentation based on optimal musical feature subsets.

The tendency towards perceptual grouping of musical and other temporal sensory information into streams of events has been well studied. This Gestalt phenomenon has been of particular interest for auditory scene analysis (ASA) psychophysical models (Bregman, 1994), as well as within music theory, for melodic expectation models (Narmour, 1992) and generative theory of tonal music (GTTM) formal descriptions (Lerdahl & Jackendoff, 1983).

Within music cognition, MIDI-based data- and model-driven methods (Wiering, de Nooijer, Volk, and Tabachneck-Schijf, 2009) have been suggested for boundary prediction in score-based monophonic musical examples. Few works have been carried out on validation of segmentation systems and rules via music listening studies (Wiering et al., 2009; Bruderer, 2008; Frankland & Cohen, 2004; Clarke & Krumhansl, 1990; Peretz, 1989; Deliège, 1987). Changes in timbre and harmonic progression are melodic description cues that listeners frequently used to justify segmentation decisions (Bruderer, 2008). Also rhythmic attributes, particularly changes in note duration, have been found to be crucial in several melodic segmentation systems (Temperley, 2007). Complex musical changes combining grouping preference rules might also be important boundary candidates, as temporal pauses of melodies are more likely to be perceived as boundaries by both musicians and non-musicians when reinforced with other determinants such as musical parallelism (Peretz, 1989).

Within music information retrieval (MIR), a number of audio-based systems for segmentation have been evaluated against perceptual ground truth, usually for polyphonic popular music. Recent studies (McFee and Ellis, 2014; Nieto & Jehan 2013) focused mainly on timbre-based features and chromagram-based (Fujishima, 1999) tonal features. 'Repetition features' are often derived from these (McFee and Ellis, 2014), yielding good results for Western popular music. Rhythmic features such as fluctuation patterns (Pampalk, Rauber & Merkl, 2002) have also shown good results in this domain (Turnbull, Lanckriet, Pampalk & Goto 2007; Jensen, 2007).

In regards to algorithms for audio-based computational modeling, the novelty approach (Foote, 2000) is still considered state-of-the-art. It is based on the computation of a feature-based self-similarity matrix, which is convolved with a Gaussian Checkerboard kernel along the diagonal to obtain a novelty curve representing transitions characterized by high dissimilarity between neighboring feature frames.

Music perception studies showed some interesting trends regarding listeners and segmentation tasks. Several studies using naturalistic stimuli (Hartmann, Lartillot & Toiviainen, 2014; Bruderer, 2008; Deliège, 1987) reported no clear effects of musicianship on segmentation, although non-musicians tend to segment more often than musicians. Effects of data

collection task were however found, as listeners marked significantly fewer boundaries in real-time contexts than in offline annotation tasks (Hartmann et al., 2014). It was also found that the perceived strength ratings of a boundary relate to the number of participants that indicated it (Bruderer, 2008).

The effects of musical training on the prediction rate of computational segmentation models are still unclear. This should be studied to improve accuracy of computational models and gain further understanding on transfer effects of musicianship. The effects of segmentation task upon prediction rate of models are also unclear, although these should be studied to understand, for example, whether computational models are more comparable to real-time or to non-real-time segmentations. Moreover, the relative contribution of distinct musical attributes on segmentation and buildup of perceptual streams awaits clarification. The interaction between different acoustic features has not been studied either, although its potential for segmentation has been stated (Turnbull et al., 2007).

In addition, we lack systematic investigation of perceptual segmentation via audio-based computational models, which are crucial because audio target stimuli can increase the ecological validity of computational models and associated findings. Also, studies generally perform analyses based on segmentation data coming from small sample sizes (McFee & Ellis, 2014; Nieto & Jehan, 2013; Jenssen, 2007; Turnbull, 2007; Clarke & Krumhansl, 1990), but larger populations are needed to improve the external validity of results and implementations. A further limitation of previous segmentation studies is that they are often limited to classical or pop music and rarely include a variety of styles. This should be considered because it could offer a more general understanding of boundary perception and increase the impact of outcomes.

