

Emotion Detection by Analyzing Voice Signal Using Wavelet

Faishal Badsha¹, Rafiqul Islam²

¹Department of Mathematics and Statistics, Bangladesh University of Business and Technology (BUBT), Dhaka, Bangladesh

²Mathematics Discipline, Science, Engineering and Technology School, Khulna University, Khulna, Bangladesh

Email: badsha@bubt.edu.bd, mrislam_66@yahoo.com

How to cite this paper: Badsha, F. and Islam, R. (2020) Emotion Detection by Analyzing Voice Signal Using Wavelet. *American Journal of Computational Mathematics*, 10, 485-502.

<https://doi.org/10.4236/ajcm.2020.104028>

Received: September 30, 2020

Accepted: November 28, 2020

Published: December 2, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Emotion is such a unique power of human trial that plays a vital role in distinguishing human civilization from others. Voice is one of the most important media of expressing emotion. We can identify many types of emotions by talking or listening to voices. This is what we know as a voice signal. Just as the way people talk is different, so is the way they express emotions. By looking or hearing a person's way of speaking, we can easily guess his/her personality and instantaneous emotions. People's emotion and feelings are expressed in different ways. It is through the expression of emotions and feelings that people fully express his thoughts. Happiness, sadness, and anger are the main medium of expression way of different human emotions. To express these emotions, people use body postures, facial expressions and vocalizations. Though people use a variety of means to express emotions and feelings, the easiest and most complete way to express emotion and feelings is voice signal. The subject of our study is whether we can identify the right human emotion by examining the human voice signal. By analyzing the voice signal through wavelet, we have tried to show whether the mean frequency, maximum frequency and L_p values conform to a pattern according to its different sensory types. Moreover, the technique applied here is to develop a concept using MATLAB programming, which will compare the mean frequency, maximum frequency and L_p norm to find relation and detect emotion by analyzing different voices.

Keywords

MATLAB Programming, Wavelet, Haar Decomposition, Voice Signal, Mean Frequency, Maximum Frequency, L_p Norm

1. Introduction

With the advancement of science, we generally became dependent on machines.

Feature as machine learning, robotics, data science etc. have made our life easier and more enjoyable day by day. Although it is not possible to give emotion and feelings to a machine, we can often accurately identify the expression of emotions and feelings of a person by machine, analyzing the voice signal. We analyze voice signal by signal processing. When we want to express an emotion or feeling through our throat or through speech or through spoken or unspoken words, it simply takes the form of a basic signal in the form of sine and cosine [1]. These signals vary according to the type of person and person's voice. By processing these voice signals, we can get an idea of the individual's characteristics, identify him by his precious work and identify his various emotion. In previous these signals were processed with Fourier series and Fourier transform but there are some problems as Fourier series is not well time localized [2]. We use Haar wavelet to process our voice signal because wavelet is well time localized [3]. Although voice identification and classification began in early 1960's, it is now a very rich research field like image processing [4]. This enrichment of research is due to some of the characteristics that determine the character of a voice signal, such as loudness, amplitude, mean frequency, maximum frequency, L_p norm, standard deviation etc. which differ from person to person. We have calculated the unique feature of signal by analyzing the voice signals, which are composed of five-level decomposition through Haar wavelet [5]. Emotion detection, one of the fields of speech classification, is a technique of extracting the speech characteristic information from people voice, and then is analyzed through machine and identifies the uniqueness of speech. Speech classification is an interconnected field of many others fields like image processing, robotics, artificial intelligence, machine learning, etc., currently a lot of work being done in machine learning with voice classification, voice identification and emotion detection, which has enriched our robotics much more. A large part of robotics today depend on speech command through which our daily electronic turns away from its button and switch. By adding speech classification to machine learning, we can look at a person's speech record and store the emotion of his speech as data to estimate his next field of emotion. With the addition of the voice detection feature to the CCTV camera, we can easily identify the culprit through emotion detection by hearing the voice without looking at the face. Emotion detection plays an important role for people with physical and mental disabilities who communicate mainly through machines. Therefore, emotion detection or voice classification has not only made our daily lives easier, it has become a more important part of our life. There are many ways of classifying voice for emotion detection; wavelet is one of the best tools among them [6].

