On the Use of Note Onsets for Improved Lyrics-to-audio Alignment in Turkish Makam Music

Georgi Dzhambazov, Ajay Srinivasamurthy, Sertan Şentürk, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
contact: georgi.dzhambazov@upf.edu

A hidden Markov model phonetic recognizer with variable-time transition matrix, guided by onset-aware phoneme transition rules

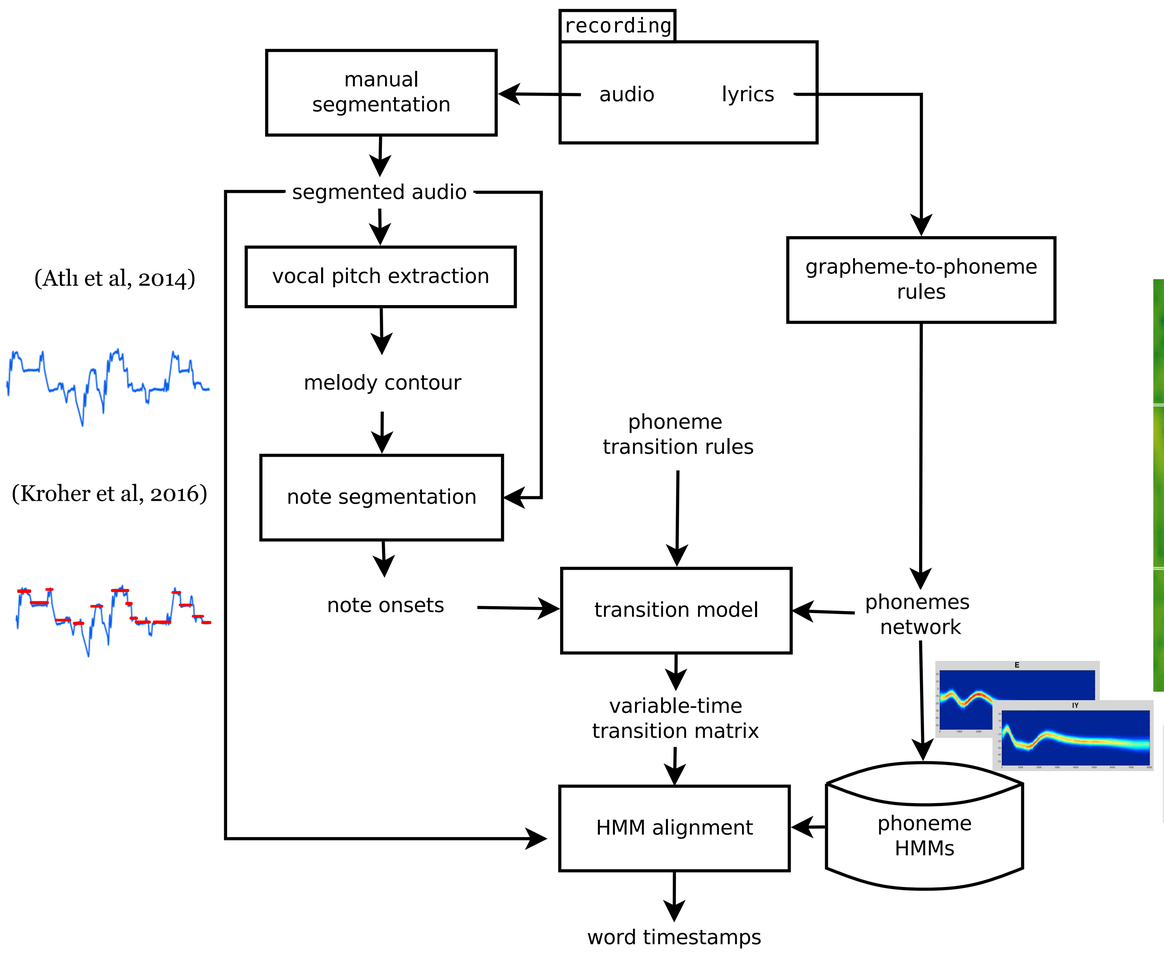
Abstract

- detect begin and end timestamps of words
- make use of note onsets detected automatically

1. phoneme transition rules aware of note onsets
2. alter transition probabilities according to the rules

