

Automatic Speech Recognition (ASR) has achieved its state-of-art performance using mainly statistical methods and large amounts of training. Although current algorithms succeed for clear carefully-produced speech signals with abundant and canonical cues, most speech is not of this type, providing strong motivation to move toward systems that can deal with more typical conversational speech signals. However, further improvement for the recognition of such casual speech is hindered by the indeterminacy introduced by “non-standard” speech productions (better expression?) in which adjacent sounds and structural context, as well as word frequency and other factors, powerfully influence the acoustic shape of a word. Motivated by evidence that human speakers identify phonemes by detecting specific sound patterns that indicate linguistic categories or “acoustic cues to distinctive features”, the Speech Communication Group studies these cues for the purpose of constructing a better acoustic model for ASR. This project aims at modeling the correlation between systematic variation in the cues and the corresponding phonemic and structural context, such as the word and syllable structure of spoken phrases and the presence of breaks and accents in casual conversation. By translating these correlations into additional constraints in the speech recognition process, this model, in combination with effective programs for detecting distinctive feature cues developed in our lab, is expected to contribute to development of new ASR systems with greater robustness for speech produced in casual settings with implications for models of human speech processing as well.