International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 5, Issue 4, 2018, PP 1-11
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
DOI: http://dx.doi.org/10.20431/2349-4859.0504001
www.arcjournals.org

Execution of Speech Recognition Model for Noise Removal

Dr. Gagandeep Jagdev

Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB).

***Corresponding Author: Dr. Gagandeep Jagdev,** Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB).*

Abstract: *The practice of converting spoken words into text by the system is referred as speech recognition. Speech recognition is one of the prominent biometric technique. The research paper is intended to develop a speech recognition system with GUI interface. The interface developed is able to hear and record the user speech without the need of microphone. The recorded speech undergoes the noise removal process via cross correlation method. The processed speech is then added to the database. The given input is then compared against the speech files in the database and appropriate results are generated.*

Keywords: *Cross correlation, frequency, noise, speech recognition, sampling rate.*

1. Introduction

Speech has been most common form of human communication. Due to advancements in speech recognition technology, the computer system today understands human languages and human voice commands. The major challenge concerning speech recognition is its alteration by accents, mannerism, and dialects [10]. In technical terms it can be said that speech recognition is the ability of a machine to recognize phrases and words in spoken language and convert them into machine-readable format [1, 2].

The main components of speech are mentioned as under [3, 4].

Utterance – An utterance refers to speaking of words that represent a single meaning to the computer. Utterances can range from single word to multiple sentences.

Speaker dependence – These systems are designed around a specific speaker. These systems are accurate for particular speaker and less accurate for other speakers. These systems give positive results when speaker speaks in a consistent voice and tempo.

Vocabularies – Vocabularies include words that are recognized by speech recognition system. Computer recognizes smaller vocabularies more accurately as compared to larger vocabularies. There is no such restriction that the word has to be of single word. It can be of a sentence or two [5].

Accuracy – Accuracy refers to identifying an utterance and also notifying whether spoken utterance is in its vocabulary or not. Efficient speech recognition systems have an accuracy of more than 98%. The acceptable accuracy depends on the application.

Training–Many speech recognizers carries the ability to adapt to a speaker. An automatic speech recognition system is trained by repeating common words and phrases and having adjusting its comparison algorithm to match the speech of particular speaker. A good training results in enhancing accuracy of the system [6, 9].

2. Objectives and Adapted Research Methodology

The objectives of the research paper are mentioned as under.

- To implement a system capable of recognizing a user’s speech and creating an audio file this can be added up to create a dynamic template or database.
- To directly record the spoken words avoiding the problems with use of microphone.

- To identify the best matched audio file from the template when an input is provided externally on the basis of graphs created by considering correlation.
- To minimize the impact of noise on the quality of actual spoken words.

The research methodology will include following phases for accomplishing the desired objective [7, 8].

Study: The study phase will emphasis on studying speech recognition as biometric in detail involving its current status, limitations, implementation, advantages and disadvantages.

Analysis: The next phase is that of analysis. The deliverable result at the end of this phase will be a document identifying the specifications to implement the most suitable algorithm in terms of accuracy.

Design and Development: This phase will use the requirement specification document and will convert the requirements into framework. The framework will define the components, their interfaces and behaviors. The deliverable design document will be the architecture that describes a plan to implement the concerned algorithm. It represents the “How” phase.

Testing: This phase will include various tests and experiments that will help to discover potential errors and limitations in the implementation phase. Testing is the process with specific intent of finding errors prior to delivery to the end user. Strategy for testing comprises of test case design methods and provides guidance describing the steps to be conducted as part of testing, plan when and where they need to be imposed and how much effort, time and resource will be required.

Implementation: In this phase the proposed implementation model will be implemented in a simulated environment to achieve the objective of the system. This phase uses the architectural document from the design phase and the requirement document from the analysis phase, to obtain and use the model. The implementation phase deals with issues of quality, performance, maintenance etc.The research makes use of Matlab as a simulation tool.

The implementation phase is distributed into two segments.

The first segment involves designing a GUI interface capable of recording user speech, removing noise from it, and storing it into the database.

The second segment deals with recognizing the given input against the available database files and plotting appropriate graphs. The graph with minimum split is considered as best match.

The flowchart in Fig. 1 describes the entire speech recognition process to be adopted in the research work.

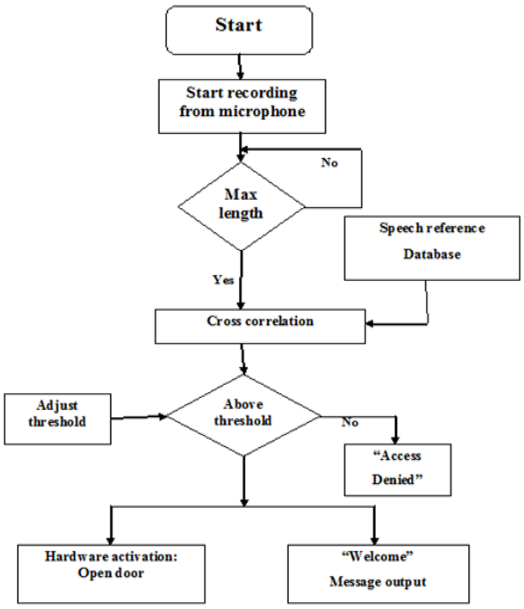


Fig1. *The figure describes the entire speech recognition process to be adopted in the research work.*

3. Implementation

Figure 2 shows the designed GUI interface for speech recognition. The left panel of the figure shows the textboxes and buttons which can be filled and adjusted as per user's preferences. The user can enter recording sampling rate, recording duration in seconds, and filename according to his/her priority. The right side shows three segments for plotting graphs mentioned as under.