- evaluate on a cappella and polyphonic Turkish Makam

Method overview



Dataset

Training Corpus

500 minutes Turkish speech

Test Corpus

- 12 acappella recordings, 18 minutes
- especially recorded for this study
- words and note onsets annotated

http://compmusic.upf.edu/turkish-makam-acapella-sections-dataset

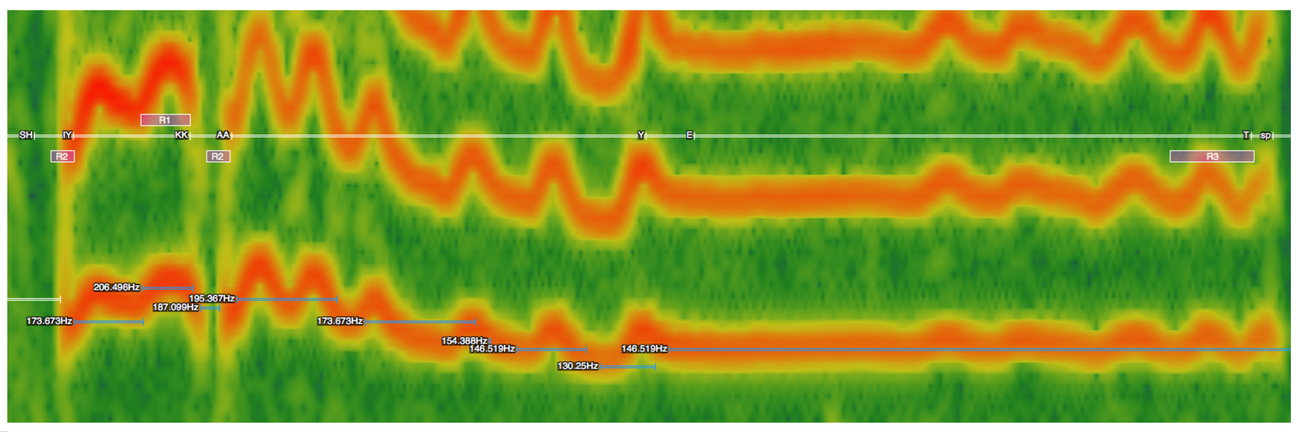
Phoneme Transition Rules

R1: $i_t = V \quad i_{t+1} = C \backslash L$
R2: $i_t = C \backslash L \quad i_{t+1} = V \text{or} L$ *
R3: $i_t = V \quad i_{t+1} = C$

} inter-syllable
} intra-syllable

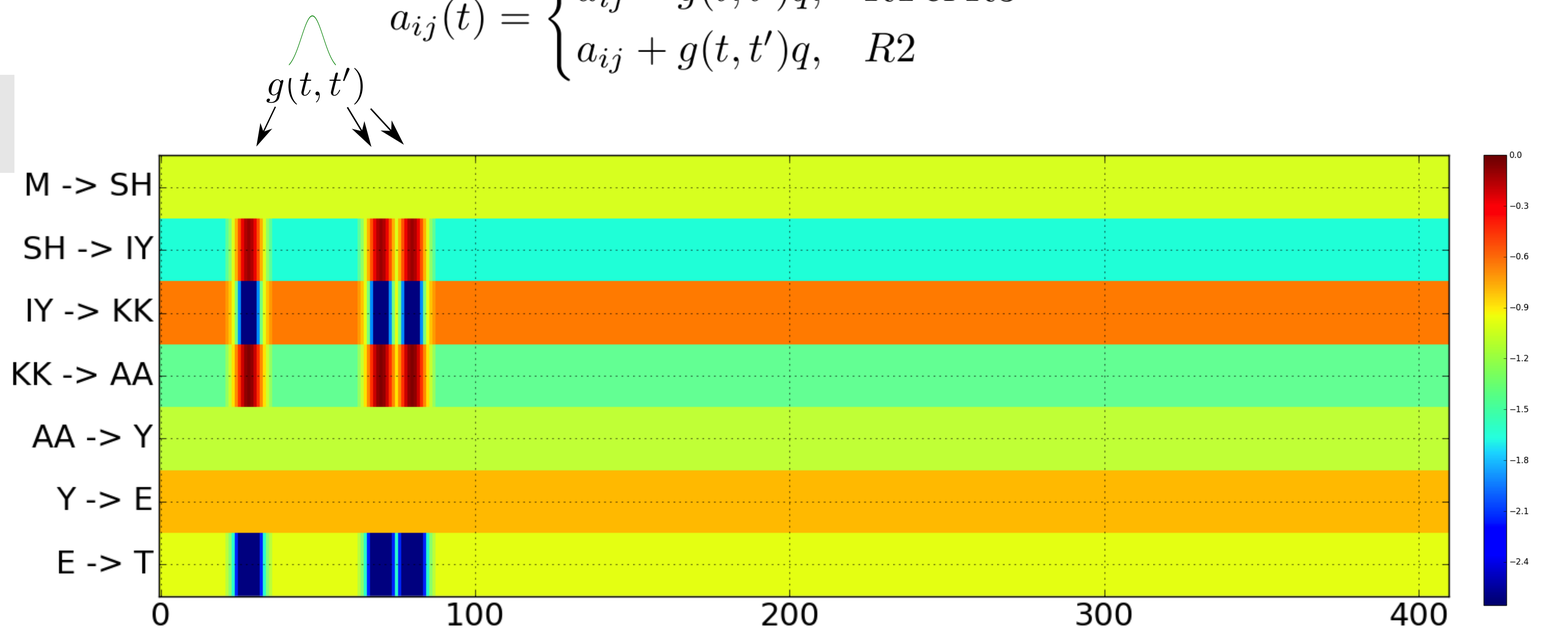
C: consonant V: vowel L: liqud(L,M,N) or semivowel Y

