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, 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
- Makes use of vocal note onsets detected automatically
  1. Introduce onset-aware phoneme transition rules
  2. Alter transition probabilities according to the rules
- Evaluate on a cappella Turkish Makam singing

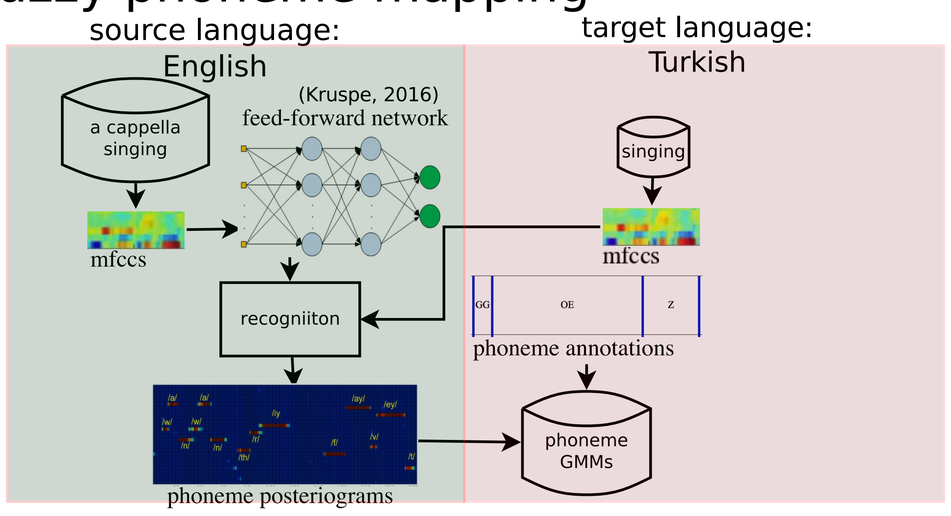
Dataset

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

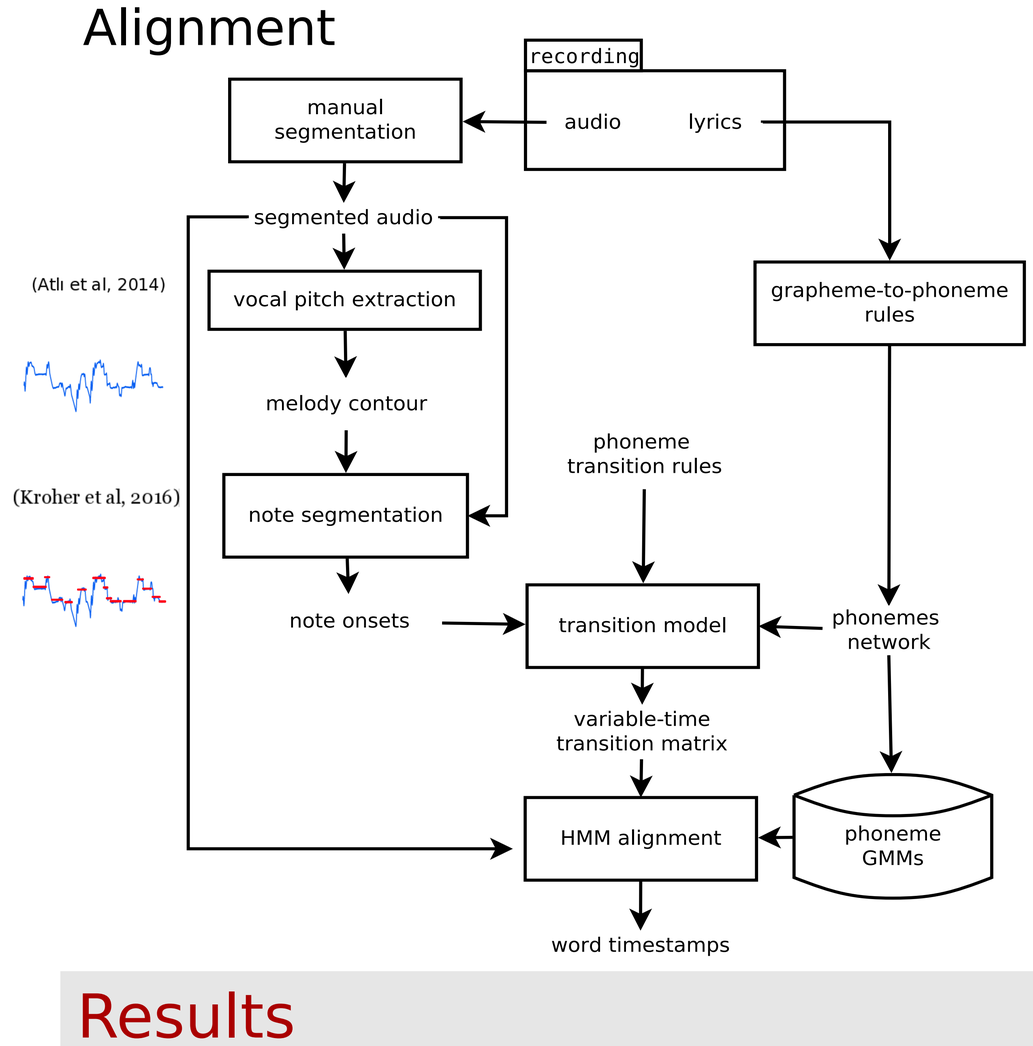
http://compmusic.upf.edu/turkish-makam-acapella-sections-dataset

