Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.7 No.3, 2018, pp.48-52
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajcst-2018.7.3.1899

Speech Recognition using Cross Correlation Algorithm Intended for Noise Reduction

Gagandeep Kaur[1] and Seema Baghla[2]
[1]PG Student, [2]Assistant Professor, Department of Computer Engineering,
Yadavindra College of Engineering, Punjabi University Guru Kashi Campus,
Talwandi Sabo, Bathinda, Punjab, India
E-Mail: gagansidhu1593@gmail.com, garg_seema238@yahoo.co.in

(Received 12 September 2018; Revised 25 September 2018; Accepted 9 October 2018; Available online 16 October 2018)

***Abstract* - Biometrics is presently a buzzword in the domain of information security as it provides high degree of accuracy in identifying an individual. Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Rudimentary speech recognition software has a limited vocabulary of words and phrases, and it may only identify these if they are spoken very clearly. The research work is intended to build a GUI environment which would provide provisions to record the speech and would assist in multiplying the database. The research work is primarily focused to implement a system capable of recognizing a user's speech and creating audio files that can be added up to create a dynamic template or database. The research work emphasizes on directly recording the spoken words avoiding the problems with use of microphone. On appropriate recording and removal of the noise, the best matched audio file from the template is recognized when an input is provided externally on the basis of graphs created by considering correlation.**

***Keywords:* Noise, speech recognition, cross correlation, biometrics, and spoken words.**

I. INTRODUCTION

Biometric systems are automated methods of verifying or recognizing the identity of a living person on the basis of some physiological characteristics, or some aspects of behavior, like patterns, speech recognition and fingerprint etc. Recognize the human speech for security purpose and remove the background noise with better accuracy. The way toward changing over talked words into content is known as speech acknowledgment. Speech Recognition makes its place among most talked about procedures of biometrics. However, before such determined objectives are achieved, a lot remains to be done at the ground level. Cross correlation algorithm is used to remove the background noise from the human speech and obtained better results. Recording the human speech without the microphone in the database and apply cross correlation algorithm for removing the background noise.

The research uses MATLAB for the implementation of my source code. The work would distribute implementation in two MATLAb files.

1. The first module depicts a GUI interface created to record the speech at certain frequency and for any particular time period dynamically. The noise removal using cross correlation is done here itself and appropriate wide spectrogram is created. The recorded file is added to the database.
2. The second module deals with recognizing and plotting graphs when exact match of the file is found in the template. Each file in the database is matched against the input file and appropriate graphs for each match are plotted. The graph which is least scattered is the best match of the inputted file which is been checked.

II. LITERATURE REVIEW

Pramanik and Raha [1] described the principle of Correlation to recognize the spoken word perfectly. Present mobile devices are having limited memory and processing capacities which are adding several challenges to ASR. Shajee *et al.* [2] discussed a survey on speech recognition application. Authors compared and summarized the well-known methods used in various stages of speech recognition system. Wu *et al.* [3] proposed an approach accomplished a promising execution for non-downplaying recuperation and misunderstanding repair and additionally an agreeable undertaking achievement rate for the exchanges utilizing the proposed technique. Chiang [4] discussed a parametric prosody coding approach for Mandarin discourse utilizing a various leveled prosodic model expressed that past research a novel parametric prosody coding approach for Mandarin discourse is proposed. Disken *et al.* [5] proposed an algorithm showed superior verification performance both with the conventional GMM- universal background model and universal background model (UBM) method, and the state of-the-art i-vector method. Farahani [6] discussed the robust features extractions using autocorrelation domain for noisy speech recognition. This paper depicted a straightforward and compelling strategy for diminishing the impact of clamor on the autocorrelation of the perfect flag. Zhang *et al.* [7] designed two segment selection approaches viz. miSATIR and crSATIR, for selecting utterance segments for use in extracting features that are based on information theory and correlation coefficients to create the purely segment-level concept of the model.

