Progress Report

During the past week we have established software infrastructure and identified necessary steps that would lead us to completion of the project.

The software tools that were made available to us will allow us to achieve our project objectives with minimal software development efforts. The HTK package makes it possible to train context-dependent HMM models by taking advantage of state clustering techniques. These models can be further processed with BMM tools to learn additional probabilistic dependencies between observation variables. We also have tools for randomly sampling BMMs and for synthesizing speech waveforms given mel cepstrum or LPC coefficients.

During this preparatory phase we have formatted the training data to satisfy HTK and “p-file” specifications. These two different format requirements are imposed by the choice of software tools that we made. We will use mel cepstra as input for training our models. The observation probability distribution will be modeled by a Gaussian with diagonal covariance.

For the next two weeks our work in this project will be organized as follows:

• Use HTK to train context-dependent HMMs with decision tree state clustering.
• Convert these models into format that can be read by BMM tools.
• Sample HMMs and synthesize speech.
• Train BMM models.
  – Starting with the HMMs, add dependencies between observation variables based on conditional mutual information between them.
  – Train statistical weights for these dependencies.
• Sample BMMs and synthesize speech.
• Do an additional experiment (see below).

We would like to accomplish all but the last step this week. This would allow us enough time for conducting an additional experiment and for preparing the report and the poster.

Additional Experiment. As we expect an improvement in the quality of synthesized speech when we go from HMM to BMM models, it is an unfair comparison, as the BMM model will have a greater number of parameters due to additional dependencies between observation variables. We would like to reduce the number of parameters in the BMM model to match the size of the HMM. This reduction can be accomplished by having fewer state clusters (i.e. pruning the decision trees). Final evaluation will be based on informal listening, since there is not enough time to conduct a perceptual experiment.