2. Methodology

2.1. Wavelet

The wavelet means small waves and in brief, a wavelet is an oscillation that decays quickly. Equivalent mathematical conditions for wavelet are:

$$\begin{aligned} \text{i) } & \int_{\mathbb{R}} |\psi(x)|^2 dx < \infty \\ \text{ii) } & \int_{\mathbb{R}} \psi(x) dx = 0 \\ \text{iii) } & \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \end{aligned}$$

where $\hat{\psi}(\omega)$ is the Fourier Transform of $\psi(x)$ [7].

2.2. Haar Wavelet

The Hungarian mathematician Alfred Haar first introduced the Haar function in 1909 in his Ph.D. thesis.

A function defined on the real line \mathbb{R} as

$$\psi(t) = \begin{cases} 1 & \text{for } t \in \left[0, \frac{1}{2}\right) \\ -1 & \text{for } t \in \left[\frac{1}{2}, 1\right) \\ 0 & \text{otherwise} \end{cases}$$

Is known as the Haar function [7].

The Haar function $\psi(t)$ is the simplest example of a Haar Wavelet. The Haar function $\psi(t)$ is a wavelet because it satisfies all the conditions of wavelet. This fundamental example has all the major features of the general wavelet theory. Haar wavelet is discontinuous at $t = 0, \frac{1}{2}, 1$ and it is very well localized in the time domain.

The Fourier transform of $\psi(t)$ is given by

$$\begin{aligned} \hat{\psi}(\omega) &= i \exp\left(-\frac{i\omega}{2}\right) \frac{\sin^2\left(\frac{\omega}{4}\right)}{\frac{\omega}{4}} \\ \therefore \text{Re}\{\hat{\psi}(\omega)\} &= \sin\left(\frac{\omega}{2}\right) \frac{\sin^2\left(\frac{\omega}{4}\right)}{\frac{\omega}{4}}. \end{aligned}$$

2.3. Signal

A signal is defined as a function $f(x)$, which has a series representation

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

Then all information about the function f is stored in the coefficients $\{a_n\}_{n=0}^{\infty}$ [8].

2.4. Wavelets and Signal Processing

The mathematical theory for wavelets deals for a great part with ways of obtain-

ing series expansions of the type $f(x) = \sum_{j \in Z} \sum_{k \in Z} d_{j,k} \psi_{j,k}(x)$ for certain functions f . Very often, the relevant functions f describe the time dependence of certain signals, *i.e.* the vibrations in a mechanical system or the current in an electric circuit. We now shortly describe a property of wavelets, which distinguish representations of the type $f(x) = \sum_{j \in Z} \sum_{k \in Z} d_{j,k} \psi_{j,k}(x)$ from the general representations $f(x) = \sum_{n=0}^{\infty} a_n f_n(x)$ [8] [9].

2.5. Frequency and Maximum Frequency

Frequency is the number of occurrences of a repeating event per unit time. In physics frequency is the number of waves that pass a fixed point in unit time; also, the number of cycle or vibration during one unit of time by a body in periodic motion. A body in periodic motion is said to have undergone one cycle of one vibration after passing through a series of events of positions and returning to its original state. If the period, or time interval, required to complete one cycle of vibration is 1/2 second, the frequency is two per second [10]. The symbol most often used for frequency are f and the Greek letters ν (nu) and omega ω . In general, the frequency is the reciprocal of the period, or time interval. If T is the time period to complete one cycle then the frequency f can be written as

$$f = \frac{1}{T}$$

When the number of cycle in a unit period is maximum then the frequency is maximum.

