Tempo Detection using Time-Warped PLP in Tempo Varying Music

Qingyang (Tom) Xi
Music Technology
NYU Steinhardt
qx244@nyu.edu

Sumanth Srinivasan
Electrical Engineering
NYU Tandon
sumanths@nyu.edu

ABSTRACT

PLP is a good indicator of local periodicity for constant-tempo music, but performs poorly when faced with expressive music with time-varying tempo. PLP with time-warped kernels in the style of Fan Chirp Transform is experimented with in this paper. The Tempo-Plane per analysis frame is introduced, and used to predict the tempo at each frame using a HMM.

1. INTRODUCTION

Rhythm is an influential character of music both in traditional western and other ethnic cultures. By establishing structure in the form of granular repeating blocks, which is then accentuated by other meter-based features, it in-forms of the cultural background and in pop-cultures, the genre. Hence, in the fields of music information retrieval and computer music, beat tracking and automated rhythm detection has been an active space of research.

A lot of work has been carried out to detect downbeats in music as they highly characterize the overall rhythm of the track, as well as establish structure to its form. Most beat tracking methods use a recurring work flow of feature extraction, estimation of periodicity in these features and phase estimation [1]. The feature used for this purpose is usually a novelty function.

Beat tracking is a particularly cumbersome task and often inaccurate in case of expressive classical music. A study on Chopin Mazurkas [2] compares the results of different beat tracking methods on each of his pieces, but also considers the consistencies of beat tracking results across a number of performances of the same piece.

2. TEMPO ESTIMATION PROBLEM

The onsets tracked in case of most beat detection algorithm is often characterized by sudden increase in energy. While this is more prominent in case of percussive music, pieces entirely based on instruments with soft onset may result in blurring of these novelties. The uncertainty makes it difficult to assign specific time to beat positions. The presence of tempo changes and expressive nuances in the piece may further aggravate this problem. Problems may also arise due to syncopation and rhythm created by silences and negative space in the piece.

Chopin Mazurkas is a great example to study the presence of such problems as it is greatly expressive in nature but also has musical embellishments that subvert general notions of rhythm. This is also supported by the availability of data-set of select performances obtained from the Mazurka Project.

Besides the Chopin Mazurka data set, a set of Beatles songs with uniform tempo that varies across sections is also used as a means to track the performance of methods used in this project.

3. METHOD

The over-arching theme of our method is to extend PLP with Fan Chirp Transform, which will now return a metric, the tempo-plane, per segment of audio. Using this metrics matrix, we then interpret its physical meanings and construct a Hidden Markov Model using the tempo-plane as observations at each time step embodied by the analysis audio segment. We start by presenting the Fan-Chirp Transform.

3.1 Fan-chirp Transform

Since tempo is a metric of periodicity just like pitch, a lot of methods formulated for pitch detection can be modified and applied to obtain tempo information in music with variable tempo.

The Fan Chirp Transform (FChT) provides both time and frequency resolution using a set of fan-geometry as an additional dimension of basis functions and is carried out by time-warping the signal in time domain and then using the FFT. This method allows it to retain the computational efficiency that FFT offers.

Our implementation is based on the descriptions in [3]. The idea is to use linear chirps that has a time varying frequency $f(t) = f_0 \cdot (1 + \alpha t)$, that starts at f_0. Using definition of phase derivative:

$$f(t) = \frac{d\phi(t)}{dt}$$

$$\phi(t) = \int f(t)dt = f_0 \cdot \int (1 + \alpha t)dt = f_0 \cdot (t + \frac{1}{2}\alpha t^2)$$

