Author(s): Hartmann, Martin; Toiviainen, Petri; Lartillot, Olivier

Title: Perception of Segment Boundaries in Musicians and Non-musicians

Year: 2014

Version:

Please cite the original version:

Hartmann, M., Toiviainen, P., & Lartillot, O. (2014). Perception of Segment Boundaries in Musicians and Non-musicians. In M. K. Song (Ed.), Proceedings of the ICMPC-APSCOM 2014 Joint Conference : 13th Biennial International Conference for Music Perception and Cognition : 5th Triennial Conference of the Asia Pacific Society for the Cognitive Sciences of Music (pp. 173-178). College of Music, Yonsei University. http://www.icmpc-apscom.org/

PERCEPTION OF SEGMENT BOUNDARIES IN MUSICIANS AND NON-MUSICIANS

Martín Hartmann [1], *Petri Toiviainen* [1], *and Olivier Lartillot* [2]

[1] Finnish Centre for Interdiscliplinary Music Research, University of Jyväskylä, Finland

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

ABSTRACT

In the act of music listening, many people break down musical pieces into chunks such as verses and choruses. Recent work on music segmentation has shown that highly agreed segment boundaries are also considered strong and are described by using multiple cues. However, these studies could not pinpoint the effects of data collection methods and of musicianship on boundary perception. Our study investigated the differences between segmentation tasks performed by musicians in real-time and non real-time listening contexts. Further, we assessed the effect of musical training on the perception of boundaries in real-time listening. We collected perceived boundaries by 18 musicians and 18 non-musicians in 9 musical examples. Musicians also completed a non real-time segmentation task for 6 of the examples. We observed high significant correlations between participant groups and between task groups at a time-scale of 10 seconds after comparing segmentation data at different resolutions. Further, musicians located significantly more boundaries in the non real-time task than in the real-time task for 5 out of 6 examples. We found a clear effect of the task but no effects of musical training upon perceived segmentation.

1. INTRODUCTION

Music listening often prompts people to spontaneously predict and detect relevant changes that demarcate the onset and offset of verses, choruses, and other parts. This skill permits musicians and dancers to break down rehearsals into logical chunks, and helps disc jockeys and music engineers to navigate through music audio files. We can describe segmentations or boundaries in a broad sense as contrasts, discontinuities, changes and repetitions (Addessi & Caterina, 2000); in this study we will specifically refer to segmentations or boundaries as *instants of significant change* in the music (Foote, 2000), and we will focus on the high level structure instead of on phrase level segmentation.

In general, people seem to share a common sense of the time locations at which musical changes become most significant, although some people systematically tend to segment more than others (Clarke & Krumhansl, 1990; Koniari, Predazzer & Mélen, 2001; Bruderer, 2008). Experimental studies related with musical boundary perception have tried to tackle the issue of how people segment music into either an unlimited or a fixed number of parts. Other issues that have been studied include how people justify their segmentations, judge their time position, and estimate their duration. To tackle these questions, perceived boundaries were compared to perceptual interpretations (Addessi & Caterina, 2000), grouping rules (Deliège, 1987; Clarke & Krumhansl, 1990; Frankland, McAdams & Cohen, 2004), cognitive and musicological theories (Peebles, 2011), and acoustic descriptions (Bailes & Dean, 2007). Further, some studies implemented automatic segmentation systems based on musical features (Hargreaves, Klapuri & Sandler, 2012; Smith, Chuan & Chew, 2013) or on sets of rules (Lartillot & Ayari, 2009; Pearce, Müllensiefen & Wiggins, 2010), and tested the performance of the systems against the perceptual ground truth. Other related work includes a study by Burunat, Alluri, Toiviainen, Numminen and Brattico (in press) in which a perceptual segmentation task was conducted to find musical triggers of working memory, whose time locations were compared with a functional magnetic resonance imaging (fMRI) dataset obtained from participants while listening to music.

Most of the studies that investigated the effect of musicianship upon boundary perception have approached this issue from the perspective of music theory. These studies focused mainly on the differences and similarities between musician and non-musician listeners in their perception of the musical structure. Deliège (1987) assessed the segmentation of short musical stimuli by music students and non-musicians, and found similar segmentation patterns between both groups. However, music students segmented significantly more in accordance with rules of the GTTM (Generative Theory of Tonal Music, see Lerdahl and Jackendoff, 1986) than non-musicians. Bruderer (2008) collected segmentation boundaries from two subsamples of around 6 musicians and 6 non-musicians each. The stimuli used for each subsample were complete examples of Western polyphonic music in MIDI and audio versions, respectively. One of his findings was that non-musicians marked significantly more boundaries than musicians.