2.6. Mean Frequency

The mean frequency of a spectrum is calculated as the sum of the product of the spectrogram intensity (in dB) and the frequency, divided by the total sum of spectrogram intensity [10]. If n is the number of frequency bins in the spectrum, f_i is the frequency of spectrum at bin i of n and I_i is the intensity (in dB) of spectrum at bin i of n then the mean frequency f_{mean} can be written as follows,

$$f_{\text{mean}} = \frac{\sum_{i=0}^{i=n} I_i f_i}{\sum_{i=0}^{i=n} I_i}$$

2.7. L_p Norm

For finite p , L_p Norm in $c[a, b]$ is defined as

$$\|f\|_p = \left[\int_a^b |f(x)|^p dx \right]^{\frac{1}{p}}; 1 \leq p < \infty$$

For discrete function it can be defined as

$$\|f\|_p = \left[\sum_{i=1}^n |f(x_i)|^p \right]^{\frac{1}{p}}$$

where $\{x_i\}$ are the components of f [11].

If we put $p = 1$ in the above equation then $\|f\|_1$ is called **L_1 Norm**.

If we put $p = 2$ in the above equation then $\|f\|_2$ is called **L_2 Norm**.

3. Experimental Sample

Here we take twelve-voice sample from four people. We collected our voice sample in four different mood (joy, sorrow and angry) from each person.

3.1. First Experimental Real Voice of First Person

First experimental real voice of first person who speaks by a microphone and say, “can you please tell me what is going on” with joy in between 0 to 5 seconds

From statistical analysis of first experimental speech signal of first person, we see that, mean frequency of signal is 347 Hz; maximum frequency of signal is 453 Hz, L_1 Norm is 2263 and L_2 Norm is 23.35.

3.2. Second Experimental Real Voice of First Person

Second experimental real voice of first person who speaks by a microphone and say, “can you please tell me what is going on” with sorrow in between 0 to 5 seconds.

From statistical analysis of second experimental speech signal of first person, we see that, mean frequency of signal is 323 Hz, maximum frequency of signal is 428 Hz, L_1 Norm is 2311 and L_2 Norm is 17.26.

3.3. Third Experimental Real Voice of First Person

Third experimental real voice of first person who speaks by a microphone and say, “can you please tell me what is going on” with anger in between 0 to 5 seconds

From statistical analysis of third experimental speech signal of first person, we see that, mean frequency of signal is 376 Hz, maximum frequency of signal is 508 Hz, L_1 Norm is 2181 and L_2 Norm is 28.02.

3.4. First Experimental Real Voice of Second Person

First experimental real voice of second person who speaks by a microphone and say, “can you please tell me what is going on” with joy in between 0 to 5 seconds.

From statistical analysis of first experimental speech signal of second person, we see that, mean frequency of signal is 366 Hz, maximum frequency of signal is 481 Hz, L_1 Norm is 2206 and L_2 Norm is 24.32.

3.5. Second Experimental Real Voice of Second Person

Second experimental real voice of second person who speaks by a microphone and say, “can you please tell me what is going on” with sorrow in between 0 to 5 seconds.

From statistical analysis of second experimental speech signal of second per-

son, we see that, mean frequency of signal is 313 Hz, maximum frequency of signal is 423 Hz, L_1 Norm is 2327 and L_2 Norm is 15.28.

3.6. Third Experimental Real Voice of Second Person

Third experimental real voice of second person who speaks by a microphone and say, “can you please tell me what is going on” with anger in between 0 to 5 seconds.

From statistical analysis of third experimental speech signal of second person, we see that, mean frequency of signal is 380 Hz, maximum frequency of signal is 497 Hz, L_1 Norm is 2049 and L_2 Norm is 27.88.

3.7. First Experimental Real Voice of Third Person

First experimental real voice of third person who speaks by a microphone and say, “can you please tell me what is going on” with joy in between 0 to 5 seconds.

From statistical analysis of first experimental speech signal of third person, we see that, mean frequency of signal is 372 Hz, maximum frequency of signal is 456 Hz, L_1 Norm is 2218 and L_2 Norm is 22.09.

3.8. Second Experimental Real Voice of Third Person

Second experimental real voice of third person who speaks by a microphone and say, “can you please tell me what is going on” with sorrow in between 0 to 5 seconds.

From statistical analysis of second experimental speech signal of third person, we see that, mean frequency of signal is 319 Hz, maximum frequency of signal is 433 Hz, L_1 Norm is 2295 and L_2 Norm is 16.43.

