Summary of audio tempo extraction presentation
Simon de Leon
McGill University, MUMT611 2/16/06

1.0 Introduction

Tempo extraction from audio is useful for several musical situations such as automatic rhythm alignment, beat driven effects, and intelligent accompaniment. Generally, tempo extraction from straightforward rhythmic music (pop, rock, rap) is mature. The main challenge is to be accurate across a wide range of genres, including the problematic areas of classical and jazz music.

We will study the winning algorithm from the MIREX 2005 audio tempo extraction competition described in (Alonso et al. 2004). Of most interest in this algorithm is the use of spectral flux to determine the beat onsets. The top performing algorithms from that competition used similar time-frequency analysis for the onset detection. Suggestions for possible improvement will also be briefly mentioned.

2.0 Algorithm

The algorithm can be decomposed into three sub-blocks: onset extraction, periodicity estimation, and temporal estimation of beat locations. Onset extraction is the process of locating the hypothetical beats of the music. The periodicity estimation extracts the tempo from these hypothetical beats. Finally, the temporal estimation of beat locations aligns a pulse train at the extracted tempo to the music.

2.1 Onset extraction

Onset extraction is concerned with the location of the most salient features of the music. Salient features typically correspond to note changes, percussive events, and harmonic shifts. To locate these features, the time derivative of the frequency component magnitudes is taken and summed at each time index. This is known as the spectral flux, and typically returns a detection function with semi-periodic spikes corresponding to the beat onsets of the music.

To reduce false beat onsets, the frequency component magnitudes are low-pass filtered according to a half-Hanning window magnitude curve and then logarithmically compressed (Klapuri 1999). These settings are informed by psychoacoustics. The derivatives are obtained using an eighth-order FIR filter differentiator as described in (Proakis and Manolakis 1995). Finally, there is a dynamic threshold applied to the spectral flux that consists of a dynamic threshold set to the median of 25 local samples.

As an alternative to the low-pass filtering, it might prove useful if the frequency component magnitudes and derivatives were weighted according to frequency masking and the critical bands. This would help eliminate the contributions of perceptually irrelevant information to the spectral flux.

2.2 Periodicity detection

The periodicity detection scheme is applied to the detection function and extracts a tempo. Two methods were studied and described in the presentation of the algorithm: autocorrelation and spectral product. The autocorrelation is the classical periodicity estimation tool that works by cross-correlating the detection function with itself. Alternatively, the spectral product takes the FFT of the detection function and

determines the frequency with the largest product sum of integer multiples. This frequency is assumed to be the frequency of the beats. In the study presented in (Alonso et al. 2004), the auto-correlation outperformed the spectral product consistently across all genres.

2.3 Temporal estimation of beat locations

Using the extracted tempo and the detection function, a pulse train is generated and mixed with the original signal to generate a metronome track. The pulse train uses the period of the locally captured tempo, and is cross-correlated with the detection function to ensure phase accuracy. Note that the correlation was performed directly in the time domain.

When beats are missing or not found within a given tolerance of their expected locations, the algorithm works to re-align the phase of the generated pulse train with the detection function. This can be a problem with music that has short silent breaks, since the algorithm will stutter and fail to produce a pulse in the empty space.

3.0 Examples

Several audio examples of the algorithm at work were demonstrated during this presentation. The algorithm works flawlessly with soul, rock, rap, and some straightforward jazz. It struggles severely with classical music, especially solo piano.

4.0 Conclusion

The winning algorithm from the MIREX 2005 audio tempo extraction competition was presented and explained with audio examples. This work focused on time-frequency analysis techniques to generate an onset detection function. By its very nature, the onsets of beats are assumed to correspond to areas of greatest spectral flux, or fastest change in frequency content. This leads to problems in music with long fading attacks and decays, several instruments playing at once, and music where the tempo varies quickly.

Possible improvements include developing a solution to the problem of brief silent breaks in music, which is prominent in classical piano. It may prove useful to give the temporal estimation of beat location section an inertia that will force it to tolerate silent sections. The difficulty is allowing this inertia without affecting the algorithm's ability to adapt to quickly changing tempos that can be found in electronic music. Furthermore, it may be useful to apply psychoacoustic principles of frequency masking, temporal masking, and critical bands to improve the spectral flux extraction.

References

Alonso, M., B. David, and G. Richard. 2004. Tempo and beat estimation of musical signals. *Proceedings of the 5th International Conference on Music Information Retrieval*.

Klapuri, A. 1999. Sound onset detection by applying psychoacoustic knowledge. *Proceedings of the IEEE International Conference of Acoustics, Speech and Signal Processing*: 3089–3092.

Proakis, J., and D. Manolakis. 1995. *Digital signal processing: Principles, algorithms and applications*. 3rd Ed. New York: Prentice Hall.