Doremalen *et al.* [8] expressed that computer-assisted language learning applications for enhancing the oral aptitudes of low-capable students need to adapt to non-local discourse that is especially testing. The primary analysis on expression determination demonstrates that the disentangling procedure can be enhanced by advancing the dialect show and the acoustic models. Himanshu *et al.* [9] described a phonetic approach, pattern acknowledgment approach and artificial insight approach to abridge and think about a portion of the notable techniques utilized as a part of different phases of discourse acknowledgment framework and recognize inquire about subjects. Hasan [10] stated that previous research correlation based method using the autocorrelation function and the YIN. The autocorrelation function and also YIN is a popular measurement in estimating fundamental frequency intimae domain. Saini *et al.* [11] proposed technique speech recognition has been taken over by a deep and smart learning method called long short-term memory. Authors simulate different algorithms of speech processing in MATLAB. Hidden Markov Model and cross-correlation. Stern *et al.* [12] presented a study using signal handling strategies that are persuaded by our comprehension of binaural observation and binaural innovation. Elavarasi and Suseendran [13] proposed a procedure by actualizing this discourse acknowledgment innovation in a vastly improved manner to deal with the gadgets through voice from anyplace and in whenever. Shankaranand *et al.* [14] examined a strategy to perceive the discourse quicker with more precision, speaker acknowledgment is trailed by discourse acknowledgment. MFCC/Autocorrelation is utilized to extricate the attributes from the information discourse motion as for a specific word articulated by a specific speaker.

III. METHODOLOGY

The methodology includes the following phases [3, 4, 7, 9] for accomplishing the objectives.

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

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

Phase III: Design and Development: This phase uses the requirement specification document and converts the requirements into framework. The framework defines the components, their interfaces and behaviors. The deliverable design document is the architecture that describes a plan to implement the concerned algorithm. It represents the “How” phase.

Phase IV: Testing: This phase includes various tests and experiments that helps 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 are required.

Phase V: Implementation: In this phase the proposed implementation model has been 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 uses MATLAB for the implementation of my source code. The work would distribute implementation in two MATLAB files. Figure 1 [1] shows the algorithm for flow of cross correlation for speech recognition

1. The first module depicts a GUI interface created to record the speech at certain frequency and for any particular time period dynamically. The noise removal using cross correlation is done here itself and appropriate wide spectrogram is created. The recorded file is added to the database.
2. The second module deals with recognizing and plotting graphs when exact match of the file is found in the template. Each file in the database is matched against the input file and appropriate graphs for each match are plotted. The graph which is least scattered is the best match of the inputted file which is been checked.

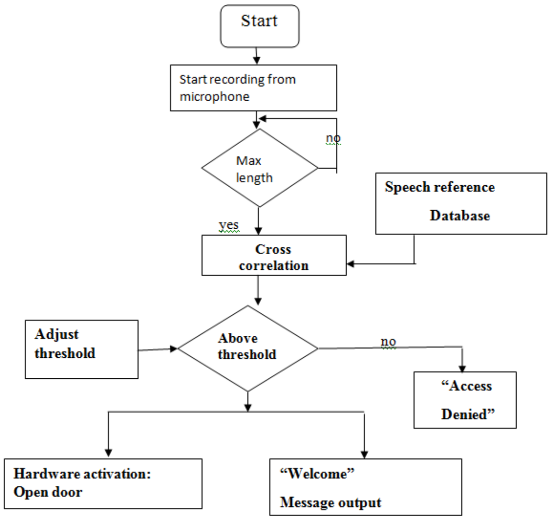

Fig. 1 Cross correlation flow for speech recognition

IV. RESULTS

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. Time in samples.

1. Frame number
2. Wideband spectrogram

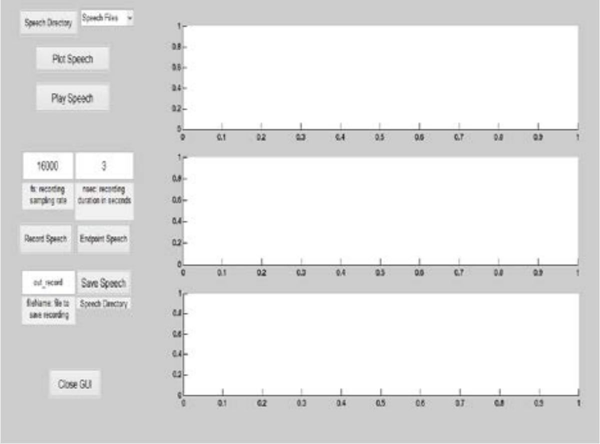

Fig. 2 Designed GUI interface for speech recognition

Figure 3 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”.