3.9. Third Experimental Real Voice of Third Person

Third experimental real voice of third person who speaks by a microphone and say, “can you please tell me what is going on” with anger in between 0 to 5 seconds.

From statistical analysis of third experimental speech signal of third person, we see that, mean frequency of signal is 396 Hz, maximum frequency of signal is 514 Hz, L_1 Norm is 2031 and L_2 Norm is 28.16.

3.10. First Experimental Real Voice of Forth Person

First experimental real voice of forth person who speaks by a microphone and say, “can you please tell me what is going on” with joy in between 0 to 5 seconds.

From statistical analysis of first experimental speech signal of forth person, we see that, mean frequency of signal is 360 Hz, maximum frequency of signal is 473 Hz, L_1 Norm is 2188 and L_2 Norm is 24.07..

3.11. Second Experimental Real Voice of Forth Person

Second experimental real voice of forth person who speaks by a microphone and

say, “can you please tell me what is going on” with sorrow in between 0 to 5 seconds.

From statistical analysis of second experimental speech signal of fourth person, we see that, mean frequency of signal is 322 Hz, maximum frequency of signal is 416 Hz, L_1 Norm is 2235 and L_2 Norm is 17.28.

3.12. Third Experimental Real Voice of Fourth Person

Third experimental real voice of fourth person who speaks by a microphone and say, “can you please tell me what is going on” with anger in between 0 to 5 seconds.

From statistical analysis of third experimental speech signal of fourth person, we see that, mean frequency of signal is 402 Hz, maximum frequency of signal is 518 Hz, L_1 Norm is 2083 and L_2 Norm is 27.12.

4. Results and Discussions

In the above experimental study, we tried to predict the approximate emotion by analyzing the four unique and basic feature (mean frequency, maximum frequency, L_1 norm and L_2 norm) of voice signal. In this purpose, we analyzed the voice signal by Haar wavelet (Shown in **Figure 1**). The signal frequency is nothing but sampling number. The number of sample of each of our signals is 40,000. In our experiment, we have taken three voices in three different mood (Joy, sorrow, anger) from each of four people (**Figures 2-13**). We calculate mean frequency, maximum frequency, L_1 norm and L_2 norm for each voice (**Tables 1-4**). We can see from the table and chart (shown in **Table 5** and **Figure 14**) the mean frequency vary from approximate 310 Hz to 400 Hz and the maximum frequency vary from approximate 410 Hz to 520 Hz for all the voice signals. Also the mean frequency vary from 340 Hz to 360 Hz approximately for joy, 310 Hz to 330 Hz approximately for sorrow, 365 Hz to 400 Hz approximately for anger and the maximum frequency vary from 450 Hz to 470 Hz approximately for joy, 410 Hz to 440 Hz approximately for sorrow, 480 Hz to 520 Hz approximately for anger. Again the L_1 Norm vary 2180 to 2250 approximately for joy, 2250 to 2330 approximately for sorrow and 2030 to 2190 approximately for anger where as L_2 Norm vary from 20 to 25 approximately for joy, 15 to 20 approximately for sorrow and 25 to 30 for anger (shown in **Table 6**, **Figure 15** and **Figure 16**). Therefore, we can see anger voice gives highest number of frequency where the lowest frequency emitted from the voice of sorrow and joy ensures the moderate type of frequency. On the other hand, L_1 and L_2 norm behave in reverse in the case of high and low frequency. In case of angry voices, L_2 norm gives the highest value where as L_1 is lowest and for the sorrow, the phenomenon is completely different also. Therefore, the mean frequency of all the happy voices in between 340 Hz to 360 Hz, the mean frequency of all the sorrow voices in between 310 Hz to 330 Hz and the mean frequency of all the angry voices in between 365 Hz to 400 Hz. On the other hand the maximum frequency of all the happy voices in

between 450 Hz to 470 Hz, the maximum frequency of all the sorrow voices in between 410 Hz to 440 Hz and the maximum frequency of all angry voices in between 480 Hz to 520 Hz. So in case of mean frequency and maximum frequency the frequency gradually increase from sorrow, joy to anger (shown in **Table 5** and **Figure 14**). The L_1 norm of all the happy voices in between 2180 to 2250, of all the sorrow voices in between 2250 to 2330 and of all the angry voices in between 2030 to 2190. Where the L_2 norm of all the happy voices in between 20 to 25, of all the sorrow voices in between 15 to 20 and of all the angry voices in between 25 to 30. It can be observed that L_1 norm and L_2 norm gradually increase and decrease respectively from sorrow, joy to angry (shown in **Table 6**, **Figure 15** and **Figure 16**).