II. AIMS

This study focused on prediction of perceptual segmentation via audio-based computational models of spectral, rhythmic, and tonal change. Our main goal was to estimate the contribution of different musical features in the prediction of boundary density using six diverse stimuli. Additionally, we aimed to understand the effect of musicianship upon perceptual segmentation in real-time listening contexts. We also aimed to shed light on the effect of task (real-time vs. non-real-time) upon perceptual segmentation.

The main hypothesis of this study is that novelty-based computational models based on multiple musical features could accurately predict boundary density, at least for highly contrasting passages (e.g., simultaneous and stark changes in dynamics, instrumentation, and key). Since perceptual segmentation is multidimensional, novelty detection should increase prediction if interactions of musical features were aggregated. Another hypothesis is that tonal and other high-level features predict segmentation better for musicians than for non-musicians. Probably both groups pay attention to the musical surface (dynamics, texture, instrumentation, register, pace), but musicians might focus relatively more on harmonic and other deeper relationships. We also assumed that high-level features predict non-real-time segmentation better than segmentation in real-time contexts, due to incomplete understanding of the musical structure during segmentation of ongoing stimuli.

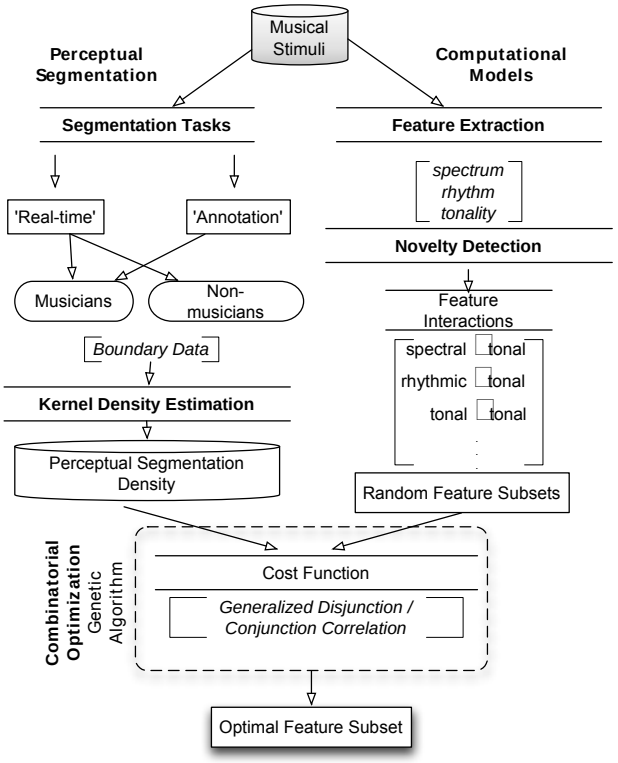

Figure 1. General design of the study.

Our approach was implemented via an assessment of model predictability for different groups, tasks, and conjoint features. We examined the predictability of boundary density using audio-based computational approaches and diverse stimuli. We aim to contribute to music perception and MIR literature via a systematic assessment of musical features for different perceptual data.

III. METHODS

We conducted two listening experiments to gather perceptual segmentation responses, and extracted musical features from the audio to computationally model the task. Figure 1 illustrates the approach described below and in the next section. The materials were six instrumental musical audio stimuli that were around two minutes in duration and of diverse styles (see Appendix). We chose these pieces because they are relatively unfamiliar and rather diverse; we searched for music whose segmentation would rely on multiple complex processes such as textural change and similarity instead of basic "Gestalt" boundaries (long inter-onset intervals, pitch jumps, etc.). For instance, some boundaries may be unexpected or perceived as blurry transition regions, delivering uncertainty and ambiguity.

