SPEECH RECOGNITION CHALLENGES AND TECHNIQUES: A REVIEW

A Report for the Evaluation 3 of
Project 2

Submitted by

ANURAG SINGH

(162210124/

16SCSE101888)

In partial fulfillment for the award of the

degree of

Bachelor of Technology

IN

Computer Science and Engineering

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision of

Mr. Soumya Ranjan Jena
Asst. Professor

APRIL / MAY- 2020

TABLE OF CONTENTS

ABSTRACT

Speech is a natural mode to interact with others. With speech, we can express our words to others. Speech recognition is a way or technology where the statements or commands of human speech to understand and react accordingly. Speech recognition allows machining system to turn the incoming speech signals into commands through the process of identifying and understanding. It also creates the natural voice communication function. Main Goal of speech recognition is to achieve better language communication between man and machine. So it is a great technology of human machine interface. The paper describes the speech recognition technology development is all basic principles, methods and classification of this technology. Accuracy of different methods of speech technology is provided to sort out methods with their performance aspect.

INTRODUCTION

Voice Recognition Framework is a framework which can perceive the voices. This can be with the end goal of words ID or with the end goal of security. Voice Recognition is the procedure of naturally perceiving who is talking or what is talking based on singular data remembered for the Speech waves. This system makes it conceivable to utilize the speaker's voice to confirm their personality and control access to administrations, for example, voice dialing, banking by phone, phone shopping, database get to administrations, data administrations, voice message, security control for secret data territories, and remote access to PCs. Some Voice Recognition Framework is planned so that they can change over the verbally expressed words into content. Voice recognition Framework or Virtual products can likewise be utilized as an option in contrast to composing on a console. Set forth plainly, you converse with the PC and your words show up on the screen. The product has been created to give a quick strategy for composing onto a PC and can help individuals with an assortment of inabilities. It is helpful for individuals with physical handicaps who frequently discover composing troublesome, excruciating or outlandish. Voice recognition programming can likewise help those with spelling troubles, incorporating clients with dyslexic, on the grounds that perceived words are in every case accurately spelled. You can also use voice recognition software in homes and businesses. A range of software products allows users to dictate to their computer and have their words converted to text in a word processing or e-mail document. You can access function commands, such as opening files and accessing menus, with voice instructions. Some programs are for specific business settings, such as medical or legal transcription. People with disabilities that prevent them from typing have also adopted voice-recognition systems. If a user has lost the use of his hands, or for visually impaired users when it is not possible or convenient to use a keyboard, the systems allow personal expression through dictation as well as control of many computer

tasks. Some programs save users' speech data after every session, allowing people with progressive speech deterioration to continue to dictate to their computers. These programs can be seen in our daily life, these programs are included in Windows XP, Windows Vista, Windows 7 operating system from Microsoft Corporation, while these software's can be installed additionally for example Dragon Naturally Speaking by Nuance, Speech Recognition by Icons, Speech Recognition by Tazti, Via Voice by IBM, iListen these are some popular voice recognition software's that are available as third party software for voice recognition.

Existing Systems

1 .Google Cloud Speech API

Best in recognizing 120 languages.Google Cloud Speech API can be used for short form and long form video. It can be used for the processing of real-time streaming and pre-recorded audio. It automatically transcribes the correct nouns, dates, and phone numbers.

Features

It can filter the inappropriate content.

It is accurate in transcribing punctuation.

It supports 120 languages.

It automatically recognizes the spoken language.

A.Voice biometrics

This technology compares the previously stored voice print or template with the utterance and produces score. It works on

voice interpretation algorithm. Biometric reduces each spoken word into some frequency segments called formants. This technology is used at various agencies like online banking, online security trading, online information services, computer access security and many more.

B. Siri technology

It is an application used by the company Apple Macintosh in their iphones. This application captures the voice from the speaker and performs the function narrated by the speaker. For example, you can ask to call a specific person from your contact list, or send him a message and can narrate the message as well

Features:

It can connect a call to someone and can send text messages.

