HUMAN PERCEPTION AND COMPUTER EXTRACTION OF BEAT STRENGTH

George Tzanetakis, Georg Essl

Computer Science Department
Princeton University
gtzan@cs.princeton.edu

Perry Cook

Computer Science and Music Department
Princeton University
gessl@cs.princeton.edu

ABSTRACT

Musical signals exhibit periodic temporal structure that creates the sensation of rhythm. In order to model, analyze, and retrieve musical signals it is important to automatically extract rhythmic information. To somewhat simplify the problem, automatic algorithms typically only extract information about the main beat of the signal which can be loosely defined as the regular periodic sequence of pulses corresponding to where a human would tap his foot while listening to the music. In these algorithms, the beat is characterized by its frequency (tempo), phase (accent locations) and a confidence measure about its detection.

The main focus of this paper is the concept of *Beat Strength*, which will be loosely defined as the rhythmic characteristics that allow to discriminate between two pieces of music having the same tempo. Using this definition, we can say that a piece of Hard Rock has a stronger sense of rhythm than a Classical Music at the same tempo. Characteristics related to *Beat Strength* have been implicitely used in automatic beat detection algorithms and shown to be as important as tempo information for music classification and retrieval. In this work, a user study exploring the perception of *Beat Strength* was conducted and the results were used to calibrate and explore automatic *Beat Strength* measures based on the calculation of *Beat Histograms*.

1. INTRODUCTION

The increasing amounts of processing power and digitally-available music enable the creation of novel algorithms and tools for structuring and interacting with large collections of music. Using techniques from Signal Processing and Machine Learning, computer audition algorithms extract information from audio signals in order to create representations that can subsequently used to organize and retrieve audio signals. A defining characteristic of musical signals, compared to other audio signals such as speech signals, is their hierarchical periodic structure at multiple temporal levels that gives rise to the perception of rhythm. Therefore, rhythmic information is an important part of any music representation used for music information retrieval (MIR) purposes.

Most automatic systems that attempt to extract rhythmic information from audio signals concentrate on the detection of the main beat of the music. Extracting rhythmic information from arbitrary audio signals is difficult as there is no explicitly available information about the individual note events as is the case in symbolic music representations such as MIDI. The main beat can be loosely defined as the regular periodic sequence of pulses corresponding to where a human would tap his foot while listening to the music. In automatic beat detection algorithms, the beat is characterized by its frequency (tempo), phase (accent locations) and a confidence measure about its detection. Some representative examples of such systems for audio signals are: [1, 2, 3, 4].

The main focus of this paper is the concept of *Beat Strength*, which is loosely defined as a one-dimensional rhythmic characteristic that allows us to discriminate between two pieces of music having the same tempo. Using this definition, we can say that a piece of Hard Rock has a stronger sense of rhythm than a Classical Music at the same tempo. Characteristics related to *Beat Strength* have been implicitly used in automatic beat detection algorithms and shown to be as important as tempo information for music classification and retrieval [5]. In this work, a user study exploring the perception of *Beat Strength* was conducted and the results were used to calibrate and explore automatic *Beat Strength* measures based on the calculation of *Beat Histograms*, which are a global representation of musical rhythm described in [5].

2. USER EXPERIMENTS

Although the concept of *Beat Strength* seems intuitive and has been shown to be useful for music information retrieval, to the best of our knowledge there has been no detailed investigation of its characteristics and perception by humans. A pilot user study was conducted with the goal of answering questions such as: how much do human subjects agree in judgements of *Beat Strength*, what characteristics of rhythm are important for these judgements, and if the human subject performance can be approximated using automatic music analysis algorithms.

2.1. Setup

The number of subjects used in the study was 32 and they were mostly undergraduate and graduate students at Princeton University. They were asked to assign 50 musical excerpts (each 15 seconds long) to 5 *Beat Strength* bins (Weak, Medium Weak, Medium, Medium Strong, Strong). A variety of different musical styles were represented. The excerpts were also preclassified into bins by the authors, to ensure that an even spread of *Beat Strength* (of course that information was not given to the subjects). The order of presentation was randomized for each subject to avoid learning order artifacts in the results. The main instructions given were: *The purpose of this study is to collect data on what attributes of songs make them seem to have a strong or weaker beat ... There are no right or wrong answers.* No definition of *Beat Strength* was provided as the purpose of the study was to determine the rhythmic attributes that correspond to the everyday verbal use of the term without biasing the results.