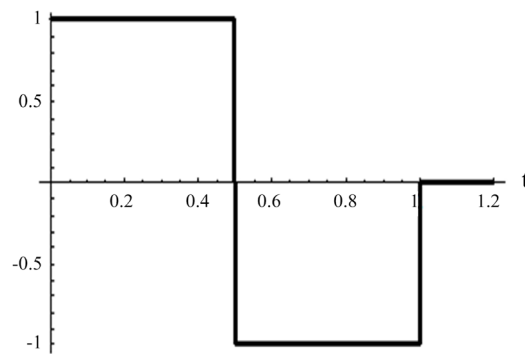
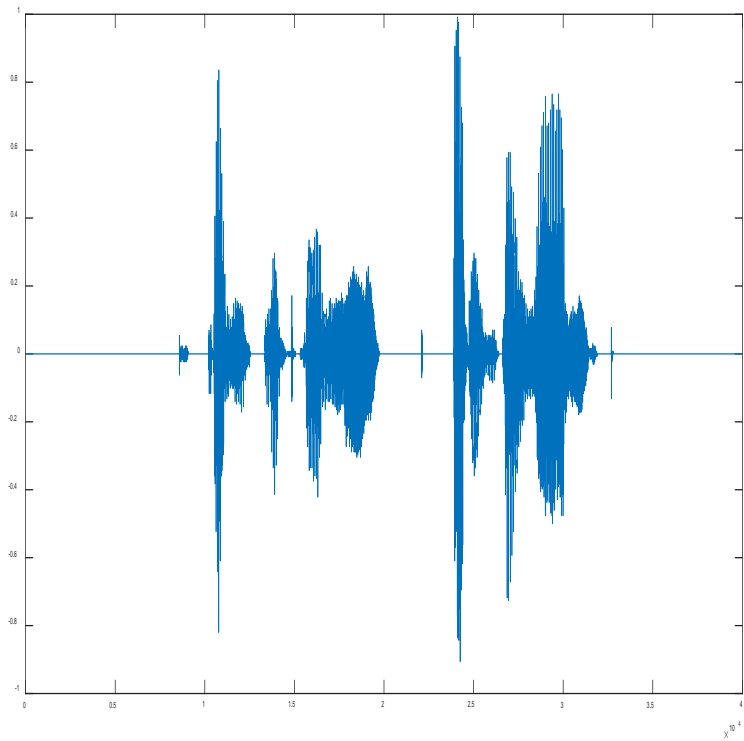
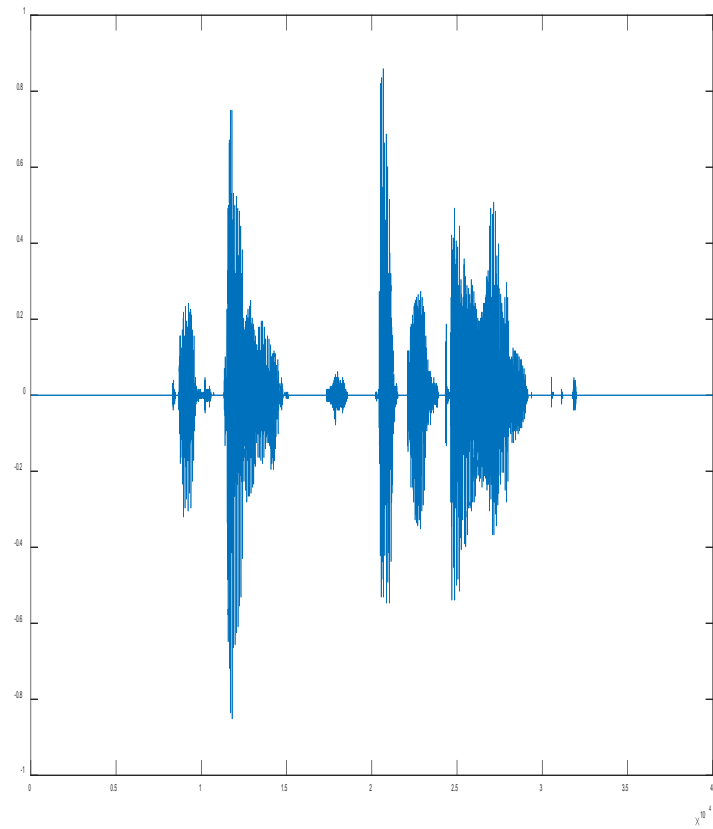


