Chapter 13
On the Inherent Segment Length in Music

Kristoffer Jensen
Aalborg University Esbjerg, Denmark

ABSTRACT
In this work, automatic segmentation is done using different original representations of music, corresponding to rhythm, chroma and timbre, and by calculating a shortest path through the selfsimilarity calculated from each time/feature representation. By varying the cost of inserting new segments, shorter segments, corresponding to grouping, or longer, corresponding to form, can be recognized. Each segmentation scale quality is analyzed through the use of the mean silhouette value. This permits automatic segmentation on different time scales and it gives indication on the inherent segment sizes in the music analyzed. Different methods are employed to verify the quality of the inherent segment sizes, by comparing them to the literature (grouping, chunks), by comparing them among themselves, and by measuring the strength of the inherent segment sizes.

INTRODUCTION
Music consists of sounds organized in time. These sounds can be understood from a rhythmic, timbral, or harmonic point of view, and they can be understood on different time scales, going from the very short (note onsets) to the medium (grouping), to the large scale with musical form. Note onsets, grouping and form are common musical terms, which can be compared to different aspects of audition, memory and grouping behavior. These terms can be compared to chunks, riffs, and other temporal segmentation terms currently used in music.

When identifying chunks, riffs, sections, forms, or other structural elements, do they really exist, or does the identification process create them? This work presents a method, based on automatic segmentation, that identifies the inherent structure sizes in music, i.e. gives indications as to what are the optimal segmentation sizes in the music. This work has implications for rhythmical and
classical music understanding, and processing. Structure is a necessary dimension in most, if not all music, and if this structure should be made visible for any purpose, the methods presented here can help identifying the optimal structure. While this fundamental research gives a method for finding the optimal segment size in music, and results using this method, more work is needed in order to assess the inherent structure with certainty for all music. Until then, research and development of automatic segmentation of music should possibly ascertain the inherent structure in the music genres that is the aim of the work, prior to performing the segmentation.

Any feature, that can be calculated from the acoustics of the music, can be presented in a manner, for instance by taking the time-derivative, so as to give indication of the local changes in the music. Such an existence of a local change is not a guarantee of an inherent structure, however. In order to assess the quality of the segmentation, the relative distance (or any measure of similarity) within a segment should be compared to the distance to the other segments. If the segment is well grouped, and far, in some sense, to the other segments, then it is a good segmentation. A method for assessing the segmentation is the silhouette (Kaufman & Rousseeuw 1990). Given a segmentation, the mean of the silhouette value for all segments is a good measure of the quality of the segmentation. Therefore, if all possible segmentations are calculated, the associated mean silhouette values can be used to ascertain the best, i.e. the inherent structure sizes.

As to the question of which feature is used for temporal perception of music, Scheirer (1998) determined in several analysis by synthesis experiments that rhythm could not be perceived by amplitude alone, but needed some frequency dependent information, which he constructed using six band-pass filters. Several other studies have investigated the influence of timbre on structure. McAuley & Ayala (2002) found that timbre did not affect the recognition of familiar melodies, but that it had importance enough to hurt recognition on non-familiar melodies. McAdams (2002) studied contemporary and tonal music, and found that the orchestration affects the perceived similarity of musical segments strongly in some cases. He also found that musically trained listeners find structure through surface features (linked to the instrumentation) whereas untrained listeners focused on more abstract features (melodic contour, rhythm).

Deliège and Mélen (1997) postulates that music is segmented into sections of varying length using cue abstraction mechanism, and the principle of sameness and difference, and that the organization of the segmentation, reiterated at different hierarchical levels, permits the structure to be grasped. The cues (essentially motifs in classical music, and acoustic, instrumental, or temporal otherwise) act as reference points during long time spans. Deliège and Mélen furthermore illustrate this cue abstraction process through several experiments, finding, among other things, that musicians are more sensitive to structural functions, and that the structuring process is used for remembering, in particular, the first and last segment. In order to ensure that at least part of the full dimensionality of music is taken into account in the work presented here, three different features are used. One feature is believed to be related to tempo and rhythm, and it is called the rhythmogram. Another feature is considered related to the timbre perception, at least the time-varying perceptive spectrum, and it is called the timbregram. Finally, another feature is related to the note values in the music, and it is called chromagram. By using three features with distinctly different content, it is the aim to further assess the results on inherent and optimal segment size presented here.

Segmentation of music is often done for thumbnailing (music summary) purposes. This is supposedly a means for presenting music, prior to selling it, for instance in online stores. Other uses of segmentation are artistic, for instance for live mixing of music, for faster navigation,
Related Content

**Computationally Assessing Expert Judgments of Freewriting Quality**
Jennifer L. Weston, Scott A. Crossley and Danielle S. McNamara (2012). *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 365-382).
[www.igi-global.com/chapter/computationally-assessing-expert-judgments-freewriting/61059?camid=4v1a](www.igi-global.com/chapter/computationally-assessing-expert-judgments-freewriting/61059?camid=4v1a)

**Sentiment Recognition from Bangla Text**
[www.igi-global.com/chapter/sentiment-recognition-bangla-text/78481?camid=4v1a](www.igi-global.com/chapter/sentiment-recognition-bangla-text/78481?camid=4v1a)

**Digital Filters for Digital Audio Effects**
[www.igi-global.com/chapter/digital-filters-digital-audio-effects/4682?camid=4v1a](www.igi-global.com/chapter/digital-filters-digital-audio-effects/4682?camid=4v1a)

**Spread Spectrum for Digital Audio Watermarking**
[www.igi-global.com/chapter/spread-spectrum-digital-audio-watermarking/8325?camid=4v1a](www.igi-global.com/chapter/spread-spectrum-digital-audio-watermarking/8325?camid=4v1a)