Time in samples Frame number Wideband spectrogram

The figure shows 16000 as recording sampling rate and 3seconds as the duration of recording. The figure shows three graphical segments having below mentioned details.

First segment

X-axis - Time in samples/second which is dynamically provided by user.

Y-axis - Indicates the scale for value

Second segment

X-axis - Indicates the number of frames which varies according to recorded

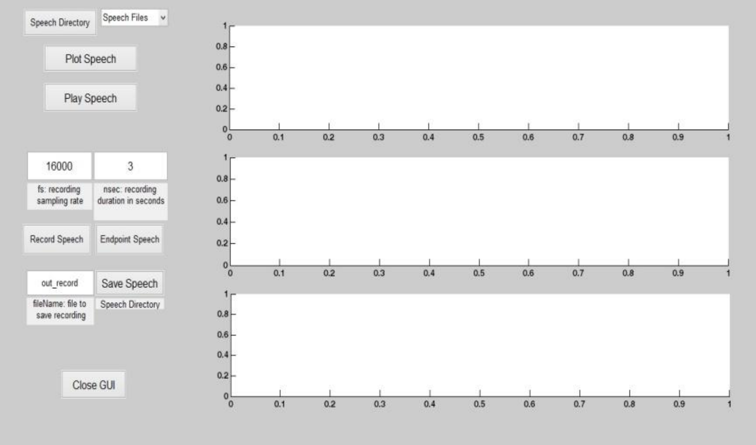
Speech.

Y-axis - Indicates the scale for value.

Third segment

X-axis - Indicates the wide spectrogram of the recorded speech

Y-axis - Indicates the frequency scale in Hertz


16000

fs: recording sampling rate

3

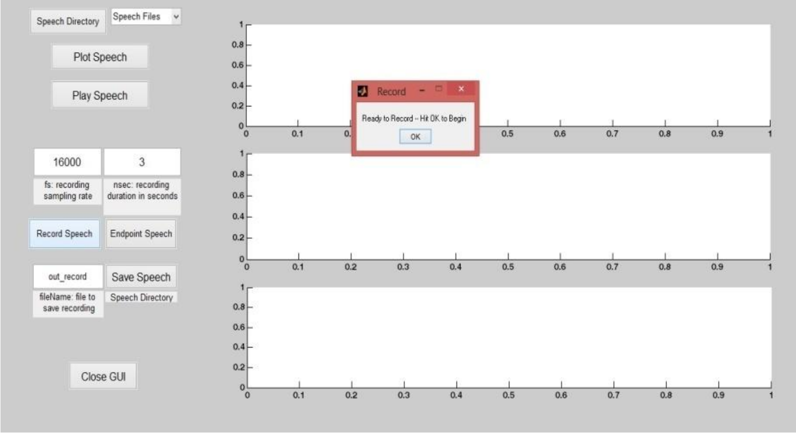
nsec: recording duration in seconds

out_record

fileName: file to save recording

Fig2. *The figure shows the designed GUI interface for speech recognition*

On pressing the "Record Speech" button, a pop appears on the screen asking to press "ok" to begin recording.


16000

fs: recording sampling rate

3

nsec: recording duration in seconds

out_record

fileName: file to save recording

Record

Ready to Record – Hit OK to Begin

Fig3. *The figure shows the pop-up message asking to begin the recording process*

Fig. 4 shows the entered name with which the recorded file is to be saved. In this case the entered file name is “gagans”.

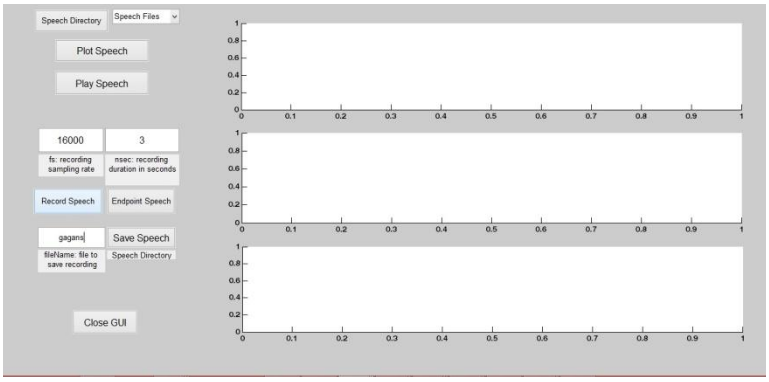
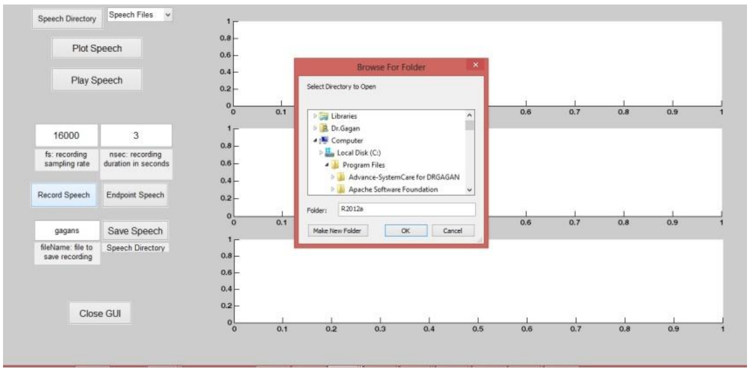

fs: recording sampling rate: 16000
nsec: recording duration in seconds: 3
fileName: file to save recording: gagans

Fig4. *The figure shows the entered name with which the recorded file is to be saved*

Fig. 5 shows the pop-up asking to save the file at an appropriate path.


fs: recording sampling rate: 16000
nsec: recording duration in seconds: 3
fileName: file to save recording: gagans

Browse For Folder
Select Directory to Open
- Libraries
- Dr.Gagan
- Computer
  - Local Disk (C:)
    - Program Files
      - Advance-SystemCare for DRGAGAN
      - Apache Software Foundation

Folder: R2012a

Fig5. *The figure shows the pop-up asking to save the file at an appropriate path*

Fig. 6 shows the plotted graph as per 16000 samples/second.

