Multi-Pitch Estimation based on Partial Event and Support Transfer

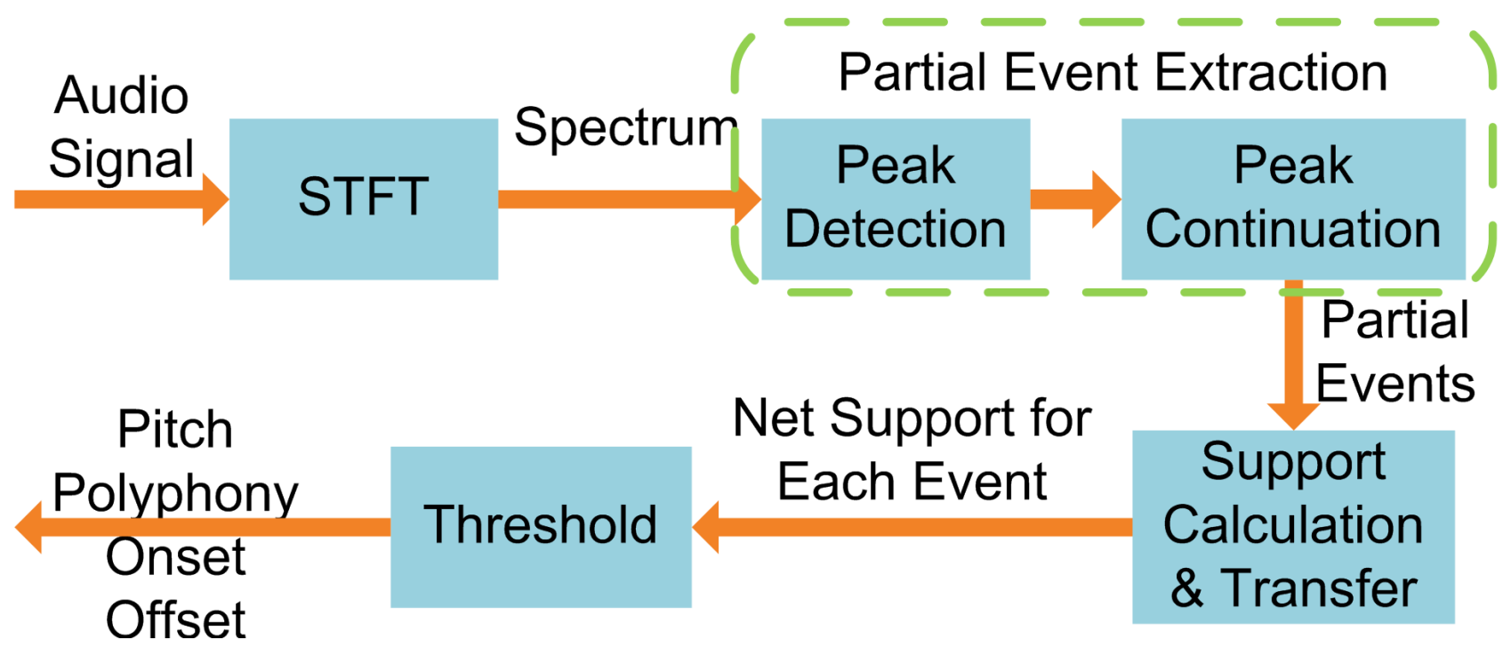
Zhiyao Duan, Dan Zhang, Changshui Zhang and Zhenwei Shi

State Key Laboratory of Intelligent Technology and Systems,
Department of Automation, Tsinghua University, Beijing, China, 100084

Summary

- The Multi-Pitch Estimation (MPE) problem is addressed on the level of "Partial Events", rather than single frames, to integrate the synchronous change cues (common onset and offset).
- For each partial event, a "Support Degree" gotten from the other events is defined to describe the likelihood that it is a fundamental.
- The support is transferred from higher frequency partial events to lower ones, and eventually is concentrated on fundamentals.
- The pitches as well as the duration of each note are estimated, without knowing the number of concurrent sounds.

The Flow Chart of the Method



Partial Event Extraction

Motivation: Synchronous changes of audio components such as onset and offset are important cues in Auditory Scene Analysis (ASA) [1]. However, most MPE algorithms [2, 3, 4] are implemented in each single frame. We intend to address the MPE problem on the level of Partial Events instead.

