

Acoustic Modelling for Speech Recognition: Hidden Markov Models and Beyond?

Mark J.F. Gales

Hidden Markov models (HMMs) are still the dominant form of acoustic model used in automatic speech recognition (ASR) systems. However over the years the form, and training, of the HMM for ASR have been extended and modified, so that the current forms used in state-of-the-art speech recognition systems are very different to those originally proposed thirty years ago. This talk will review two of the more important extensions that have been proposed over the years: discriminative training; and speaker and environment adaptation. The use of discriminative training is now common with forms based on minimum Bayes' training and minimum classification error being applied to systems trained on many hundreds of hours of speech data. The talk will describe these current approaches, as well as discussing the current trends towards schemes based on large-margin training approaches. Linear transform based speaker adaptation is the dominant form for speaker adaptation. Current approaches, including extensions to linear transforms and model-based noise robustness techniques, and trends will also be described. Details of the various forms of the adaptation/noise transformation, training criterion and approaches for adaptive training will be given. The final part of the talk will discuss research beyond the current HMM framework. Schemes based on both discriminative models and functions, as well as non-parametric approaches will be described.