ESTIMATING ONSET AND OFFSET ASYNCHRONIES IN POLYPHONIC AUDIO-TO-SCORE ALIGNMENT

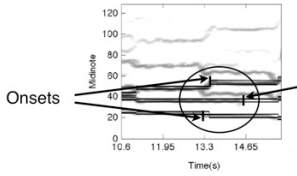
Johanna Devaney

SCHOOL OF **MUSIC**
THE OHIO STATE UNIVERSITY

- Identifying onset and offset asynchronies in sounded simultaneities is important for studying expressive performance and guiding signal processing algorithms
- This poster describes a hybrid dynamic time warping (DTW)/hidden Markov model (HMM) algorithm to improve the accuracy of DTW-based alignment for polyphonic recordings

PROBLEM

- Dynamic time warping (DTW)-based alignment algorithms are not sufficient by themselves to estimate asynchronies in polyphonic performances (Devaney and Ellis, 2009)
- Example of errors in estimating onset and offset asynchrony:



boxes indicate the DTW estimates

vertical lines and arrows identify the locations of the true onsets and offsets

RELATED WORK

- Devaney, Mandel and Ellis (2009)
  - Used periodicity, power, and fundamental frequency estimates of the singing voice to guide an HMM to refine a DTW alignment algorithm
  - The algorithm identified the transient and steady state sections of the note
  - Decreased median alignment error in the initial DTW alignment from 52 to 28 ms
- Niedermeyer and Widmer (2010)
  - Used non-negative matrix factorization (NMF) to refine onset estimates obtained through DTW alignment
  - Algorithm evaluated against two accuracy thresholds
    - 10 ms (Friberg and Sundberg1993) – 40% to 50%
    - 50 ms (other onset detection evaluations) – 85% to 89%
  - Note offsets were neither estimated nor evaluated.

REFERENCES

Devaney, J., & Ellis, D.P.W. (2009). Handing asynchrony in audio-score alignment. In *Proceedings of the ICMC*.

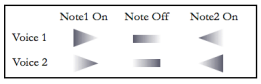
Devaney, J., Mandel, M.I., & Ellis, D.P.W. (2009). Improving MIDI-audio alignment with acoustic features. *Proceedings of WASPAAo*.

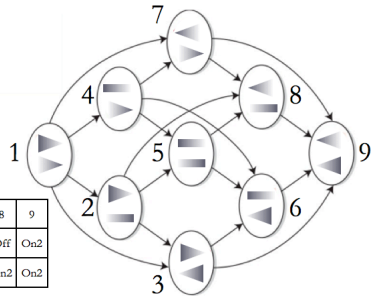
Friberg, A., & Sundberg, J. (1993). Perception of just noticeable time displacement of a tone presented in a metrical sequence at different tempos. In *Proceedings of the Stockholm Music Acoustics Conference*.

Niedermayer, B., & Widmer, G. (2010). A multi-pass algorithm for accurate audio-to-score alignment. In *Proceedings of ISMIR*

HYBRID DTW/HMM MODEL

- The offset-onset transitions identified by the DTW alignment are refined by a 3-state HMM
  - number of states is 3^N, where N is the number of voices





State	1	2	3	4	5	6	7	8	9
Voice 1	On1	Off	On1	On2	Off	On1	On2	Off	On2
Voice 2	On1	On1	Off	On1	Off	On2	Off	On2	On2

- Observations calculated from a constant-Q filter bank decomposition of the signal with one filter per semitone
  - a power measurement (in decibels) was summed over a 3-semitone span around the fundamental of the MIDI note for both the ending and starting note
- Evaluation
  - four multi-tracked recordings of a three-part piece
  - multi-tracked recordings were hand annotated to generate ground truth
  - tests were run on the composite signals
- Metric 1: number of onsets and offsets within a fixed time of the ground truth

	10 ms (228 notes)		50 ms (228 notes)	
	On	***Off***	***On***	***Off***
DTW	14	**38**	88	102
DTW/HMM	**20**	25	**119**	**109**

- Metric 2: 2.5th, 25th, 50th, 75th, and 97.5th percentiles of the difference between the predictions and the ground truth

Percentiles		**2.5**	**25**	**50**	**75**	**97.5**
Ons	**DTW**	**3.1**	29.2	66.7	122.5	759.5
	DTW/HMM	3.8	**19.2**	**47.7**	**100.1**	**653.2**
Offs	**DTW**	**1.0**	**21.8**	64.8	141.9	752.9
	DTW/HMM	1.5	25.1	**53.4**	**131.3**	**677.0**