There is no established listening experiment method for the collection of perceived musical boundaries. However, it is usual to present the stimulus to the participant in one or more “listening only” trials, and to afterwards collect segmentation responses as participants listen again to the stimulus. This is done to ensure that the participants are familiar with the structure of the stimulus before they mark the musical boundaries. With this respect, the GTTM postulates that a complete hierarchical mental representation is only achieved after the whole example has been heard (Koniari & Tsougras, 2012). For example, Clarke & Krumhansl (1990) asked participants to segment two complete piano pieces by initially listening the complete stimulus, then marking boundaries as they listened to the music, and finally making changes or deletions of the previously marked boundaries. Bruderer (2008) asked participants to first listen the complete example to ensure familiarity with the stimulus, and then to segment as they listened once again to the stimulus. The segmentation was performed three times in a row to obtain multiple trials from the same participant. He found that the number of marked boundaries remained similar across the segmentation trials. Also Deliège (1987) utilized a familiarization phase in the listening experiments, which was followed by a segmentation task. The segmentation was offline, in the sense that participants had to mark the boundaries only after listening to the complete stimulus. In later studies a different approach was utilized; Deliège, Mélen, Stammers and Cross (1996) asked non-musician participants, in an online task, to segment a 30-second piano piece as it was listened for the first time and found that they analyzed the music based mostly upon rhythmic and metric characteristics of the music and less upon harmonic functions. Burunat et al. (in press) asked musicians to segment large chunks of a contemporary tango piece (*Piazzolla*, which is used in the present study); each chunk was presented twice to the participants and in randomized order. Burunat has reported in a personal communication that there was a high within-subject consistency with respect to the indicated boundaries.

Our study contributes to the presented literature on some accounts; first, it introduces a *perceptual* data collection task for the comparison between musicians and non-musicians. We collected data in a real-time segmentation task that, in contrast to most of the previous studies, is not preceded by a "listening only" trial or by practice trials using the same stimuli. In this sense, it expands on the segmentation approaches by Deliège et al. (1996) and Burunat et al. (in press) since it incorporates segmentation by musicians and non-musicians, as well as diverse musical stimuli. The present study also proposes a *cognitive* data collection task; this task resembles the approach by Clarke and Krumhansl (1990) but replaces the tools utilized for fine-tuning the position of the marked boundaries. Instead of a score-based annotation, we obtained precise annotations using audio editing software, similarly to the approach by Wiering, de Nooijer, Volk and Tabachneck-Schijf (2009) for segmentation of rendered MIDI melodies. In short, our contribution to the state of the art with respect to the data collection task could help to understand what are the differences between an immediate hierarchical representation of the music and a perhaps more meaningful description obtained after a complete listen of the stimulus and the possibility to reconsider the location and marking of the boundaries. In addition, we chose a subsample that is almost half the size of the original sample because we think that the dividing line between musicians and non-musicians is fuzzy, and therefore we wanted to look at the extremes of the distribution. For example, we included in the subsample only those musicians who self-reported themselves as semiprofessional or professional musicians. Bruderer (2008) also utilized a subsample of the participants, but his findings are based only on relatively small groups of non-musicians and non-professional musicians. Other studies did not provide enough information about the musicianship levels of the subjects. Moving to a more technical viewpoint, our data analysis is based on Kernel Density Estimation (KDE, Silverman, 1986) matrices for data representation, extends the number and range of considered time-scale resolutions with respect to previous work and presents an alternative approach for the calculation of an optimal time-scale.

We believe that the study of music segmentation can deepen our knowledge of temporal processing of perceptual streams, which are found in music, speech and movement, and that the study of the effects of musicianship upon musical boundary perception can help us gain a better understanding of the possible transfer effects of music learning. In addition, systematic studies on the perception of instants of significant musical change can encourage developments in automatic segmentation tools to facilitate music editing and playback for, among others, everyday life tasks such as adding music to family videos.