1) Perceptual segmentation experiment

The subjects of the study were 36 participants, 18 of whom (11 males, 7 females) were musicians with an average of 14 years of training (SD = 7.49). All the musicians considered

themselves either to be semi-professional or professional musicians specialized in classical (12 participants) or other

Table 1. Correlations between segmentation density and basic features.

Type	Basic Feature	NMrt	Mrt	Ma	Maw
Spectral	Subband Flux	.16	.18	.13	.23
Rhythmic	Fluctuation Patterns	**.32**	**.25**	.29	**.36**
Tonal	Chromagram (1s)	.24	.22	**.32**	.29
	Chromagram (3s)	.25	.24	.30	.29
	Key Strength (1s)	.15	.13	.19	.21
	Key Strength (3s)	.12	.12	.17	.17
	Tonal Centroid (1s)	.17	.14	.22	.24
	Tonal Centroid (3s)	.14	.14	.23	.22

Table 2. Best correlations between segmentation density and feature interactions.

Segmentation	Feature Interaction	*r*
NMrt	Fluctuation Patterns ◦ Chromagram (3s)	**.37**
	Fluctuation Patterns ◦ Chromagram (1s)	.35
	Fluctuation Patterns ◦ Tonal Centroid (3s)	.32
Mrt	Subband Flux ◦ Chromagram (3s)	.29
	Fluctuation Patterns ◦ Chromagram (3s)	**.30**
	Fluctuation Patterns ◦ Chromagram (1s)	.29
Ma	Fluctuation Patterns ◦ Chromagram (1s)	.39
	Fluctuation Patterns ◦ Chromagram (3s)	.39
	Fluctuation Patterns ◦ Tonal Centroid (3s)	**.41**
Maw	Fluctuation Patterns ◦ Chromagram (1s)	**.45**
	Fluctuation Patterns ◦ Chromagram (3s)	.44
	Fluctuation Patterns ◦ Tonal Centroid (3s)	.44

(6 participants) styles. The remaining 18 participants (10 females, 8 males) reported being musically untrained, and none of them reported having skills in dance or sound engineering. The subjects were local or exchange students and graduates from the University of Jyväskylä and Jyväskylä University of Applied Sciences. The groups were matched in terms of their age distribution; the mean age was 27 years ($SD = 4.5$) for both musicians and non-musicians.

Two listening experiments were conducted to collect sets of boundary indications from 18 non-musicians and 18 musicians via different tasks.

2) Real-time task

Participants were asked to indicate instants of significant change as they listened to the music by pressing the space bar key of a computer. After reading instructions and completing a trial, they segmented each of the musical stimuli presented in randomized order. The listeners were requested to offer their "first impression" as they did not have a chance to listen to the whole stimulus beforehand or change their choice afterwards. The interface included a playbar that offered the beginning, current and end time position of the ongoing stimuli as visual-spatial cue. The real-time task segmentation density is abbreviated in Tables 1, 2, and 3 as NMrt for non-musicians and as Mrt for musicians.

3) Annotation task

We conducted a second experiment with the purpose of obtaining a more comprehensive and precise set of segmentations from participants. Audio editing skills were needed for this task, so we collected data only from musicians as they reported familiarity with this software. The same 18 musicians of the first task took part in this experiment, which we call Annotation task as it resembles structure annotation.

We collected boundaries and perceived boundary strength via an editing interface that allowed playback, marking, reposition, and labeling (Sonic Visualizer, see Cannam, Landone & Sandler, 2010). Participants were requested to listen to the complete stimulus, and at the same time mark instants of significant change over a waveform. The next step was to freely playback the music from desired time points and reposition or remove boundaries that were added by mistake. Finally, listeners were asked to mark the perceived strength of each boundary with a value between 1 (not strong at all) and 10 (very strong). We abbreviated segmentation density of the annotation task as Ma, and segmentation density considering perceived strength weights as Maw for Tables 1, 2, and 3.