Partial Event: It is defined as a vector, similar to the concept of a note event in MIDI.

$$e_i = (f_i, A_i, t_{ia}, t_{ib}) \qquad (1)$$

where f_i is its average frequency, A_i is its average logarithm amplitude, t_{ia} is its onset time and t_{ib} is its offset time. Partial events are extracted from the audio signal dynamically along with the process of the STFT.

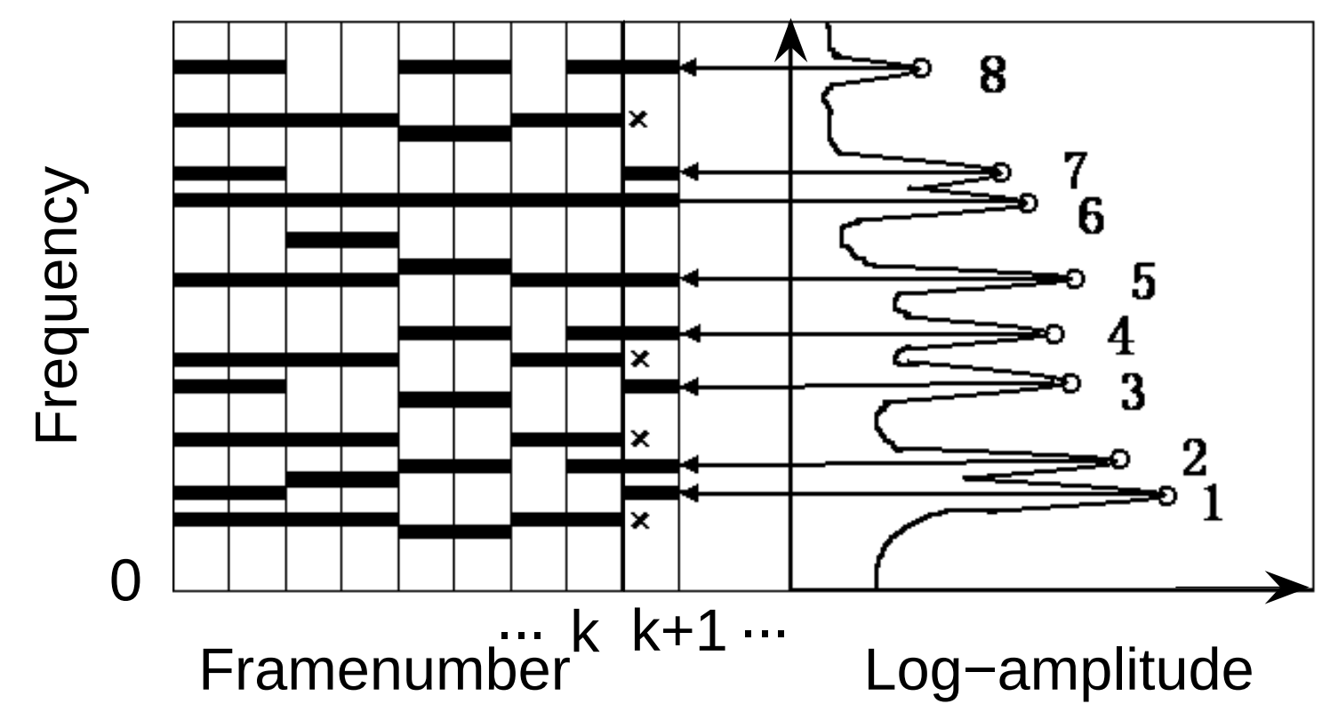


FIGURE 1: Partial events extraction process demonstration. The left part is the time-frequency plane, depicting the partial events, which have been extracted from the audio signal, as horizontal bars. The right part is the spectrum of the current processed frame, with peaks detected.

After the growth of the time-frequency plane is finished, two morphological operations (a close and an open) in image processing are employed on the plane, to eliminate the noise and get the final partial events time-frequency plane, as shown in Fig. 2.