People seem to intuitively understand music as a conglomerate with relatively precise boundaries. Further, many subjects may indicate musical boundaries at similar locations (Bruderer, Mckinney & Kohlrausch, 2009). There even seems to be a relationship between the number of subjects that assigned a boundary to a given time location and its rated salience (Bruderer, 2008). Musical boundaries cannot be arbitrary, and might instead emerge as a complex interplay between distinct musical structures that iterate or variate throughout a piece and our psychological mechanisms of perception and cognition (Deliege, 2007). There should be manifest and contrasting events in the music that prompt people to perceive beginnings and ends of musical segments. However, such events could differ in their number and time locations based upon multiple factors. For example, there can be differences in the musical training level of the listeners and in the data collection method that is used to gather segment boundaries. We primarily aimed to reach further insights regarding these two aforementioned factors. Our aims could be condensed into the following research questions:

1. What is the effect of musicianship on the perception of musical segment boundaries?

2. What are the differences between a spontaneous first impression of the musical structure as it unfolds over time and a deeper, more knowledge-driven impression?

We predicted that musicians would segment differently from non-musicians, perhaps because of differences in the perception of e.g. harmonic changes in the music. We also estimated that participants would segment differently in each task, since they would indicate more surface changes, such as rhythmic changes, in the perceptual (or real-time) task than in the cognitive task. We estimated the cognitive (or non real-time) task data to parse the musical structure based upon deeper changes.

2. EXPERIMENT I – PERCEPTUAL TASK

We conducted two listening experiments on perceived segmentation, where the first was a prerequisite for the second. In the first experiment, participants were asked to indicate significant musical boundaries at the same time as they listened to unfamiliar musical examples. The aim was to capture a fresh, "live" description or first impression of the music as it unfolded over time.

We collected real-time segmentation responses from participants using computers with a Max/MSP Patch. The stimuli comprised 9 musical pieces of a variety of Western musical styles. We included an Appendix to this paper with a glossary of the abbreviations used for the stimuli and with information about their duration. The music was played back to the participants using headphones at a comfortable volume level. We originally collected segmentation data from 74 participants, and later chose a final sample that comprised 18 non-musicians (11 males, 7 females) and 18 musicians (10 females, 8 males). The mean age of the participants was 27.45 years. They were all students or graduates from different faculties of the University of Jyväskylä and of the JAMK University of Applied Sciences. The musicians had an average of 14.39 years of musical training. All the non-musicians reported having had no musical training, whereas all of the selected musicians considered themselves either as semiprofessional or professional musicians at the time of the data collection.

The experiment took place with a computer in a sound-attenuated room. The participants were instructed to mark instants of significant change as they listened to the music by pressing the space bar of the computer keyboard. After completing a trial, they listened and marked each of the musical stimuli, which were presented in a randomized order. Participants were instructed to give their "first impression" because they would not have a chance to listen to the whole example before they started marking. The interface included a play bar that offered basic visual-spatial cues regarding the beginning, current time position and end of the examples. On average, it took 47 minutes to complete the whole experiment for the 18 non-musicians chosen for the study and 50 minutes for the 18 musicians. The participants filled a questionnaire at the end of the experiment, which included demographic and musicianship questions.

3. EXPERIMENT II – COGNITIVE TASK

The second experiment was conducted in order to obtain, from the musician participants, a segmentation that would be more comprehensive and precise than the first one. We prepared an interface in Sonic Visualiser to collect segmentation boundaries and perceived strength from musical examples. We intended to keep the duration of this experiment at around one hour, so we chose 6 examples from Experiment I that lasted around 2 minutes each for Experiment II. Headphones were used to playback the music at a comfortable listening level.

The final sample consisted of 18 musician participants (10 females, 8 males); they were selected among 36 subjects who completed the second experiment. All the participants had previously taken part of the first experiment. We did not recruit non-musicians for Experiment II because only a few of them reported experience in audio editing.

