

Improvement of An Online Music Alignment Algorithm Based on Onset Detection

Bochen Li, Zhiyao Duan

Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York, 14627

ABSTRACT

Most polyphonic audio-score alignment methods are offline algorithms, and alignment accuracy drops significantly with online algorithms. But most application based on alignment need real-time results, such as Automatically Accompaniment System or Music Tutor. So we aim at improving the alignment result of an online alignment algorithm.

In general online methods, a polyphonic music audio is segmented into time frames and they are fed to the score follower in sequence. Then the algorithm outputs a score position for each frame right after it is processed. In this paper, we find that not all audio frames give us the same level of reliability within the evolution of one note, especially in staccato performing or decay of piano sound when real sound will last shorter compared with our expectation from the score. Then the online follower in previous algorithm will get lost.

So in this paper, we propose to utilize note onsets information to find the “faithful” frames as observation to do the alignment. A weighting function is introduced to assign different weights to the frames as the measure of their reliability. The value is dropping with evolution of one note and score position tends to move forward evenly excluding observations from audio, while when a new onset is detected, it will be dragged to right position quickly due to high trust of observation.

Experiments shows that proposed improvements can get better alignment results than previous online methods, especially in piano music.

CONTACT

<Bochen Li>
<Audio Information Lab>
Email: bli23@ur.rochester.edu
Phone: 917-969-5428
Website: <http://www.ece.rochester.edu/projects/air>

INTRODUCTION ABOUT MUSIC ALIGNMENT

Music alignment (Audio-score alignment) is the association of events in a musical score with frames in the time axis of an audio signal.