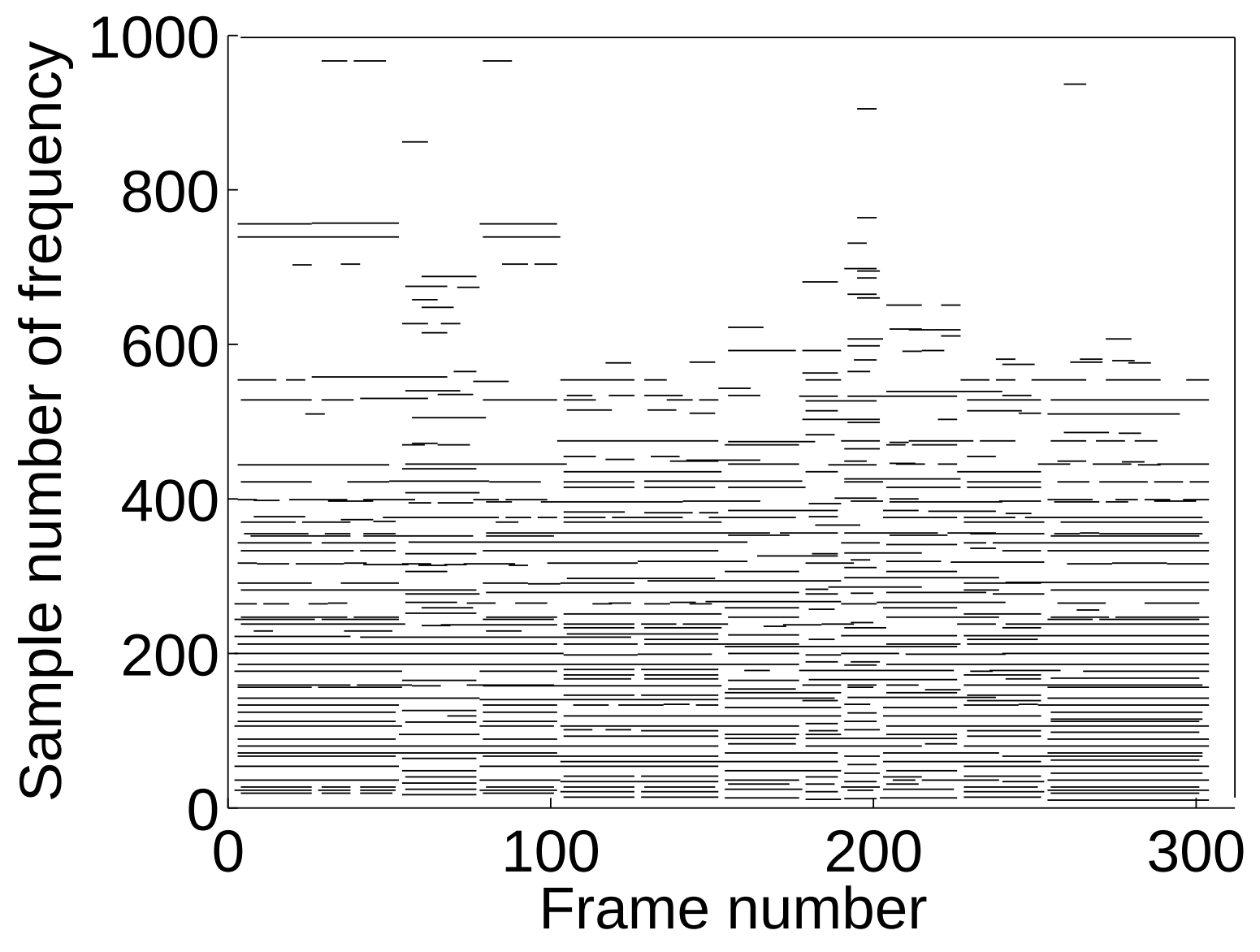


FIGURE 2: Time-frequency plane with partial events depicted.

Support Degree Calculation and Transfer

Motivation: From the extracted partial events, we want to find the fundamental frequency events. For each partial event, a "Support Degree" of being fundamental is defined. This support is acquired from the other partial events based on ASA cues, such as harmonicity, common onset etc.

Support Degree for partial event e_i, gotten from e_j:

$$s_{ij}(f_i') = \begin{cases} R_{ij} \cdot P_{f_i'} \cdot Q_{f_i' f_j} \cdot A_i \cdot A_j & e_j \in \mathscr{E}_i \\ 0 & e_j \notin \mathscr{E}_i \end{cases} \qquad (2)$$

where f_i' is the supposed frequency of e_i. $\mathscr{E}_i$ is the set of events, whose frequencies are higher than e_i, and onset times are the same as e_i.

$$R_{ij} = \frac{\min(t_{ib}, t_{jb}) - \max(t_{ia}, t_{ja})}{t_{ib} - t_{ia}} \qquad (3)$$

$$P_{f_i'} = \exp\left(-\frac{(\frac{f_i'}{f_i} - 1)^2}{\sigma^2}\right) \qquad (4)$$

$$Q_{f_i' f_j} = \exp\left(-\frac{(\frac{f_j}{f_i'} - [\frac{f_j}{f_i'}])^2}{\sigma^2}\right) \qquad (5)$$

where $[\cdot]$ denotes rounding to the nearest integer. R_{ij} represents the overlap ratio between e_i and e_j. $P_{f_i'}$ defines the proximity between the average frequency f_i and the supposed exact frequency f_i'. $Q_{f_i' f_j}$ represents the weight caused by the harmonic relationship between f_i' and f_j. σ is set to 0.015.