4) Segmentation density

For each participant, we concatenated the obtained boundaries across all six stimuli in order to investigate general segmentation principles across stimuli. Subsequently, we constructed an estimate of the boundary indications within each task and group with normalized Kernel Density Estimation to obtain a smooth curve of boundary density over time. Following previous work (Bruderer 2008, Hartmann et al., 2014), we used a kernel width of 1.5 seconds. The upper plot of Figure 2 shows indications by non-musicians as rug marks and the perceptual segmentation density as a curve for *Aus Böhmens Hain und Flur* (B. Smetana).

A. Computational segmentation models

In addition, we obtained computational segmentation profiles via detection of novelty points over time based on local changes of 36 musical descriptors.

We extracted musical features describing spectral (Subband Flux, see Alluri & Toiviainen, 2010), rhythmic (Fluctuation Patterns, see Pampalk, et al., 2002) and tonal (Chromagram; Key Strength, see Krumhansl, 1990; Tonal Centroid, see Harte, Sandler & Gasser) attributes of the audio stimuli. We utilized conventional extraction parameters for rhythmic and spectral features (Fluctuation Patterns: 1 s and hop size of .1 s; Subband Flux: .025 s and hop size of .0125 s). As regards tonal features, we utilized two different window lengths to capture the chord-level (1 s, hop size .1 s) and model the tonal context (3 s, hop size .1 s). We computed novelty curves with a kernel of 16 s from these eight features to represent spectral, rhythmic and tonal dissimilarity over time. This kernel size was found to provide temporal smoothness comparable to the perceptual segmentation density.

Since we also focused on the interaction of musical features, we merged each pair of basic novelty curves to obtain all 28 possible combinations. Each interaction feature was computed as pairwise multiplication of two novelty curves, symbolized as ◦ and illustrated in Figure 1.

Table 3. Correlations between segmentation density and percentile-based computational models

	NMrt	Mrt	Ma	Maw
Subset	Fluct. Pat.	Sub. Flux	Fluct. Pat.	Fluct. Pat.
	Tonal Centr. (1s)	Fluct. Pat.	Tonal Centr. (1s)	Tonal Centr. (1s)
	Sub. Flux ◦ Fluct. Pat.	Sub. Flux ◦ Fluct. Pat.	Sub. Flux ◦ Chromag. (1s)	Sub. Flux ◦ Fluct. Pat.
	Sub. Flux ◦ Tonal Centr. (3s)	Sub. Flux ◦ Key Strength (3s)	Sub. Flux ◦ Chromag. (3s)	Sub. Flux ◦ Key Strength (3s)
	Fluct. Pat. ◦ Chromag. (3s)	Sub. Flux ◦ Tonal Centr. (3s)	Sub. Flux ◦ Tonal Centr. (3s)	Sub. Flux ◦ Tonal Centr. (3s)
		Fluct. Pat. ◦ Chromag. (1s)	Fluct. Pat. ◦ Chromag. (3s)	Fluct. Pat. ◦ Chromag. (1s)
		Fluct. Pat. ◦ Chromag. (3s)	Fluct. Pat. ◦ Tonal Centr. (1s)	Fluct. Pat. ◦ Chromag. (3s)
			Fluct. Pat. ◦ Tonal Centr. (3s)	
Type	Rhythmic	Spectral	Rhythmic	Rhythmic
	Tonal	Rhythmic	Tonal	Tonal
	Spectral ◦ Rhythmic	Spectral ◦ Rhythmic	Spectral ◦ Tonal (3x)	Spectral ◦ Rhythmic
	Spectral ◦ Tonal	Spectral ◦ Tonal (2x)	Rhythmic ◦ Tonal (3x)	Spectral ◦ Tonal (2x)
	Rhythmic ◦ Tonal	Rhythmic ◦ Tonal (2x)		Rhythmic ◦ Tonal (2x)
r	.41***	.38***	.44***	.52***

*****: $p < .001$

IV. RESULTS

The perceptual segmentations were compared with the novelty curves and also with computational segmentation models derived from optimal feature subsets.

