Review on: Jarvis, Digital Life Assistant

Shrutika Khobragade

Department of Computer

Vishwakarma Institute of InformationTechnology Pune,INDIA

Introduction:

Speech is an effective and natural way for people to interact with applications, complementing or even replacing the use of mice, keyboards, controllers, and gestures. A hands-free, yet accurate way to communicate with applications, speech lets people be productive and stay informed in a variety of situations where other interfaces will not. Speech recognition is a topic that is very useful in man applications and environments in our daily life. Generally speech recognizer is a machine which understands humans and their spoken word in some way and can act thereafter. A different aspect of speech recognition is to facilitate for people with functional disability or other kinds of handicap. To make their daily chores easier, voice control could be helpful. With their voice they could operate the light switch turn off/on or operate some other domestic appliances. This leads to the discussion about intelligent homes where these operations can be made available for the common man as 5well as for handicapped

SPEECH REPRESENTATION:

The speech signal and all its characteristics can be represented in two different domains, the time and the frequency domain A speech signal is a slowly time varying signal in the sense that, when examined over a short period of time (between 5 and 100 ms), its characteristics are short time stationary. This is not the case if we look at a speech signal under a longer time perspective (approximately time T>0.5 s).
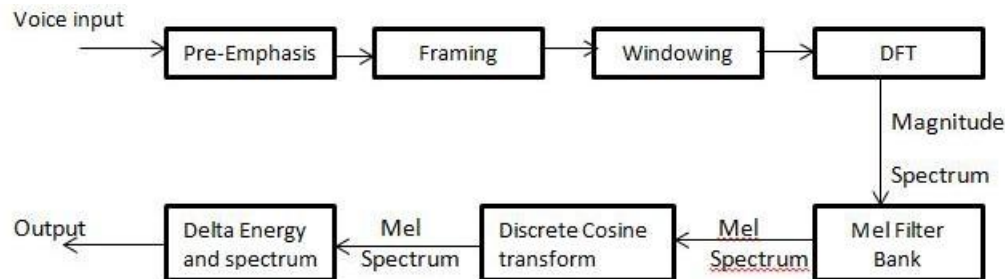
THREE STATE REPRESENTATION

The three-state representation is one way to classify events in speech. The events of interest for the three-state representation are

- Silence (S) - No speech is produced.

- Unvoiced (U) - Vocal cords are not vibrating, resulting in an aperiodic or random speech waveform.

- Voiced (V) - Vocal cords are tensed and vibrating

periodically, resulting in a speech waveform that is quasiperiodic.

FEATURE EXTRACTION (MFCC):

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior. MFCC is based on human hearing perceptions which cannot perceive frequencies over 1Khz. In other words, in MFCC is based on known variation of the human ear's critical bandwidth with frequency [8-10].

Voice input
Pre-Emphasis → Framing → Windowing → DFT
Magnitude Spectrum
Output ← Delta Energy and spectrum ← Mel Spectrum ← Discrete Cosine transform ← Mel Spectrum ← Mel Filter Bank

METHODOLOGIES:-

Process	Description
1) Speech	2Female(age=20,age=53) 2 Male(age=22,age=45)
2) Tool	Mono Microphone Microsoft Speech software
3) Environment	College Campus
4) Utterance	Twice each of the following word 1) Volume Up 2) Volume Down 3) "Jarvis there" 4) Introduce yourself 5) Show date.
5) Sampling Frequency	16000 KHz
6) Feature Computational	39 double delta MFCC coefficient

Flow chart:
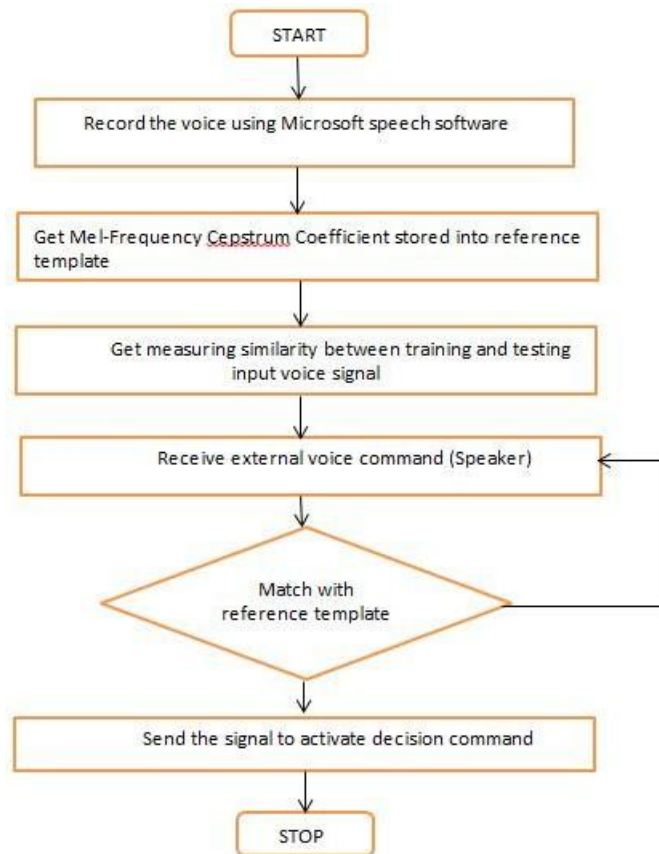
Conclusion:

This paper has discussed voice recognition algorithms which are important in improving the voice recognition performance. The technique was able to authenticate the particular speaker based on the individual information that was included in the voice signal. The results show that these techniques could use effectively for voice recognition purposes. Several other techniques such as Liner Predictive Coding (LPC), Dynamic Time Wrapping (DTW), and Artificial Neural Network (ANN) are currently being investigated. The findings will be presented in future publications.

Signature:_____