Total support for e_i got from the other events:

$$S_i(f_i') = \sum_{j=1}^{N} s_{ij}(f_i') \qquad (6)$$

$$\hat{S}_i = \max(S_i(f_i')) \qquad (7)$$

Support Transfer: Each partial event gets some support from higher frequency events, and also gives support to lower ones. The support is transferred from higher events to lower events, and finally concentrated on fundamental events.

Net Support for e_i:

$$NS_i = \sum_{j=1}^{N} \hat{s}_{ij} - \alpha \sum_{k=1}^{N} \hat{s}_{ki} \qquad (8)$$

where α is the tradeoff between received and given support.

After the calculation of the net support for each partial event, the events whose net supports exceed a threshold are selected as the fundamental frequency events.

Threshold:

$$\tau = mean(NS_i) + \beta \cdot std(NS_i) \qquad (9)$$

where β is set to 1.2 typically.

Experimental Results on Randomly Mixed Chords

There are in total 2600 mixed chords of different polyphony, which are generated from the Iowa instrument database [5]. In the left panel of Fig. 3, the predominant F0 error rate (white), precision (grey) and recall (black) are calculated. In the right panel the Average Overlap Ratio (AOR) for both predominant estimation (white) and Multiple-F0 (black) are calculated.

Average Overlap Ratio (AOR):

$$AOR = mean\left(\frac{\min(offsets) - \max(onsets)}{\max(offsets) - \min(onsets)}\right) \qquad (10)$$

