A Speech Processing Based System for Detection of Suspected User on GSM Network

Dharmendra Gurve[1], Mahendra Gurve[2]

[1] ECE Department, Dr. B. R. A. National Institute of Technology, Jalandhar, Punjab, India

[2] Computer Science & IT Department, Jaypee Institute of Information Technology, Noida, UP, India

Abstract: ***With rapid development of mobile technology the criminal activity has also increased. Harassing, obscene or threatening phone calls are illegal and are punishable. Anyone can be the victim of such calls, however existing technology have limitation in terms of identification of mobile user when user will not use SIM card or destroy it after making calls. Anyone can be the victim of harassing, annoying, obscene, or threatening telephone calls. This paper proposed an idea for detection of such suspected user on GSM network with speech recognizer. Speech recognizer is placed on global system for mobile communication (GSM) network more specifically at gateway mobile switching center (GMSC). At GMSC, speech recognizer continuously scan for suspected user's voice, once voice matches suspected user can be easily identify. Advantage of this method is that we do not need to wait for next call. Suspected user can be detecting during the call with someone else on same GSM network. Proposed system can be useful to find out the suspected users which are related with threatening or harassing phone calls.***

Keywords: Average pitch, Formant analysis, GMSC, GSM, MSISDN

1. Introduction

With rapid development of mobile technology the criminal activity has also increased. Now days we go through different news related to threatening or harassing phone calls and in such cases it is not possible to identify the criminal (suspected user) in existing technology [1]. Because it may be possible that after making such calls the suspected user will not use SIM card again or destroy it. In the present paper a unique solution to identify the suspected mobile user who is making such calls is proposed. The frequency of each user's voice is unique. Each person has vocal cords that vary in shape and size resulting in different tones and frequencies. In this paper by using this property off speech a new idea to detect a suspected mobile user through user's voice is explained.

If the calls are frequent or particularly threatening, the telecommunication company can set up a trap on the user's mobile call line. Trap allows the telecommunication company to record the conversation. Once conversation recorded it is possible to extract the voice of the suspected mobile user who is making threatening call and save it as '.wav' file. Now a speech recognizer which is able to recognize the voice of suspected user among several users with the help of MATLAB programming and tools is designed is used [2]. Design of speech recognizer involves three main steps. In first step speech editing a MATLAB code is developed which can read '.wav' file and create proper time vector. This vector is divided into two equal parts and saved into reverse order. This will create a new '.wav' file. In speech degradation and enhancement again the code can read the '.wav' file say original signal or reference file. The Gaussian noise is added into original signal. In order to filter recorded voice fast Fourier transform (FFT) of both signals (original signal and noisy signal) is used, and shifted FFT of both signals calculated. Butterworth filter is applied on shifted FFT of noisy signal and filtered signal is scaled to compare with original signal. In pitch and format analysis we extract of pitch information from speech files ('.wav' file) [2], [7]. Average pitch can be determined by the peak of autocorrelation, (usually the original speech file is segmented into frames and pitch contour can be derived by plot of peaks from frames). Further first three formants present in a speech file is calculate and the vector differences between peak positions of the formants is determined. The speech recognizer is placed at gateway mobile switching centre (GMSC) (since all call within a cell is routed through GMSC)[3], code is prepare which continuously compare voice files of all user using network at same time along with suspected user. This code first compares the reference '.wav' file to all other files of different users based on average pitch. The top most likely matches are then compared by the differences in their average pitch and formant peak vectors [7]. The resulting closest matches file is the required file of suspected user. As soon as suspected user voice matches, GMSC find mobile station international subscriber directory number (MSISDN) of suspected user and contact with home location register (HLR) to find the user location [3].

2. Proposed Method

Proposed technique is divided into three main sections explained as follows:

2.1 Design of Speech Recognizer

Design of speech recognizer is further divided in to three sub-parts:

2.1.1 Speech Editing

In the first step a set of the speech signal in '.wav' (dot) wave format is recorded and by taking any one of speech signal from the set of recorded speech waves the speech editing is performed. The length of the vector of the reference file must have a magnitude of 30,000. However, this vector is then divided into two separate vectors having equal length and in opposite order [7]. Now with the help of MATLAB a code is developed to read the given wave file and then the same file is played in reverse order. The representation of the speech editing waveform is shown in fig.1.