In first segment, X-axis indicates the “Time in samples per second” and Y-axis indicates the dynamically recorded “value”.

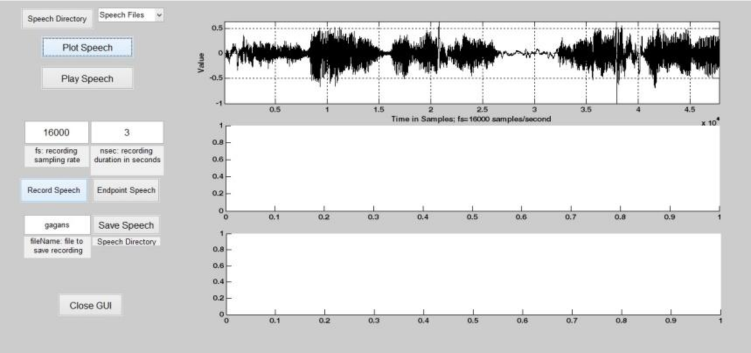

fs: recording sampling rate: 16000
nsec: recording duration in seconds: 3
fileName: file to save recording: gagans

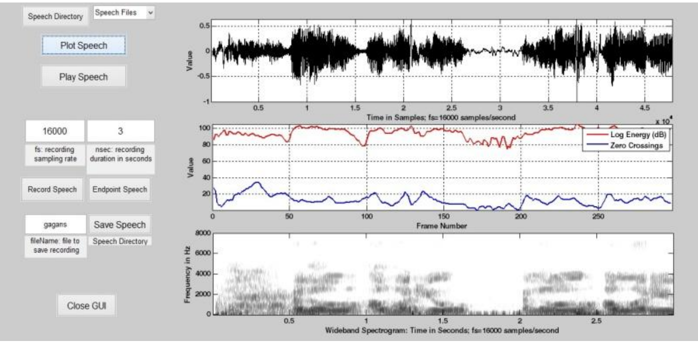
Fig6. *The figure shows the plotted graph as per 16000 samples/second*

Fig. 7 shows the plotted lines for involved Log Energy (red color) and Zero Crossings (blue color) against Frame number. The bottom section depicts the wideband spectrogram.

In first segment, X-axis indicates the “Time in samples per second” and Y-axis indicates the dynamically recorded “value”.

In second segment, X-axis refers to the “frame number” of the recorded speech (red color refers to Log energy and blue color line refers to Zero crossings) and Y-axis shows the value scale.

In third segment, X-axis refers to the wide spectrogram of the recorded speech and Y-axis indicates the frequency scale in Hertz.


16000
fs: recording sampling rate

3
nsec: recording duration in seconds

gagans
fileName: file to save recording

Fig7. *The figure shows the plotted lines for involved Log Energy (red color) and Zero crossings (blue color) against Frame number and wideband spectrogram*

The file to be checked is entered in the code file “speech recognition.m” .On running the code, the entered file is checked against all the speech files present in the template or database. The entered file is compared with each file in the database and the subsequent graphs are plotted. The file which matches the most with the entered file shows the minimum scattered graph. The greater the graph is scattered, the lesser the entered file matches with the particular file in the database. The file whose graph is minimum scattered is the best match with the entered file to be checked.

Fig. 8 to Fig. 17 shows the appropriate plotted graphs against the entered file to be checked and it is found that the best match is made with Fig. 9 titled Figure 2.

The plotted figure shows the scattered graph when the first audio file in the database titled “one.wav” is matched with the inputted audio file “two.wav”.

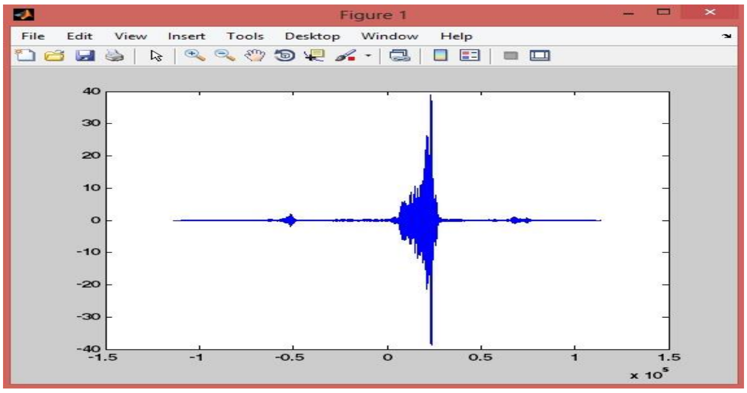


Fig8. *The figure shows the graph plotted after matching the entered file with the first file present in the database*

The plotted figure shows the scattered graph when the second audio file in the database titled “two.wav” is matched with the inputted audio file “two.wav”.

