Experiments on the perception of segment boundaries in popular music

Michael Bruderer[1], Armin Kohlrausch[1,2], Martin McKinney[2]

[1] *Human-Technology Interaction, Technische Universiteit Eindhoven, P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands*

[2] *Philips Research Europe, High-Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands, Email: armin.kohlrausch@philips.com*

Abstract

Two experiments are described in which the temporal position and strength of segment boundaries in 6 popular music songs were measured. In a first experiment, listeners were asked to indicate segment boundaries by pressing a keyboard while listening to a song. All 21 subjects repeated this segmentation experiment three times. From the in total 63 trials, an overall perceptual boundary profile was constructed, revealing how often a boundary was indicated within a temporal window of 1.25 s. In a second experiment, about 15 to 20 boundaries were selected per song and subjects had to rate the perceived salience for each boundary on a seven-point scale. Furthermore, they had to describe the perceptual and musicological cues for each boundary. A comparison between the two experiments revealed a significant correlation between the height of a boundary in the perceptual profile and the salience rating. Furthermore, the rated salience was highly correlated with the number of terms used: the higher the salience, the more terms were used by the listeners to describe the boundary.

Introduction

This paper deals with the perception of musical segmentation, i.e., how a musical piece can be partitioned into its constituent entities. The segmentation process is possible through the organization of the musical events in the piece. The listener perceives the organisation through the presence of cues in the music which can be used to segment the music. The seminal work of Lerdahl and Jackendoff, "A generative theory of tonal music" (GTTM) [1], is a musicological theory for the analysis of structure in music. Part of GTTM is the set of "grouping preference rules", which indicate where segment boundaries should be placed in the score. Although these rules are musicological constructs, they have also been tested for their perceptual relevance [2,3,4]. All in all the results from these studies support the perceptual validity of the grouping preference rules proposed in GTTM. It is, however, not clear which rules, or combination of rules, have the strongest contribution to the perceived salience of a boundary.

The work described in this paper is part of a larger study [5] which investigated a number of open problems in the area of music structure perception and modeling. The work was partly inspired by applications in the area of music information retrieval and therefore, the music examples were taken from the popular music genre. Furthermore, different representations of the stimuli, ranging from melody lines rendered from the MIDI score to full audio, were compared for resulting boundary profiles. And finally, the experimentally determined boundary patterns were compared with the predictions of a number of formal musicological models. In the present paper, the focus is on some methodological aspects of the segmentation experiments.

Segmentation experiment

The goal of this experiment was to measure, where in a song subjects perceive segment boundaries. The stimuli consisted of the melody lines of six popular music songs. Subjects were asked to listen to and segment the songs into smaller pieces by pressing the space bar on the computer keyboard each time they perceived a boundary. The exact task description was: "Please press the space bar when you hear a segment boundary (phrase, section, passage)". When subjects did not understand the task based on this description, they were asked to divide the song into smaller meaningful pieces.

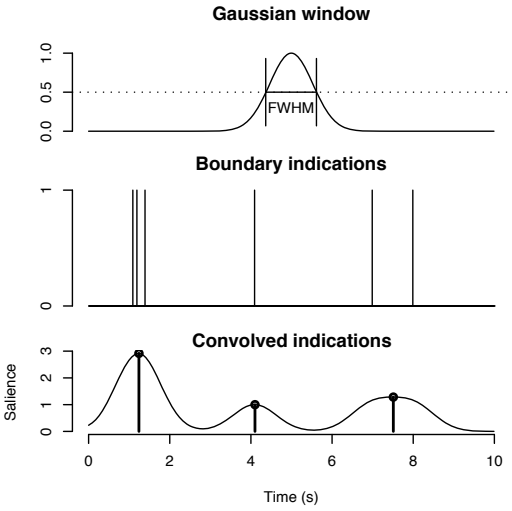

Figure 1: Visual representation of the Gaussian smoothing. The top row shows the Gaussian function with the full width at half maximum of 1.25 s, the second row some examples of indicated boundaries, and the third row shows the indicated boundaries convolved with the Gaussian function resulting in the smoothed boundary profile with the relative maxima, which are taken as the time points of a boundary.

The six songs were presented to the subjects in random order and they listened to each song four times in a row. Initially, subjects listened to a given song without pressing any key to get familiar with the piece, and in the following three presentations they were asked to indicate segment boundaries by pressing the space bar on

the computer keyboard. Subjects could not pause the playback of the song during listening and there was no visual information about the song (i.e., no score or timeline) provided to the subjects.