Method overview

Fuzzy phoneme mapping



Alignment



Phoneme Transition Rules

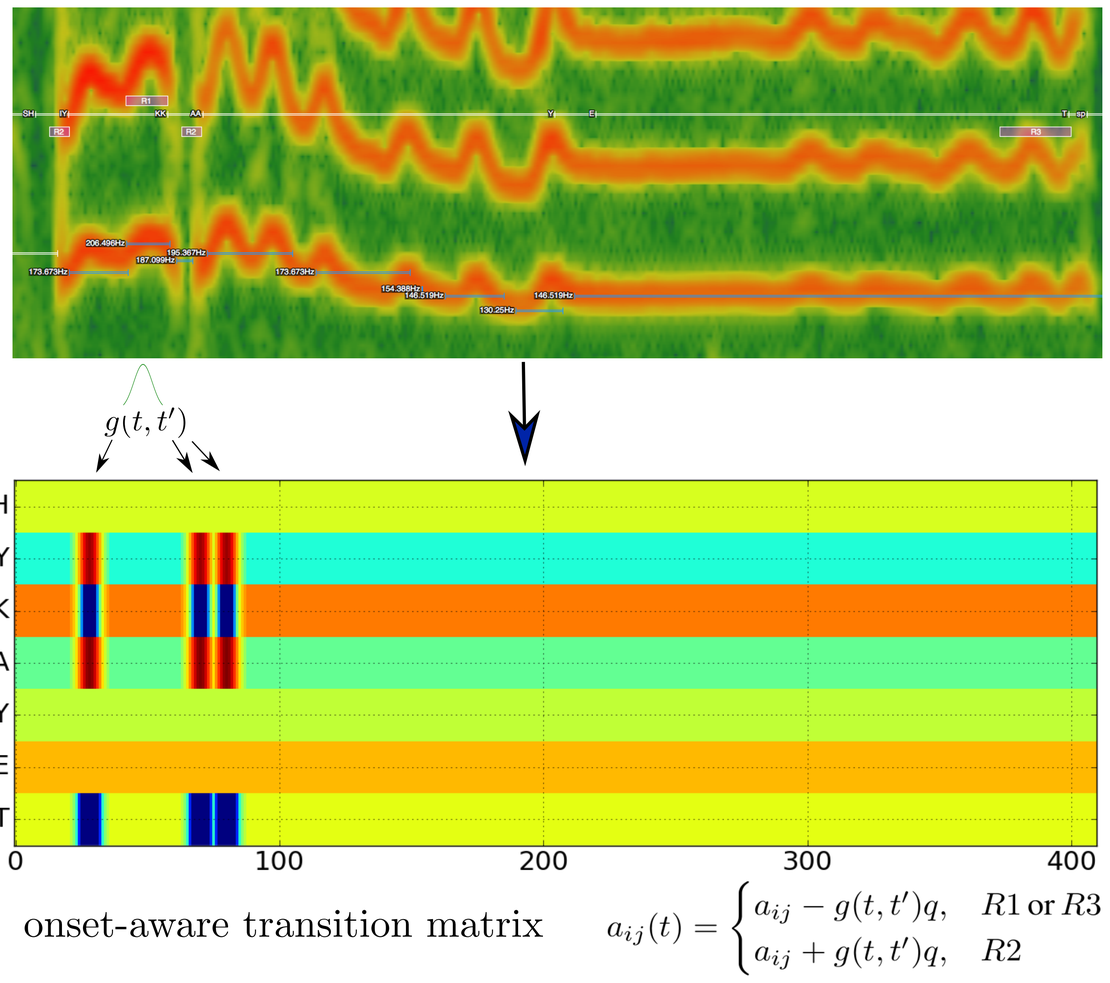
R1: $i_t = V \quad i_{t+1} = C \backslash L$

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

R3: $i_t = V \quad i_{t+1} = C$ } intra-syllable

(Dzhambazov et al. 2016)

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



onset-aware 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 %)

Results

Alignment accuracy Vs recall of automatic note onset detection

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

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

- Phoneme model mapping works relatively well for English-to-Turkish
- Onset-aware phoneme transition rules help improve alignment
- Glass ceiling with oracle onsets indicate fitness of the approach

References

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.