A probabilistic model for lyrics-to-audio alignment based on knowledge of singing voice specific characteristics



Georgi Dzhambazov, Ajay Srinivasamurthy, Sertan Şentürk, Xavier Serra

Music
Technology
Group

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

contact: georgi.dzhambazov@upf.edu

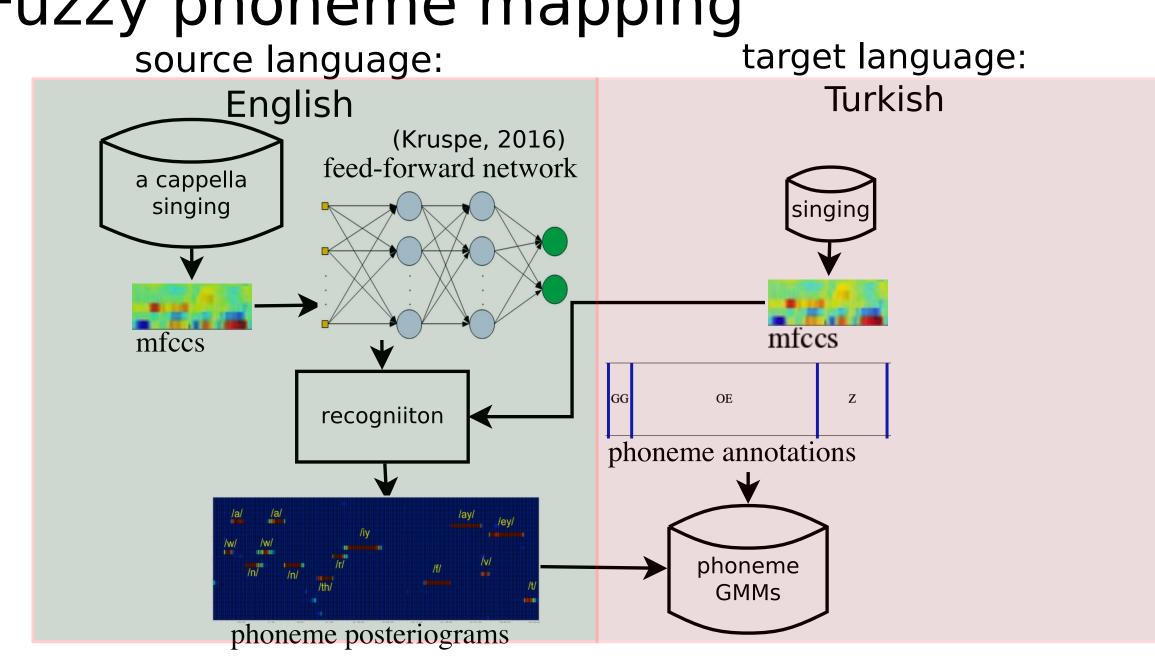
A hidden Markov model phonetic recognizer with variable-time transition matrix, guided by onset-aware phoneme transition rules