A. Baseline: Perceptual segmentation vs. novelty

Each perceptual segmentation density curve was correlated with each basic novelty feature and each interaction feature. Table 1 shows the correlations for each of the eight basic novelty features; values in bold show the best correlation obtained for each perceptual segmentation density. Correlations ranged from weak to moderately low; the features yielding the highest similarity with perceptual segmentation density curves were rhythmic (Fluctuation Patterns) and tonal (Chromagram). Tonal features were better predictors in the Annotation task than in the Real-time task.

The highest correlations between segmentation density and interaction features are presented in Table 2. The three highest correlations obtained for each perceptual segmentation density are shown; the highest correlation for each segmentation density is indicated in bold font. The interaction features also exhibited weak to moderately low correlations, which peaked for rhythmic-tonal interactions regardless of the perceptual task.

B. Perceptual segmentation vs. multidimensional novelty

Next, we investigated how the perceptual data could be predicted using combinations of novelty curves. We deemed multiple regression to be inadequate for this purpose, since it would assume a constant contribution of each feature across stimuli and time. Therefore, we combined the novelty features via ranking-based aggregation. Roughly, our computational modelling approach involved obtaining a percentile across an optimal subset of novelty features for each time point.

1) Combining novelty curves

The used model is inspired by soft computing and describes musical change based on a flexible operation to aggregate features. The features are integrated using a percentile measure, which can be considered as a generalized conjunction/disjunction function (Dujmović, 2007). This can be understood as a 'majority voting' that is neither based on all the features nor on only one feature. For example, the 50th percentile across features will be high if at least half of the considered features exhibit high musical change. We found that the 50th percentile (median ordinal position) yielded computational segmentation models that provided the best fit to the perceptual segmentations.

2) Optimal feature subset via combinatorial optimization

Based on the correlation between perceptual and computational models, we selected an optimal feature subset to compute the aggregate feature. Due to the high number of possible feature combinations per perceptual segmentation (2^{36}), we used Genetic Algorithm optimization to find the optimal subset. The optimization cost function was initialized with random subsets of all 36 features and evaluated using correlation as criterion.

The middle plot of Figure 2 displays the optimal set of novelty features for non-musicians, and the respective aggregate feature.

We found that the feature aggregation method increased the prediction rate over the individual novelty features. Table 3 shows the best correlations found via the percentile-based computational model, and their p-values (obtained via Monte Carlo simulation). The correlations were moderately high, reaching $r = .52$ for the prediction of segmentation by musicians in the Annotation task (with strength weights). The lowest plot of Figure 2 compares the computational model with the perceptual segmentation density obtained for

non-musicians. Notably, Tables 1, 2, and 3 show increased computational model prediction rates for segmentation for non-musicians over musicians. Moreover, prediction rates are overall higher for the annotation task than for the real-time task.