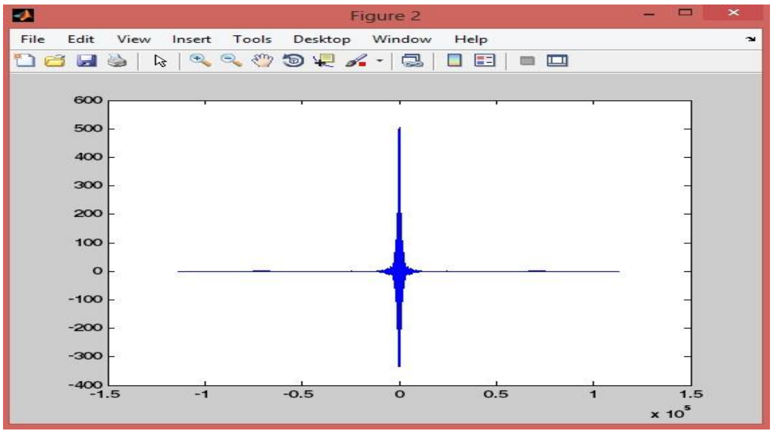


Fig9. *The graph is plotted after matching the entered file with the second file present in the database*

The plotted figure shows the scattered graph when the third audio file in the database titled “three.wav” is matched with the inputted audio file “two.wav”.

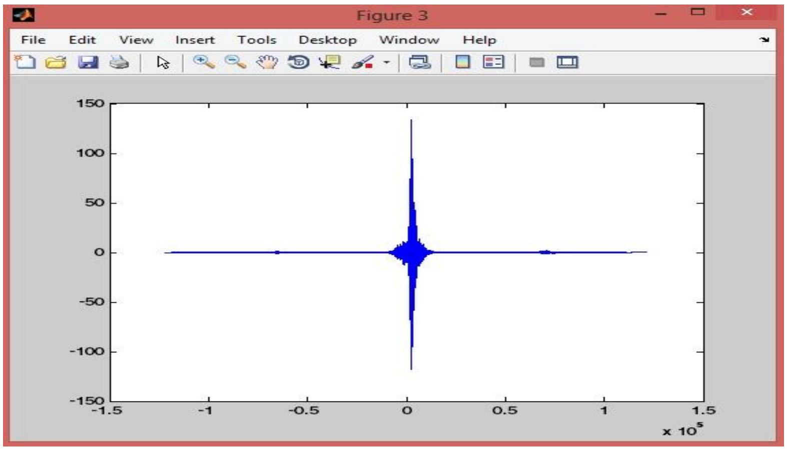


Fig10. *The graph is plotted after matching the entered file with the third file present in the database*

The plotted figure shows the scattered graph when the fourth audio file in the database titled "four.wav" is matched with the inputted audio file "two.wav".

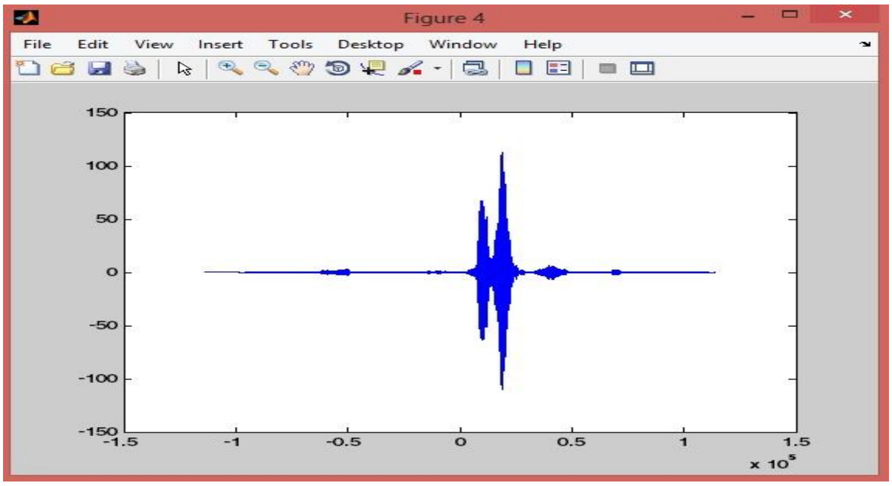


Fig11. *The graph is plotted after matching the entered file with the fourth file present in the database*

The plotted figure shows the scattered graph when the fifth audio file in the database titled "five.wav" is matched with the inputted audio file "two.wav".

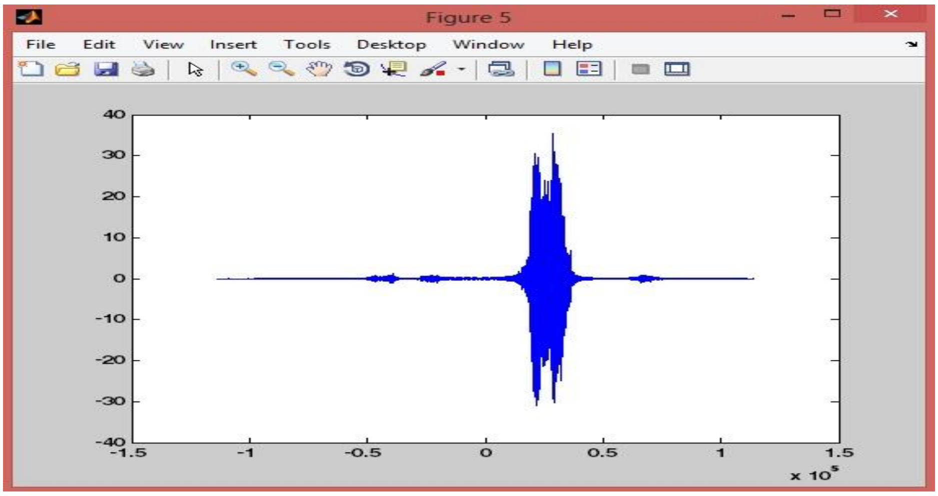


Fig12. *The graph is plotted after matching the entered file with the fifth file present in the database*

The plotted figure shows the scattered graph when the sixth audio file in the database titled "mytest1.wav" is matched with the inputted audio file "two.wav".