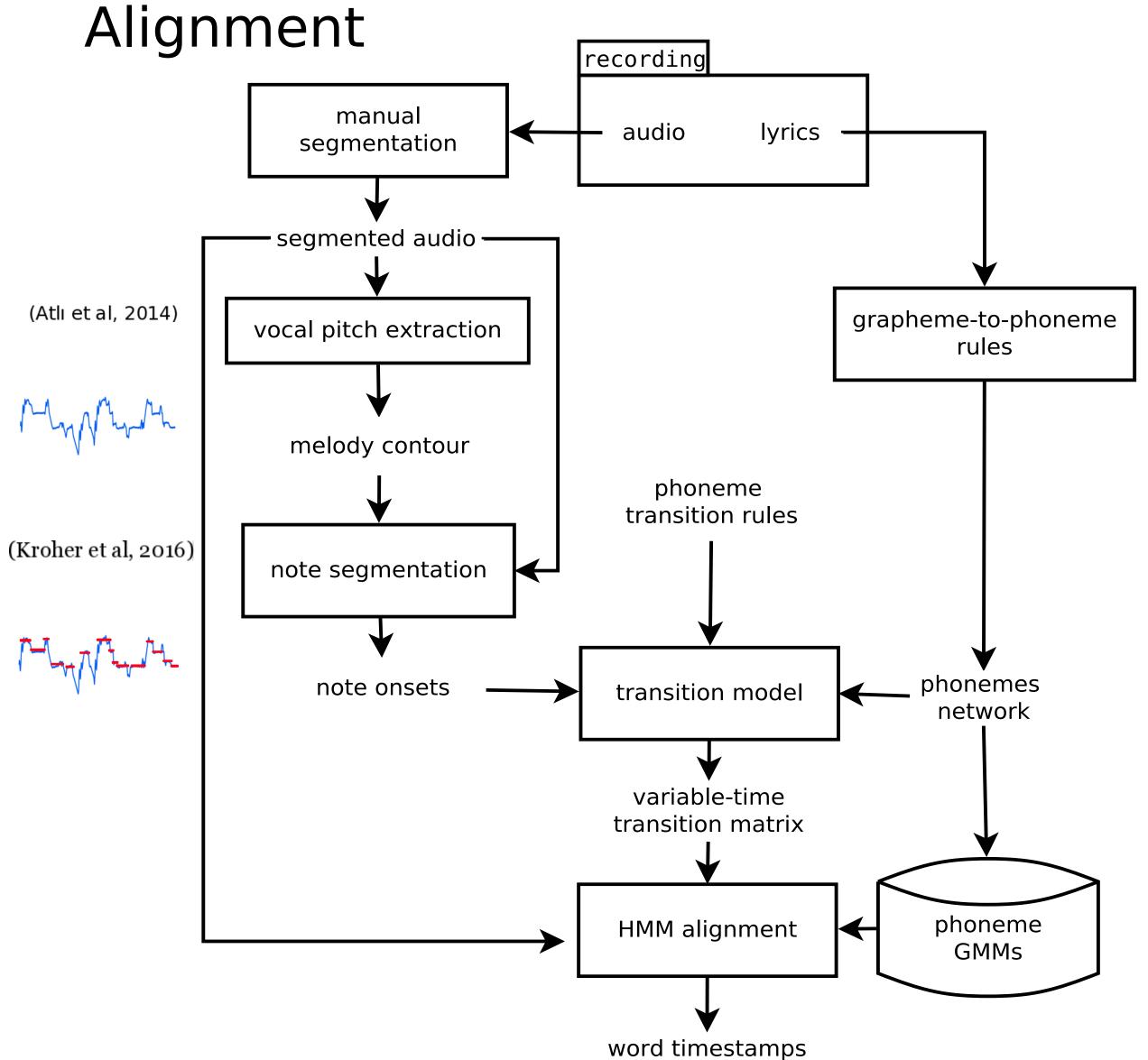
Abstract

- goal: detect begin and end timestamps of words
- novel fuzzy cross-language phoneme mapping
- 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 Turkish Makam singing

