David Li
ISE 575
2/3/2005

Reviewing Tzanetakis and Essl's "Human Perception and Computer Extraction on Musical Beat Strength"

Tzanetakis and Essl's paper on extracting the musical beat strength did a comparison between what the computer decided and what the human perceived. They took a large variety of music from classical, rock, jazz, and hip-hop and presented the music to the test subjects. One group received CD samplers and the other group did online surveys. Both were supposed to categorize the music into five different categories (Weak, Medium Weak, Medium, Medium Strong, or Strong). To avoid changing "learning order artifacts" as the paper called it, the subjects were asked to listen to all the pieces at least once before categorizing the music.

First they had to define the beat strength. Since each category of music has a different style, the beats are different. A presorted selection was compared to the human experiments. The human experiments had two trials. One listened to audio CDs and the other downloaded links from a web-page. It turns out people have the tendency to avoid extremes. The data varied greatly, but on average there was a trend showing some songs have stronger beats than others.

The human results were followed up by the automated beat and strength extraction. The calculation uses the Discrete Wavelet Transform (DWT). This changes the amplitude in time-based domain to amplitude in the frequency-based domain. First the musical piece is broken down into separate octave frequency bands. To find out which beat best fits, it goes through an envelope extraction. It is rectified, passed through a low pass filter, down sampled, and the mean is removed to shift it down to zero. Then by taking the sum of the frequency bands, a beat histogram is built. The optimal beat is the highest peak in the beat histogram, which should have the units of beats-per-minute.

A comparison was made to show that the DWT method were not random. It also shows that the SUM and PEAK characterizations were able to distinguish the beat and the beat strength. The SUM is made from taking the sum of all the histograms to use as a reference for how this music signal is similar to the other music pieces. This helped find the beat. The PEAK is the average amplitude of the piece and is used to calculate the ratio of the dominant beat to the average beat, which can be quantified as the beat strength. The results show that there were barely a distinguishable beat in classical music and jazz. However for rock and hip-hop, the beats were strong and measurable.

The paper avoids syncopation, and changing tempos. However, if the algorithm they wrote was used there would be false peaks for syncopation, and multiple peaks for tempo changes. Also, the results from the SUM and PEAK seemed to vary greatly across all 50 songs. Perhaps the person who arranged the pre-sort was wrong, or the algorithm needs more tuning, but it does seem to match the spread in the human results, but I am not sure that is a good thing.

This paper also avoids beat track, so they do not know the phase of where the beat would start. It focuses on beat strength, which I believe is a relative term. They introduced the SUM and PEAK characterizations, help define beat strength. But they also did a human sampling experiment for how people define beat strength. Sadly they did not tie the two experiments together, so it feels like two separate papers.