HUMAN PERCEPTION AND COMPUTER EXTRACTION OF MUSICAL BEAT STRENGTH

George Tzanetakis, Georg Essl

Computer Science, CISE Department
Carnegie Mellon University, University of Florida
gtzan@cs.cmu.edu, gessl@cise.ufl.edu

Perry Cook

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

ABSTRACT

Musical signals exhibit periodic temporal structure that create 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 one rhythmic characteristic that could allow to discriminate between two pieces of music having the same tempo. Using this definition, we might say that a piece of Hard Rock has a higher beat strength than a piece of 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 the work presented in this paper, 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 music available digitally 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 be 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, 5]. They can be broadly classified into two catagories: event based and self-similarity based. In event based algorithms, transient events such as note onsets or percussion hits are detected and their InterOnset Arrival Intervals (IOI) are used to estimate the main tempo. In self-similarity based algorithms, the periodicity (self-similarity) of amplitude envelopes usually of multiple bands is calculated and used to detect the tempo.

The main focus of this paper is the concept of *Beat Strength*, which will loosely be defined as the rhythmic characteristic(s) 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 higher beat strength 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 [6]. 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 based on self-similarity described in [6]. The results of this paper should also be applicable to other global representations such as the *Beat Spectrum* 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 published 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. The subject pool constisted of 16 undergraduates (ages 18-21), 15 graduate students (ages 21-30) of Princeton University, and one professional adult (age 40-50). The undergraduates consisted of a wide variety of majors including engineering, social and natural sciences as well as humanities. No note of formal musical training was taken. The graduate students were either in the Computer Science or Music doctoral programs. 10 graduate students had formal musical training (learned instrument, music theory, composition). Formal training showed no effect on test outcome and hence was discarded as a biasing category.

Subjects were asked to assign 50 musical excerpts (each 15 seconds long) to 5 *Beat Strength* categories (Weak, Medium Weak, Medium, Medium Strong, Strong). A variety of different musical styles are represented in the selection of excerpts. Although there is some variability in the excerpts's tempo, it is mainly medium (70-100bpm) without any extremes. The excerpts were also pre-classified into the given categories by the authors, to ensure an even spread of *Beat Strength* (of course that information was not given to the subjects). Two forms of presentation were used. One form consisted of Audio CDs with 50 audio tracks containing the listening excerpts. The second form was a web-page containing links to CD quality audio-files. The order of presentation was randomized for each subject to avoid learning order artifacts in the results. Eight sets of two randomized presentation orders were prepared by the authors. Five Audio CDs and three web-pages were created. The CDs were used for the graduate student subjects and were randomly assigned to each. The web pages were used by undergraduates and the professional, and assignment to a particular set was also random. The random assignment of subjects to sets guarantees ignorance of the researcher to the presentation. No effect of presentation type or population was observed.

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. The subjects were asked to put each musical excerpt into one of the five categories as well as to choose one excerpt as the strongest and one excerpt as the weakest. For CD presentation they were asked to write the track number, in the case of the web presentation they were asked to write down the link name. Link names consisted of two letters followed by a number. The number referred to the label of the randomized set presented and remained the same for one subject but changed among subjects (for example BD1 would refer to random excerpt “BD” of randomized set 1). The letters were a alphabetical coding of the randomized list of tracks and were unrelated to the content of the track or its precategories beat strength.

The subject were ask to listen through the excerpts in two passes. First they were asked to listen to all to determine which example they assumed to be strongest and weakest. The purpose of this task was also to familiarize the subjects with the range of examples and the variety of styles and genres, and to help calibrate the subjects' notion of strongest and weakest beats. The second pass asked them to listen to all again and put them into the appropriate category of beat strength and the subject were encouraged to use the whole range.

