Thesis Seminar

Articulatory Speech Processing

Sam Roweis

Computation and Neural Systems

Wednesday January 13, 1999

1:00 PM   Room 114 East Bridge

Abstract

When difficult computations are to be performed on sensory data it is often advantageous to employ a model of the underlying process which produced the observations. Because such *generative models* capture information about the set of possible observations, they can help to explain complex variability naturally present in the data and are useful in separating signal from noise. In the case of neural and artificial sensory processing systems generative models are learned directly from environmental input although they are often rooted in the underlying physics of the modality involved. One effective use of learned models is made by performing *model inversion* or *state inference* on incoming observation sequences to discover the underlying state or control parameter trajectories which could have produced them. These inferred states can then be used as inputs to a pattern recognition or pattern completion module.

In the case of human speech perception and production, the models in question are called *articulatory models* and relate the movements of a talker's mouth to the sequence of sounds produced. Linguistic theories and substantial psychophysical evidence argue strongly that articulatory model inversion plays an important role in speech perception and recognition in the brain. Unfortunately, despite potential engineering advantages and evidence for being part of the human strategy, such inversion of speech production models is absent in almost all artificial speech processing systems.

My thesis work has involved a series of experiments which investigate articulatory speech processing using real speech production data from a database containing simultaneous audio and mouth movement recordings. In this seminar I will show that it is possible to learn simple low dimensionality models which accurately capture the structure observed in such real production data. I will discuss how these models can be used to learn a forward synthesis system which generates sounds from articulatory movements. I will also describe an inversion algorithm which estimates movements from an acoustic signal. Finally, I will demonstrate the use of articulatory movements, both true and recovered, in a simple speech recognition task, showing the possibility of doing true articulatory speech recognition in artificial systems.