It will help you to know 'Who is calling you?'

It can set alarms, timers, and reminders

.C. Games and toys

voice driven toys and games are available in market. The simulation of these types of products is based on the voice command given to them.

D. Fighter aircrafts

The fighter jets are controlled with the voice system and are being given commands from the base. The speech recognition systems are helpful for controlling the hands free weapons system .

ARCHITECTURAL DIAGRAMS
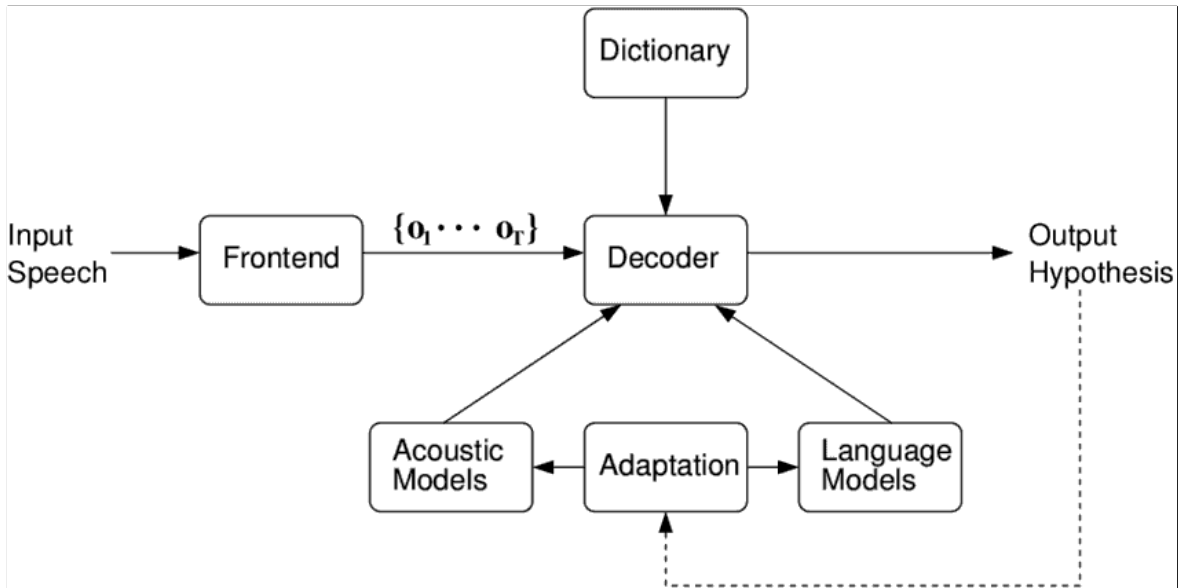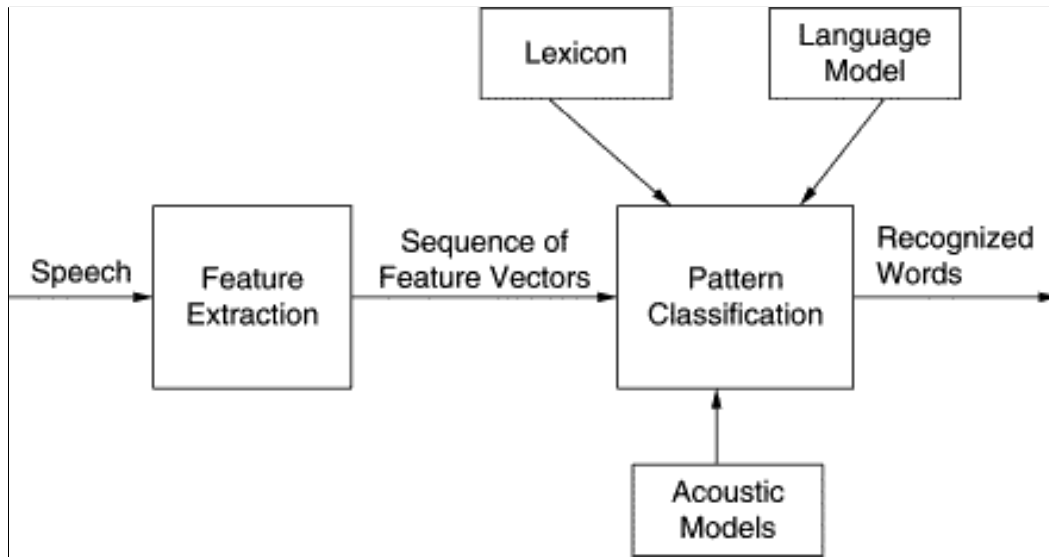
TECHNIQUES

