# Saarush the Intelligence System

**Mrs. A. A. Patkar[1] Aftab Mulani[2] Rutuja Galande[3] Samruddhi Taru[4] Yashodeep Mahangare[5]**
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]AISSMS College of Polytechnic, Maharashtra, India

*Abstract*— In this Project Saarush could be a Digital Life supporter which uses mainly human communication means such Twitter, instant message and voice to make two-way connections between human assist in cooking, notify him of breaking news, Facebook's Notifications and plenty of more. In our project, we mainly use voice as communication means therefore the Saarush is that the Speech recognition application. The concept of speech technology encompasses two technologies: Synthesizer and recognizer. A speech synthesizer takes as 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 turns it into text transcription. The voice could be a signal of infinite information. Direct analysis and synthesizing the complex voice signal is because of an excessive amount of information contained within the signal.

***Keywords:*** Saarush the Intelligence System

## I. Introduction

The conception of speech technology extremely encompasses 2 technologies: Synthesizer and recognizer. A speech synthesizer takes as input associate degreed produces an audio stream as output. A speech recognizer on the other hand will opposite. It takes associate degree audio stream as input and therefore turns it into text transcription. The voice is also a symptom of infinite info. a right away associate degreealysisand synthesizing the complicated voice signal is because of an excessive quantity of data contained at intervals the signal. so the digital signal processes like Feature Extraction and have Matching area unit introduced to represent the voice signal. In this project we tend to directly use speech engine that use Feature extraction technique as Mel scaled frequency cepstral. The melscaled frequency cepstral coefficients (MFCCs) derived from Fourier remodel and filter bank analysis area unit maybe the foremost wide used front ends in progressive speech recognition systems. Our aim to make a lot of and a lot of functionalities which could facilitate human to assist in their means of life and conjointly reduces their efforts. In our check we tend to check all this practicality is functioning properly. we tend to check this on a pair of speakers(1 feminine and one Male) for accuracy purpose. The speech signal and each one its characteristics is also drawn in 2 completely different domains, the time and conjointly the frequency domain A speech signal is also a slowly time varied signal at intervals the sense that, once examined over a short amount of it slow (between five and a hundred ms), its characteristics area unit shorttime stationary. this may be not the case if we tend to look at a speech signal below a extended time perspective (approximately time T>0.5 s). throughout this case the signals characteristics area unit nonstationary, that means that it changes to replicate the varied sounds spoken by the verbaliser To be able to use a speech signal and interpret its characteristics throughout a correct manner some moderately illustration of the speech signal area unit most popular.

## II. 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 a periodic or random speech waveform.
- Voiced (V)- Vocal cords are tensed and vibrating periodically, resulting in a speech waveform that is quasiperiodic.

Quasi-periodic means that the speech waveform can be seen as periodic over a short-time period (5-100 ms) during which it is stationary.

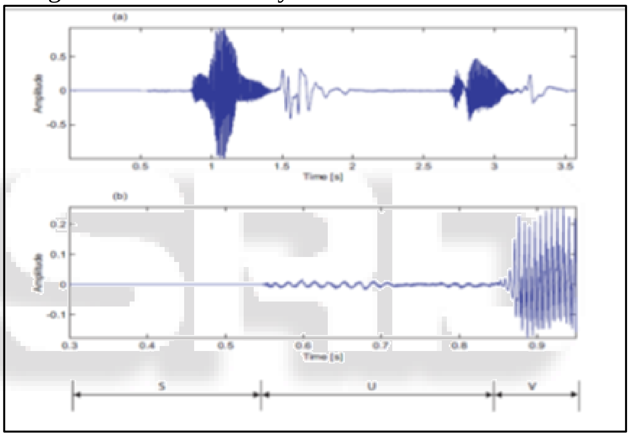


Fig. 1: Three State Representation

The upper plot (a) contains the whole speech sequence and in the middle plot (b) a part of the upper plot (a) is reproduced by zooming an area of the speech sequence. At the bottom of Fig. 1 the segmentation into a three-state representation, in relation to the different parts of the middle plot, is given. The segmentation of the speech waveform into well-defined states is not straight forward. But this difficulty is not as a big problem as one can think.