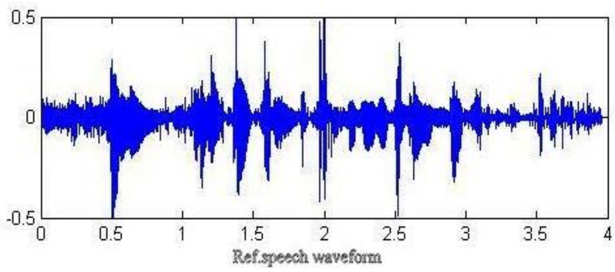

Figure 1: The representation of the speech editing waveform.

2.1.2 Speech Degradation & Enhancements

Speech enhancement is the improvement of the quality of speech signals by removing the background interference or noise to avoid speech distortion. Noise plays an important role in speech enhancement & degradation. Thus noise estimation is one of the key factors while performing speech recognition. If the estimated noise is high then speech signal will get distorted. So to remove the noise there are two steps: Speech degradation and speech enhancement [6].The speech degradation technique involves the addition of Gaussian noise to the original '.wav' format file. The representation of the degraded speech wave is shown in fig.2.

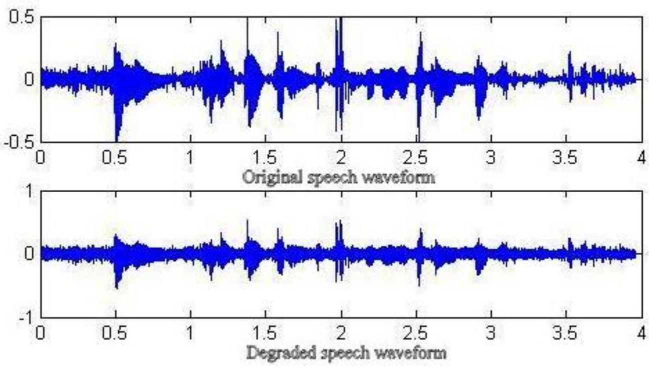

Figure 2: The representation of the degraded speech wave.

In speech enhancement technique the degraded signal which is mixed with Gaussian noise is firstly translated into frequency domain [2]. The higher frequency noise components are then removed with the help of Butterworth low pass filter, according to the following equation:

$$HB\ (u, v) \cong 1/\ (1+ (\sqrt{2}-1)\ (D\ (u, v)\ /Do))$$

Where,
D (u, v) is the rms value of u and v
Do determines the cut-off frequency
n is the filter order.

The reason to choose Butterworth filter is that it has the capability to filter the Gaussian noise more closely and it is an approximation to an ideal low pass filter as the order 'n' is increased. The resulting filtered signal is scaled and plotted with the original noisy signal to compare the filtering result and the representation of speech enhanced waveform is shown in fig.3.

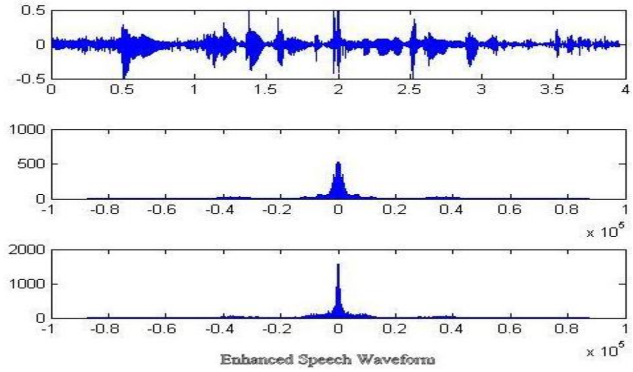

Figure 3: Representation of speech enhancement waveform.

2.1.3 Pitch & Formant Analysis

Pitch which contains speaker-specific information is an important attribute of speech signal. Pitch represents the perceived fundamental frequency of a sound and is one of the major fundamental properties of sound wave along with loudness and quality [09-10]. There are a numerous methods [4], [5], [6] developed in the speech processing area for the estimation of pitch. Among them the autocorrelation of speech is mostly used methods [2]. In this paper a technique that involves the extraction of basic parameters of pitch analysis through autocorrelation has been described. Now the average pitch of the entire '.wav' format speech files of different speakers that are recorded in data base is calculated and can be used in voice recognition. Average pitch can be obtained from the peak of autocorrelation. Usually the original speech file is segmented into frames and pitch contour can be mapped from the plot of peaks [7]. The pitch contour versus time frame is shown in fig.4.