MFCC- Mel Recurrence Cepstral Coefficients (MFCC) is the most unmistakable technique utilized during the time spent element extraction in speech recognition. It on the recurrence area which depends on Mel scale dependent on human ear scale. MFCCs, being recurrence area highlights, are more exact than time area highlights. MFCC speaks to the genuine cepstral of windowed brief timeframe signal which is determined from Quick Fourier Change (FFT). These coefficients are strong and solid for varieties of speaker and activity condition.

LPC- Straight Prescient Coding (LPC) is an instrument most broadly utilized for medium or low piece rate coder. Computerized signal is compacted for productive transmission and capacity. Calculation of parametric model dependent on least mean squared blunder hypothesis is known as direct forecast (LP). The signal is communicated as a direct mix of past tests. Configuration frequencies are where reverberation top happens

CONCLUSION

This survey attempt to gather the arrangements of various methodologies in the two stages in highlight extraction and discourse acknowledgment framework. It centers around the determination of the past examinations and utilized strategies so as to know the advantages of knowing the systems and contrasting

its exhibition concurring with the outcomes. The paper have introduced numerous procedures of the periods of highlight extraction. Moreover it has introduced numerous apparatuses of voice acknowledgment. Discourse Acknowledgment Framework (SRS) is developing step by step and has boundless applications. The study has indicated the diagram of the discourse acknowledgment process, its fundamental model, and applications. Right now various methodologies which are generally utilized for SRS have been examined and after relative investigation of these approaches it is presumed that Shrouded markov method (HMM) is best appropriate methodology for a SRS since it productive, powerful, and lessens time and multifaceted nature

References

[1] Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, "Neural Network used for Speech Recognition", Journal of Automatic Control, University of Belgrade, Vol. 20, pp. 1-7,2010.://dx.doi.org/10.2298/JAC1001001G

[2] Reynolds DA (1995) Speaker identification and verification using Gaussian mix-ture speaker models. Speech Commun 17:91−108

[3] Bimbot F, Bonastre J, Fredouille C, Gravier G, Chagnolleau MI, Meignier S,Merlin T, Garcia OJ, Delacretaz P, Reynolds DA (1997) A tutorial on text-independent speaker verification. EURASIP J Appl Sig Proc 2004(4):430−451

[4] Zahi N.Karam,William M.Campbell "A new Kernel for SVM

[5] MIIR based Speaker recognition "MIT Lincoln Laboratory, Lexington, MA, USA

[6] Bahl, L. R., Brown, P. F., de Souza, P. V., Mercer, R. L., Picheny, M. A. (1993). A method for the construction of acoustic Markov models for words. Speech and Audio Processing, IEEE Transactions on, 1(4), 443-452

[7] Butzberger, J., Murveit, H., Shriberg, E., Price, P. (1992, February). Spontaneous speech effects in large vocabulary speech recognition applications. In Proceedings of the workshop on Speech and Natural Language (pp. 339-343). Association for Computational Linguistics.

[8] Charles, A. H., Devaraj, G. (2004). Alaigal-A Tamil Speech Recognition. Tamil Internet.

[9] Dumitru, C. O., Gavat, I. (2006, June). A Comparative Study of Feature

Extraction Methods Applied to Continuous Speech Recognition in Romanian Language. In Multimedia Signal Processing and Communications, 48th International Symposium ELMAR-2006 focused on (pp.115-118).