The Capacity of Mel Frequency Cepstral Coefficients for Speech Recognition

Authors : Fawaz S. Al-Anzi, Dia AbuZeina

Abstract : Speech recognition is of an important contribution in promoting new technologies in human computer interaction. Today, there is a growing need to employ speech technology in daily life and business activities. However, speech recognition is a challenging task that requires different stages before obtaining the desired output. Among automatic speech recognition (ASR) components is the feature extraction process, which parameterizes the speech signal to produce the corresponding feature vectors. Feature extraction process aims at approximating the linguistic content that is conveyed by the input speech signal. In speech processing field, there are several methods to extract speech features, however, Mel Frequency Cepstral Coefficients (MFCC) is the popular technique. It has been long observed that the MFCC is dominantly used in the well-known recognizers such as the Carnegie Mellon University (CMU) Sphinx and the Markov Model Toolkit (HTK). Hence, this paper focuses on the MFCC method as the standard choice to identify the different speech segments in order to obtain the language phonemes for further training and decoding steps. Due to MFCC good performance, the previous studies show that the MFCC dominates the Arabic ASR research. In this paper, we demonstrate MFCC as well as the intermediate steps that are performed to get these coefficients using the HTK toolkit.

Keywords : speech recognition, acoustic features, mel frequency, cepstral coefficients
Conference Title : ICNMLKD 2017 : International Conference on Network, Machine Learning and Knowledge Discovery
Conference Location : Bangkok, Thailand
Conference Dates : October 26-27, 2017