We calculated a perceptual segmentation profile by converting the indicated boundary time stamps for each trial to a pulse train, and then smoothing the pulse train by convolving it with a Gaussian window resulting in a smoothed boundary profile. The procedure of smoothing is shown in Figure 1. The temporal positions and heights of the maxima in the smoothed boundary profile were taken as the accumulated boundary indications. The main findings of this experiment were that there is a variety in the number of times a boundary was indicated across subjects. Some boundaries were indicated by nearly all subjects and others only once by few subjects. Furthermore, different subjects have different segmentation patterns for boundaries with an intermediate salience. These results were found consistently for all six songs.

Salience rating experiment

In previous studies, it has been assumed that the number of subjects indicating a specific boundary represents a measure of the salience of that boundary [e.g., 2,4]. This implicit measure of a boundary salience, however, is not necessarily correlated with the *perceived* boundary salience. For this reason subjects were asked in the second experiment to explicitly rate the salience of a subset of boundaries. The resulting boundary salience ratings were then correlated with the boundary profiles of the segmentation experiment to see how well the implicit measure of salience corresponds to the explicit measure.

The task description of the salience rating experiment was: "In the second experiment you are asked to rate the salience of the given boundaries and write down the cues that make it a boundary." Subjects rated the boundary salience on a seven point scale labeled from "no boundary" (0) to "very strong boundary" (6). For each boundary, subjects were asked to write down the cues in the musical signal responsible for the boundary. The selection of the boundaries for the salience rating experiment was based on the boundary profiles from the segmentation experiment. The selection included all of the strong boundaries, i.e., the 10 % of all indicated boundaries with the highest number of indications, and two to three moderate as well as two to three lower boundaries.

We were particularly interested in the comparison between the number of boundary indications and the salience rating because if these two measures are correlated, the number of indications within a time window can be directly used as a measure of the boundary salience. Figure 2 shows the relation between these two measures for all boundaries included in the salience rating experiment. The different symbols indicate the six different songs. The figure shows that the number of boundaries within the optimal window size is indeed correlated with the mean boundary salience rating. The correlation between these two measures was r=0.68 (p<0.001).

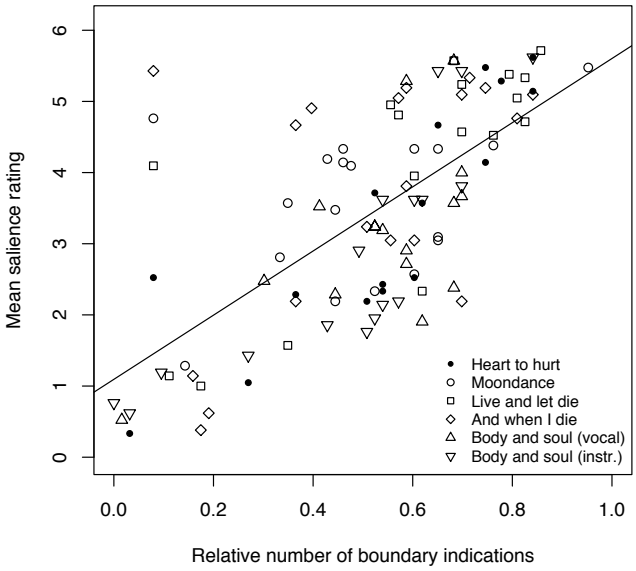

Figure 2: The relative number of boundary indications from the segmentation experiment on the abscissa and the mean salience rating of the selected boundaries from the salience rating experiment on the ordinate. The line is the linear regression of the mean salience rating and the relative number of boundary indications across all songs.

The terms used to describe the perceptual cues were classified into a number of cue classes (see [5]), which showed that the most often used terms came from the class "timbre change". The use of terms differed between songs and reflected differences in their composition, e.g. the presence or absence of tempo and rhythm changes. Furthermore, there was a high correlation between the salience rating of a boundary and the number of terms used to describe the boundary: the higher the salience rating, the more terms were used in the description.

In summary the experiments revealed a number of different and correlated measures which indicate the salience of perceived boundaries.

References

[1] Lerdahl, F., and Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge, MA, 1983.

[2] Deliège, I.: Grouping conditions in listening to music. Music Perception 4 (1987), 325-360.

[3] Clarke, E. F., and Krumhansl, C. L.: Perceiving musical time. Music Perception 7 (1990), 213-252.

[4] Frankland, B. W., and Cohen, A. J.: 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 (2004), 499-543.

[5] Bruderer, M. J.: Perception and modeling of segment boundaries in popular music. Ph.D. thesis, TU Eindhoven, 2008.