Abstract

Speech is the output of a time varying excitation excited by a time varying system. It generates pulses with fundamental frequency F0. This time varying impulse trained as one of the features, characterized by fundamental frequency F0 and its formant frequencies. These features vary from one speaker to another speaker and from gender to gender also. In this paper the effect of gender on improving speech recognition is considered. Variation in F0 and formant frequencies is the main features that characterize variation in a speaker. The variation becomes very less within speaker, medium within the same gender and very high among different genders. This variation in information can be exploited to recognize gender type and to improve performance of speech recognition system through modeling separate models based on gender type information. Five sentences are selected for training. Each of the sentences are spoken and recorded by 20 female’s speakers and 20 male speakers. The speech corpus will be preprocessed to identify the voiced and unvoiced region. The voiced region is the only region which carries information about F0. From each voiced segment, F0 and the first three formant frequencies and also MFCC features are computed. Each forms the feature space labeled with
the speaker identification: i.e., male or female. This information misused to parameterize the model for male and female. K-means algorithm is used during training as well as testing. Testing is conducted in two ways: speaker dependent testing and speaker independent testing. SPHINX-III software by Carnegie Mellon University has been used to measure the accuracy of speech recognition of data taking into account the case of gender separation which has been used in this research.

References


Index Terms

Computer Science  Pattern Recognition
Keywords

Speech recognition (SR), Linear Prediction Coding (LPC), Accent, Speakers.