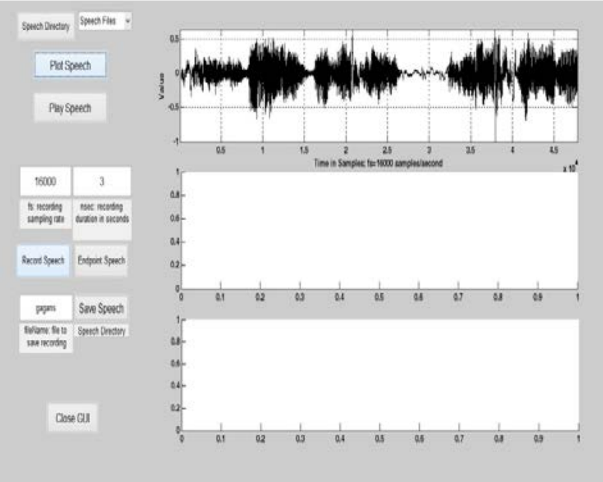
Speech Directory
Speech Files
Plot Speech
Play Speech
16000
3
fs: recording sampling rate
nsec: recording duration in seconds
Record Speech
Endpoint Speech
pgpms
Save Speech
filename: file to save recording
Speech Directory
Close GUI
Value
Time in Samples; fs=16000 samples/second

Fig. 3 Graph at 16000 samples/second

Figure 4 shows the plotted lines for involved Log Energy (red color) and Zero Crossings (blue color) against Frame number.

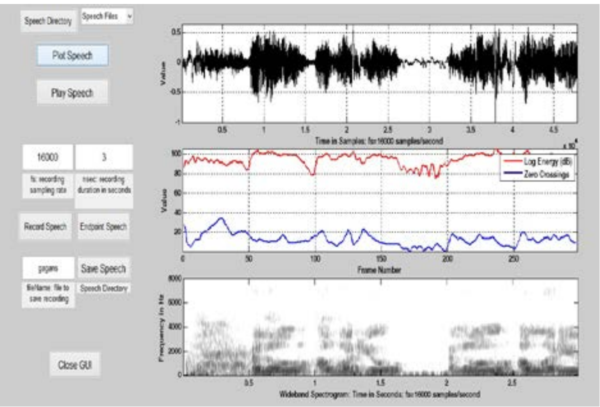
Speech Directory
Speech Files
Plot Speech
Play Speech
16000
3
fs: recording sampling rate
nsec: recording duration in seconds
Record Speech
Endpoint Speech
pgpms
Save Speech
filename: file to save recording
Speech Directory
Close GUI
Value
Time in Samples; fs=16000 samples/second
Log Energy (dB)
Zero Crossings
Frame Number
Frequency in Hz
Wideband Spectrogram: Time in Seconds; fs=16000 samples/second

Fig. 4 Log Energy (red color) and Zero crossings (blue color) against Frame number and wideband spectrogram

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. On running the code, the entered file is checked against all the speech files present in the template or database as shown in figure 5.

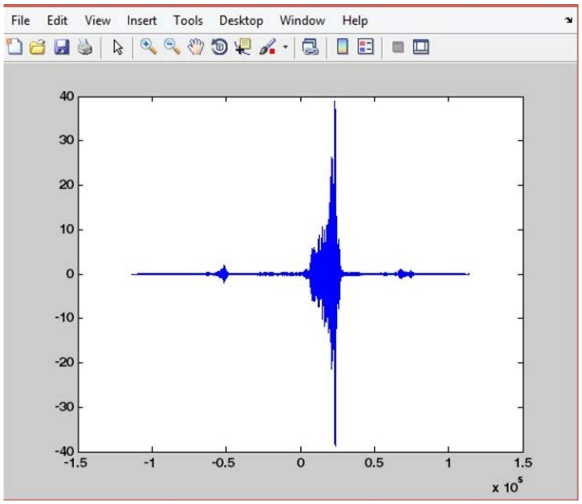

Fig. 5 Graph after matching the input file with the first file present in the database

The entered file is compared with each file in the database and the subsequent graphs are plotted. The file whose graph is minimum scattered is the best match with the entered file to be checked. Figure 6 shows the appropriate plotted graphs against the entered file to be checked and it is found that the best match is made.