Let $\psi_\alpha(t)$ be the time-warping function:

$$\psi_\alpha(t) = (1 + \frac{1}{2}\alpha t)t$$

$$\phi(t) = f_0 \cdot \psi_\alpha(t) + \phi_0$$

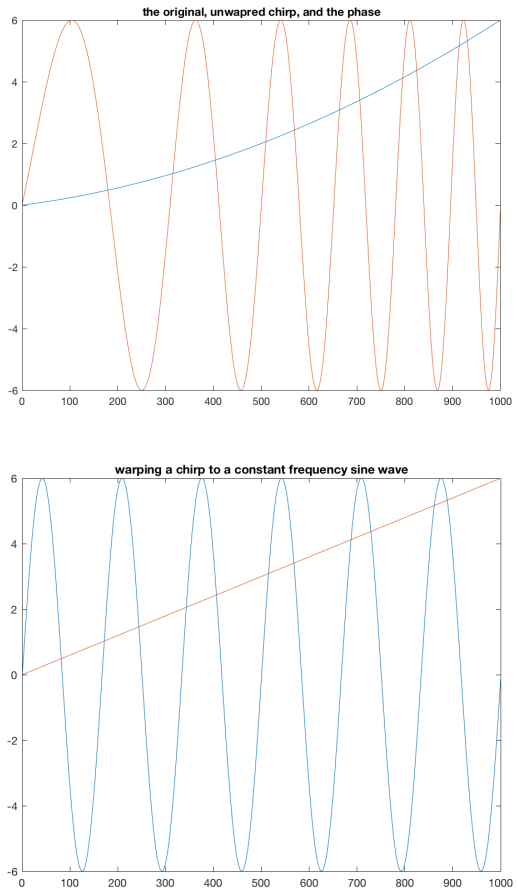

Figure 1. Chirp before and after warping

Using the warped time $\tau = \psi_\alpha(t)$, the Fan Chirp transform can be described as the following:

$$X(\alpha, f) = \int_{-\infty}^{\infty} x(\psi_\alpha^{-1}(\tau)) \cdot e^{-j2\pi f\tau} d\tau$$

With the substitution of letters, this is not insignificant. This equation can be interpreted as the following: a FChT is nothing but a normal DFT with a α dependent warping function $\psi_\alpha(\cdot)$. The concept is the following: by warping, potential chirps in the original signal that is hard to analyze by FFT is now warped according to become constant frequency in time domain τ. The basis functions of the analysis are now in the warped time domain, they are $f_{basis}(\tau)$. They are constant frequency sine waves in this warped domain. To get these basis functions in the unwarped domain, one simply needs to find the unwarping function $\psi_\alpha^{-1}(\cdot)$. The warping and unwarping functions can be found analytically, and if we constrain the warping to be constant frequency to linear chirps, or quadratic phase, then

$$\psi_\alpha(t) = (1 + \frac{1}{2}\alpha t)t$$

$$\psi_\alpha^{-1}(t) = -\frac{1}{\alpha} + \frac{\sqrt{1 + 2\alpha t}}{\alpha}$$

However, any arbitrary monotonous $\psi(\cdot)$ that permits an inverse is permissible as a warping function. In our implementation, the inverse function $\psi_\alpha^{-1}(\cdot)$ is generated numerically by flipping the y and x axis and resampling.

Figure 1 shows such a warping operation applied to a chirp, which warps it to a constant frequency linear phase sine wave for easy analysis.

Both Spectral Flux (Rectified) and Log-Energy Derivative were used as candidates for the input onset function in this step. It was seen that the latter performed much better in case of percussive music (such as The Beatles) while the former was marginally better in case of the Chopin Marzurkas.

3.2 WPLP: Warped Predominantly Local Pulse

We proceed to combine the above idea of warping before frequency analysis with the PLP method of tempo detection that is introduced by [4].

In summary, the PLP method included doing a frequency analysis on the novelty function computed from the audio. The creation of the novelty curve is not the focus of this paper, and there are many choices over which novelty curve to use. For this study, the novelty curves are constructed according to [5].

The basis functions for this analysis are complex exponential with frequencies in the typical tempo range. In our study only integer tempo between 44BPM and 144BPM are considered:

$$s_\omega[n] = e^{-2\pi j\omega n}, \text{where } \omega \in \{k/60 | k \in [44 : 144]\}$$

Combining this list of basis frequencies ω and a user defined range of α's, The PLP would produce a Tempo Plane $X(\alpha, f)$ for each analysis frame in time. A sample tempo plane is shown in figure 2.

The classical PLP method would have produced a 1d array as opposed to a 2d matrix as the result of its analysis, and would've picked the peak magnitude index as it's best frequency for the frame. It then continues to construct a sine wave kernel that best match the novelty function during that frame using the 1d array of analysis result. Now, with an additional analysis parameter, picking the best f and α require some insights.

