ARABIC SPEECH RECOGNITION USING HIDDEN MARKOV MODELS

BY

Iman Abuel Maaly Abdel Rahman

{ B.sc. (Cairo University) 1982,
M.Sc.(University of Khartoum) 1993}

A thesis submitted to the University of Khartoum for the degree of Doctor of Philosophy (Ph.D) in Electrical Engineering.

Department of Electrical Engineering
Faculty of Engineering and Architecture
University of Khartoum

September, 1997

ABSTRACT

The objective of this research is to introduce an Arabic phoneme recognition system using Hidden Markov Models. This system is one of the important steps in the continuous speech recognition systems for Arabic. As each natural language has its distinctive features, we presented a preformance test to evaluate the best observation vectors (features) that can represent Arabic speech signals.

First, in the front end feature extraction of this system, different sets of observation vectors are extracted. Most of these observations are based on the Covariance Lattice Method of Linear Prediction Coding (LPC) analysis technique [ ]and power calculation. These sets of observation vectors are the following :

1. The weighted Cepstral coefficients.
2. The delta-cepstral coefficients.
3. The delta-delta cepstral coefficients.
4. The area function.
5. Log area ratios.
6. Power, delta- power, and delta-delta power.

Second, the prewhitening transformation is implemented on some combinations of the above sets of observations to remove correlation between their parameters. This is because correlation implies redundancy, and to achieve a normalized observation vector that can give real distance measures when comparing between them.

Third, we implemented the Vector Quantization (VQ) on these sets by modeling the observation vectors using a finite set of discrete values with vanishingly small errors. VQ is implemented as a compression technique to reduce the computational complexity [ ] [ ] .

Finally, the discrete HMM is implemented for each of the 32 Arabic phonemes. The Viterbi reestimation method is used to estimate model parameters. Recognition is also based on Viterbi algorithm. Performance tests are accomplished on different observation sets using this final phoneme recognition system.

After a comprehensive study of the phonetic properties of modern Arabic speech, we introduced methods for improving Arabic speech recognition systems

1. Articulating each Arabic phoneme preceded by (hamzah) and ended by a silence (as discribed by Saibawaihi for studying the place of articulation of phonemes) produces the ideal phonemes that are extracted and trained to obtain Hidden Markov Models. These units are then used in tests of at the final recognition stage. Another sets of phonemes extracted from continuous speech are modeled and tested using the models trained using the ideal phonemes.
2. A new method is introduced for automatic segmentation of continuous speech to the level of individual syllables. This method is based on classifying Arabic phonemes according to their level of perception. Power is the measurement unit used in this algorithms.

Results showed that the best features to represent Arabic phonemes are the weighted cepstral coefficients plus their differenced values in one observation vector (performacne 74%).

While the area function is the worst features to represent Arabic phonemes (performance 48%).

The proposed ideal phoneme units use in the training and recognition of HMMs for all the experiments showed good performance. Nevertheless, using another units extracted from continuous speech for the HMM recognition with training using the ideal ones could not be accomplished. This due to different lengths of context dependent phonemes. The program may need modification to deal with very short phonemes.

This system is constructed using Object Oriented Programming. The main advantage of using this method of programming is reducing the computational complexity and reducing memory allocated.