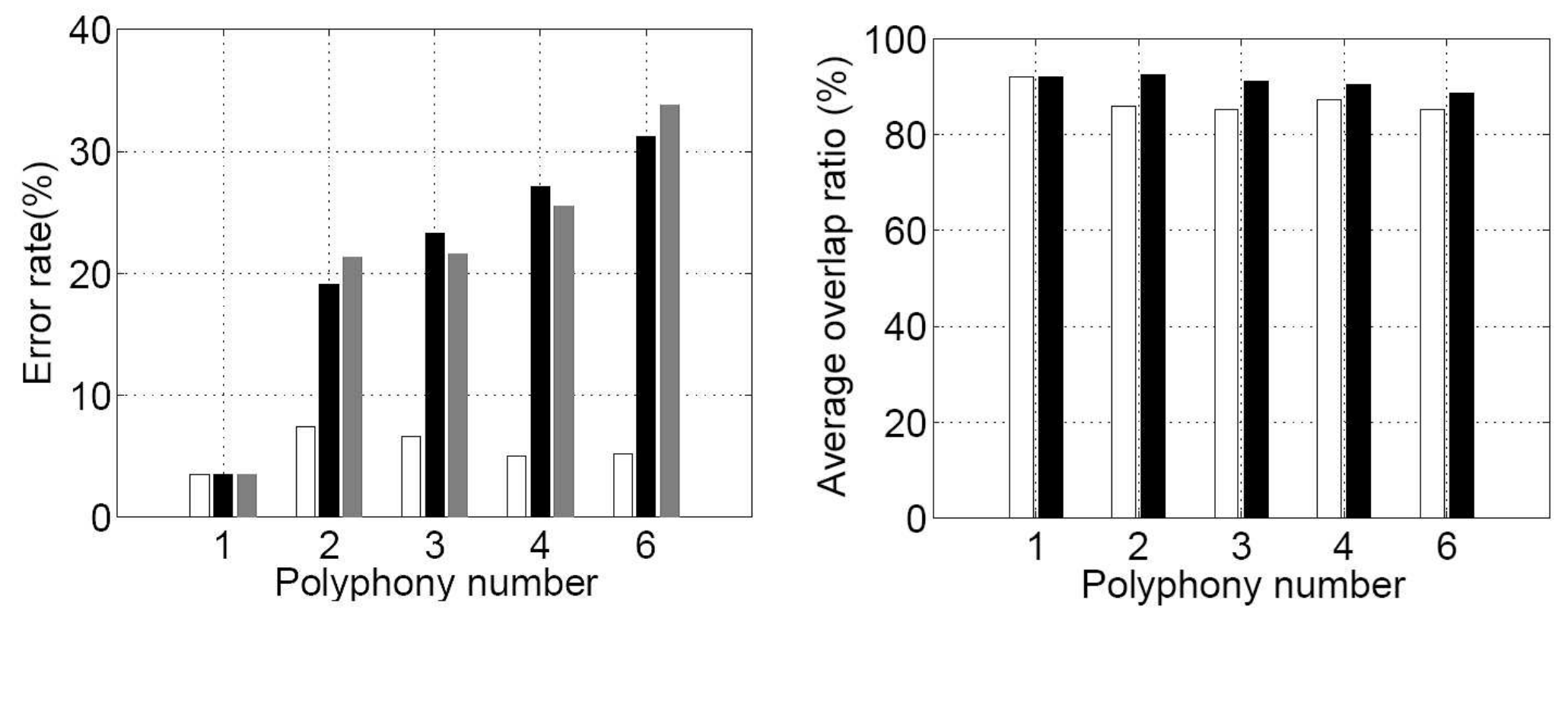


FIGURE 3: MPE Results. Note that the polyphony information is not given.

The polyphony is also estimated in Fig. 4.

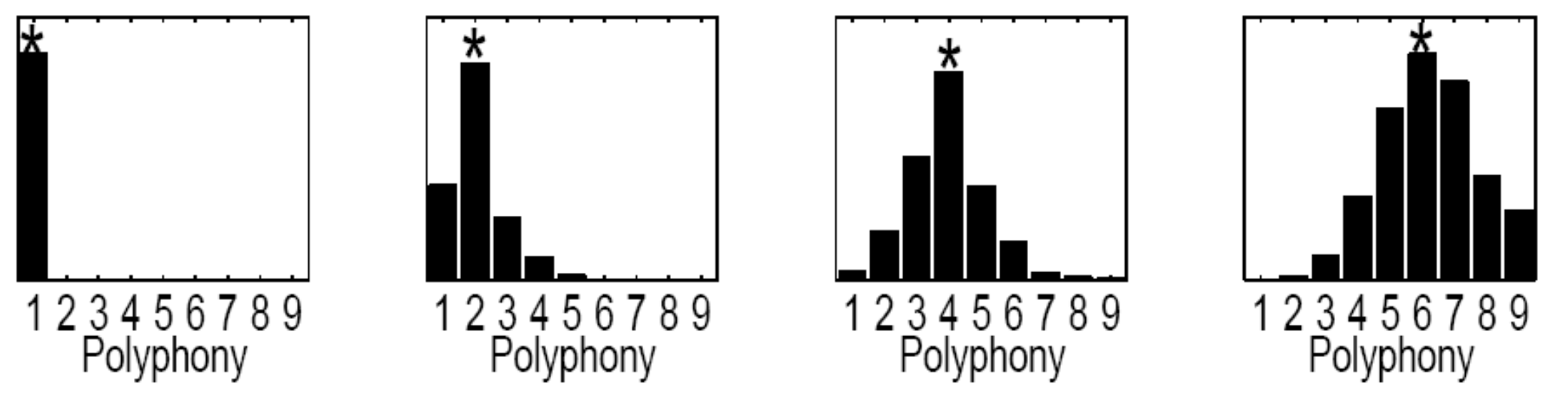


FIGURE 4: Histograms of polyphony estimates. The asterisks indicate the true polyphony(1,2,4 and 6, from left to right)

Experimental Results on A Synthesized Music Piece

This piece of synthesized ensemble music is a four part chamber music played by flute, oboe, clarinet and bassoon respectively.