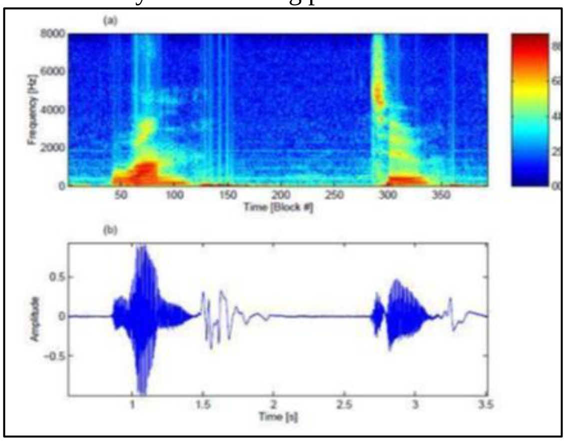


Fig. 2:Spectrogram using Welch's Method(a)andspeech amplitude (b)Here the darkest (dark blue) parts represents

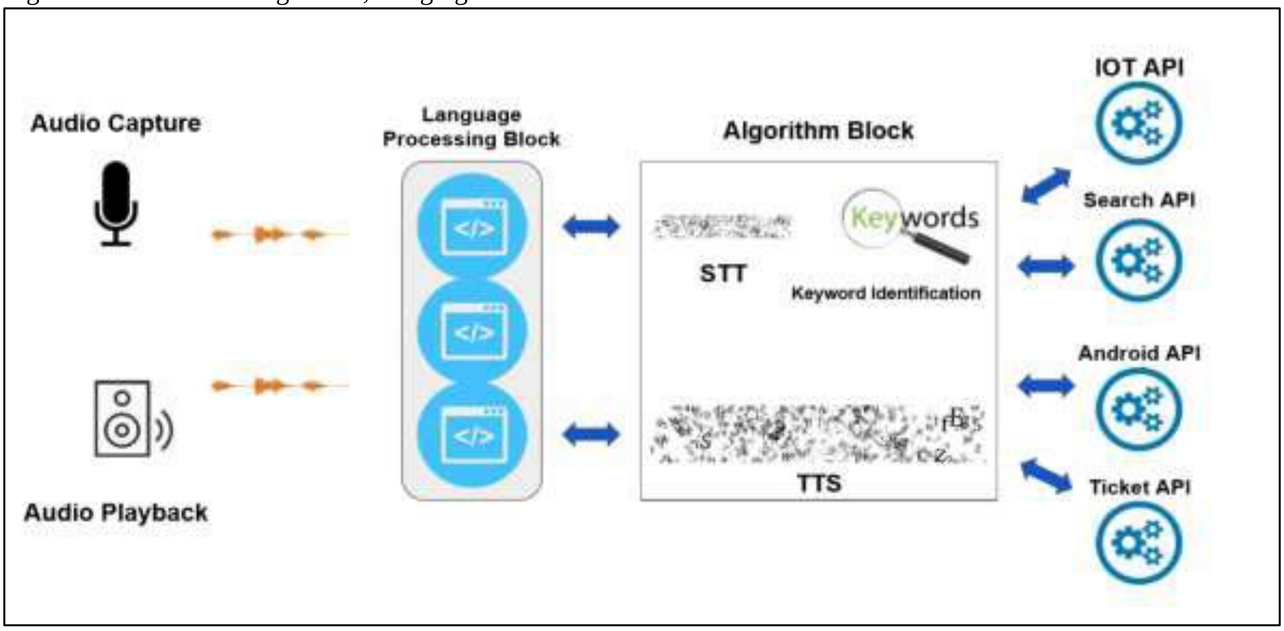
 

the parts of the speech waveform2 where no speech is produced and the lighter (red) parts represents intensity if speech is produced. speech waveform is given in the time domain. For the spectrogram Welch's method is used, which uses averaging modified periodograms [3]. Parameters used in this method are block size K= 320, window type Hamming with 62.5% overlap resulting in blocks of 20 ms with a distance of 6.25ms between block.

