ABSTRACT

COMPUTER RECOGNITION OF V/CV SPEECH UNITS BASED ON LINEAR AND NON LINEAR DYNAMICAL SYSTEM MODELS USING BRAIN LIKE COMPUTING AND STATISTICAL LEARNING ALGORITHMS

By

THASLEEMA. T. M

This thesis focuses on the use of Automatic Speech Recognition (ASR) for Malayalam Vowel/Consonant-Vowel (V/CV) speech unit recognition. Malayalam language has the richest phonemic utterances among Indian languages. V/CV speech unit occur frequently in normal speech and the recognition of these units are crucial for development of an accurate speech recognition system. Accurate feature extraction techniques based on linear and non-linear dynamical system models are discussed in this thesis. Two new feature parameters are introduced in this work. The proposed representations are proved to be robust to differentiate V/CV speech units in speaker independent, clean and noisy environments.

A Malayalam V/CV speech database of 14 long and short vowel(V)units and 36 Consonant-Vowel (CV) units, uttered by both male and female native, 96 Malayalam speakers of age between 21 to 35 years is created for the recognition studies. Each V/CV speech waveform, band limited to 4 kHz and digitized using a 16 bit A/D converter at a sampling rate of 8 kHz is stored in the database for recognition experiments.

A preliminary analysis on the recognition of Malayalam V/CV speech units using traditional basic speech features viz., Linear Predictive Coding
Coefficient(LPCC) and Mel Frequency Cepstral Coefficient(MFCC) are discussed in this thesis for the purpose of comparison.

A novel feature extraction method using Multi Resolution Analysis (MRA) based Wavelet Transform is investigated and presented in this thesis. To develop the Normalized Wavelet Hybrid Feature (NWHF) feature vector two decomposition algorithms namely Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition (WPD) are combined, and then normalized using z-score normalization technique. The extracted hybrid feature vector gives better classification result.

Another accurate time domain approach based on the non-linear dynamical system model for V/CV speech unit recognition is also investigated and presented in this thesis. This technique is based on statistical models of Reconstructed State Space(RSS) from the time domain waveform. The State Space Map (SSM) for each V/CV speech unit is generated and State Space Point Distribution (SSPD) features are extracted from SSM.

The later part of the thesis reports the effectiveness of the features (LPCC, MFCC, NWHF and SSPD feature vector) by conducting experimental studies using various classification algorithms. Among the pattern recognition approach, the connectionist approach using Artificial Neural Networks (ANN) based Feed Forward Multi Layer Neural Network Back Propagation (FFMLNNBP) learning algorithm and statistical learning approach using k - Nearest Neighbor (k -NN) & Support Vector Machines (SVM) based Decision Directed Acyclic Graph Support
Vector Machine (DDAGSVM) algorithm are used for this purpose. Simulation results are provided to evaluate the performance of these three classifiers. A comparatively good recognition accuracy obtained using DDAGSVM algorithm with the recognition accuracy of 51.59% for LPCC, 55.6% for MFCC, 65.32% for NWHF and 73.82% for SSPD feature method by dividing each V/CV speech sequences into 256 sample blocks and its multiples. Grouping the V/CV speech database into six phonetic classes average recognition accuracy of 61.01% for LPCC, 67.08% for MFCC, 81.08% for NWHF and 90.31% for SSPD feature methods is obtained using DDAGSVM.