

Human perception and computer extraction of musical beat strength

By Tzanetakis, Essl & Cook

Automatic extraction of rhythmic information of a music signals, is very important for modeling and analyzing and generally, working with that kind of signals. The focus of this paper is on beat strength which is considered as rhythmic characteristic that can discriminate between two pieces of music with the same tempo.

In general, main beat of a music piece is when a human listener, who is musically trained and is listening to the music, would tap his/her foot. For the purpose of automatic beat detection algorithms, the beat is characterized by its frequency (tempo), phase (accent locations), and a confidence measure about the detection. These algorithms are classified into two categories: event-based and self-similarity-based. In the first category, transient events are detected and their InterOnset Arrival Intervals (IOI) is used to extract the main tempo. In self-similarity algorithms, the periodicity of amplitude envelopes is used for detecting the tempo.

Extracting the main beat of the music signal, is a way of simplifying the problem when working with music signals. A defining characteristic of a musical signal is its structure which is hierarchical periodic at temporal levels and helps to catch the rhythm.

A pilot user study was conducted to see how human subjects agree in beat strength of different pieces, and what characteristics affect their decision, and whether their performance can be approximated by an automatic algorithm. In this experiment 32 subjects participated: 16 undergraduate students from different majors, 15 graduate students with formal musical training, either in Computer Science or Music departments, and one professional adult. Subjects were given 50 15-seconds excerpts, of different music styles to categorize based on their beat strength. Two forms of presentation were used, audio CDs with 50 tracks and web pages containing 50 links to audio quality music files. The excerpts were randomly distributed and ordered in the CDs and the web pages and were identified by a code containing subjects' code and the CD/web page code.

No definition of the beat strength was given to the subjects but they were asked to find the excerpt with the strongest beat and also to categorize the excerpts into the 5 categories. They were asked to first listen to all of the pieces to become familiar with the range of the beats and then in the second pass categorize the excerpts.

The results showed that the presentation type didn't have any effects. They also indicated that there is a significant agreement on the beat strength judgments, and somehow close to the authors categorizations. The average standard deviation across subjects is 1.25. The results showed that the subjects agree more on the strong range of the beats and show greater variability on the weak side.

Since the user study showed there are significant agreements in beat strength judgments, it makes sense to develop an automatic algorithm to do the same. Calculation of the beat strength is based on beat histogram (BT) which is a global representation of

Engineering Approaches to Music Perception & Cognition

Baharak Zali
8679928115

Homework # 3
Feb. 3, 2005

rhythmic information and it developed for music retrieval and automatic genre classification. BH contains statistical information about the amplitude envelope periodicities of multiple frequency bands. For BH calculations, Discrete Wavelet Transformation (DWT) is applied in a window. This window is advanced by a larger window to capture the signal repetition at the beat and sub-beat levels. The final histogram has bins corresponding to tempos in beats per minute and the amplitude of each bin corresponds to the strength of repetition of the amplitude envelopes of each channel for that particular tempo.

Next, two measures of beat strength derived from BH were tested. The first measure is sum of all histogram bins (SUM). This measure indicates how strong the self-similarity of the signal is at different tempos. The second measure is the ratio of the highest peak of the BH to the average amplitude (PEAK) and shows how dominant the main beat is.

For the purpose of comparing the performance of these measures with the results of the user study, the average beat strength of the excerpts across the subjects were calculated and considered as a base for evaluating the algorithm's performance (ground truth). The comparison was done by taking the absolute difference of the subjects' results from the automatically assigned value calculated for each excerpt. The average difference is 1.12 for the SUM measure and 1.08 for the PEAK measure. To get an idea about the comparison, the average absolute difference is approximately 3.0 for random assignment and is 0.77 for the original category assignment performed by the authors.