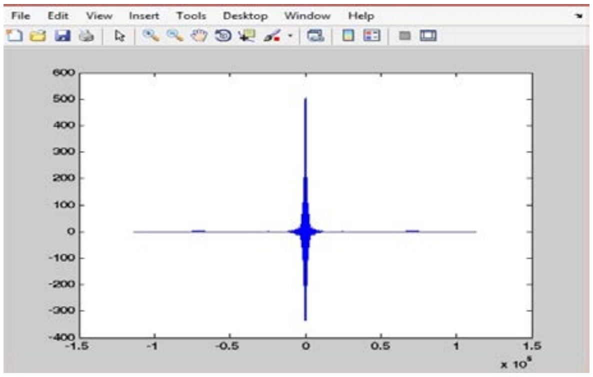

Fig. 6 Graph after matching the input file with the second file present the database

The different readings obtained from the designed interface are detailed in the Table I to Table IV. Table I depicts the maximum log energy attained by different wav files when executed against particular sampling rate for a defined duration in seconds and figure 7 depicts the maximum log energy attained by different wav files when executed against particular sampling rate.

TABLE I MAXIMUM LOG ENERGY DB.

File Name	Sampling Rate	Duration (sec)	Max.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
Punjabi1	30000	5	110
Ycoe1	30000	8	110

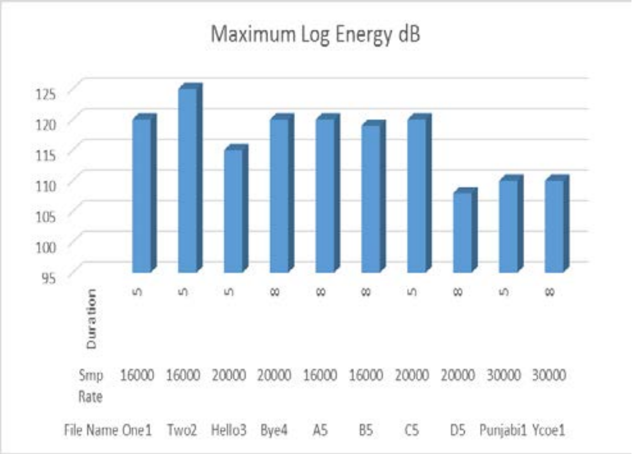

Fig. 7 Maximum Log Energy

Table II depicts the zero crossing attained by different wasv files when executed against particular sampling rate for a defined duration in seconds. Figure 8 depicts the zero crossing attained by different wav files when executed against particular sampling rate for a defined duration.

TABLE II ZERO CROSSING RESULT

File Name (.wav)	Sampling Rate	Duration (sec)	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
Punjabi1	30000	5	78
Ycoe1	30000	8	79

Zero Crossing
30000
25000
20000
15000
10000
5000
0
One1 Two2 Hello3 Bye4 A5 B5 C5 D5 Punjabi1 Ycoe1
■ Sampling Rate ■ Duration ■ Zero Crossing

Fig. 8 Reading with respect to zero crossing

Table III depicts the concentrated frequency attained by different wav files when executed against particular sampling rate for a defined duration in seconds. Figure 9 depicts the concentrated frequency in Hz attained by different wav files when executed against particular sampling rate for a defined duration in seconds.

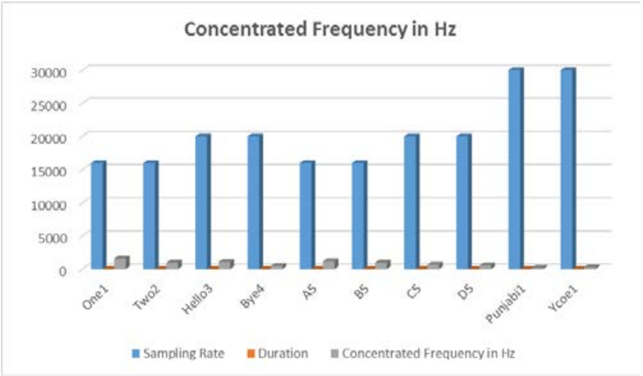

Fig. 9 The concentrated frequency in Hz

TABLE III CONCENTRATED FREQUENCY IN HZ

File Name	Sampling Rate	Duration (sec)	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
Punjabi1	30000	5	200
Ycoe1	30000	8	300

Table IV depicts the maximum frequency attained by different wav files when executed against particular sampling rate for a defined duration in seconds. Figure 10 depicts the maximum frequency attained by different wav files when executed against particular sampling rate for a defined duration in seconds.