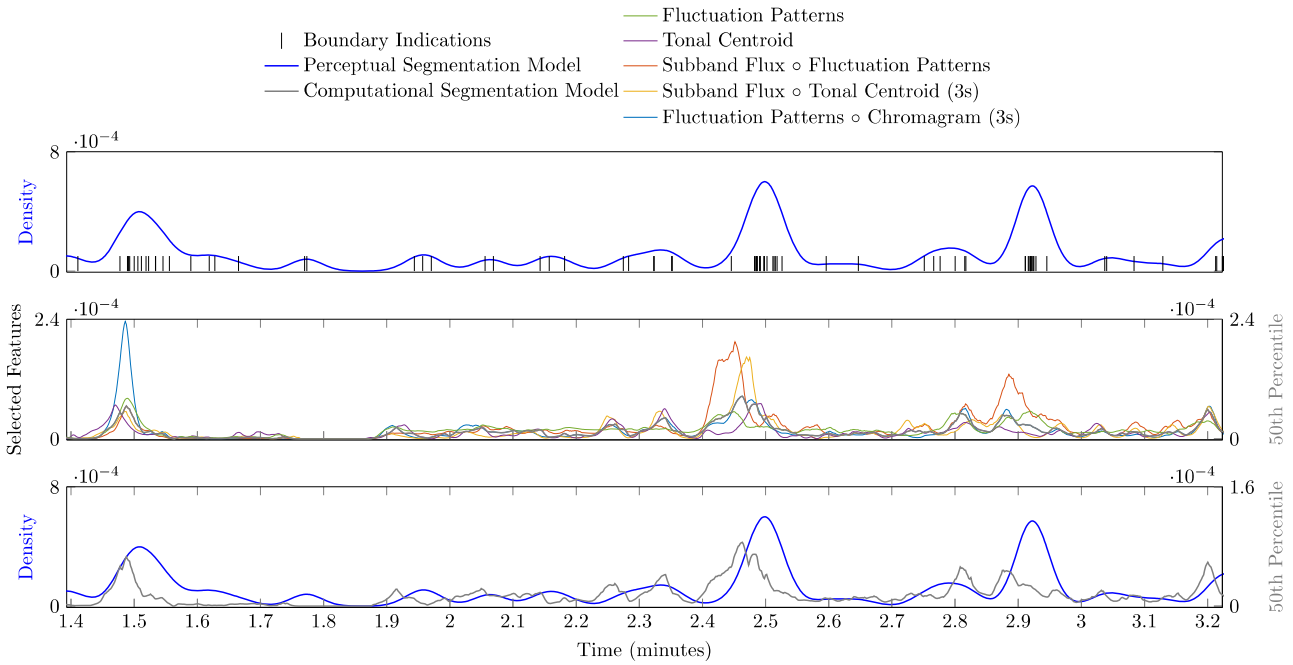

Figure 2. Perceptual segmentation density and computational segmentation model for non-musicians in the Real-time task (*Aus Böhmens Hain und Flur*, B. Smetana). Upper plot: Perceptual boundary data and segmentation density. Middle plot: Optimal feature subset and computational model. Lower plot: Segmentation density and computational model.

In regards to selected features, we found a general trend with rhythm and rhythmic-tonal interactions contributing to higher correlations. For both participant groups, rhythmic (Fluctuation Patterns) and rhythmic-tonal interactions (Fluctuation Patterns ∘ Chromagram 3s) were included in the optimal model. The computational model of the segmentation by musicians (Table 3), however, involved more features, especially feature interactions. For both segmentation tasks, rhythmic-tonal interactions as well as rhythmic and tonal basic features exhibited the highest correlations. The number of aggregated features, particularly feature interactions, was higher in the optimal computational model of the Annotation task (Table 3).

V. DISCUSSION

Our results indicate that, despite differences between groups and tasks, rhythm and tonality are the most important features in segmentation modeling. In particular, we found that spectral-tonal and rhythmic-tonal interactions were crucial for segmentation prediction. The role of high-level features in prediction via computational modeling increased both for musicians and for the annotation task.

One general finding is that the prediction rate of the computational models does depend on the musicianship level and segmentation task. The obtained correlations suggest that computational segmentation models can yield better prediction for non-musicians than for musicians. Perhaps this is because segmentation by musicians relies on more complex musical knowledge and involves conceptually driven processing. Moreover, the results show increased prediction of computational segmentation models for the Annotation task than for the Real-time task. One explanation for this could be that perceptual delays were corrected in the Annotation task since participants had the possibility to reposition their indications. Boundary density weighted with strength ratings further increased the prediction rate in the Annotation task, suggesting that the height of novelty peaks is predictive of the perceived salience.