This experiment took also place in a sound-attenuated room with a computer. Exceptionally, five subjects participated at the same time in a sound-attenuated classroom with computers. Two of them were chosen for the final sample of 18 participants. Compared to Experiment I, the second experiment required the training to be completed while the experimenter was in the room. The experimenter read the instructions together with the participant and presented the interface. He asked the participant to perform the task upon two short trial stimuli by following the experiment instructions. Once the trial concluded, the experimenter left the room and the participant could start with the task. Participants were asked to: 1) Listen to the complete musical example; 2) Listen again to the complete example, and at the same time mark instants of significant change by pressing the Enter key; 3) Freely playback the musical example from different time points and correct marked positions to make them more precise, or remove them if these were added by mistake; they were also asked not to add any new markings at this stage; 4) Mark the strength of the significant change for each instant with a value ranging from 1 to 10; 5) Move to the next musical example and start over from the first step. The interface showed participants the waveform of the musical examples, over which they would play back and segment the stimuli, correct the boundaries and mark their strength. The participants were asked to focus on the music and not on the visual content. It took the 18 chosen participants one hour in average to complete the second experiment.

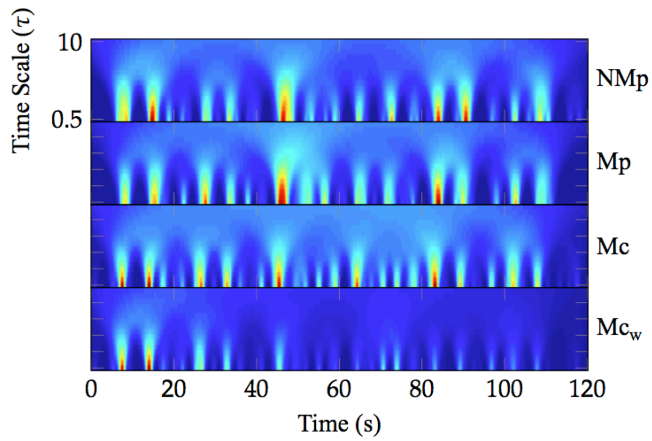

Figure 1: The segmentation sets of data marked by participants are visualized as multi-resolution KDE matrices for the musical example *Ragtime*. The density function over time is represented within each of the four matrices and for each time-scale of the KDE considered. Warm colors denote high values while cool colors denote low values.

4. RESULTS

We organized the segmentation responses into three main groups based on the level of musicianship of the participants and the corresponding segmentation task. We allocated 162 segmentations to the group of musicians in the perceptual task, 162 segmentations to the group of non-musicians in the perceptual task, and 108 segmentations to the group of musicians in the cognitive task. We abbreviated these groups as *NMp* for non-musicians in the perceptual task, *Mp* for musicians in the perceptual task, and

Mc for musicians in the cognitive task. We chose a method to visualize the segmentations that would summarize the segmentation data in a precise and hierarchical way. We added together the responses of each group and task using KDE's with different smoothing bandwidths. For each song, we obtained 16 KDE curves that were organized into matrices following a multi-resolution approach that has been previously utilized upon musical descriptors (Martorell Dominguez, 2013) and novelty curves (Kaiser & Peeters, 2013). Since we had collected ratings from participants on perceived boundary strength in the second experiment, we included a fourth group of segmentation responses. This group corresponded to the responses by musicians in the cognitive task with added boundary strength weights, and it is abbreviated as *Mcw*. The boundary indications in the fourth group were at the same time positions as in Mc; this would allow us to estimate the effect of adding perceived boundary strength to the cognitive segmentations. We obtained 30 KDE matrices, since the perceptual task data (NMp and Mp) was based on 9 stimuli and the cognitive task data (Mc and Mcw) was based on 6 stimuli. Figure 1 shows the 4 KDE matrices that were obtained for the example Ragtime. We constructed the matrices after obtaining KDE curves at different KDE time-scales (τ). We considered 16 time-scales logarithmically ranging from .5 seconds to 10 seconds, based upon the time-span of the working memory. We did not choose linearly spaced KDE time-scales since we followed the assumption, in agreement with Weber's law, that time is perceived on a logarithmic-like scale.

5. ANALYSIS

We found a higher mean number of indicated boundaries across participants in Mc (11.33) than in Mp (5.8) for these musical examples. We computed paired samples, two-tailed *t*-tests to estimate the significance level of the difference between Mp and Mc with respect to the number of indicated boundaries by musicians. The differences reached significance for 5 out of 6 examples: at $p < .01$ for the example *Couperin* (paired $t(17) = 3.3$, $p < .01$), and at $p < .05$ for the examples *Genesis* (paired $t(17) = 2.3$, $p < .05$), *Smetana* (paired $t(17) = 2.1$, $p < .05$), *Ragtime* (paired $t(17) = 2.8$, $p < .05$) and *Ravel* (paired $t(17) = 2.2$, $p < .05$).

