

Project Title: Automatic Tracking of Vocal Tract Resonances of Speech Signals

Project Description:

Vocal tract resonances/formants are essential information-bearing elements in speech signals which can potentially provide an efficient and compact representation of their time-varying acoustic characteristics. This project is designed to automatically track the first four vocal tract resonances ($F_1 - F_4$) of speech signals from the TIMIT database. This will provide a very valuable database for the speech community to quantitatively measure the performance of any formant tracking algorithm (which is lacking now).

According to the speech production model, the observed speech signal is generated as the output of a vocal tract filter given the excitation source. This model decomposes the speech signal into two independent components - source and filter. In terms of power spectrum, the vocal tract resonances/formants correspond to the major peaks of the spectrum envelope. In this project, the vocal tract resonances/formants are estimated from the spectrum of each speech frame based on a Gaussian mixture model. Given the continuous nature of vocal tract resonances, a dynamic programming technique is applied later on to smooth the trajectories of estimated values by taking into account the "spectral cost" and "transition cost". To make the tracking as accurate as possible, manual error correction (especially during CV/VC transitions) is needed as the final step using acoustic knowledge. The manual correction can be facilitated by a Matlab GUI tool from Microsoft. One of the goals of the project is to produce the corrected VTR values to be released to the research community.

Data Sources:

The project will use speech data from the TIMIT database which is a well-known and widely-used acoustic-phonetic speech corpus in the community. It contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States. In addition to a waveform file, the corpus also provides its corresponding text transcription, word and phoneme level alignments.

Project Objectives:

- Survey literatures on formant analysis and formant tracking algorithms
- Get familiar with commonly-used speech processing techniques
- Develop a Gaussian mixture model based algorithm to estimate vocal tract resonances
- Develop mathematically solid methods (such as dynamic programming) that are effective in tracking the resonances.

- Learn some basics of speech acoustics
- Provide qualified vocal tract resonance labeling for the TIMIT speech data
- Produce the corrected VTR values to be released to the research community.

References:

- L. Rabiner and R. Schafer. *Digital Processing of Speech Signals*, Prentice Hall, 1978.
- L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- L. Deng and D. O’Shaughnessy. *Speech Processing - A Dynamic and Optimization-Oriented Approach*, Marcel Dekker Inc, 2003.
- L. Deng, L. Lee, H. Attias, and A. Acero. “A structured speech model with continuous hidden dynamics and prediction-residual training for tracking vocal tract resonances,” *ICASSP* 2004.
- P. Zolfaghari and T. Robinson. “Formant Analysis Using Mixtures Of Gaussians,” *ICSLP* 1996.
- P. Zolfaghari, S. Watanabe, A. Nakamura and A. Katagiri. “Bayesian Modelling of the Speech Spectrum Using Mixture of Gaussians,” *ICASSP* 2004.