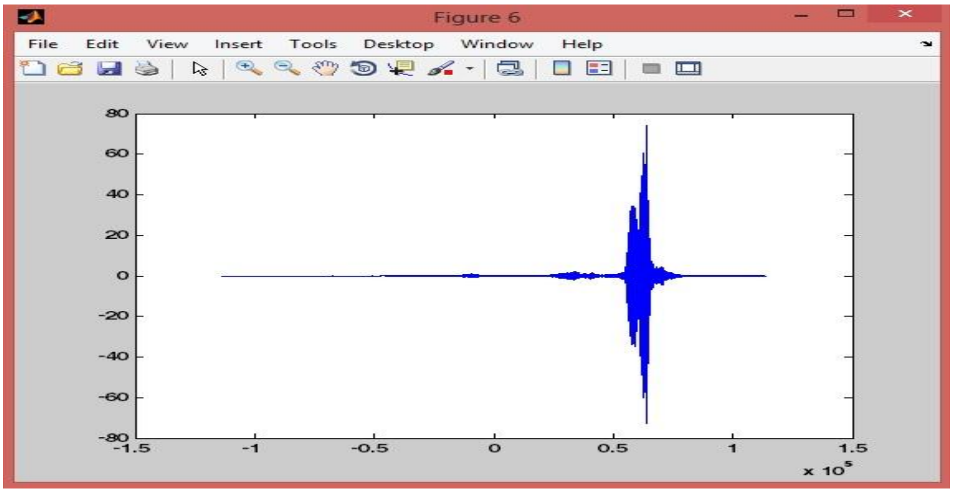


Fig13. *The graph is plotted after matching the entered file with the sixth file present in the database*

The plotted figure shows the scattered graph when the seventh audio file in the database titled “nine.wav” is matched with the inputted audio file “two.wav”.

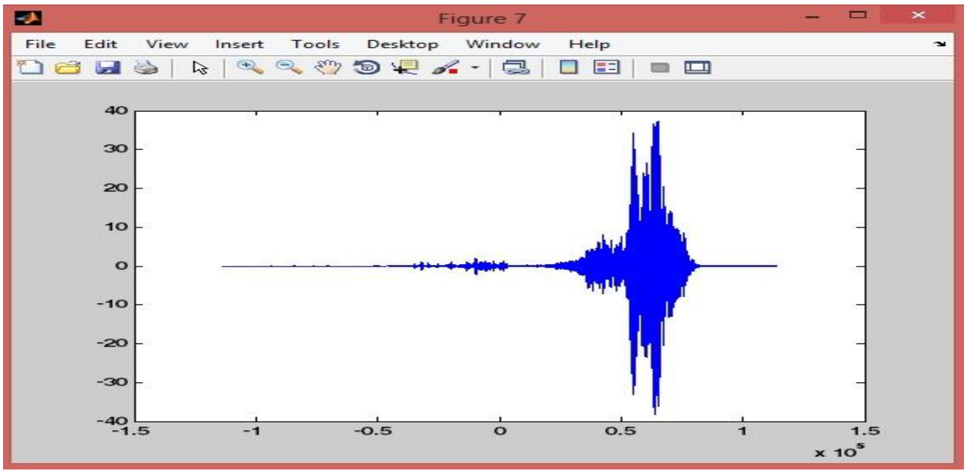


Fig14. *The graph is plotted after matching the entered file with the seventh file present in the database*

The plotted figure shows the scattered graph when the eight audio file in the database titled “gagan.wav” is matched with the inputted audio file “two.wav”.

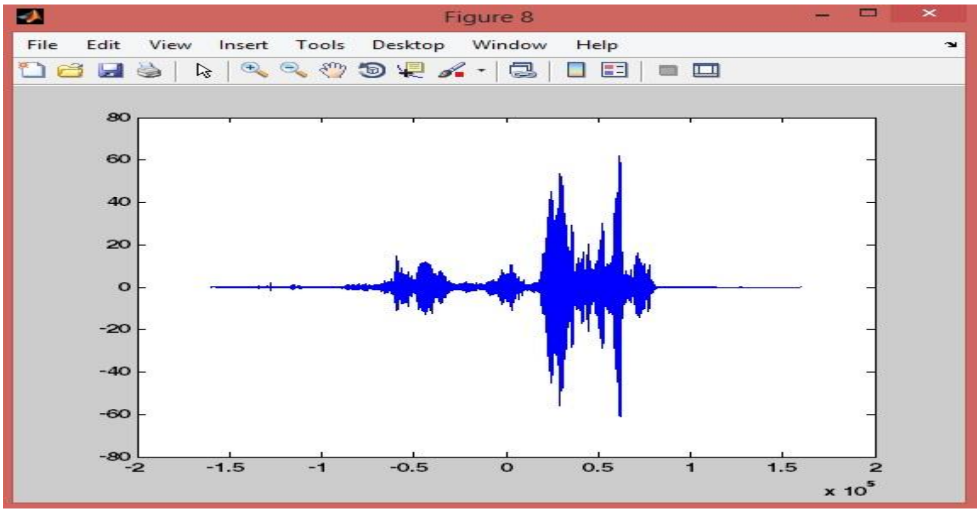


Fig15. *The graph is plotted after matching the entered file with the eight file present in the database*

The plotted figure shows the scattered graph when the ninth audio file in the database titled “july1.wav” is matched with the inputted audio file “two.wav”.

