Malay Language Speech Recognition for Preschool Children using Hidden Markov Model (HMM) System Training

ABSTRACT

This paper aims to discuss the implementation of Malay Language Speech Recognition (SR) using Hidden Markov Model (HMM) system training for Malay preschool children. The system is developed by implementing the architecture of HMM-based Recognizer with different feature extraction algorithm. The system is trained for 16 Malay words by collecting 704 speech samples (640 for training samples and 64 for testing samples). Data is collected from 20 preschool children aged between five to six years old in real time environments. The paper also describes the process flow to develop the architecture of the system. The experimental results show that the highest overall system performance is 94.86% - Train and 76.04% - Testing which is using MFCC (Mel-Frequency Cepstral Coefficient) with 39 extracted feature vectors (MFCC39).