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

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, rather than single-frames, to describe the likelihood that it is a fundamental.

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

\[ e_i = \{ f_i, A_i, t_{on}, t_{off} \} \]  

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

There are in total 2000 mixed chords of different polyphony, which are generated from the four instrument database [5]. In the left panel of Fig. 3, the predominant F0 error rate (white), precision (grey) and recall (black) are calculated. The overlap ratio between each partial event and the resultant event is defined to describe the likelihood that it is a fundamental.

Average Overlap Ratio (AOR):

\[ AOR = \frac{\min\left\{ f_i \right\} - \max\left\{ f_i \right\}}{\max\left\{ f_i \right\} - \min\left\{ f_i \right\}} \]  

where \( f_i \) is the set of events, whose frequencies are higher than \( f_j \), and onset times are the same as \( t_j \).

Support Degree Calculation and Transfer

Support Degree for partial event \( e_i \) is defined as a vector, similar to the concept of a note event in MIDI:

\[ v_i(f_j) = \begin{cases} R_{ij} Q_{ij} P_{ij} & f_j \in \mathcal{D}_i \setminus \mathcal{D}_j \neq \emptyset \vspace{-1em} \\ 0 & \text{otherwise} \end{cases} \]  

where \( \mathcal{D}_i \) is the set of events, whose frequencies are higher than \( f_j \), and onset times are the same as \( t_j \).

Threshold:

\[ \tau = \text{mean}(NS) + \beta \cdot \text{std}(NS) \]  

where \( \beta \) is set to 1.2 typically.

Experimental Results on Randomly Mixed Chords

Our algorithm addresses the MPE problem on the level of partials, to integrate the frequency 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: