International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 55-57

JARVIS - AI Based Personal Desktop

Joney Kumar[1], Anant gaur[2], Arjun singh[3],Manmeet kaur[4]
[1]*Assistant Professor & IT Department & Meerut Institute of Engineering and Technology* [2,3,4]*Students & IT Department & Meerut Institute of Engineering and Technology Meerut, India*

Abstract—"*Jarvis" was one in every of the main characters of Tony Stark's life assistant in Marvel Cinematic Universal's Iron Man. Unlike the first comic within which Jarvis was Stark's human servant, the movie version of Jarvis is an intelligent computer that converses with Stark, monitors his household, and helps to create and program his iron man suit.*

In this project, Jarvis is a Digital Life Assistant who uses mainly human communication means like Whatsapp, instant messaging, and voice notes to ascertain a two-way conversation between a user and his query. During this project, we mainly use voice as a communication means specified Jarvis is sort of a Speech recognition application that acts as a private desktop assistant. The concept of speech technology comprises two technologies: Synthesizer and Recognizer. A speech synthesizer takes an input and produces an audio stream as output. A speech recognizer on the opposite hand does the alternative. It takes an audio stream as input and thus

Male) for accuracy.

Keywords: *Speech Synthesizer, Recognizer, Mel Frequency Cepstral turns it into text transcription as output. We test this on 2 voices(1 Female and 1 Coefficients*

Date of Submission: 21-06-2021 Date of acceptance: 06-07-2021

I. INTRODUCTION

Speech is often an efficient and natural way for people to interact with applications, complementing or perhaps replacing the utilization of mice, keyboards, controllers, and gestures. A hands-free, yet extremely accurate thanks to communication with applications, speech lets people be productive and stay informed during a type of situations where other interfaces won't. Speech recognition is a topic that's extremely useful in many applications and environments in our everyday life. Generally, speech recognizer is a machine that understands humans and their spoken word in some way and can act afterward. A different aspect of speech recognition is to facilitate for people with functional disability or other kinds of handicap and mental disabilities. To make their daily chores easier, voice control could be very helpful. With their voice, they could operate the light switch turn off/on or operate some other domestic appliances and also many more functions. This leads to the discussion about intelligent homes where these operations

II. CLASSIFICATION OF SPEECH[1]

Several parameters define the capability of speech recognition and classification

i) Isolated word: The Isolated word have can be made available for the common man as well as for the handicapped and lives can be made easier.

With the information presented so far one question comes naturally: how is speech recognition done and what's the procedure? To get knowledge of how speech recognition problems can be approached today, a review of some research highlights has been presented. The earliest attempts to devise systems for automatic speech recognition by machine came in the 1950s when various researchers tried to exploit the fundamental ideas of acoustic phonetics. In 1952, at Bell Laboratories, Davis, Biddulph, and Balashek built a system for isolated digit recognition for a single speaker [11]. The system relied heavily on measuring spectral resonances during the vowel region of each digit. As of 1959, another attempt was made by Forgie , constructed at MIT Lincoln Laboratories. Ten vowels embedded in a /b/-vowel-/t/ format were recognized in a speaker-independent manner [12]. In the 1970s speech recognition research achieved several significant milestones and kept on growing. First, the area of isolated word or discrete utterance recognition became a viable and usable technology based on the fundamental studies by Velichko and Zagoruyko in Russia, Sakoe and Chiba in Japan, and Itakura in the United States and many more things as well.

single utterances or occurrences at a time. Isolated utterance might be a better name for this work.

[2]

ii) Connected word: The Connected word system is quite the same to isolated sample windows. Itaccepts single word or words but allow separate utterances to be "run together by a minimum pause between them".

iii) Continuous speech: It allows the user to speak almost freely, while the computer will examine the content and context. There are special methods used to determine
utterance boundaries and various difficulties occurred in it as well.

III. SPEECH RECOGNITION TECHNIQUES

The goal of speech recognition is to analyze, extract, characterize and recognize information about the speaker identity and execute a certain query-related action.A variety of techniques and methods are used for determining the speech characteristics. The speech data contains various types of information that show the speaker's identity. This includes speaker-specific information due to vocal tract, excitation as well [3]

i) **Segmentation analysis**:- In this work, speech is analyzed using the frame size and shift in the range of 10-30 ms to extract speaker information in certain parts. This method is used to gather vocal tract information on speaker recognition.

ii) **Sub segmental analysis**:- Speech analyzed utilizing the frame size and shift in the range 3-5 ms is known as Sub segmental analysis. This technique is used to mainly analyze and extract the characteristics of the excitation state. [4]

iii) **Supra segmental analysis**:- In this work, speech is also analyzed
using the frame size. This technique Is mainly used to examine and characterize the behavior characteristics of the speaker.

IV. FEATURE EXTRACTION

The extraction of the features of the parameters which represent an acoustic signal is an important task to produce a better recognition behavior. The efficiency of this method is important for the next method since it affects its behavior. Different feature extraction methods are available with their
extraction method, it has a Supervised linear map and is fast and eigenvector-based. This method is better than PCA for classification [5]

iii) **The Linear Predictive**:- This coding uses the Static feature extraction method which has 10 to 16 lower-order coefficients which are quite significant. It is used for gathering features at the lower order.

iv) **In Mel-frequency cepstrum (MFCCs)**:- It has the property that the Power spectrum is computed by performing Fourier Analysis.[6] features.

i) **In Principal Component Analysis (PCA)**:- It uses the Nonlinear feature extraction method and gives Linear map and is quite acute and eigenvector-based.

ii) **In Linear Discriminate Analysis(LDA)**:- It depends on the Nonlinear feature