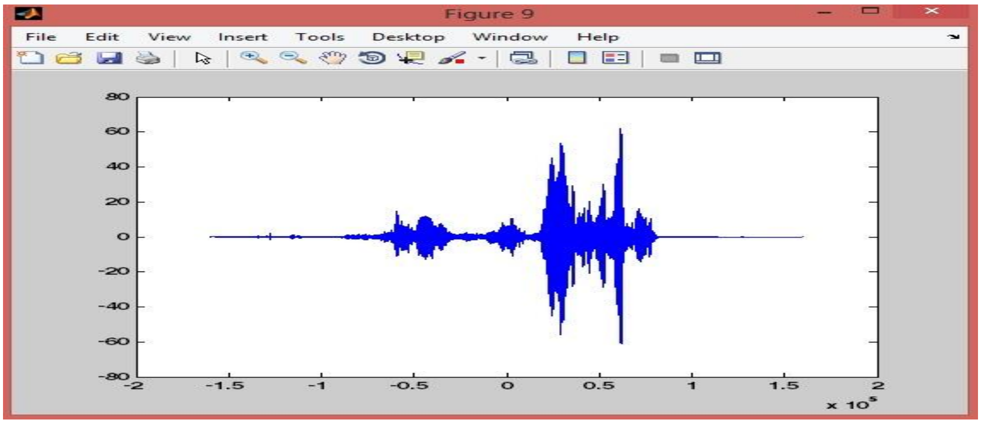


Fig16. *The graph is plotted after matching the entered file with the ninth file present in the database*

The plotted figure shows the scattered graph when the tenth audio file in the database titled “july2.wav” is matched with the inputted audio file “two.wav”.

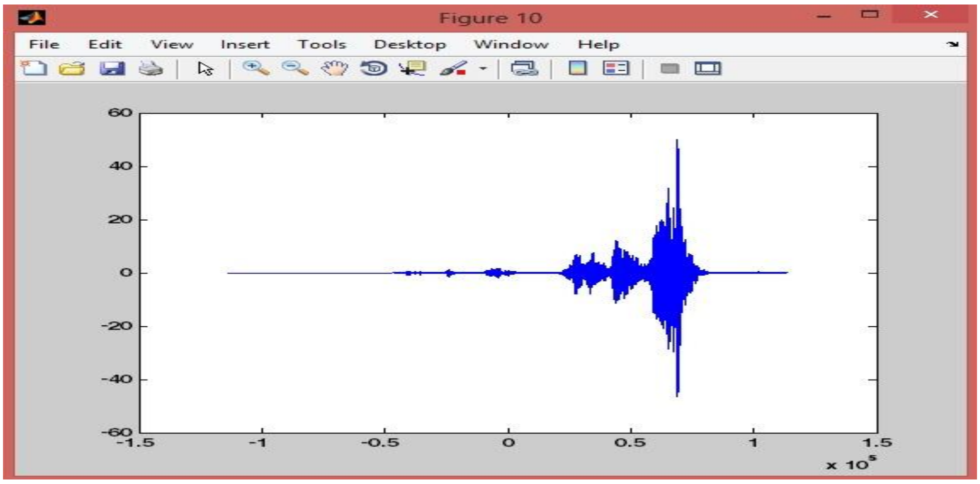


Fig17. *The graph is plotted after matching the entered file with the tenth file present in the database*

The different readings obtained from the designed interface are detailed in the Table 1 to Table 4 mentioned as under.

Table1. *Maximum log energy dB*

File Name	Sampling Rate	Duration	Maximum Log Energy dB
One1	16000	5	120
Two2	16000	5	125
Hello3	20000	5	115
Bye4	20000	8	120
A5	16000	8	120
B5	16000	8	119
C5	20000	5	120
D5	20000	8	108
Gagandeep	30000	5	110
Jagdev	30000	8	110