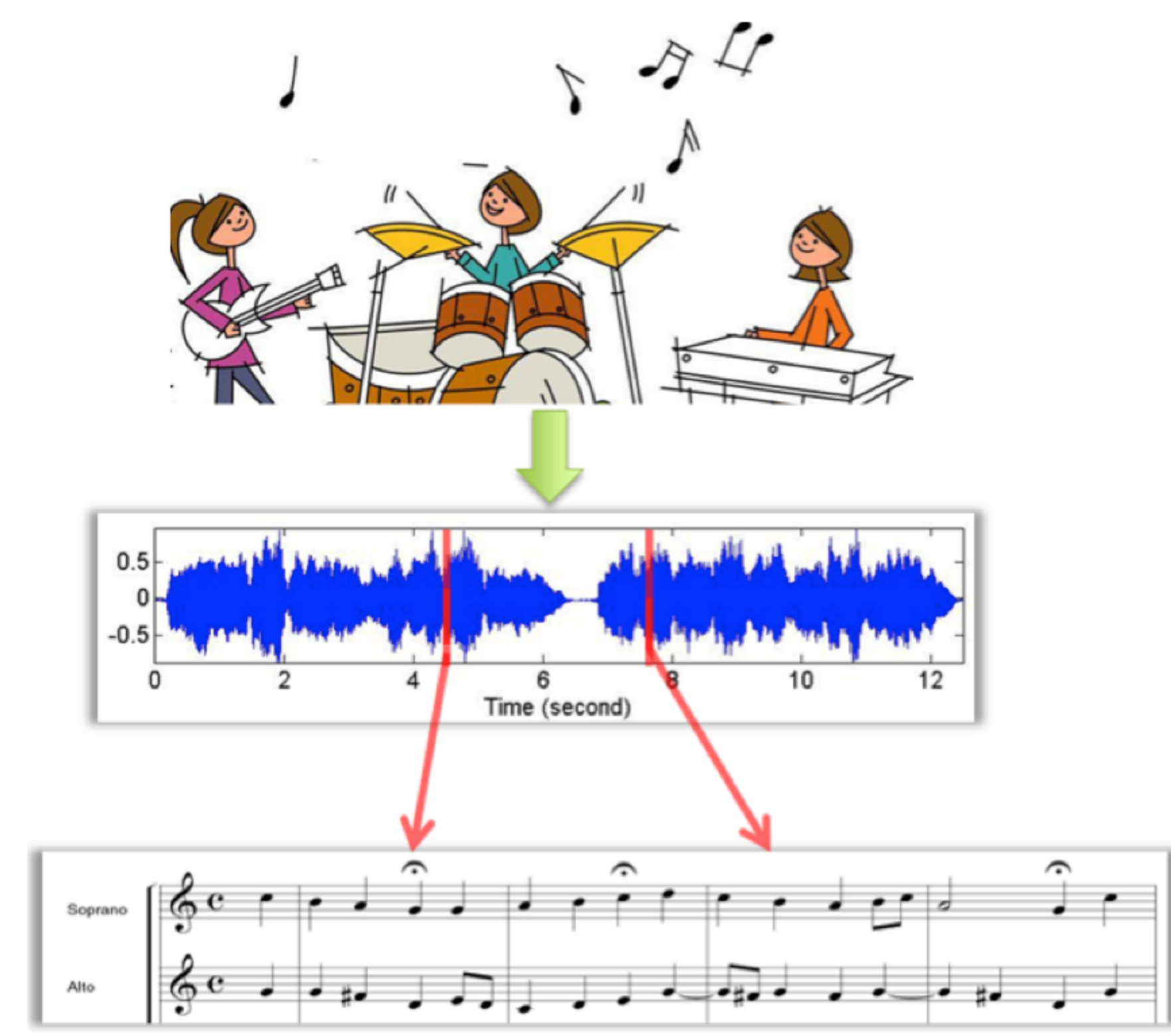


Figure 1. Graphical show of audio-score alignment

MODEL STRUCTURE

A hidden Markov process model is illustrated in figure 2.

