A Study on Modeling of Asynchrony and Context Dependency of Individual Feature Vector Components for Speech Recognition

Shigeki Matsuda
School of Information Science,
Japan Advanced Institute of Science and Technology

January 10, 2003

Abstract

This thesis presents our proposals to improve speech recognition systems. The state-of-the-art hidden Markov model (HMM) based systems usually treat the acoustic features as a chain of stationary signal sources. The observed values of these features are represented by vectors. We assume that they might be better modeled by individual vector components. We discuss two methods based on this assumption.

In the first method, we try to model asynchronous changes of individual acoustic vector components. Conventional HMM implicitly assumes that individual components change their statistical properties simultaneously. This assumption may not be true. Temporally changing patterns of individual acoustic components do not necessarily synchronize with each other. We propose a new HMM that allows asynchronous state transitions between individual vector components. We demonstrate that this new HMM outperforms the conventional HMM in speaker-dependent speech recognition task.

In the second method, we try to model phoneme context dependency of individual acoustic vector components. Conventional parameter tying techniques provide a common tying structure for all vector components, no matter how different is their individual components complexity and phoneme context dependency. In this discussion, we propose a new parameter tying technique that allows to have distinct tying structures for each component. Our experimental results show that proposed HMM with feature-depended tying works better than conventional HMM with a common tying.

Both proposed methods are based on treating the observed feature vector as vector of individual components. Moreover, we discuss time characteristics and phoneme context dependencies of individual components, and develop the new HMM structures and the new training techniques. All these methods are evaluated in continuous phoneme recognition task.

Key Words: **Speech Recognition, Acoustic Model, Asynchronous Transition Hidden Markov Model, Temporal Tying Technique, Feature-Dependent Phoneme Environment Clustering, Feature-Dependent Successive State Splitting**