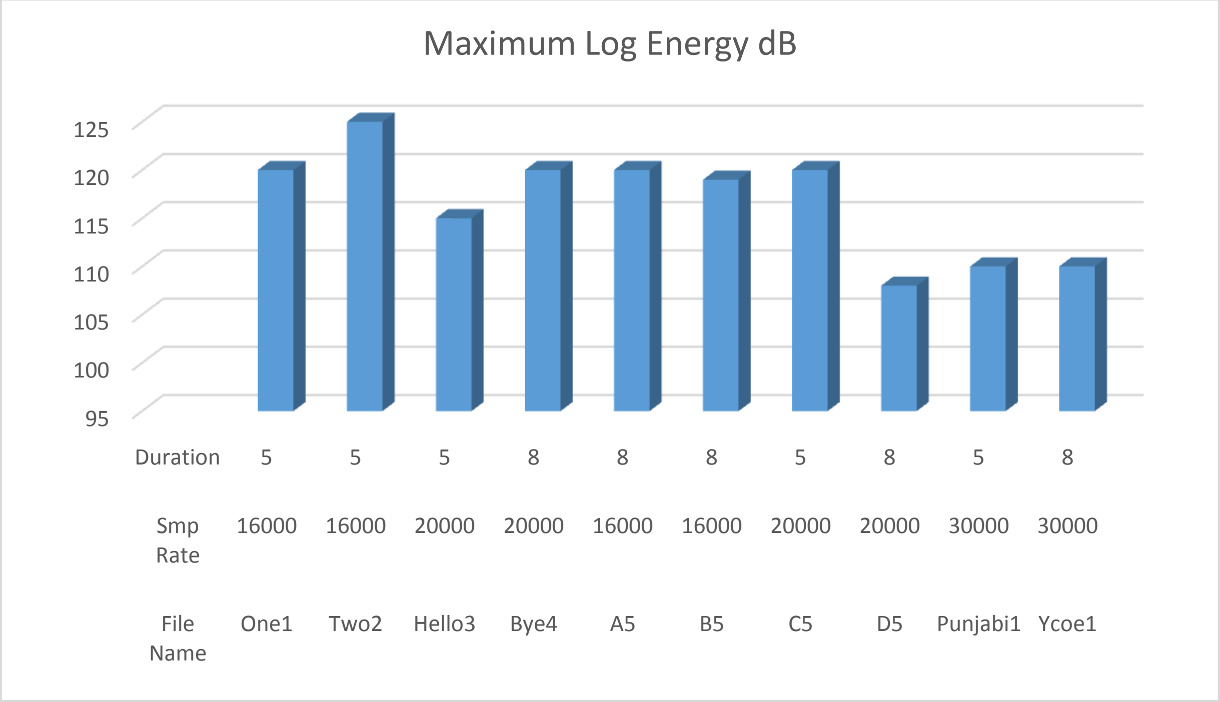


Fig18. *The figure depicts reading with respect to Maximum Log Energy*

Table2. *Zero Crossing*

File Name(.wav)	Sampling Rate	Duration	Zero Crossing
One1	16000	5	22
Two2	16000	5	18
Hello3	20000	5	83
Bye4	20000	8	82
A5	16000	8	102
B5	16000	8	82
C5	20000	5	90
D5	20000	8	93
Gagandeep	30000	5	78
Jagdev	30000	8	79

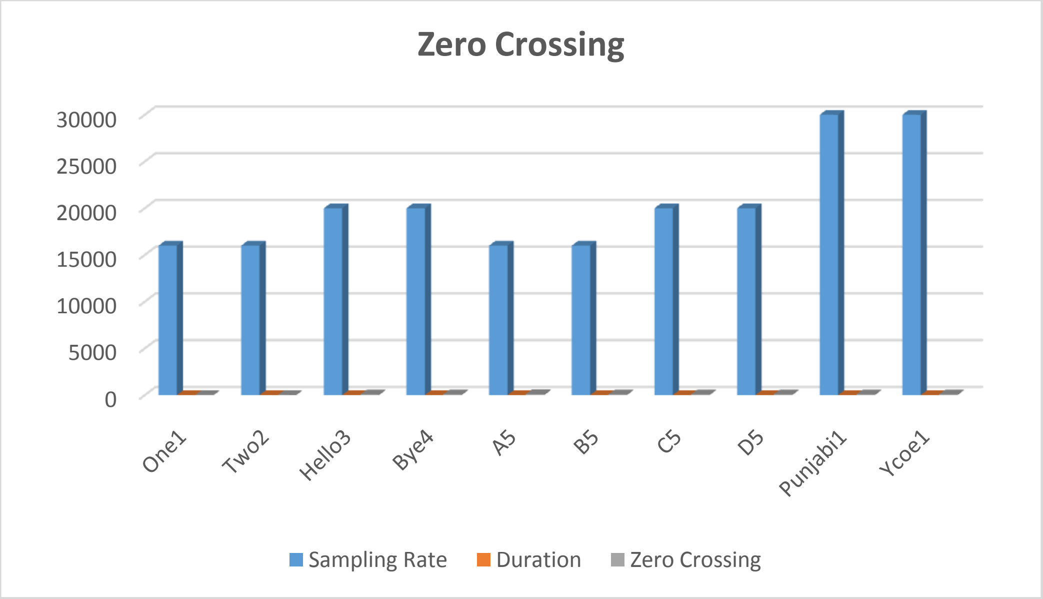


Fig19. *The figure depicts reading with respect to Zero Crossing*

Table2. *Concentrated Frequency in Hz*

File Name	Sampling Rate	Duration	Concentrated Frequency in Hz
One1	16000	5	1600
Two2	16000	5	1000
Hello3	20000	5	1100
Bye4	20000	8	500
A5	16000	8	1200
B5	16000	8	1000
C5	20000	5	700
D5	20000	8	600
Gagandeep	30000	5	200
Jagdev	30000	8	300