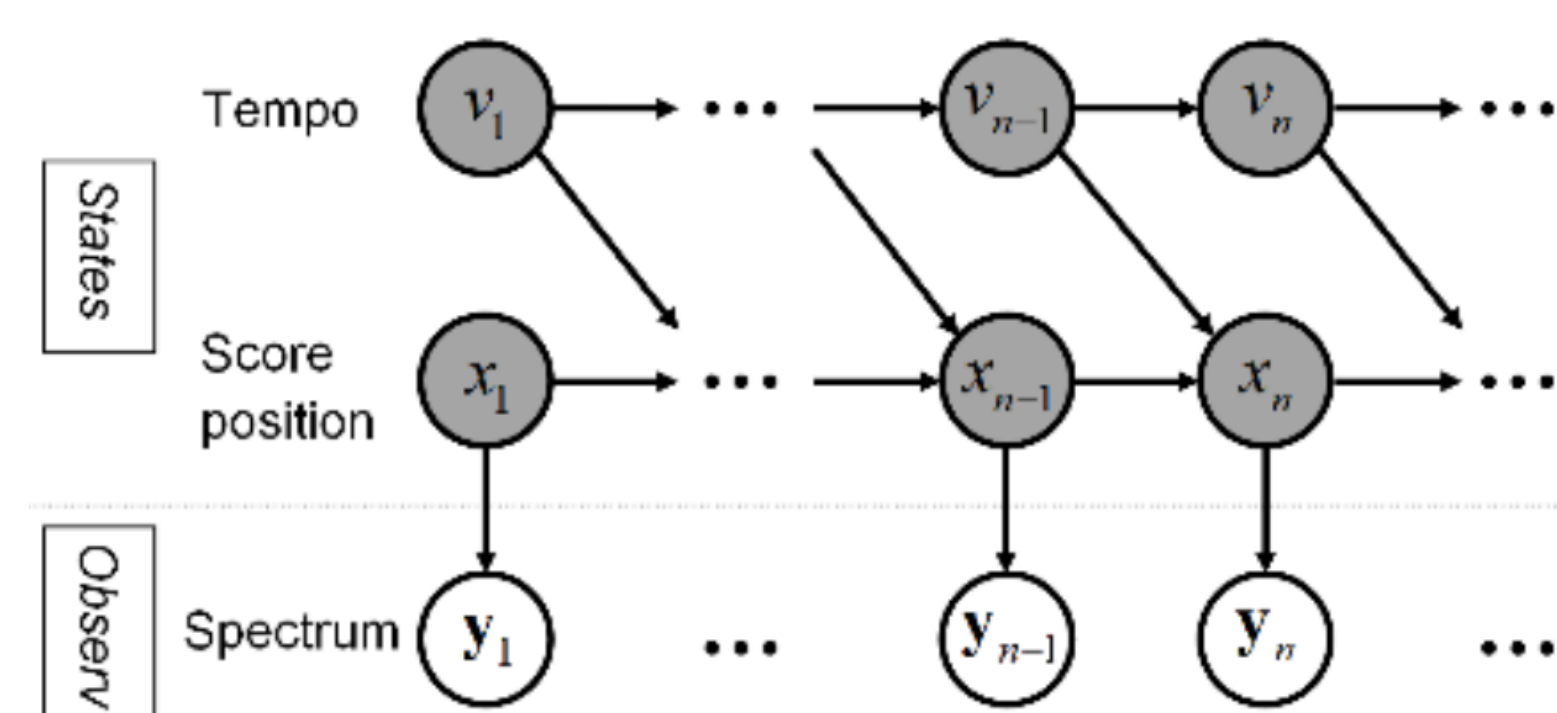


Figure 2. Model Structure

The n -th frame is associated with a 2-dimensional hidden state vector $\mathbf{s}_n = (x_n, v_n)^T$

Each audio frame is associated with an observation \mathbf{y} .

Process Model: $p(\mathbf{s}_n | \mathbf{s}_{n-1})$

defines the transition probability from the previous state to the current state

$$\mathbf{s}_n = (x_n, v_n)^T \quad x_n = x_{n-1} + l \cdot v_{n-1}$$

$$v_n = \begin{cases} v_{n-1} + n_v & \text{if } z_k \in [x_{n-1}, x_n] \text{ for some } k \\ v_{n-1} & \text{otherwise} \end{cases}$$

Observation Model: $p(\mathbf{y}_n | \mathbf{s}_n)$

is to evaluate whether a hypothesized state can explain the observation

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{peak region}}(\theta) \cdot \mathcal{L}_{\text{non-peak region}}(\theta)$$

Method: **Multi-pitch observation likelihood**

Inference

$$p(\mathbf{s}_n | \mathbf{Y}_{1:n}) = \frac{p(\mathbf{Y}_{1:n} | \mathbf{s}_n) \cdot p(\mathbf{s}_n)}{p(\mathbf{Y}_{1:n})}$$

$$= C_n \cdot p(\mathbf{y}_n | \mathbf{s}_n) \cdot \int p(\mathbf{s}_{n-1} | \mathbf{Y}_{1:n-1}) \cdot p(\mathbf{s}_n | \mathbf{s}_{n-1}) d\mathbf{s}_{n-1}$$

