**Audio-visual noise removal in speech using a binary mask: exploiting correlations between lip contours and modulation envelopes**

*Arun Palghat Udayashankar*

*Oregon Health and Science University*

Showing a speaker’s face is known to improve speech recognition in noise performance compared with the audio-alone condition. We explore several approaches to demonstrate the potential benefit of using visual features to augment a hearing aid noise removal signal processing algorithm. Firstly, we investigate how modulation frequency envelopes correlate with lip motion. Modulation filtering of speech involves decomposing the signal into frequency bands and then dividing each band into a high frequency carrier and a low frequency modulation envelope. The modulation envelope temporally shapes the carrier frequency in a way that is similar to the way the lips shape sound energy during speech production. We present the relationship between lip contour movements and modulation envelopes as recorded during audio-visual speech at the sentence, word, and simple-sound level.

Secondly, we modify an established acoustic-based hearing aid signal processing algorithm that uses a non-linear classification technique (Gaussian mixture model - Bayes) to differentiate speech from noise. This algorithm makes use of a time-frequency approach to generate a multi-band binary mask that identifies speech vs. noise over time windows. We evaluate whether including the lip contours plus acoustic features can yield a more accurate binary mask than when acoustic features are used alone. A generalized linear model was trained using the data corrupted with wide band noise at various signal to noise ratios. Following pruning of the features, the feature matrix was reduced to a subset of audio and video features. This subset was then used to train a Gaussian mixture model for each class (signal and noise). The resulting probability density functions were used to classify the test data as signal or noise using a Bayesian classifier. During the course of the talk results from the modulation filtering analysis and the signal-noise classification techniques will be discussed.