* liquids might bear onsets (Sundberg, 2006)



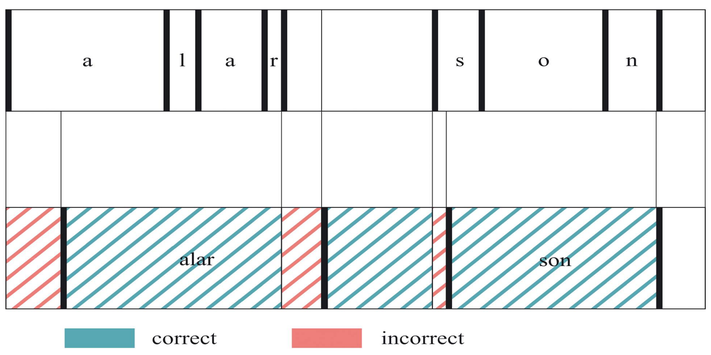
Variable-time Transition Matrix

$$a_{ij}(t) = \begin{cases} a_{ij} - g(t,t')q, & R1 \text{ or } R3 \\ a_{ij} + g(t,t')q, & R2 \end{cases}$$



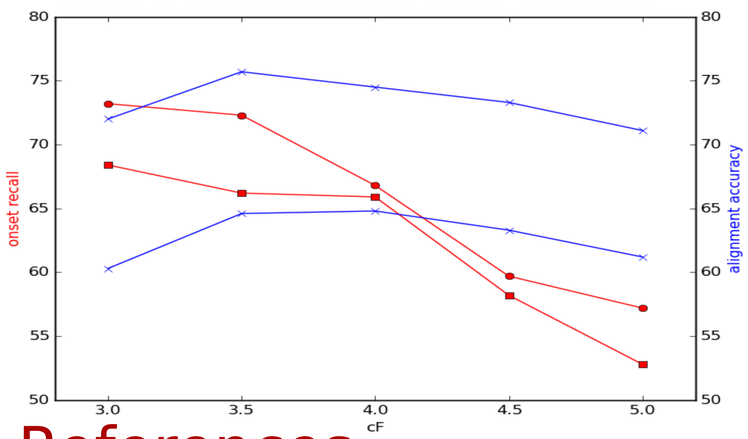
Evaluation

Alignment accuracy (in %) for phrases of 1-4 words



Results

Alignment accuracy Vs recall of automatic note onset detection



Comparison between baseline HMM, proposed Variable-time HMM (VTHMM) and VTHMM with annotated oracle onsets

	HMM	VTHMM	oracle
a cappella	70.2	75.7	83.5
polyphonic	61.5	64.8	67.1

available in python at:
https://github.com/georgid/AlignmentDuration/tree/noteOnsets

demo available at:
http://dunya.compmusic.upf.edu/makam/lyric-align/727cff89-392f-4d15-926d-63b2697d7f3f

Conclusion

- Onset-aware transition probabilities improve alignment
  - Rules R1 and R3 discourage premature transitions of vowels
  - No contribution for onsets not complying with rules and at sustained vowels
- Lightweight model
- Could be easily combined with other modeling concepts

References

Kroher, N. and Gómez, E Automatic transcription of flamenco singing from polyphonic music recordings. IEEE/ACM TASLP, 24(5):901–913, 2016.
Atlı H.S. et al. Audio feature extraction for exploring Turkish makam music. In 3rd International Conference on Audio Technologies for Music and Media, Ankara, Turkey, 2014.
Sundberg, J. The KTH synthesis of singing. Advances in Cognitive Psychology, 2(2-3):131–143, 2006.