3.3 Framewise Tempo Estimation using HMM

A more robust method of picking the best kernel estimate is by considering the tempo plane obtained using WPLP as an observation corresponding to a hidden state i.e., the tempo at that instant. For each frame, this gives us a 2-d observation plane. This observation and related probabilities are then implemented as a Viterbi model to predict the tempo at each frame. [5]

The distribution of energy across α values for each ω is used to build a transition matrix that informs change from one state

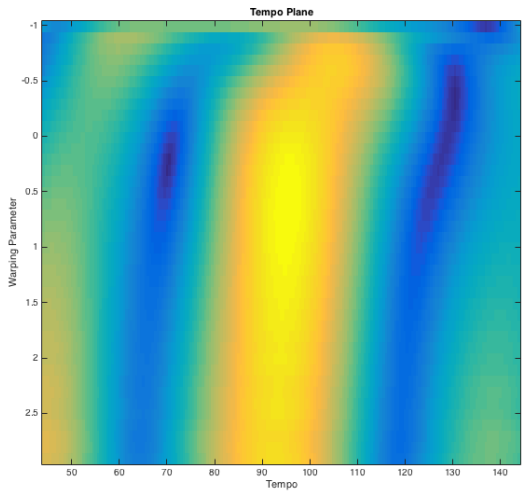

Figure 2. Tempo Plane for one frame

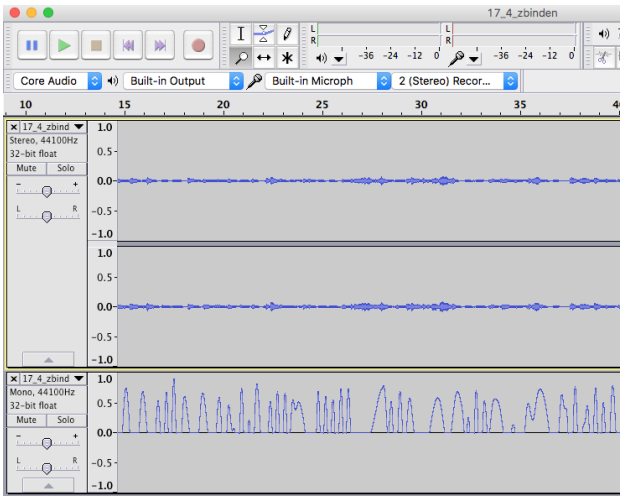

Figure 3. PLP result for a segment of Chopin Mazurka Recording.

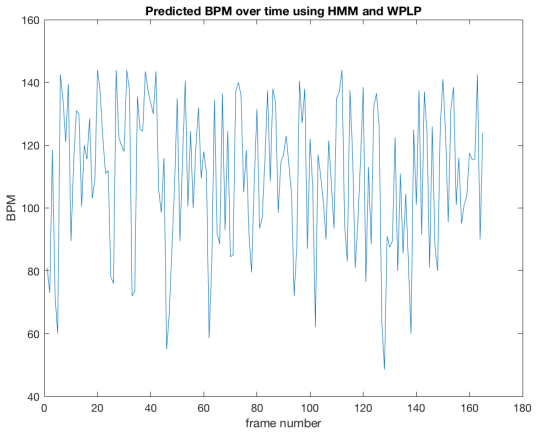

Figure 4. Framewise tempo prediction using WPLP and HMM .

(tempo) to another. This is done by locating each source w_i from the transition matrix in the tempo plane, and extracting a vector of its distribution across α values. This vector is interpolated to find the target frequency for each α, and then resampled to estimate the probabilities of transitioning from the given $w\omega_i$ to ω_j. For each frame of the signal, a new transition matrix is generated.

The emission probability for each frame is PLP of that frame for an un-warped input signal. This is normalized across frequencies for each frame, hence giving the probability for each state ω given the observation.

This allows us to treat the signal with variable tempo as a Markov process and make informed predictions of hidden states.

```
[value_wt, i_hat] = ...
    max(value_tl_i+log(a_ij), [], 1);
```

The best path obtained at the end of this process is a vector of ω values for each frame. Localized by ω, we pick the best α, and hence the kernel estimate, simply by looking for the a peak in the tempo plane for the given tempo value.