Method overview

Fuzzy phoneme mapping





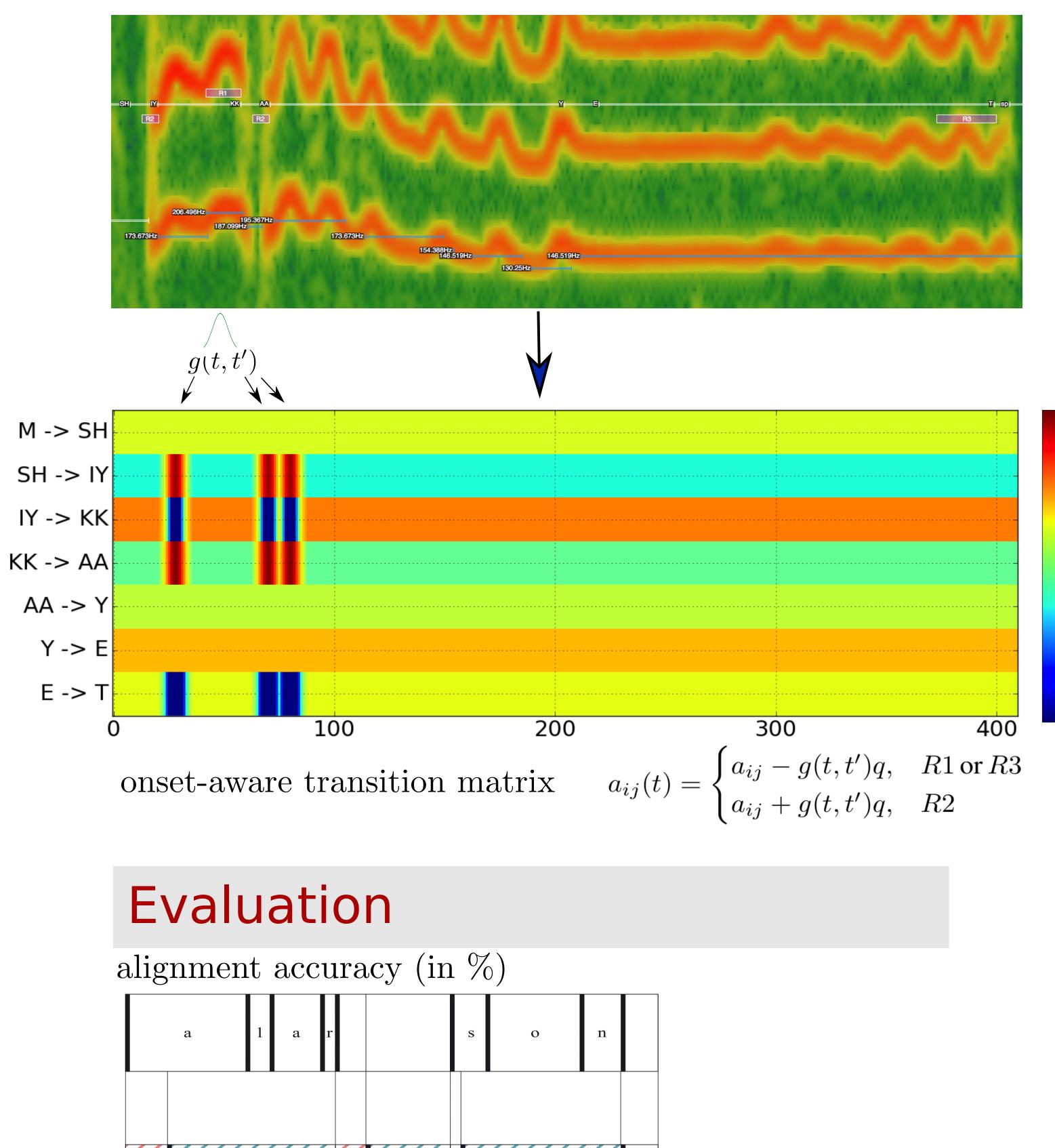
Dataset

- 12 acappella recordings, 25 minutes
- especially recorded for this study
- words, phonemes and note onsets annotated

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

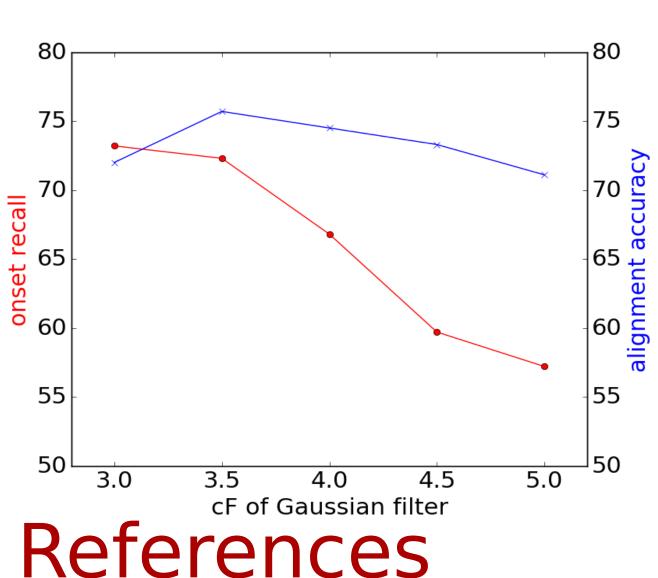
Phoneme Transition Rules

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



Results

Alignment accuracy Vs recall of automatic note onset detection



Comparison of baseline HMM and onset-aware HMM with different phoneme mapping strategy

correct

phoneme mapping	HMM .	onset-aware HMM	
		automatic onsets	oracle onsets
direct	79.2	81.7	82.5
fuzzy	83.5	84.8	86.1

available in python at: https://github.com/georgid/AlignmentDuration/tree/noteOnsets demo available at: http://dunya.compmusic.upf.edu/makam/lyric-align/

Conclusion

- English-to-Turkish phoneme model mapping works
- onset-aware phoneme transition rules improve alignment
- glass ceiling with oracle onsets indicate approach fitness

Dzhambazov G. et al "On the Use of Note Onsets for Improved Lyrics-to-audio Alignment in Turkish Makam Music", in 17th International Society for Music Information Retrieval Conference (ISMIR), New York, NY, USA, 2016 Kruspe, A. M. "Bootstrapping a system for phoneme recognition and keyword spotting in unaccompanied singing", In 17th International Conference on Music Information Retrieval (ISMIR), New York, NY, USA, 2016 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.