2.2. Experimental Results

The results indicate that there is significant subject agreement about *Beat Strength* judgements. Figure 1 shows the average beat strength chosen across subjects for each musical excerpt and compares it to random agreement (the flat line at 3.0) and the pre-test categories chosen by the authors (seen as the staircase function). Figure 2 shows the intersubject variance for each listening task. The average standard deviation across subjects is 1.25. The limited range of the average is mainly caused by disagreemnent between subjects

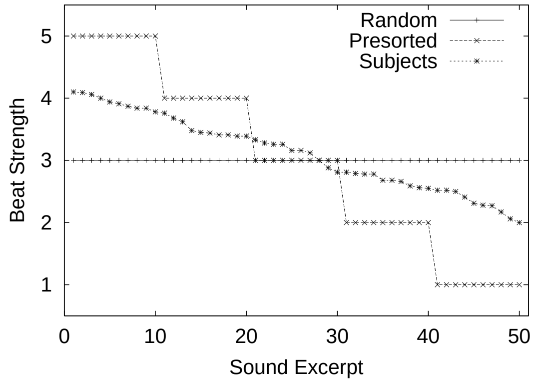

Figure 1: *Beat Study Subject Agreement.*

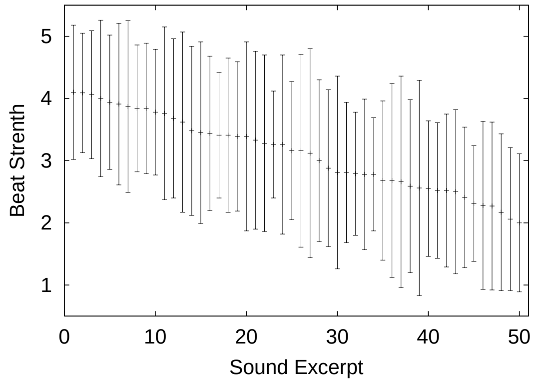

Figure 2: *Beat Study Subject Variance.*

and not the fact that the subjects avoid extremes as can be seen in Figure 2. Figure 3 shows which excerpt the subjects picked as the strongest and Figure 4 shows which ones they picked as having the weakest beat. The ordering of the excerpts on the x axis is the same as in Figures 1 and 2. Hence picks are ordered by average perceived beat strength. As can be seen, subjects agree more on the strong range of the spectrum and show greater variability on the weak side. Also within the strong distribution the strongest average beat strength and the most likely strongest pick overlap. This is clearly not the case for the weakest picks. This may indicate various causes: weak and no beat are less differentiable. Genre and style may be more important so subjective perception of beat strength or may otherwise be more influencial on the individual categories of beat strength in the weak range. Also the study was not geared towards finding a solid measure of the extreme ranges and the first exposure to the data-set may also be responsible for some of the variability. As this is not of immediate concern for our purpose, we delay these questions for future studies.

