Tempo Tracking With a Periodicity Comb Kernel

Ian Leue Ozgur Izmirli
Center for Arts and Technology
Connecticut College
270 Mohegan Ave, New London, CT. USA.
ipleu,oizm@conncoll.edu

Abstract

Automatic tempo extraction and beat tracking from audio is an important ability, with many applications in music information retrieval. This paper describes a method for tempo tracking which builds on current research in the field. In this algorithm, an autocorrelation surface is calculated from the output of a spectral energy flux onset novelty function. The most salient repetition rate is calculated by cross-correlating dilations of a comb-like prototype spanning multiple frames and the autocorrelation surface. The method addresses tempo tracking through time to account for pieces with variable tempos. In order to compare the performance of our method on music with strong and weak percussive onsets we have evaluated it on both classical music with and without percussion and popular music with percussion. Additionally, beats are phase-aligned and superimposed on the signal for aural evaluation. Results show the comb kernel to be a useful feature in determining the correct beat level.

Keywords: beat, tempo tracking, onset detection.

1. Introduction

Much work has been done in onset detection, tempo extraction and beat tracking. However, the problem of metrical hierarchy identification has proven to be difficult. It is non-trivial to find which multiple of the tatum is perceived as the beat. This paper describes a tempo tracking algorithm that emphasizes the most salient multiple of the tatum. It works locally, tracking both static and changing tempos. We have evaluated the algorithm on both popular and classical music, which is harder to track.

Tempo is a fundamental aspect of western music, and its recognition is considered imperative for computer understanding of music [1]. Advances have been made in recent years towards onset detection and effective tempo tracking from audio [2,3]. See [1,4] for overviews.

Our algorithm works directly from uncompressed audio and creates a time-variable tempo curve. It performs periodicity estimation on spectral onset features reported in previous work [1,5,6,7]. Specifically, we detect onsets using a spectral energy flux feature [1,7,8], perform an autocorrelation on the onsets, then cross-correlate a range of dilations of a comb-like prototype spanning multiple frames with the autocorrelation for an estimate of tempo.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2006 University of Victoria

2. Method

2.1 Onset Detection and Periodicity Estimation

We separate the audio into 7 logarithmic frequency bands and perform a differentiation on energy in each band. The sum of the differentiated energy outputs across all bands is a signal representing the level of onset versus time.

After detecting onsets, we calculate the autocorrelation surface $A(\tau, n)$ from the onset signal to find the repetition pattern over time. The autocorrelation is performed on 8 second windows and is hopped in 1 second increments. Here, τ represents the lag and n is the time index. We calculate autocorrelation by sliding windows over the original full signal. The result is a "trend-corrected" autocorrelation surface (see Figure 1) that favors all lag times equally, and is primed for the comb prototype.

2.2 Beat-level Estimation

While autocorrelation can effectively find self-similarity at various lag times, it still leaves open the question of beat-level. Past algorithms picked the highest non-zero lag in the autocorrelation [9] or the highest lag with a multiplicity relationship with other high lags [1]. We sought to build on this groundwork by utilizing a tented comb-like prototype.

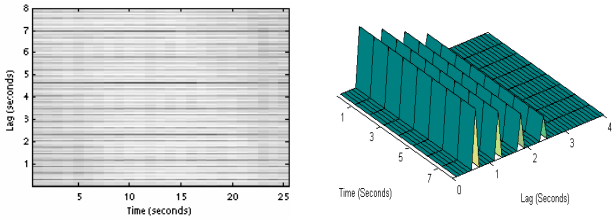

Figure 1. Left: Autocorrelation surface $A(\tau, n)$. Right: Tented Comb-like Prototype $C(\tau, j)$, graphed at a specific dilation in the lag-domain. Height indicates prong weight.

We construct a comb prototype corresponding to the slowest allowable tempo, T_s. Then, for each time index, we

cross-correlate between dilations of this prototype and the autocorrelation surface. Prototypes $C(\tau, j)$ are designed to highlight the salient beat level (see Figure 1). Each comb has M prongs (typically 4) at multiples of its current lag value. This is expected to reveal the highest similarity to the beat-level lag when cross-correlated with the autocorrelation surface. These prongs are tented in the lag-domain to allow for irregularities in the performance and time-quantization, and index j extends through N time frames to ensure that the repetition found in that frame is not temporary. Consecutive prongs have decaying weights.