We also found differences between groups and segmentation tasks in the size and composition of feature subsets selected for the optimal computational models. Our results show that more features were needed to predict musical change indicated by musicians, suggesting that they pay attention to more features. Compared to non-musicians, musicians followed a more complex pattern, as their optimal models were derived from more feature interactions. Since musicians relied on more interaction features, they might process musical structure with more emphasis on simultaneous change of multiple attributes. Interaction features can be considered high-level or structural features, because they represent simultaneous change in two dimensions. Previous findings (Hartmann et al., 2014; Bruderer, 2008; Deliège, 1987) showing fewer boundary indications by musicians than non-musicians are in the same vein, suggesting that musicians pay attention to higher levels of the structural hierarchy.

Comparing tasks, we found that the optimal models for the Annotation task are larger in feature subset size and more diverse in composition than for the Real-time task. Probably the Annotation task involved more high-level features because non-real-time contexts prompt deeper structure representations and include retrospective aspects of segmentation.

In regards to our proposed percentile-based computational model, it provided better prediction than correlation with

individual novelty features. The 'majority voting' logic described musical change as a trend across features, whose relative contribution varied over time and stimuli.

Our results expand previous evidence (Pearce and Wiggins, 2006) on the influence of harmonic, metrical, and rhythmic pattern changes on melodic boundary perception. We suggest the importance of simultaneous change of these aspects in phrase-level segmentation of polyphonic audio. Hence, chord boundaries that are isochronous with rhythmic or metrical pattern change might constitute important cues for boundary perception.

VI. CONCLUSIONS

This study focused on the contribution of spectral, rhythmic, and tonal features for prediction of segmentation using six diverse stimuli. Moreover, we estimated the effects of musicianship and task upon perceived segmentation of naturalistic stimuli in real-time and non-real-time listening contexts. Using a novel approach, we built computational segmentation models based on optimal subsets of basic and interaction of musical features. We found that simultaneous change in rhythmic patterns and tonal context had an important role in prediction of perceptual segmentation. More features, particularly high-level interactions, were important for prediction of segmentation by musicians compared to non-musicians. Similarly, optimal prediction of segmentation in a non-real-time task required more features, mainly high-level, than in a real-time task. Implications for music education include development of listening and expressive skills regarding simultaneous rhythmic and tonal changes. Our results also make an impact on digital music retail for music streaming services and on other applications such as audio software. Our bottom-up model, however, did not take into consideration top-down aspects, such as violations of musical expectation, and our focus on instants of change disregarded the contribution of other aspects of segmentation such as repetition. Another shortcoming is the lack of qualitative analysis of the stimuli, which would allow a better understanding of the segmentation process. We consider completing the block design in future work by collecting segmentation indications by non-musicians in the Annotation task. In regards to the stimuli, the repertoire is biased towards common practice piano music. It is expected that the outcomes of this study will encourage work in music perception and MIR on the contribution of high-level interaction for music segmentation.

ACKNOWLEDGMENT

The authors would like to thank Birgitta Burger and Emily Carlson. This work was financially supported by the Academy of Finland (project numbers 272250 and 274037).

REFERENCES

Alluri, V. and Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception*, 27(3): 223–241.

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, (17): 289–308.

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Bruderer, M. J. (2008). *Perception and Modeling of Segment Boundaries in Popular Music*. PhD thesis, JF Schouten School for User-System Interaction Research, Technische Universiteit Eindhoven, Netherlands.

Cannam, C., Landone, C., and 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*, pages 1467–1468, Firenze, Italy.

Clarke, E. and Krumhansl, C. (1990). Perceiving musical time. *Music Perception*, pages 213–251.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, 7(3): 325–359.

Dujmović, J. J. and Larsen, H. L. (2007). Generalized conjunction/disjunction. *International Journal of Approximate Reasoning*, 46(3):423–446.

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

Frankland, B. W. and Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music. *Music Perception*, 21(4):499–543.

Fujishima, T. (1999). Realtime chord recognition of musical sound: A system using Common Lisp Music, in *Proceedings of the International Computer Music Conference*, Beijing, China, 1999, pp. 464–467.