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 mu-

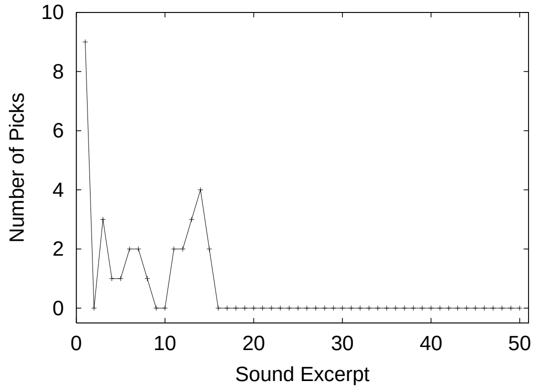

Figure 3: *Beat Study Strongest Picks.*

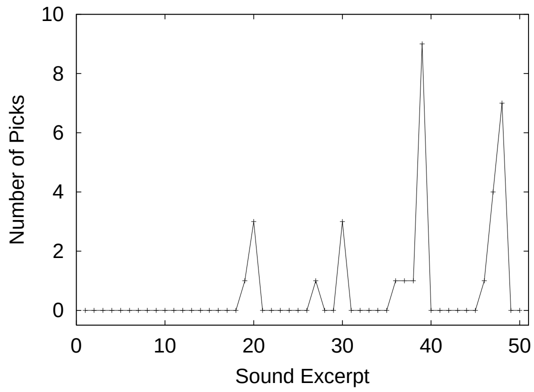

Figure 4: *Beat Study Weakest Picks.*

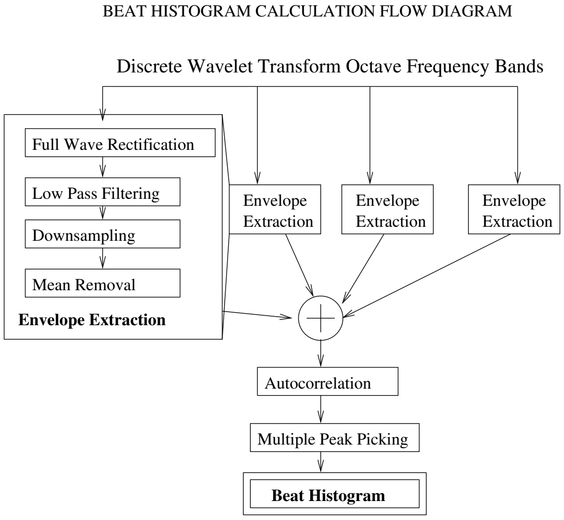

Figure 5: *Beat Histogram Calculation.*

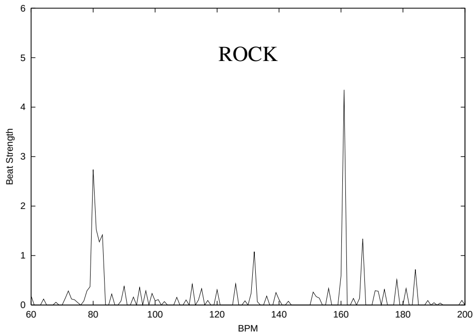

Figure 6: *Beat Histogram Example.*

sic 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 [6]. The main idea behind the calculation of BH is to collect statistics about the amplitude envelope periodicities of multiple frequency bands. A specific method for their calculation based on using a Discrete Wavelet Transform (DWT) filterbank as the the front-end, followed by multiple channel envelope extraction and periodicity detection was initially described in [7] and later used for deriving features for automatic musical genre classification in [6]. This method is shown schematically in Figure 5. For the BH calculation, the DWT is applied in a window of 65536 samples at 22050 Hz sampling rate which corresponds to approximately 3 seconds. This window is advanced by a hop size of 32768 samples. This larger window is necessary to capture the signal repetitions at the beat and subbeat levels. 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.

Figure 6 shows a beat histogram for a 30 second excerpt of the song “Come Together” by the Beatles. The two main peaks of the *Beat Histogram (BH)* correspond to the main beat at approximately 80 bpm and its first harmonic (twice the speed) at 160 bpm. Figure 7 shows four beat histograms of pieces from different musical genres. The upper left corner, labeled classical, is the BH of an excerpt from “La Mer” by Claude Debussy. The beat histogram is flat because of the absence of a clear rhythmic structure. More strong peaks can be seen at the lower left corner, labeled Jazz, which is an excerpt from a live performance by Dee Dee Bridgewater. The two peaks correspond to the beat of the song (70 and 140 bpm). The BH of Figure 6 is shown on the upper right corner where the peaks are more pronounced because of the stronger beat of rock music. The highest peaks of the lower right corner indicate the strong rhythmic structure of a HipHop song by Neneh Cherry.

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

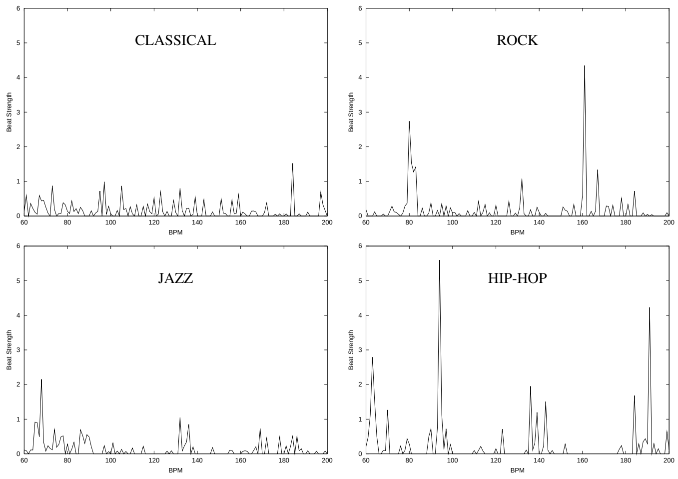

Figure 7: *Beat Histogram Examples.*

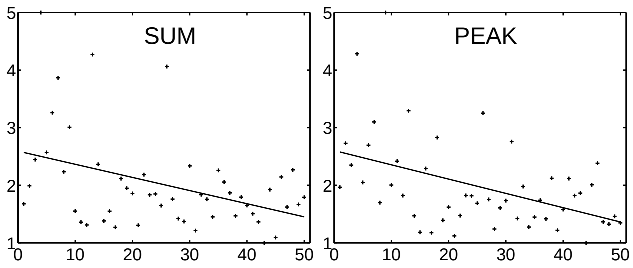

Figure 8: *Beat Strength Measure Results*

amplitude (PEAK) and indicates how dominant the main beat is.