4. OBSERVATIONS AND DISCUSSION

When using a naive global peak picking method for determining the best $f$and α in each frame, the resulting gamma function that is traditional to the PLP approach is shown in figure 3. The resulting PLP perdiction is reasonable in case of The Beatles tracks, but does not work in case of Chopin Mazurkas.

In our special formation of the HMM using the frequency of each frame as the state variable, the resulting output from the HMM method is a list of tempos per analysis frame, and this is shown in figure 4.

As can be gather from the plot, the prediction varies wildly between 44 and 144 BPM, the allowable limits of the system. However, for tempo value, the best alpha was picked from the tempo plane corresponding to that frame and basis functions were generated using the resulting kernels. The γ vector resulting upon overlap add of these basis functions performs better for both The Beatles as well as the Chopin Mazurkas. Figure 5 shows the beat predictor with and without using a HMM model.

5. CONCLUSIONS AND FUTURE WORK

In this project, two methods for tempo detection in variable tempo music were employed and their performance was tested. The first method involves using Fan Chirp Transform and PLP to obtain tempo-plane for each frame from which global peak picking was done to obtain the kernel. In the second method, the tempo-plane is used as an observation for the hidden tempo state and is treated as a Hidden Markov Model upon which Viterbi algorithm is used.

The peak picking method works as per intuition but per-

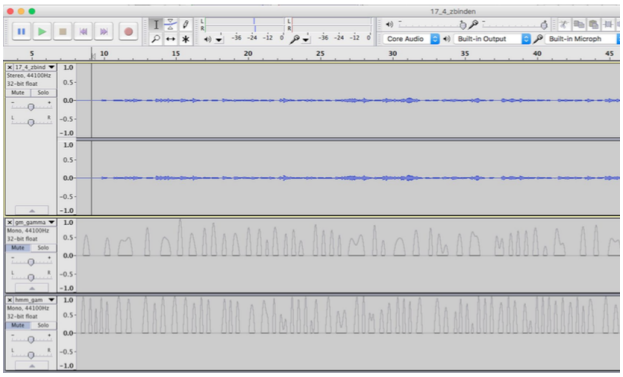

Figure 5. Comparing onset detection for Chopin Mazurka track.

forms poorly as an onset function. The usage of viterbi greatly improves onset detection and works for both constant tempo and temp variant music very well.

Much can be done in case of method 2 to make it a better predictor. This can involve employing a machine learning model that learns to pick the correct α and kernel values once there is a list of ω states across time.

Further experiments may also be conducted in shifting the parameters used in the Viterbi model itself - particularly in building the transition and emission probability matrices. Several other optimizations may also be done such as choosing a better novelty detection function that is more sensitive to non-percussive sounds, and using non-linear warping functions to detect changes in tempo more accurately.

Thus there seems to be ample opportunity to apply the usage of some labeled data, which we didn't get a chance to do due to the time frame of the project. Besides, the warping functions $\psi(\cdot)$ can be used this already in creative work as an aesthetic in interesting ways.

6. REFERENCES

[1] S.Durand, B.David, and G.Richard, "Enhancing downbeat detection when facing different music styles," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*. IEEE, 2014, pp. 3132–3136.

[2] P. Grosche and M. Mller, "What Makes Beat Tracking Difficult? A Case Study on Chopin Mazurkas," in *in Proceedings of the 11th International Society for Music Information Retrieval Conference, 2010*, 2010, pp. 649–654.

[3] P. Cancela, E. López, and M. Rocamora, "Fan chirp transform for music representation," in *International Conference on Digital Audio Effects, 13th. DAFx-10. Graz, Austria*, 6-10 Sep 2010.

[4] P. Grosche and M. Mller, "A mid-level representation for capturing dominant tempo and pulse information in music recordings," in *In Proc. of ISMIR*, 2009, pp. 189–194.

[5] J. P. Bello, "Chroma and Tonality," in *Notes for Music Information Retrieval Class, NYU Steinhardt*, 2016.