Figure 1. Haar wavelet.



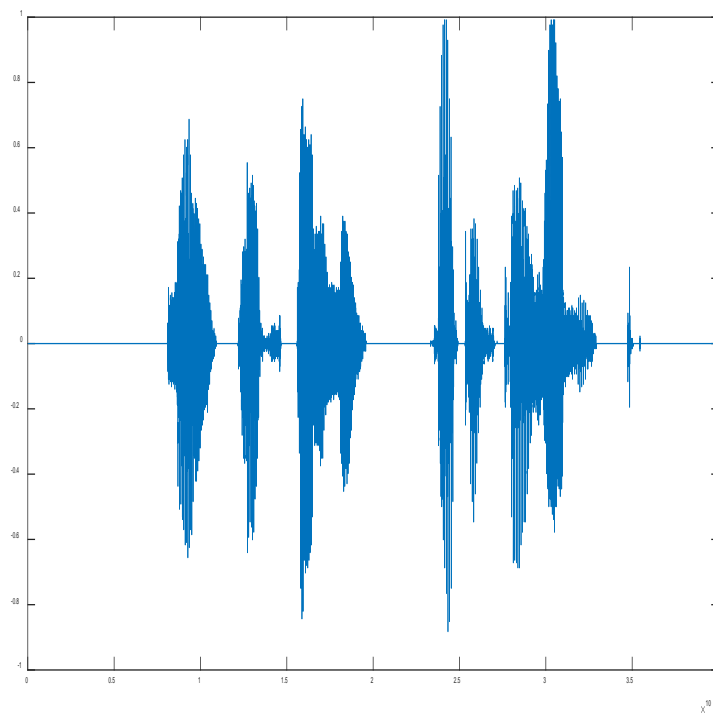
X-axis → time, Y-axis → amplitude

Figure 2. First experimental speech signal of first person.



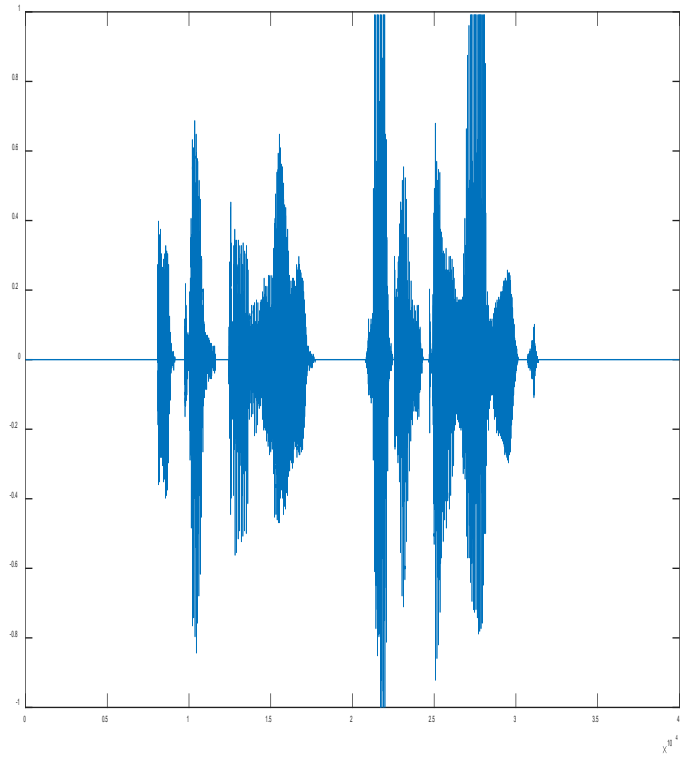
X-axis → time, Y-axis → amplitude

Figure 3. Second experimental speech signal of first person.