To look further at the similarity between the segmentation groups, we calculated the correlation between pairs of KDE matrices for each stimulus. As shown in Figure 2, we found strong correlations for all stimuli between Mp and NMp, and these correlations were significant at $p < .001$ based on a Montecarlo simulation with 10000 iterations. Similarly, we found strong significant correlations ($p < .001$) between KDEs corresponding to the Mc and Mcw for all musical examples. In addition, we found moderately strong correlations between the KDEs of Mp and Mc, but these only reached statistical significance ($p < .001$) for the *Dvorak* stimulus. We also found moderately strong significant correlations ($p < .001$) between Mp and the Mcw for the same stimulus.

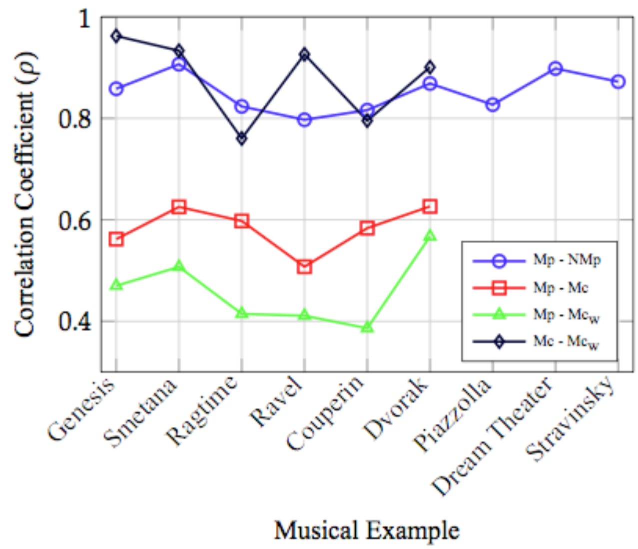

Figure 2: Profiles showing the correlation between multi-resolution KDE matrices.

We also compared, for each musical example, the KDE curves that were obtained with different bandwidths in order to find an optimal time-scale for segmentation responses. The KDE's were compared between pairs of sets of data using the identical smoothing parameters. The correlation between musicians and non-musicians for the perceptual task (Mp-NMp) was at least high for all the considered time-scales. In contrast, the comparisons between Mp and Mcw mostly exhibited moderate correlations. The comparisons between Mp and Mc indicated a general tendency to increase from low correlations at the lowest time-scales to high correlations at time-scales higher than 2 seconds. We also obtained very high correlations at all time-scales with a tendency to gradually decrease at higher time-scales for the comparison between Mc and Mcw. This decreasing tendency was much more important for *Couperin* and *Ragtime*, which correspond to solo piano performances.

We calculated a mean curve across all stimuli and considered the time-scale with the maximum correlation value of the mean curve as an optimal KDE time-scale for comparison between groups. We found a strong ($\rho = .89$, $t < .001$) overall maximal correlation coefficient for the comparison of Mp and NMp at an optimal time-scale of 10 seconds.

6. DISCUSSION

We obtained a notably higher mean number of perceived boundaries by musicians in the cognitive task than in the perceptual task. The participants marked more boundaries in the cognitive task for all six musical examples, and for five of them, the trend reached statistical significance, suggesting an effect of the data collection task upon the number of indicated boundaries. We could argue that the differences between task groups are related with the progressive familiarization with the stimuli, since the participants had already listened to the stimuli once in the perceptual task and they were asked to listen to the complete stimuli once again before the segmentation task of the cognitive experiment. This would give support to the idea that a thorough hierarchical mental representation of a musical piece can only be reached once it has been listened in its entirety (Koniari & Tsougras, 2012). Participants might have also noticed more boundaries in the perceptual task than those that they actually marked. Alternatively, some boundaries could have been perceived but left unmarked in order to avoid markings located after the occurrence of the boundary. It could have also been the case that the cognitive task gave margin to indicate boundaries after these were perceived since these could have been later repositioned to previous time instants.