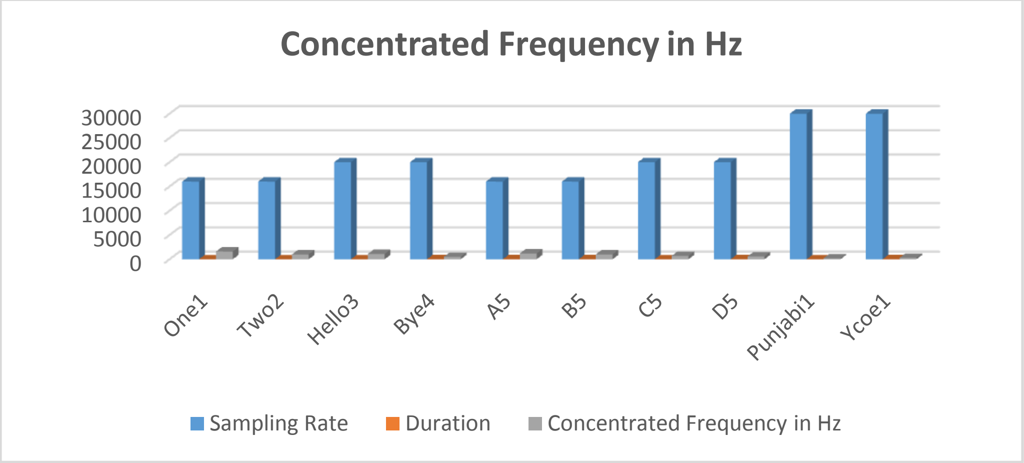


Fig20. *The figure depicts reading with respect to Concentrated Frequency in Hz*

Table3. *Maximum Frequency in Hz*

File Name	Sampling Rate	Duration	Maximum Frequency in Hz
One1	16000	5	7900
Two2	16000	5	7000
Hello3	20000	5	7600
Bye4	20000	8	7800
A5	16000	8	7900
B5	16000	8	7800
C5	20000	5	7900
D5	20000	8	7900
Punjabi1	30000	5	8000
Ycoe1	30000	8	7000

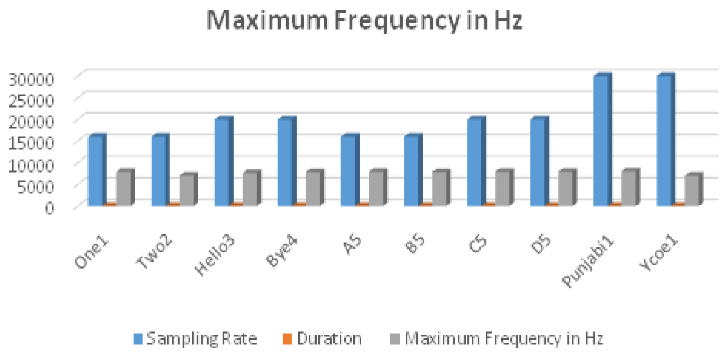


Fig21. *The figure depicts reading with respect to Concentrated Frequency in Hz*

4. Conclusion

Speech recognition is an emerging technique that helps in recognizing the human speech by the machine. There are numerous researches going on in building a model for recognizing speech and converting into text. The research work successfully dealt with the problem of removal of noise from the recorded speech. The enhanced speech is then matched against the appropriate file in the template. The research work described the speech recognition technology and application development towards implementing a user interface that can respond to anything spoken by the user.

References

[1] A. Pramanik, R. Raha, (2018), "Automaticspeech recognition using correlation analysis", World Congress on Information and Communication Technologies, pp. 670-674.

[2] A. Shajee, D. Patel, R. Mishra, H. Saikia, Narendran. M, (2017), "A survey: Speech recognition application", International Journal of Advances in Electronics and Computer Science, 4 (11), pp. 56-59.

[3] C. H. Wu, M. H. Su, W. B. Liang, (2017), "Miscommunication handling in spoken dialog systems based on error-aware dialog state detection", Journal on Audio, Speech and Music processing, 9, pp. 1-17.

[4] C. Y. Chiang, (2018), "A parametric prosody coding approach for Mandarin speech using a hierarchical prosodic model", Journal on Audio, Speech and Music processing, 5, pp. 1-24.

[5] G. Disken1, Z. Tufekci, U. Cevik, (2017), "A robust polynomial regression-based voice activity detector for speaker verification", Journal on Audio, Speech and Music processing, 23, pp. 1-16.

[6] G. Farahani, (2017), "Robust feature extraction using autocorrelation domain for noisy speech recognition", Signal & Image Processing: An International Journal, 8 (1), pp- 23-44.

[7] G. Farahani, (2017), "Autocorrelation-based noise subtraction method with smoothing, overestimation, energy, and cepstral mean and variance normalization for noisy speech recognition", Journal on Audio, Speech and Music processing,13, pp. 1-16.

[8] H. Zhang, S. Warisawa, I. Yamada (2014), "An Approach for Emotion Recognition using Purely Segment-Level Acoustic Features", International conference on kansei engineering and emotion research, pp. 1-11.

[9] https://www.tutorialspoint.com/biometrics/voice_recognition.html. Last accessed at 9:10pm on 27march, 2018.

[10] J.V. Doremalen, C. Cucchiarini, H. Strik, (2009), "Optimizing automatic speech recognition for low-proficient non-native speakers", Journal on Audio, Speech and Music Processing, 2010, pp. 1-13.

Author's Biography

Dr. Gagandeep Jagdev, is a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). His total teaching experience is more than 12 years and has 125 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of several international peer-reviewed journals and has been active Technical Program Committee member of several international and national conferences conducted by renowned universities and academic institutions. His field of expertise is Big Data, Data Mining, Image processing in Diabetic Retinopathy, ANN, Biometrics, RFID, Cloud Computing, Cloud security, Cryptography, and VANETS.

Citation: *Jagdev, G. (2018). Execution of Speech Recognition Model for Noise Removal. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(4), pp.1-11. http://dx.doi.org/10.20431/2349-4859.0504001*