TABLE IV MAXIMUM FREQUENCY IN HZ

File Name	Sampling Rate	Duration (sec)	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

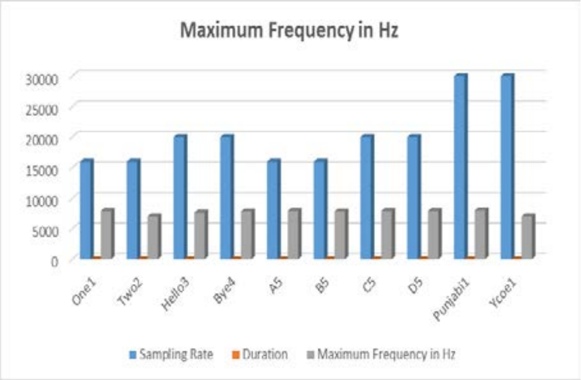

Fig. 10 The frequency in Hz

V. 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.

VI. FUTURE SCOPE

The obtained results can be improved by fine tuning the system with larger training databases. The next step would be to recognize live speech, which would require more resources including larger speech databases, acoustic models and exhaustive vocabularies to produce good recognition results. Very soon the speech recognition may become speech understanding. The speech of person may soon be grasped to learn the meaning behind the words. But still there is a long way to go in terms of software complexity and computational power. The researchers are confident that computers will be based on the concept of artificial intelligence. In the span of 25 years, the computers will surely talk back to the persons.

REFERENCES

[1] A. Pramanik, R. Raha, "Automatic speech recognition using correlation analysis", *World Congress on Information and Communication Technologies*, pp. 670-6742018,.

[2] A. Shajee, D. Patel, R. Mishra, H. Saikia, M. Narendran, "A survey: Speech recognition application", *International Journal of Advances in Electronics and Computer Science*, Vol. 4, No. 11, pp. 56-59, 2017.

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

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

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

[6] G. Farahani, "Robust feature extraction using autocorrelation domain for noisy speech recognition", *Signal & Image Processing: An International Journal*, Vol. 8, No. 1, pp. 23-44, 2017.

[7] H. Zhang, S. Warisawa, I. Yamada, "An Approach for Emotion Recognition using Purely Segment-Level Acoustic Features", *International Conference on Kansei Engineering and Emotion Research*, pp. 1-11, 2014.

[8] J.V. Doremalen, C. Cucchiarini, H. Strik, "Optimizing automatic speech recognition for low-proficient non-native speakers", *Journal on Audio, Speech and Music Processing*, Vol. 22, pp. 1-13, 2009.

[9] M. Himanshu, S. Kaur, V. Chaudhary, "Literature survey on automatic speech recognition system", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, No. 7, pp. 398-402, 2014.

[10] M. A. Hasan, "Correlation based fundamental frequency extraction method in noisy speech signal", *International Journal of Computer Science, Engineering and Information Technology*, Vol. 7, No. 1, pp. 1-12, 2017.

[11] N. K. Saini, A. M. Laxmi, N. Balai, "Data extraction from web using speech recognition", *International Journal for Scientific Research and Development*, Vol. 5, No. 11, pp. 175-177, 2018.

[12] M. Stern, C. Kim, A.R. Moghimi, A. Menon, "Binaural technology and automatic speech recognition", *International Congress of Acoustics*, pp. 1-10, 2016.

[13] S. Elavarasi, G. Suseendran, "Speech Recognition on Handling Device", *Jour of Adv Research in Dynamical & Control Systems*, Vol. 9, No. 6, pp. 97-103, 2017.

[14] S. Shankaranand, S. Manasa, M. Sharma, A.S. Nithya, K.S. Roopa, K.V. Ramakrishan, "An enhanced speech recognition system", *International Journal of Recent Development in Engineering and Technology*, Vol. 2, No. 3, pp. 78-81, 2014.