In addition, we found strong and significant correlations between the cognitive task with and without the addition of boundary strength. We found however that *Couperin* and *Ragtime*, two solo piano performances, yielded lower correlations than the other examples. Musicians rated the boundaries of these examples with relatively low strength, which increased the difference between the cognitive task sets. These two musical examples are characterized by relatively less timbral contrasts than other musical examples, which probably prompted lower boundary strength ratings. Other musical examples have strong changes in harmony, dynamics and rhythm that are accompanied with percussion instruments or with changes in instrumentation. Even *Ravel*, which is not multi-instrumental but a piano piece instead, sounds very different from *Couperin* and *Ragtime* with its contrasting melodic passages and sudden changes in register, harmony, dynamics and rhythmic patterns. We also noticed that, for both *Ragtime* and *Couperin*, participants indicated relatively important strength in the beginning as the main themes were introduced, but the boundaries that they indicated for variations of these themes were rated with lower strength. We assume that this contrast between parts in the overall strength of the examples led to a lower correlation than if the strength markings had been more homogeneous.

We found a strong significant correlation between KDE matrices corresponding to segmentation by musicians and non-musicians for the perceptual task, contradicting previous findings by Bruderer (2008) who found effects of musical training upon the indication of musical boundaries. Another finding was that the optimal time-scale for comparison of segmentations by musicians and non-musicians corresponds to the maximal time-scale considered (10 s). This raises the question of whether the range of time-scales considered could be extended in order to find out if the optimal time-scale of the examples exceeds 10 seconds. We doubt, however, that the perception of segment boundaries would be appropriately represented if we computed KDEs at smoothing bandwidths that exceeded the temporal span of working memory.

7. CONCLUSIONS

We compared perceived segmentation between musicians and non-musicians in a real-time segmentation task. We did not find evidence of an effect of musical training upon music segmentation using the proposed approach, since we found similar segmentation profiles between both groups and a similar number of marked boundaries. We also found an overall maximal correlation between these two groups at a time-scale of 10 seconds. In addition, we found that musicians marked significantly more boundaries for the same stimuli in the cognitive segmentation task than in the perceptual segmentation task, which might be due to an increased familiarity with the stimuli or to other differences between the data collection tasks. Our results showed that relatively large time-scales, corresponding to a high-level hierarchy of the musical structure, are optimal for comparison of segmentation responses between musicians and non-musicians, and may be appropriate parameters for representation of perceived musical change. To gain more understanding on optimal KDE time-scales, further work could focus on alternatives to fixed smoothing bandwidths such as variable KDE estimation methods. Future work could focus on which specific boundaries are indicated in one task but not in the other one, and explore preliminary and final time positions of boundaries in the cognitive task. We will attempt to gain more insights regarding the issue of segmentation in future work by assessing the relationship between perceived segmentation and quantitative musical descriptions extracted from the stimuli.

8. REFERENCES

1. Addessi, A. R. and Caterina, R. (2000). Perceptual musical analysis: Segmentation and perception of tension. *Musicæ Scientiæ*, 4(1):31–54.
2. Bailes, F. and Dean, R. T. (2007). Facilitation and coherence between the dynamic and retrospective perception of segmentation in computer-generated music. *Empirical Musicology Review*, 2(3).
3. 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.
4. Bruderer, M. J., Mckinney, M. F., and Kohlrausch, A. (2009). The perception of structural boundaries in melody lines of western popular music. *Musicæ Scientiæ*, 13(2):273–313.
5. Burunat, I., Alluri, V., Toiviainen, P., Numminen, J., & Brattico, E. (in press). Dynamics of brain activity underlying working memory for music in a naturalistic condition. *Cortex*. DOI:10.1016/j.cortex.2014.04.012
6. Clarke, E. and Krumhansl, C. (1990). Perceiving musical time. *Music Perception*, pages 213–251.
7. Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception*, pages 325–359.
8. Deliège, I., Mélen, M., Stammers, D., and Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception*, pages 117–159.

9. Frankland, B. W., McAdams, S., 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): pp. 499–543.

10. Hargreaves, S., Klapuri, A., and Sandler, M. (2012). Structural segmentation of multitrack audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(10).

11. Kaiser, F. and 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.

12. Koniari, D., Predazzer, S., and M´elen, M. (2001). Categorization and schematization processes used in music perception by 10-to 11-year-old children. *Music Perception*, 18(3):297–324.