In order to compare the performance of these measures with the user study results, the excerpts were sorted according to average beat strength as determined by the test subject and reassigned to a scale from 1 to 5 by equal division. The resulting assignment was then used as ground truth and compared with the assignments by sorting and equal division of the two computed *Beat Strength* measures. The raw measure results can be seen in Figure 8. The solid line indicates a linear fit of the data to illustrate more clearly the overall trend. The comparison was done by taking the absolute difference of the ground truth value from the automatically-assigned value calculated for each excerpt. The average absolute difference is 1.12 for the SUM measure and 1.08 for the PEAK measure. For 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. This can be seen in Figure 9. It is likely that humans utilize both self-similarity (SUM) and main beat dominance (PEAK) in order to characterize *Beat Strength* therefore it will be interesting to combine these two measures.

4. CONCLUSIONS AND FUTURE WORK

A user study exploring the concept of *Beat Strength* was conducted and there appears to be significant agreement about this concept among the subjects. This indicates that it can be utilized as another descriptor of music content for classification and retrieval

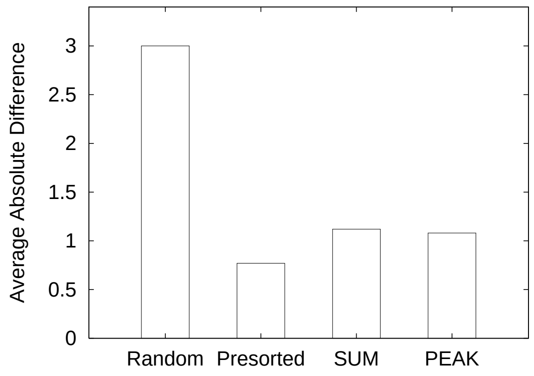

Figure 9: *Performance Comparison Between Random, Human Subjects, SUM and PEAK Measures.*

purposes. Two measures of *Beat Strength* based on the calculation of *Beat Histograms* were proposed and evaluated by comparing their performance with the results of the user study and it was shown that human beat strength judgements can be approximated automatically. The software used to calculate the Beat Histograms and Beat Strength measures is available as part of Marsyas [8] a free software framework for computer audition research available at: `www.cs.princeton.edu/˜gtzan/marsyas.html`.

There are several directions for future work we are exploring. A comparison of alternative automatic beat detection front-ends such as event-based algorithms and the *Beat Spectrum* for the purpose of calculating *Beat Strength* measures is planned for the future. Although the measures proposed in this paper are intuitive and provide good performance it is possible that other measures or combinations will have better performance. Separating the dimensions of *Tempo* and *Beat Strength* allows the creation of 2D rhythm-based browsing interfaces for musical signals. Another interesting possibility is the use of the *Beat Strength* concept in the characterization of other audio signals such as sound effects. Obviously it would mostly be applicable to repetitive sounds such as walking, running, clapping or striking a nail.

Another possibility we are exploring is to train statistical pattern recognition algorithms such as Gaussian or Nearest-Neighbor classifiers to do the bin assignment [9] instead of dividing the beat strength manually or by equal division.

5. 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] Jonathan Foote and Shingo Uchihashi, "The Beat Spectrum:a new approach to rhythmic analysis," in *Int. Conf. on Multimedia & Expo*. IEEE, 2001.

[6] George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, July 2002.

[7] George Tzanetakis, Georg Essl, and Perry Cook, "Audio Analysis using the Discrete Wavelet Transform," in *Proc. Conf. in Acoustics and Music Theory Applications*. WSES, Sept. 2001.

[8] George Tzanetakis and Perry Cook, "Marsyas: A framework for audio analysis," *Organised Sound*, vol. 4(3), 2000.

[9] Richard Duda, Peter Hart, and David Stork, *Pattern classification*, John Wiley & Sons, New York, 2000.