Speech Recognition with Dynamic Time Warping using MATLAB

**Abstract**

Speech recognition has found its application on various aspects of our daily lives from automatic phone answering service to dictating text and issuing voice commands to computers, we present the historical background and technological advances in speech recognition technology over the past few decades. More importantly, we present the steps involved in the design of a speaker-independent speech recognition system. We focus mainly on the pre-processing stage that extracts salient features of a speech signal and a technique called Dynamic Time Warping commonly used to compare the feature vectors of speech signals. These techniques are applied for recognition of isolated as well as connected words spoken. We conduct experiments on MATLAB to verify these technique.

**INTRODUCTION**

Language is man's most important means of communication and speech its primary medium. Speech provides an international forum for communication among researchers in the disciplines that contribute to our understanding of the production, perception, processing, learning and use. Spoken interaction both between human interlocutors and between humans and machines is inescapably embedded in the laws and conditions of Communication, which comprise the encoding and decoding of meaning as well as the mere transmission of messages over an acoustical channel. Here we deal with this interaction between the man and machine through synthesis and recognition applications. The paper dwells on the speech technology and conversion of speech into analog and digital waveforms which is understood by the machines. Speech recognition, or speech-to-text, involves capturing and digitizing the sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analysing the words to ensure correct spelling for words that sound alike. Speech Recognition is the ability of a computer to recognize general, naturally flowing utterances from a wide variety of users. It recognizes the caller's answers to move along the flow of the call. Early attempts to design systems for automatic speech recognition were mostly guided by the theory of acoustic phonetics, which describes the phonetic elements of speech (the basic sounds of the language) and tries to explain how they are acoustically realized in a spoken utterance. These elements include the phonemes and the corresponding place and manner of articulation used to produce the sound in various phonetic contexts

**Block Diagram**

