Final Project Definition – Audio Tempo Extraction

Simon de Leon
McGill University, MUMT611 2/9/06

Introduction

The goal of this project is to implement the winning algorithm from the Audio Tempo Extraction competition for MIREX 2005 (Alonso, David, and Richard 2004). In brief, the algorithm works by taking the derivative of the frequency energy content to obtain a detection function consisting of "spikes" that correspond to the most salient features of the song. The classic autocorrelation method for periodicity estimation is then applied to this detection function to obtain a tempo. Finally, a pulse train at the extracted tempo is generated and aligned to the beats of the song by cross-correlating the pulse train with the detection function and adjusting the phase of the pulse train as needed.

Project Milestones

This project consists of three main MATLAB-based milestones and a fourth "cherry-on-the-top" STK-based milestone. Monophonic .WAV formatted songs will be used exclusively for simplicity.

Milestone 1) I will implement the onset extraction algorithm as described in (Alonso, David, and Richard 2004) using MATLAB. This requires me to operate on STFT blocks that are 4ms long with 50% overlap, based on a 128-point FFT. The absolute value of these STFT blocks will be low-pass filtered using the half-Hanning magnitude response curve, followed by a logarithmic compression as described in (Klapuri 1999). Finally, the logarithmically compressed data will pass through an 8^{th}-order FIR filter differentiator as described in (Proakis and Manolakis 1995). The result of this processing is the detection function that corresponds to the onsets of the most salient features of the song (note/harmonic changes, percussive events, etc.). A simple dynamic threshold which uses the median of a block of 25 samples is used to eliminate the weakest peaks.

Milestone 2) I will implement the periodicity detection algorithm as described in (Alonso, David, and Richard 2004) using MATLAB. Although both spectral product and autocorrelation were explored in the paper, only the autocorrelation method will be implemented since it clearly outperformed the spectral product method.

Milestone 3) I will impose a metronome "tick" track onto the song by generating a pulse train at the tempo extracted and continually phase-aligning it with the beat onsets by cross-correlating it with the detection function.

Milestone 4) Port the tempo extraction algorithm to a realtime STK implementation. This is the most difficult milestone and is not expected to be completed within the timeframe of this class due to its complexity. This requires the development of vectorized classes for the STFT, logarithmic compression, and autocorrelation functions.

Comments

There are a few minor parameters not clearly indicated in (Alonso, David, and Richard 2004) that I will need to contact the authors about. Specifically, the STFT window used (Hanning, Blackman, rectangular, etc.) is not specified. Furthermore, the test songs they used were downsampled and it is not clear how the parameters should be scaled for higher sampling rates, i.e. should the FFT be increased to a 256-point FFT for 44.1 kHz audio? These are not mission critical issues and are expected to be resolved empirically in the worst case that the authors are unreachable.

References

Alonso, Miguel, Bertrand David, and Gael Richard. 2004. Tempo and Beat Estimation of Musical Signals. *Proceedings of the 5th International Conference on Music Information Retrieval*.

Klapuri, Anssi. 1999. Sound Onset Detection by Applying Psychoacoustic Knowledge. *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*: 3089-3092.

Proakis, John G., and Dimitris K. Manolakis. 1995. *Digital Signal Processing: Principles, Algorithms and Applications*. 3rd Ed. New York: Prentice Hall.