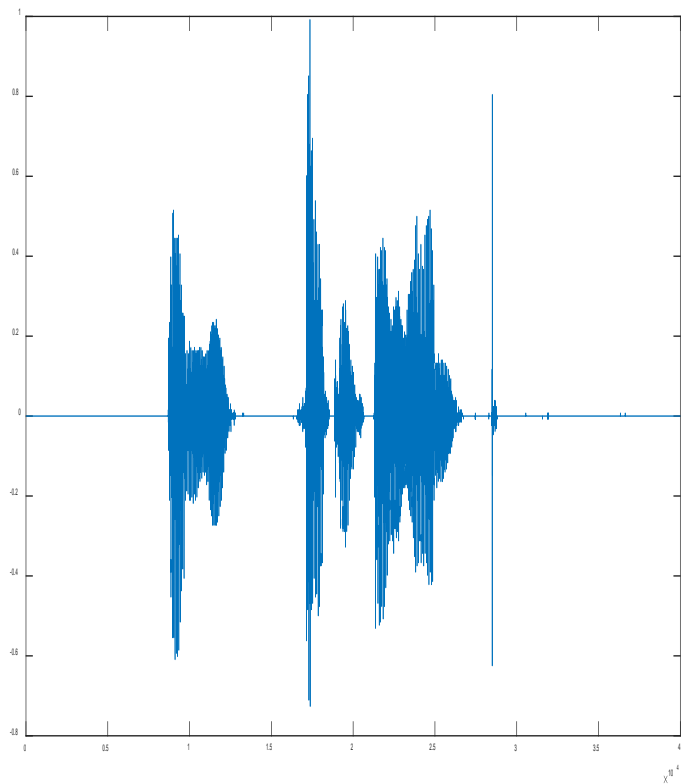
X-axis → time, Y-axis → amplitude

Figure 4. Third experimental speech signal of first person.



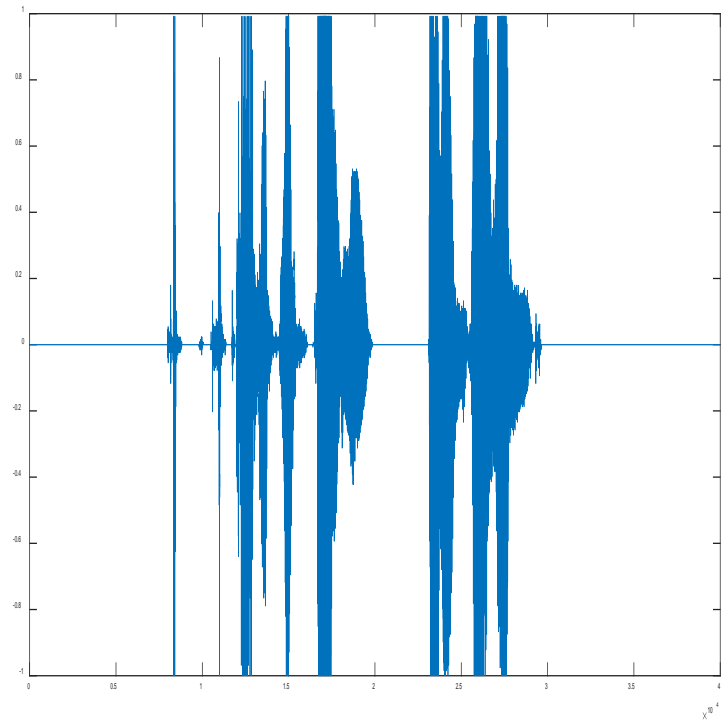
X-axis → time, Y-axis → amplitude

Figure 5. First experimental speech signal of second person.