Harte, C., Sandler, M., and Gasser, M. (2006). Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pages 21–26. ACM.

Hartmann, M., Toiviainen, P., and Lartillot, O. (2014). Perception of segment boundaries in musicians and non-musicians. In Song, M. K., editor, *Proceedings of the ICMPC-APSCOM 2014 Joint Conference*, pages 165–170, Seoul, South Korea. College of Music, Yonsei University.

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

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*, volume 17. Oxford University Press New York.

Lerdahl, F. and Jackendoff, R. (1983). *A generative theory of tonal music*. The MIT Press, Cambridge, M.A.

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

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

Nieto, O. and Jehan, T. (2013). Convex non-negative matrix factorization for automatic music structure identification. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), pages 236–240.

Pampalk, E., Rauber, A., and Merkl, D. (2002). Content-based organization and visualization of music archives. In *Proceedings of the tenth ACM international conference on Multimedia*, pages 570–579. ACM.

Pearce, M. and Wiggins, G. (2006). The information dynamics of melodic boundary detection. In *Proceedings of the Ninth International Conference on Music Perception and Cognition*, pages 860–865.

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology*, 24(1-5):157–178.

Temperley, D. (2007). *Music and probability*. The MIT Press.

Turnbull, D., Lanckriet, G. R., Pampalk, E., and Goto, M. (2007). A supervised approach for detecting boundaries in music using difference features and boosting. In *Proceedings of the 5th International Conference on Music Information Retrieval* (ISMIR), pages 51–54.

Wiering, F., de Nooijer, J., Volk, A., and Tabachneck-Schijf, H. (2009). Cognition-based segmentation for music information retrieval systems. *Journal of New Music Research*, 38(2):139–154.

APPENDIX

Musical Stimuli

Banks, T., Collins, P. & Rutherford, M. (1986). The Brazilian. [Recorded by Genesis]. On Invisible Touch [CD]. Virgin Records. (1986).
Spotify link: ttp://open.spotify.com/track/7s4hAEJupZLpJEaOel5SwV
Excerpt: 01:10.200-02:58.143.

Smetana, B. (1875). Aus Böhmens Hain und Flur. [Recorded by Gewandhausorchester Leipzig - Václav Neumann]. On Smetana: Mein Vaterland [CD]. BC - Eterna Collection. (2002).
Spotify link: http://open.spotify.com/track/2115JFwiNvHxB6mJPkVtbp
Excerpt: 04:06.137-06:02.419.

Morton, F. (1915). Original Jelly Roll Blues. On The Piano Rolls [CD]. Nonesuch Records. (1997).
Spotify link: http://open.spotify.com/track/6XtCierLPd6qg9QLcbmj61
Excerpt: 0-02:00.104.

Ravel, M. (1901). Jeux d'Eau. [Recorded by Martha Argerich]. On Martha Argerich, The Collection, Vol. 1: The Solo Recordings [CD]. Deutsche Grammophon. (2008).
Spotify link: http://open.spotify.com/track/27oSfz8DKHs66IM12zejKf
Excerpt: 03:27.449-05:21.884

Couperin, F. (1717). Douzième Ordre / VIII. L'Atalante. [Recorded by Claudio Colombo]. On François Couperin : Les 27 Ordres pour piano, vol. 3 (Ordres 10-17) [CD]. Claudio Colombo. (2011).
Spotify link: http://open.spotify.com/track/6wJyTK8SJAmtqhcRnaIpKr
Excerpt: 0-02:00

Dvořák, A. (1878). Slavonic Dances, Op. 46 / Slavonic Dance No. 4 in F Major. [Recorded by Philharmonia Orchestra - Sir Andrew Davis]. On Andrew Davis Conducts Dvořák [CD]. Sony Music. (2012).
Spotify link: http://open.spotify.com/track/5xna3brB1AqGW7zEuoYks4
Excerpt: 00:57.964-03:23.145