## III. RELATED WORK

An Intelligent Voice Assistant system on robot platform was developed as Associate in Nursing robot application that demonstrate the employment of linguistic communication process that helps to send messages and even use the in-build mobile application by victimization voice commands. this technique was surveyed to use the mailing and calendar wherever user were ready to mail and conjointly produce their event victimization voice command. Home Automation system supported net of Things was tested to figure satisfactorily by connecting straightforward appliances thereto and therefore the application appliances were with success controlled remotely through net. The designed system not solely monitors the detector knowledge like temperature, gas, light, motion sensors however conjointly actuates a method per the necessity. as an example, switching on the lights once it gets dark. It also stores the detector parameters within the cloud during a timely manner. this can facilitate user to analyse the condition of assorted parameters within the home anytime anyplace.Everyone should be aware of Siri, Cortana, Google currently or Watson or with any of the infinite fictional virtual assistant. These real-life virtual assistants are not as good as Ironman's Jarvis, however their meant perform is basically constant, voiceactivated computing power driven by computing. raise an issue, get a solution. provides a command, get results. Here is that the insight of a number of the non-public Assistant devices. Ivee Sleek could be a voice activated timepiece by Interactive Voicethat produces setting Associate in Nursing alarm,changing Associate in Nursing alarm sound, Assoiate in Nursing turning an alarm on or off utterly hands-free. in addition,Ivee Sleek encompasses a form of attainable voice commands starting from the present date and time to uncategorizable inquiries. Homey could be a voice-activated home automation hub created by Netherlands-based start-upAthom.It comes with multiplatform Smartphone app and it can still communicate with a bunch of differentlyconfigured gadgets promptly. It's multi-lingual and understands English, Dutch, Spanish, and French. it's compatible with a bunch of app enabled good home product. Amazon's Alexa could be a "Virtual-Assistant" manages to line itself apart. not like mobile-based virtual assistants like Siri, Alexa is centralized among dedicated, in-home Amazon devices -- most notably the Amazon Echo, Associate in Nursing always-on, always-listening Internet-connected speaker. Here area unit some options of Alexa.
Stream music: raise Alexa to play a song and it'll stream it from the Amazon Prime Music Library.

- Read the headlines: Alexa will scan out headlines from the news shops of your selection on thetopicsyoucare regarding.
- Keep tabs on traffic and therefore the weather: Alexa can gayly scan off the forecast, or let you know if there is Associatein Nursing accident electronic jamming up your morning commute.
- Set timers and alarms: you'll be able to tell Alexa to wake you up each weekday morning at seven a.m.or raise it to line a timer.
- Answer your questions: Alexa will search for basic facts, solve mathematical issues, or maybe tell you a joke .Alexa, Sleek, homy are not one in all the primary home automation systems that responds to voicecommands, though. CastleOS has been aroundsincel ate 2012, however its central hub and voicecontrol app will solely carefor a Windows computer. House Logix's VoicePad is another such voice-activated hub.



## ACKNOWLEDGEMENTS

It gives us a great pleasure to finalize this project on "Artificial Intelligence"domain with title. "Saarush the Intelligent System"We are thankful to our guide Mrs.A.A.Patkar for the encouragement and support that they have extended and also for their effort to the final Capstone project..We would also like to thank our coordinator, Mrs.A.A.Patkar for her valuable guidance and support throughout the processes. We sincerely express thanks to our HOD Mr.V.N. Kukre of Computer Engineering Department for allowing us and providing support for approving our project. We would also thank to all Staff members of Computer Engineering Department for showing us the way for finalizing project proposal.

## IV. CONCLUSION

Thus, we have studied about Python programming and the architecture of Visual Studio Code. We also learnt about working principle of an AI System as well as Machine Learning. We learn from AI how to pamper to people, and aware them from future technologies.

## REFERENCES

[1] Rabiner Lawrence, Juang Bing-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 JohnG., Discrete-Time Processing of Speech Signals, IEEE Press, ISBN 0-7803-5386-2

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

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

[5] AshishJain,HohnHarris,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 2 Oct 2012.

[7] http://web.science.mq.edu.au/~cassidy/comp449/html/ch11s02.html , downloaded on 2 Oct 2012.

[8] Hiroaki Sakoe and Seibi Chiba, Dynamic Programming algorithm Optimization for spoken word Recognition, IEEE transaction on Acoustic speech and Signal Processing, Fecruary 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 BalashekS, Automatic Recognition of Spoken Digits, J. Acoust. Soc. Am., 24 (6):637-642, 1952