13. Koniari, D. and Tsougras, C. (2012). The cognition of grouping structure in real-time listening of music. a GTTM-based empirical research on 6 and 8-year-old children. In *Proceedings of the 12th International Conference on Music Perception and Cognition*, Thessaloniki, Greece.

14. Lartillot, O. and Ayari, M. (2009). Segmentation of Tunisian modal improvisation: Comparing listeners' responses with computational predictions. *Journal of New Music Research*, 38(2):117–127.

15. Lerdahl, F. and Jackendoff, R. (1986). *A generative theory of tonal music*. The MIT Press, Cambridge, M.A.

16. Martorell Dominguez, A. (2013). *Modelling Tonal Context Dynamics by Temporal Multi-Scale Analysis*. PhD thesis, Universitat Pompeu Fabra, Barcelona.

17. Pearce, M. T., Müllensiefen, D., and Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, 39(10):1367–1391.

18. Peebles, C. (2011). *The Role Of Segmentation And Expectation In The Perception Of Closure*. PhD thesis, Florida State University, Florida.

19. Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

20. Smith, J. B. L., Chuan, C., and Chew, E. (2013). Audio properties of perceived boundaries in music. *IEEE Transactions on Multimedia*.

21. 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.

9. APPENDIX – LIST OF STIMULI

Genesis – Banks, T., Collins, P. & Rutherford, M. (1986). The Brazilian. [Recorded by Genesis]. On *Invisible Touch* [CD]. Virgin Records. (1986) Excerpt: 01:10.200-02:58.143

Smetana – Smetana, B. (1875). Aus Böhmens Hain und Flur. [Recorded by Gewandhausorchester Leipzig - Václav Neumann]. On *Smetana: Mein Vaterland* [CD]. BC - Eterna Collection. (2002) Excerpt: 04:06.137-06:02.419

Ragtime – Morton, F. (1915). Original Jelly Roll Blues. On *The Piano Rolls* [CD]. Nonesuch Records. (1997) Excerpt: 0-02:00.104

Ravel – Ravel, M. (1901). Jeux d'Eau. [Recorded by Martha Argerich]. On *Martha Argerich, The Collection, Vol. 1: The Solo Recordings* [CD]. Deutsche Grammophon. (2008) Excerpt: 03:27.449-05:21.884

Couperin – 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) Excerpt: 0-02:00

Dvorak – 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) Excerpt: 00:57.964-03:23.145

The following examples were used for Experiment I only. We trimmed these ~8 minute examples into sections of ~2 minutes each for a more even length distribution across the pool of stimuli and to avoid fatigue to the participants. The sections were overlapped by 3 seconds, which corresponds to the duration of the echoic memory store. We later concatenated the segmentation data in order to obtain a set of indicated boundaries for the complete stimulus. We corrected the overlapping segmentation data by discarding the first 3 seconds of each non-initial chunk.

Piazzolla – Piazzolla, A. (1959). Adiós Nonino. [Recorded by Astor Piazzolla y su Sexteto]. On *The Lausanne Concert* [CD]. BMG Music. (1993) Excerpt: 0-08:07.968. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-08:07.968

Dream Theater – Petrucci, J., Myung, J., Rudess, J. & Portnoy, M. (2003). Stream of Consciousness (instrumental). [Recorded by Dream Theater]. On *Train of Thought* [CD]. Elektra Records. (2003) Excerpt: 0-07:50.979. Trimmed sections: 0-02:00, 01:57-03.57, 03:54-05:54, 05:51-07:50.979

Stravinsky – Stravinsky, I. (1947). The Rite of Spring (revised version for Orchestra) Part I: The Adoration of The Earth (Introduction, The Augurs of Spring: Dances of the Young Girls, Ritual of Abduction). [Recorded by Orchestra of the Kirov Opera, St. Petersburg - Valery Gergiev]. On *Stravinsky: The Rite of Spring / Scriabin: The Poem of Ecstasy* [CD]. Philips. (2001) Excerpts: 00:05-03:23, 0-03:12, 0-01:16 - total duration: 07:47.243. Trimmed sections: 00:05-02:05, 02:02-04:02, 03:59-05:59, 05:56-07:52.243