Implementation: **Particle filtering**

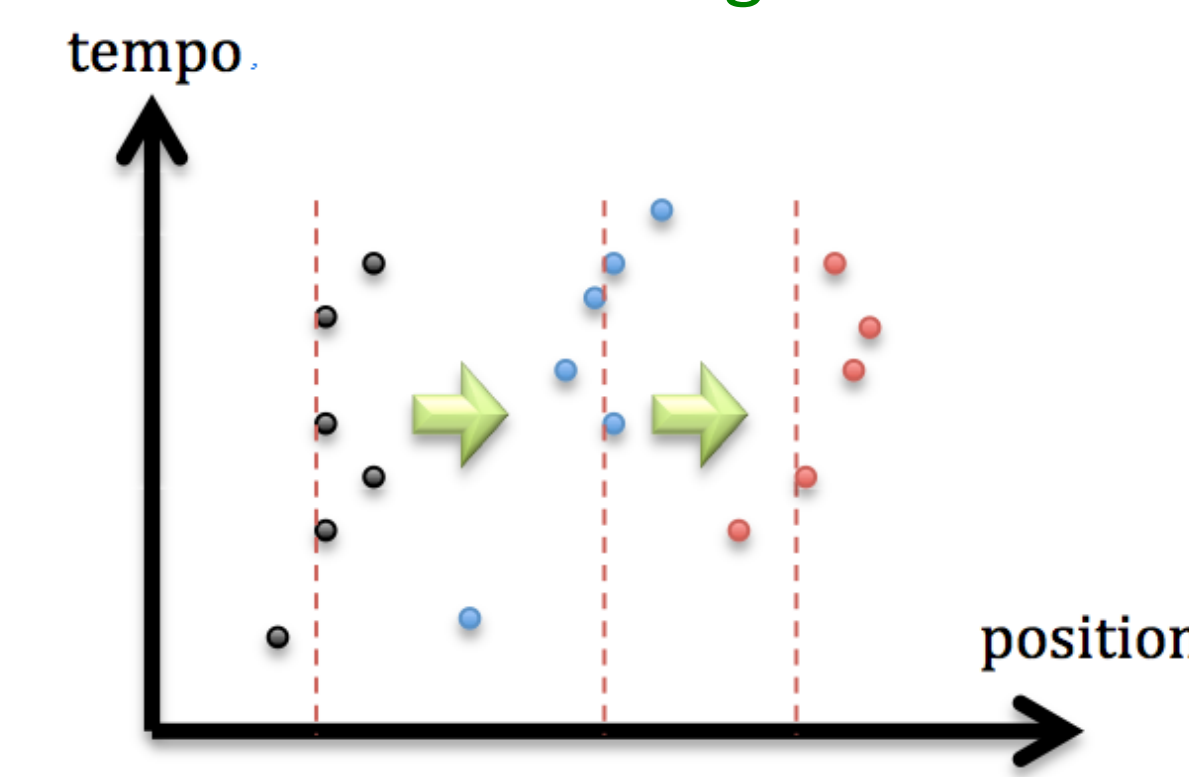


Figure 3. Particle Filtering

NOTE ONSET INFORMATION

We propose to use the note onset information to find faithful frames.