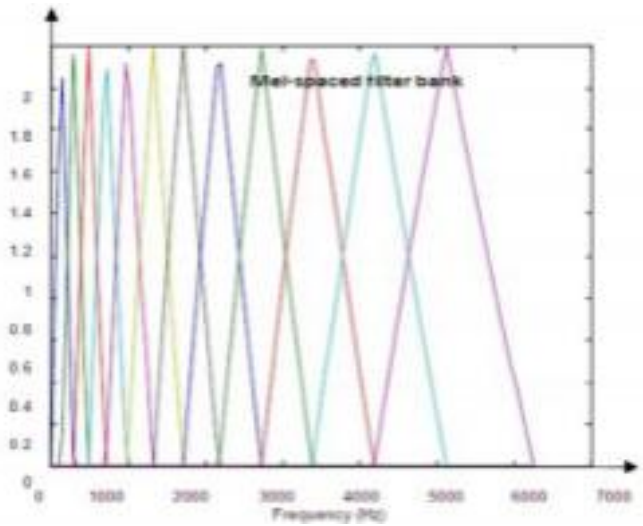

FIG I MEL FREQUENCY CEPSTRAL COEFFICIENTS V. FEATURE

MATCHING

Different gathering methods here are Dynamic Time Wrapping (DTW), Vector Quantization (VQ), LBG etc. Each methodology has its own feature matching function and specification

i) **DTW:** Dynamic time warping is an algorithm for measuring the same characteristics between two temporal sequences which may vary in time or speed. DTW is a methodology that calculates an optimal match between

two given sequences and matches them. DTW has been applied to temporal sequences of video, audio, and graphics data and many more — indeed, any data which can be turned into a linear sequence that can be applied with DTW. Applications include speaker recognition and online signature forger.

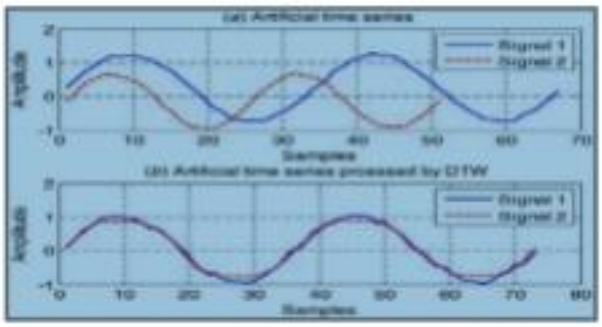

FIG II DYNAMIC TIME WRAPPING OF TWO SIGNALS

ii) VQ:- Vector quantization (VQ) is an optimal quantization technique from signal processing. It was originally used for data compression. It works by dividing a large set of points (vectors) into smaller groups of acute sizes having approximately the same number of points closest to them. Each group is represented by its centroid point which in some cases may also be the center point, as in k-means and some other clustering algorithms. The density matching property of vector quantization is very powerful for big quantities and high-dimensional data. Hence VQ is extremely suited for lossy data compression. It can also be used for lossy data correction and how dense is the estimation.

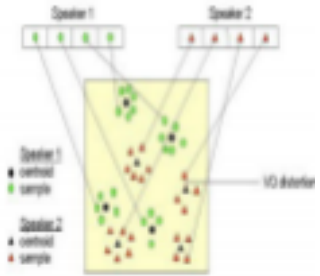

FIG III VECTOR QUANTIZATION OF TWO SPEECH SIGNALS

V. CONCLUSION

In this Literature survey, various techniques are discussed about speech recognition systems. This paper also presents the list of techniques with their properties of Feature extraction and Feature matching. Through this review paper, it is found that MFCC is widely used for feature Extraction and VQ is better than DTW.

REFERENCES

[1]. Rabiner Lawrence, Juang Biing-Hwang.Fundamentals of Speech Recognition Prentice Hall, New Jersey, 1993, ISBN 0-13-015157-2

[2]. Deller John R., Jr., Hansen John J.L., Proakis John G. ,Discrete-Time Processing of Speech Signals, IEEE Press, ISBN 0-7803-5386-2

[3]. Hayes H. Monson,Statistical Digital Signal Processing and Modeling, John Wiley & Sons Inc. ,Toronto, 1996, ISBN 0-471-59431-8

[4]. Proakis John G., Manolakis Dimitris G., Digital Signal Processing, principles, algorithms, and applications, Third Edition, Prentice-Hall , New Jersey, 1996, ISBN 0-13-394338-9

[5]. Ashish Jain, Hohn Harris, Speaker identification using MFCC and HMM-based techniques, university Of Florida,April 25,2004.

[6]. http://www.cse.unsw.edu.au/~waleed/phd/ html/node38.html downloaded on 1February 2021.

[7]. http://web.science.mq.edu.au/~cassidy/co mp449/html/ch11s02.html.

[8]. Hiroaki Sakoe and Seibi Chiba, Dynamic Programming algorithm Optimization for spoken word Recognition, IEEE transaction on Acoustic Speech and Signal Processing, February 1978.

[9]. Young Steve, A Review of Large-vocabulary Continuous-speech Recognition, IEEE SP Magazine, 13:45-57, 1996, ISSN 1053-5888 .

[10]. http://www.microsoft.com/MSDN/speech. html,downloaded on 2Oct 2012.

[11]. Davis K. H., Biddulph R. and Balashek S., Automatic Recognition of Spoken Digits, J. Acoust. Soc.Am., 24 (6):637-642, 1952

[12]. Mammone Richard J., Zhang Xiaoyu, Ramachandran Ravi P., Robust Speaker Recognition, IEEE SP Magazine, 13:58-71, 1996, ISSN 1053-5888.

[13]. Jarvis, Digital Life Assistant by – Shrutika Khobragade, Department of Computer, Vishwakarma Institute of Information Technology, Pune