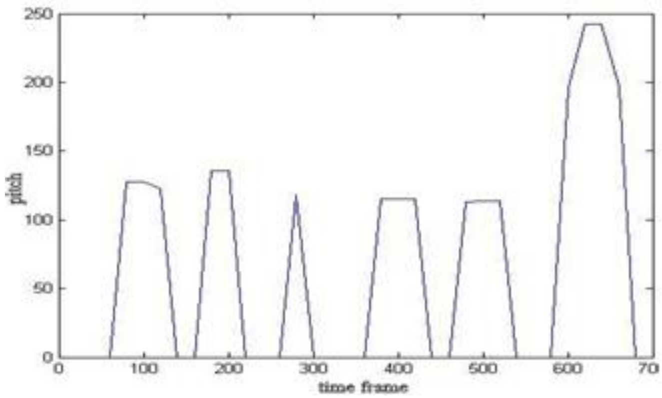

Figure 4: Representation of pitch analysis waveform

Formants are the meaningful frequency components of speech signal [8]. The information that humans require to differentiate between vowels can be shown by the frequency content of the vowel sounds. In formant analysis formant routine which contain calculation of PSD is perform on reference file taken from recorded .wav speech signal. By applying formant routine on reference .wav which return PSD and normalized frequency here PSD is calculated by Yule-Walker's method and thus position of the peaks is determined[2],[7]. Vector difference between peak positions of the first five formants is calculated. Waveform of formant analysis is shown in fig.5.

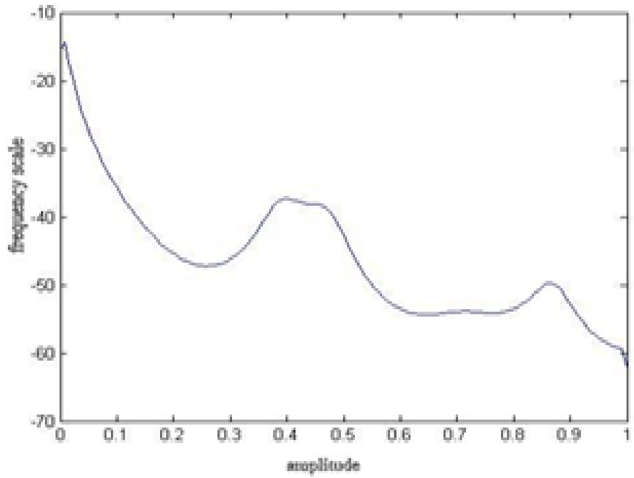

Figure 5: Representation of Formant Analysis waveform

2.2 Implementation of speech recognition system at GSM network

The Gateway Mobile Switching Centre is a special kind of mobile switching centre that is used to route calls outside the mobile network [3]. In this section all the steps as described above are applied on reference recorded ".wav" file and calculate the average pitch of the entire ".wav" file and the first five formants that are present in ".wav" speech file. All the extracted features of reference .wav file are stored at GMSC database. Speech recognition system feed in between GSM network as shown in fig.6

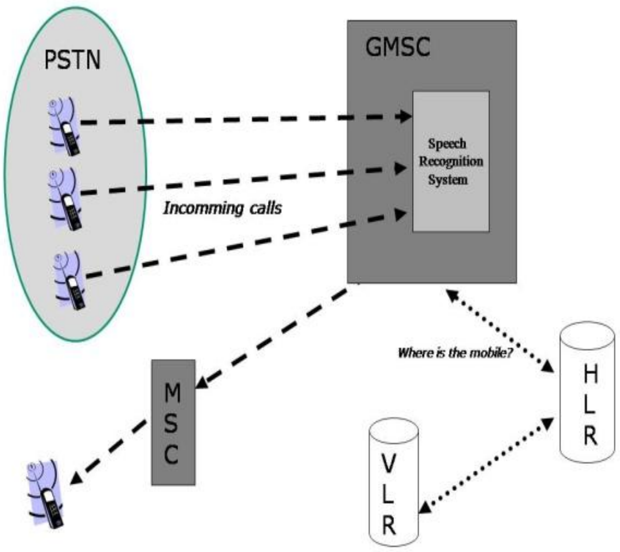

Figure 6: Speech recognition system feed in between GSM network.