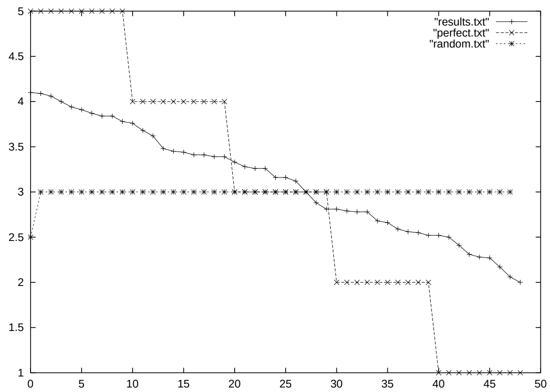

Figure 1: *Beat Study Subject Agreement.*

2.2. Experimental Results

The results indicate that there is significant subject agreement about *Beat Strength* judgements. Figure 1 shows the average bin across subjects for each musical excerpt and compares it to random agreement (the flat line at 3.0) and perfect agreement (the staircase function). The average bin standard deviation across subjects is 1.25 bins. More detailed experimental analyses will be provided in the full version of this paper.

3. AUTOMATIC BEAT STRENGTH EXTRACTION

The results of the user study indicate that there is significant subject agreement in *Beat Strength* judgements. Therefore it makes sense to try to develop automatic algorithms to extract these beat attributes from music signals in audio format and use them for music information retrieval purposes.

3.1. Beat Histogram Calculation

The calculation of *Beat Strength* measures is based on *Beat Histograms* (BH) a global representation of rhythmic information developed for the purposes of music retrieval and automatic musical genre classification [5]. Their calculation is based on a Discrete Wavelet Transform (DWT) filterbank as the the front-end, followed by multiple channel envelope extraction and periodicity detection shown schematically in Figure 2. The resulting histogram has bins corresponding to tempos in beats per minute (bpm) and the amplitude of each bin corresponds to the strength of repetition of the amplitude envelopes of each channel for that particular tempo.

3.2. Beat Strength Measures

Two measures of *Beat Strength* derived from the BH were explored. The first measure is the sum of all histogram bins (SUM). Because of the autocorrelation calculation used for periodicity detection in the BH this measure indicates how strong the self similarity of the signal is at various tempos. The second measure is the ratio of the amplitude of the highest peak of the BH to the average amplitude (PEAK) and indicates how dominant the main beat is.

In order to compare the performance of these measure with the user study results, the excerpts were sorted according to average subject bin and reassigned bins from 1 to 5 by equal division. The resulting bin assignment was then used as ground truth and compared with the bins assigned by sorting and equal division of the two *Beat Strength* measures. The comparison was done by taking the absolute difference of the ground truth bin from the automatically-assigned bin assigned for each excerpt. The average absolute bin difference is 1.12 for the SUM measure and 1.08 for the PEAK measure. For comparison the average absolute bin difference is approximately 3.0 for random assignment and is 0.77 for the original bin assignment performed by the authors. More detailed experimental analyses, descriptions of the algorithms and directions for future work will be provided in the full version of this paper.

To conclude there appears to be significant agreement about the concept of *Beat Strength* between human subjects and their performance can be approximated automatically.

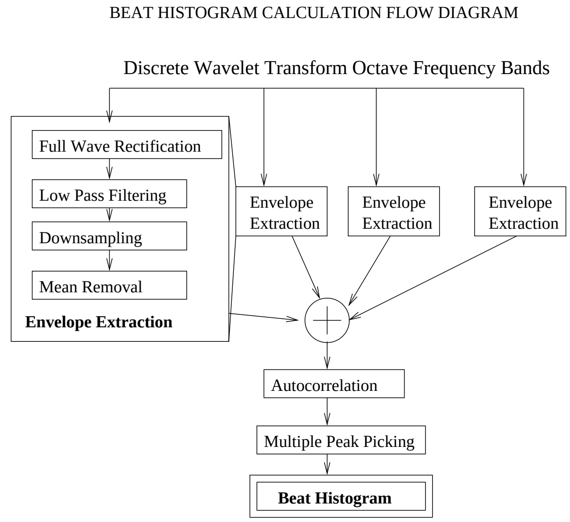

Figure 2: *Beat Histogram Calculation.*

4. REFERENCES

[1] Masataka Goto and Yoichi Muraoka, "Music Understanding at the Beat Level: Real-time Beat Tracking of Audio Signals," in *Computational Auditory Scene Analysis*, David Rosenthal and Hiroshi Okuno, Eds., pp. 157–176. Lawrence Erlbaum Associates, 1998.

[2] Eric Scheirer, "Tempo and beat analysis of acoustic musical signals," *Journal of the .Acoustical Society of America*, vol. 103, no. 1, pp. 588,601, Jan. 1998.

[3] Jean Laroche, "Estimating Tempo, Swing and Beat Locations in Audio Recordings," in *Proc. Int. Workshop on applications of Signal Processing to Audio and Acoustics WASPAA*, Mohonk, NY, 2001, IEEE, pp. 135–139.

[4] Jarno Seppänen, "Quantum Grid Analysis of Musical Signals," in *Proc. Int. Workshop on applications of Signal Processing to Audio and Acoustics WASPAA*, Mohonk, NY, 2001, IEEE, pp. 131–135.

[5] George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, 2002, (accepted for publication).