Each comb dilation is cross-correlated with the autocorrelation surface to calculate the most salient lag:

$$B(r,n) = \sum_{k=n}^{n+N-1} \sum_{\tau=\tau_{min}}^{\tau_{max}} A(\tau,k)\, C(r\tau, k-n) \qquad (1)$$

where r is the dilation factor of the prototype ($1..R_{max}$). R_{max} is the ratio of the fastest tempo to the slowest tempo. τ_{min} and τ_{max} are limits on the lag axis corresponding to the span of the comb for the fastest and slowest tempos. n is the time index of the tempo estimate, N is the number of time frames used in the cross-correlation.

Studies have shown that listeners prefer a beat centered around 120 beats per minute (BPM)[10]. To model this, we apply a preference curve using the Parncutt function, $w(r)$, as formulated in [10] but with a β value of 0.25:

$$w(r) = \exp\left[-\beta \log_2^2\left(120/\left(T_s\left(R_{max}+1-r\right)\right)\right)\right] \qquad (2)$$

As formulated, the spontaneous tempo is 120 BPM and T_s is the slowest tempo of interest. The tempo estimate for each time index, n, is found by first determining the value of r that maximizes $w(r)B(r,n)$. The tempo estimate is given by rT_s in BPM. This operation, performed on all frames, results in a time-variable tempo curve with estimates 1 second apart.

3. Results and Evaluation

In order to compare the performance of our method on music with strong and weak percussive onsets we constructed two test sets: one containing percussive and non-percussive classical music, and one containing percussive popular music. These test sets contained 45 and 30 tracks respectively and were picked randomly.

We annotated each piece and compared that to the longest-lasting tempo in the tempo curve. We then calculated a performance score for each test set based on the evaluation criteria used for MIREX 2005. Within an 8% tolerance zone, 1 point was awarded to the correct tempo, .8 for twice or half, and .6 for thrice or one third. These scores were averaged over the entire test set to create a single performance score between 0 and 1. For comparison purposes, we also implemented a very simple beat-level estimation scheme in which we picked the peak non-zero lag from the autocorrelation surface for each time frame. Our method found 80% (score = 0.96) of the tempos correctly for popular music and 63% (score = 0.81) of the tempos correctly for classical music. The simple method performed poorly with 16% (score = 0.22) on the popular set and 7% (score = 0.12) on classical set.

In addition, we phase-aligned and superimposed synthetic beats onto the original signal for aural evaluation. This consisted of constructing a local 4-second beat train from the tempo curve, cross-correlating that beat train with the onset output, picking the cross-correlation's peak as the time of a beat and repeating this for the entire piece. This made it easier to aurally evaluate the algorithm's performance with variable tempos. The piece "Alphabet Aerobics" by Blackalicious for instance, has a steadily increasing tempo, and we were able to listen to the algorithm successfully track the changing tempo.

4. Conclusion

The results are promising with MIREX-based scores of 0.96 and 0.81 for the two test sets. They also indicate clearly that utilizing a comb kernel as formulated in this paper can be more effective than other (admittedly primitive) beat-level estimation schemes. We tested and evaluated our method on both popular and classical music in order to have a baseline benchmark for both genres. Results show that classical music remains an area with room for improvement, and future work will focus on new onset features targeted towards detecting tonal onsets.

References

[1] M. Alonso, B. David, and G. Richard. "Tempo and Beat Estimation of Musical Signals," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc*., 2004, pp. 158-163.

[2] F. Gouyon, "A Computational Approach to Rhythm Description," *Ph.D. Thesis, Univ. Pompeu Fabra*, 2005.

[3] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M.E. Davies and M.B. Sandler, "A Tutorial on Onset Detection in Musical Signals," *IEEE Tran..on Speech and Audio Processing*, Feb. 2004.

[4] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle and P. Cano. "An Experimental Comparison of Audio Tempo Induction Algorithms," *IEEE Tran. on Speech and Audio Proc.*, vol. 14, No. 5, 2006.

[5] Goto, M, "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds," *J. New Music Research*, vol. 30, No. 2, pp. 159–171, 2001.

[6] A. Klapuri, "Musical Meter Estimation and Music Transcription," *Cambridge Music Processing Colloquium*, Cambridge University, March 2003.

[7] C. Uhle and J. Herre. "Estimation of Tempo, Micro Time and Time Signature From Percussive Music," in *DAFx-03 6th Int. Conf. Of Digital Audio Effects*, 2003.

[8] J. Laroche, "Efficient Tempo and Beat Tracking in Audio Recordings," *J. Audio. Eng. Soc.*, vol. 51, No. 4, pp. 226–233, April 2003.

[9] M. McKinney and D. Moelants. "Extracting The Perceptual Tempo From Music," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc*., 2004.

[10] K. Frieler, "Beat and Meter Extraction Using Gaussified Onsets," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc*., 2004.