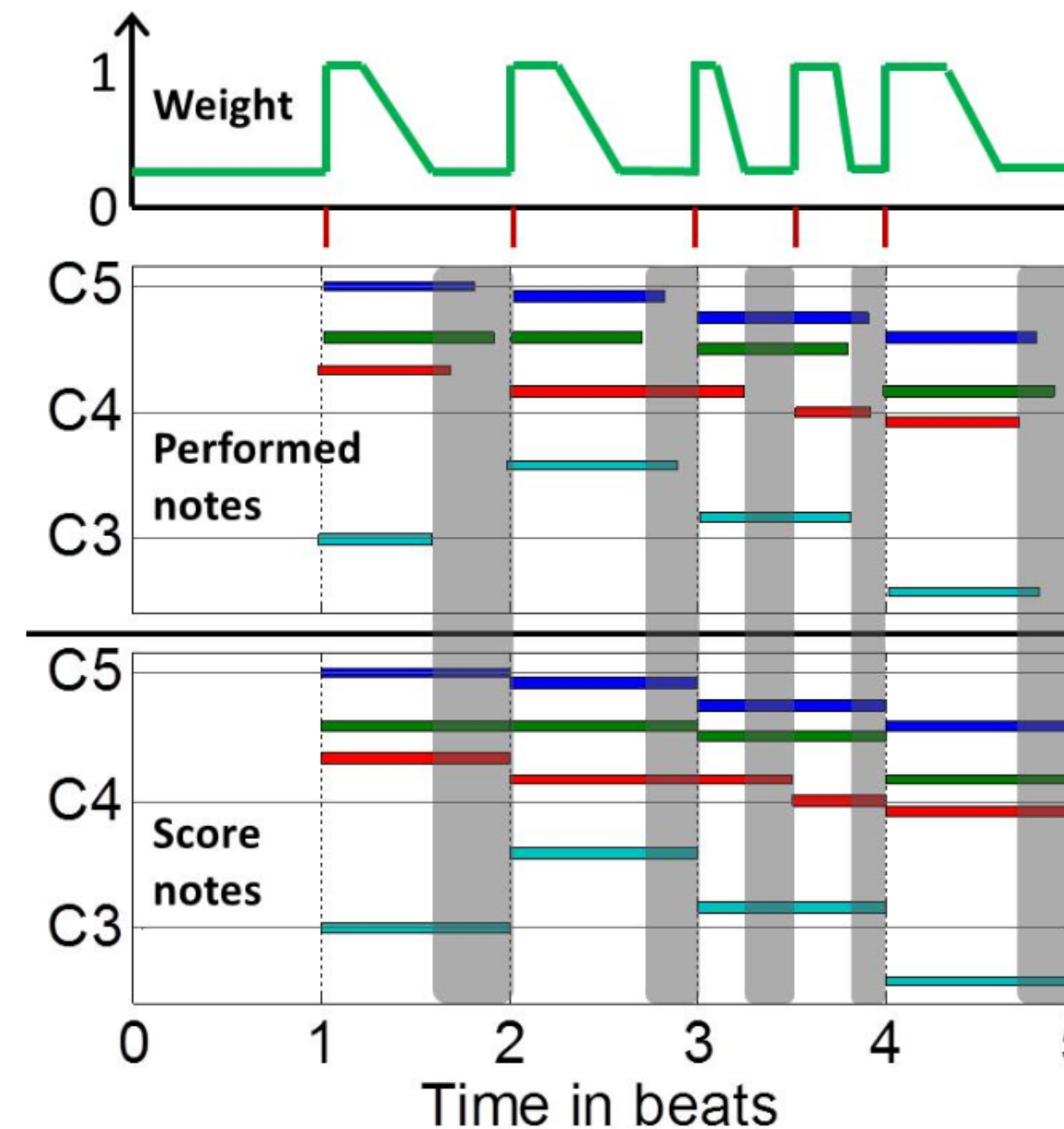


Figure 4. Note segments show with faithful frames (white) and unreliable frames (gray).

