Speech Processing

subodh wandile[@??]

[@]subodhwandile77g@gmail.com

Introduction

Speech is a natural mode of communication for people. We learn all the relevant skills during early childhood, without instruction, and we continue to rely on speech communication throughout our lives. It comes so naturally to us that we donâĂŹt realize how complex a phenomenon speech is. Speech recognition is basically making a computer understand spoken language. By understand we mean react appropriately or convert the input speech into another medium. Speech recognition is more and more useful now a days. Various interactive softwares are available in market today but they are useful for general-purpose computers. With the growth in the needs for embedded computing and the demand for embedding platforms, it is required that speech recognition systems are available on them too. Quantization: is the operation which allows the change of signal with continuous variable to signal With finite number of values. This is as shown in (Fig 1) 1.

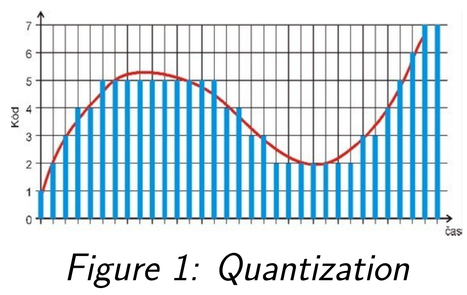

Figure 1: Quantization

Speech recognition basically means talking to a computer, having it recognise what we are saying, and lastly doing it in real time. This process fundamentally functions as a pipeline that converts PCM (Pulse Code Modulation) digital audio from a sound card into recognised speech. Speech recognition is basically making a computer understand spoken language. By understand we mean react appropriately or convert the input speech into another medium. We humans have natural speech recognition. Articulation produces sound waves, which the ear conveys to the brain for processing. The basic question is how might a computer do it? It does it in three ways-Digitization, acoustic analysis of speech signal and linguistic interpretation.

Steps in speech processing

A.Digitization

Digitization is basically analog to digital conversion of speech signal, followed by sampling and quantising the signal.

Using filters to measure energy levels for various points on the frequency spectrum does this. Knowing the relative importance of different frequency bands (for speech) makes this process more efficient.

B.Dynamic time warping

Dynamic time warping (DTW) is an algorithm for measuring similarity between two temporal sequences, which may vary in speed. In general, DTW is a method that calculates an optimal match between two given sequences (e.g. time series) with certain restriction and rules. The optimal match is denoted by the match that satisfies all the restrictions and the rules and that has the minimal cost, where the cost is computed as the sum of absolute differences, for each matched pair of indices, between their values.

History

Early attempts at speech processing and recognition were primarily focused on understanding a handful of simple phonetic elements such as vowels. In 1952, three researchers at Bell Labs, Stephen. Balashek, R. Biddulph, and K. H. Davis, developed a system that could recognize digits spoken by a single speaker.

One of the first commercially available speech recognition products was Dragon Dictate, released in 1990. In 1992, technology developed by Lawrence Rabiner and others at Bell Labs was used by ATT in their Voice Recognition Call Processing service to route calls without a human operator. By this point, the vocabulary of these systems was larger than the average human vocabulary.

By the early 2000s, the dominant speech processing strategy started to shift away from Hidden Markov Models towards more modern neural networks and deep learning.

Conclusion

First and very important step when recognizing speech is the signal processing. It creates output for classificators. In order of faster classifying these information must be reduced to lowest possible rate with insignificant loss of information content. This is very important especially for embedded systems in cars, which have less memory and operating output than PC. following are the techniques for speech recognition:

- Artificial neural networks
- Hidden Markov models
- Dynamic time warping

Please compile this document with lualatex to see a QR code here:
`lualatex output`