2.3 Detection of Suspected User

Whenever a call for a mobile subscriber comes from mobile network or the subscriber wants to make a call to somebody the call is routed through the GMSC [3]. In fig 7 Let there be N number of users accessing network at a time including the suspected user. Since all the calls are routed through GMSC, at that time speech recognition system which is placed at GMSC scan voice of all the users. Now with the help of above discussed pitch and formant analysis, a waveform

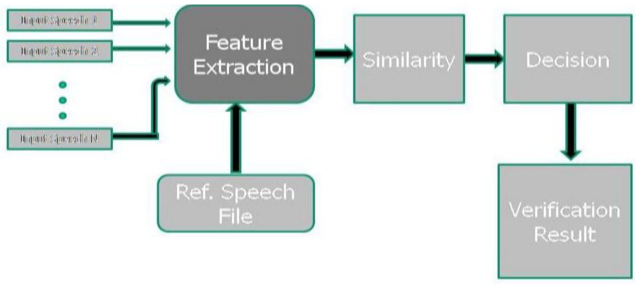

Figure 7: The general block diagram of Speaker identification

Comparison code is developed. Thus, based on this code all speech waveform files can be easily characterized. In this process a reference .wav file (user03.wav) used which is then compared with the remaining speech files. Moreover, a sorting routine is performed which allows sorting and comparison of the average pitch of the reference file with all the other .wav files. This technique further includes the comparison of formant vector of the reference file to all speech files, sorting for the top average pitch correlations, sorting again these files by formant vectors correlations and sorts these by average pitch [7]. The general block diagram of Speaker Identification is shown in fig.7.

Using the information obtained from Fig 7, the result of this system could easily be found. The ID of speaker that has the minimum formant difference should be the best matched speaker for the unknown speaker. The next best matching speakers are found easily from the sorted formant difference vector between "user03.wav" file and other selected trained files. We checked out the trained file with other files and found that two voices are of the same speaker.

3. Results

To verify the performance of the proposed method, the speech signals of 20 speakers are recorded on the GSM environment. For identification phase some speech signals also recorded in laboratory and in noisy environment as well. We got good accuracy for normal voices. Pitch contour versus time frame for reference file (i.e. user 3) is shown in fig.4 and results of the average pitch of 20 speech files are summarized in Table (1)

Table 1

Wave file Name	Average Pitch (Hz)
User01.wav	143.1346
User02.wav	138.4857
User03.wav (Ref. file)	145.6279
User04.wav	189.0452
User05.wav	147.0045
User06.wav	151.4460
User07.wav	144.6573
User08.wav	145.0023
User09.wav	154.4526
User10.wav	144.5682
User11.wav	140.6902
User12.wav	135.5689
User13.wav	140.5098
User14.wav	145.5279
User15.wav	155.5623
User16.wav	139.6309
User17.wav	149.5652

User18.wav	155.5680
User19.wav	147.6622
User20.wav	158.6690

In order to create a speech recognition algorithm, criteria to compare speech files must be defined hence by combining both the steps (average pitch and formant analysis) it is possible to correctly pick out the suspected user voice i.e. here user03.wav as shown in Table(2). As discussed before all voice files are compares with reference wav file based on average pitch. The top most likely matches in our case user03.wav user08.wav and user14.wav are then compared by the differences in their formant peak vectors (here user03.wav, user07.wav and user18.wav) in Table 2. The speech file which is common in both steps is suspected user's voice.