Onset Detection

Method: Spectral Flux

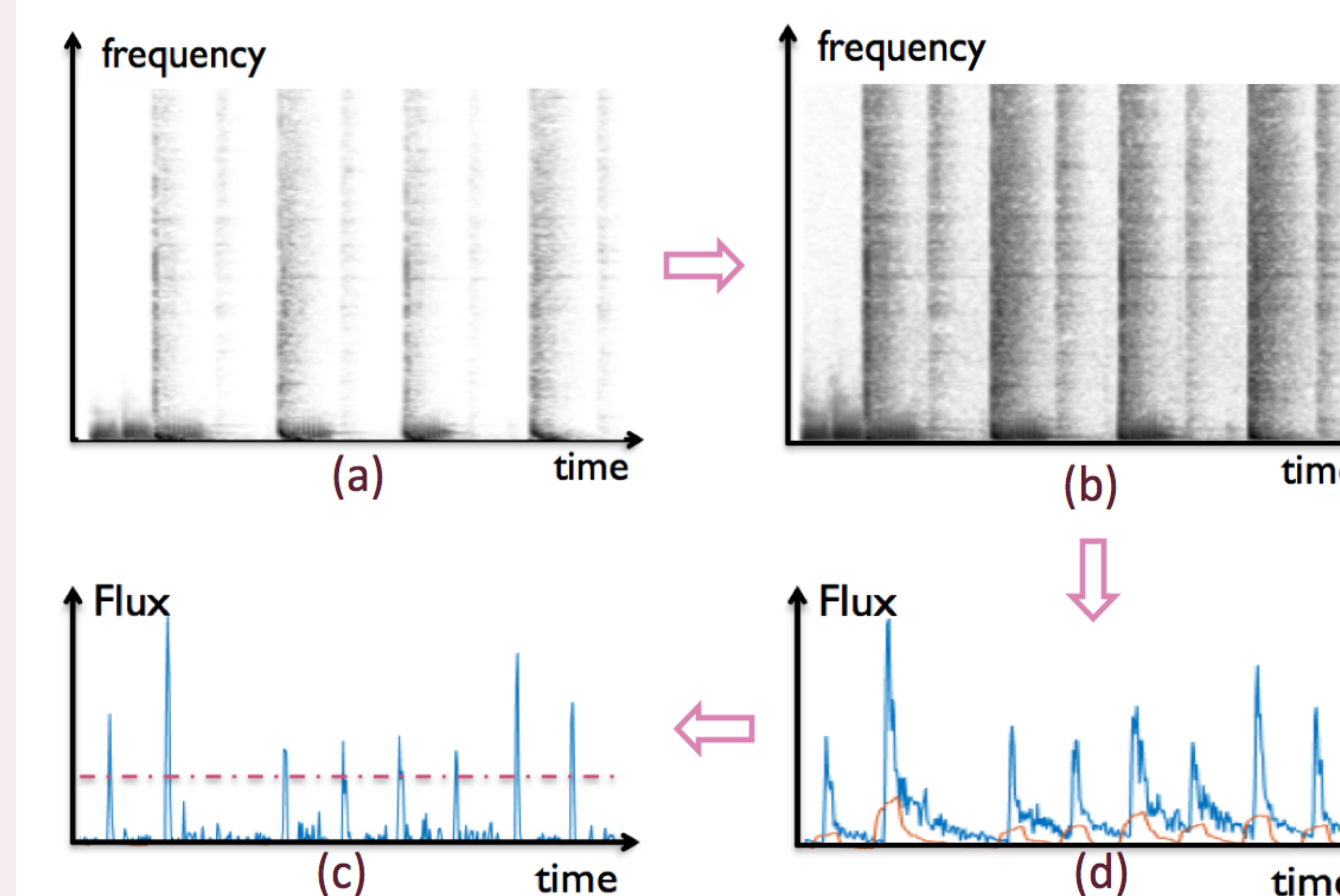


Figure 5. Flowchart of onset detection by spectral flux

SEPTS:

1. Get spectrogram (Figure 5.a)
2. Logarithmic compression (Figure 5.b)
3. Calculate Spectral Flux (Figure 5.d)

$$\tilde{S}(n, k) = \log(1 + \gamma \cdot |S(n, k)|)$$

$$\Delta S(n) = \sum_{k=0}^K |\tilde{S}(n+1, k) - \tilde{S}(n, k)|_{\geq 0}$$

4. Subtracted by local mean (Figure 5.d red line) and set threshold to find peaks (Figure 5.c)

Weighting Function

Shape the decay of piano notes as the weighting function get better results on piano music and improve the results of staccato notes.

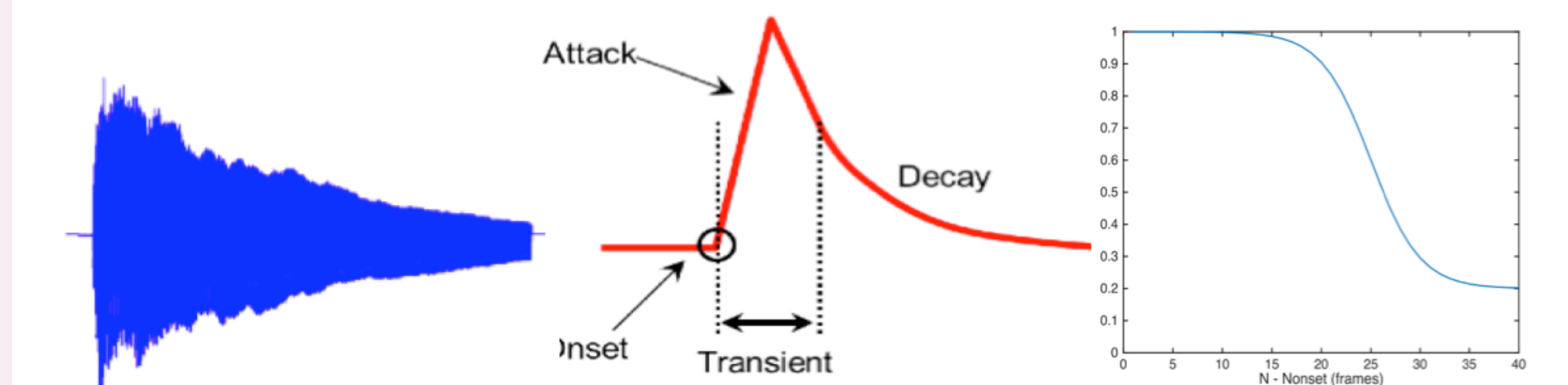


Figure 6. Waveform of one piano note (left), envelope (middle) and one possible function simulation (right)

EXPERIMENTS

Materials:

1. Bach's Chorales
2. RWC Classical Music Dataset

Results:

Since there is no ground-truth alignment in most music database, we can only measure the result by hearing the warped audio or MIDI.

The alignment results are improved when dealing with staccato performing and piano music. But it doesn't work well for strings music with soft onsets. More music database need to be tested.

FUTURE WORK

1. The current method highly relies on the multi-pitch likelihood model. We need more models based on different timbre and volume distribution of different pitch.
2. The robust of onset detection is also important, so we need to apply some more accurate onset detection algorithms, especially for soft onsets.