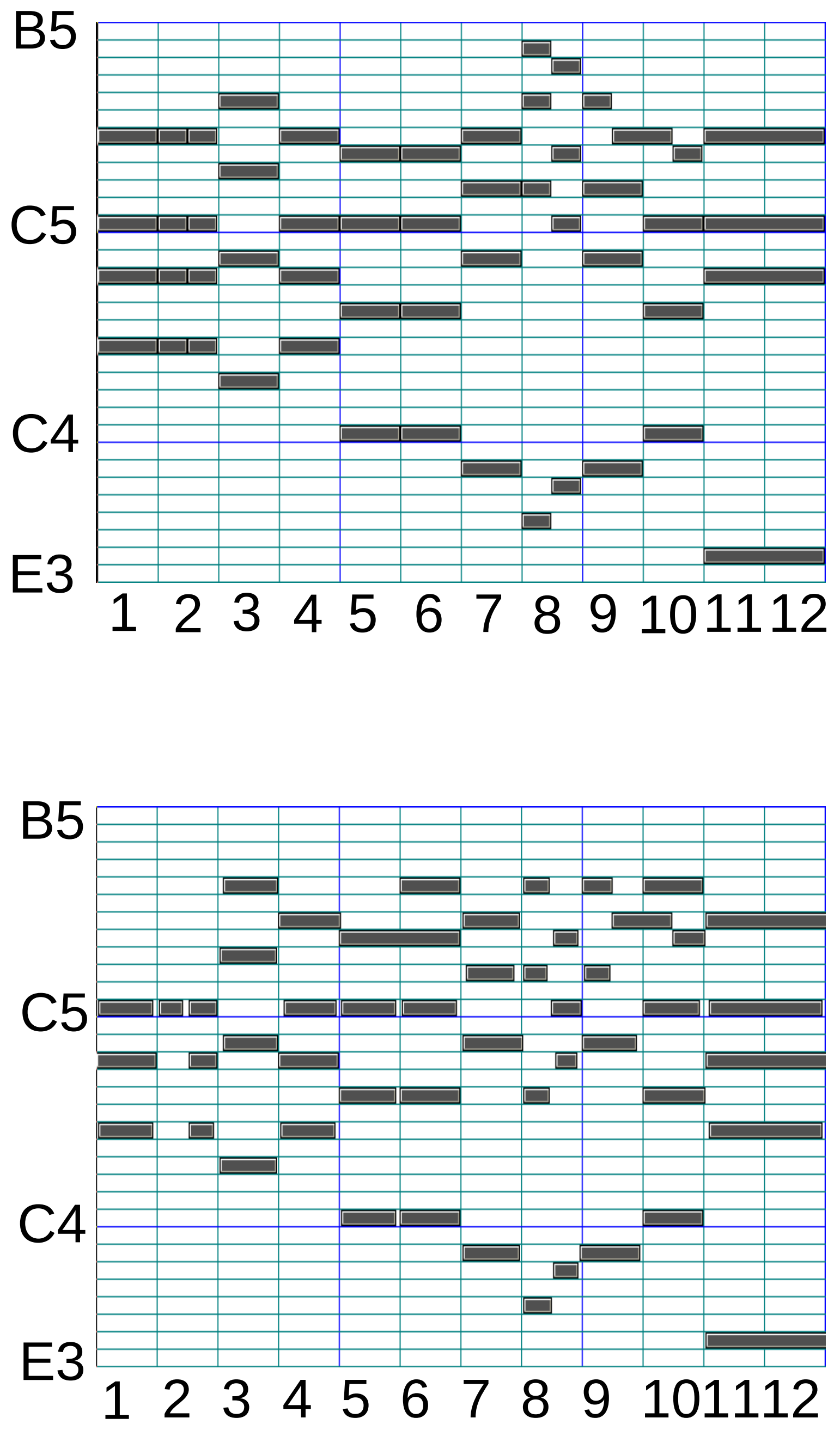


FIGURE 5: Pianorolls of synthesized music (upper) and the transcribed result (lower). The horizontal axis refer to beat.

Future Work

Our algorithm addresses the MPE problem on the level of partials, to integrate the time information cues such as common onset and offset. However, the modeling of the partial is still inadequate, because only the average value is used to model the frequency and amplitude of each partial event. Our work can be extended by calculating the instantaneous frequency and amplitude of each partial, to integrate the synchronous change cues of frequencies and amplitudes.

References:

[1]A. S. Bregman, *Auditory Scene Analysis*. The MIT Press, Cambridge, Massachusetts, 1990.

[2] M. Goto, "A Real-Time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals," *Speech Communication*, vol. 43, no. 4, pp. 311-329, 2004.

[3] A. Klapuri, "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes," In *Proc. International Conference on Music Information Retrieval (ISMIR2006)*, Oct. 2006.

[4] G. Poliner and D. Ellis, "A Discriminative Model for Polyphonic Piano Transcription," *Eurasip Journal on Applied Signal Processing*, to appear, 2007.

[5] The University of Iowa Musical Instrument Samples. *http://theremin.music.uiowa.edu/* [Online]