Table 2

Pitch analysis		Formant analysis	
Wave file Name	Is Suspected user's pitch match ?	Wave file Name	Is Suspected user's formant match ?
User01.wav	No	User01.wav	No
User02.wav	No	User02.wav	No
User03.wav (Ref. file)	Yes √	User03.wav (Ref. file)	Yes √
User04.wav	No	User04.wav	No
User05.wav	No	User05.wav	No
User06.wav	No	User06.wav	No
User07.wav	No ✕	User07.wav	Yes √
User08.wav	Yes √	User08.wav	No ✕
User09.wav	No	User09.wav	No
User10.wav	No	User10.wav	No
User11.wav	No	User11.wav	No
User12.wav	No	User12.wav	No
User13.wav	No	User13.wav	No
User14.wav	Yes √	User14.wav	No ✕
User 15.wav	No	User 15.wav	No
User16.wav	No	User16.wav	No
User17.wav	No	User17.wav	No
User18.wav	No ✕	User18.wav	Yes √
User19.wav	No	User19.wav	No
User20.wav	No	User20.wav	No

4. Conclusion

Anyone can be the victim of harassing, annoying, obscene, or threatening telephone calls. This paper proposed an idea for detection of such suspected user on GSM network and discussed about the speech recognizer. The analysis and implementation of proposed system was carried on MATLAB. System design followed the major sub-parts: Speech Editing, Speech Degradation & Enhancements, Pitch & Formant Analysis and Detection of suspected subscriber/user. The implemented speech recognition system at GSM network generated favorable results. Proposed system can be useful to find out the suspected users which are related with threatening or harassing phone calls.

Time required for continuous scanning and sorting for reference speech file is one of the limitations of system. Since implementation of speech recognition system is not easy at academic level. Hence this issue that needs to be tackled by the industrial organization in this field, which will help to make investigation easier than existing technologies.

References

[1] Emilio Ferrara , Pasquale De Meo, Salvatore Catanese Giacomo Fiumara "Detecting criminal organizations in mobile phone networks" Expert Systems with Applications 41 pp-5733–5750,2014.
[2] Aseem Saxena, Amit Kumar Sinha, Shashank Chakarwarti, Surbhi Charu, "Speech Recognition Using MATLAB," International Journal of Advances In Computer Science and Cloud Computing, ISSN: 2321-4058 Volume-1, Issue- 2, Nov-2013.
[3] Jörg Eberspächer, Hans-Joerg Vögel, Christian Bettstetter, Christian Hartmann, Handbook on"GSM - Architecture,Protocols and Services" 3rd edition WILEY,2009.
[4] Xiaopeng Wei,Lasheng Zhao, Qiang Zhang, Jing Dong "Robust pitch estimation using a wavelet variance analysis model'' Signal Processing 89, pp 1216–1223,2009.
[5] Johan Xi Zhang , Mads Græsbøll Christensen, Søren Holdt Jensen, Marc Moonen, "An iterative subspace-based multi-pitch estimation algorithm" Signal Processing 91,pp 150– 154,2011.
[6] Anuprita P. Pawar, Kirtimalini B. Choudhari,Madhuri A Joshi, "Review of Single Channel Speech Enhancement Methods in Spectral Domain", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 7 No.11,2012.
[7] E. Darren Ellis Department of Computer and Electrical Engineering – University of Tennessee, Knoxville Tennessee 37996 topic on "Design of a Speaker Recognition Code using MATLAB".
[8] Wikipedia. http://en.wikipedia.org/wiki/.
[9] D. Gerhard. "Pitch Extraction and Fundamental Frequency: History and Current Tech-niques", technical report, Dept. of Computer Science, University of Regina, 2003.
[10] Dmitry Terez, "Fundamental frequency estimation using signal embedding in state space". Journal of the Acoustical Society of America, 112(5):2279, November 2002.

Author Profile

Dharmendra Gurve received B.E. degree in Electronics & Communication Engineering from J.N. College of Technology, Bhopal in 2012. Currently he is pursing M.Tech in ECE from Dr. B. R. Ambedkar National Institute of Technology (NIT) Jalandhar. His research interests include speech signal processing, cellular communication and biomedical signal processing.

Mr. Mahendra Gurve received B.E. degree in Computer Science and Engineering from Patel College of Science & Technology, Bhopal in 2010 and M. degree in in GIS and Remote Sensing from Maulana Azad National Institute of Technology (NIT), Bhopal in 2013. In the same year, he joined Jaypee Institute of Information Technology Noida at the Department of Computer Science & IT as a Assistant Professor. His research interests include Web technology, Signal Processing and Data mining.