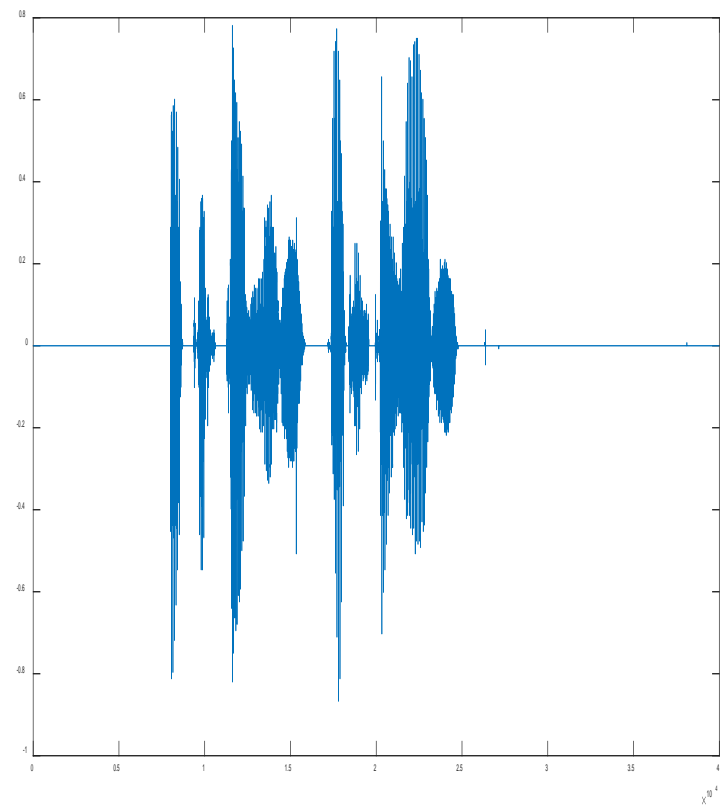
X-axis → time, Y-axis → amplitude

Figure 6. Second experimental speech signal of second person.



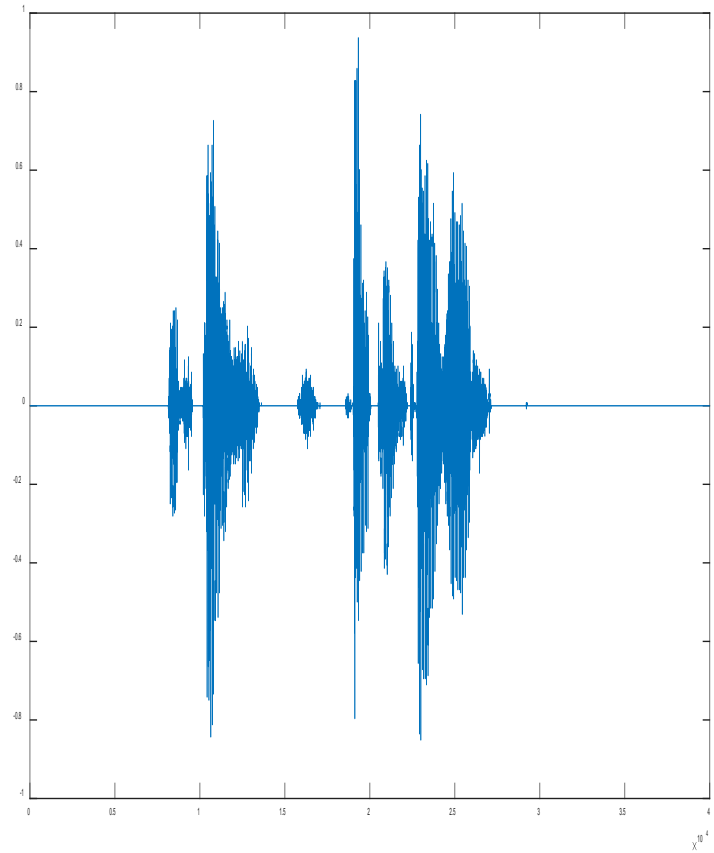
X-axis → time, Y-axis → amplitude

Figure 7. Third experimental speech signal of second person.



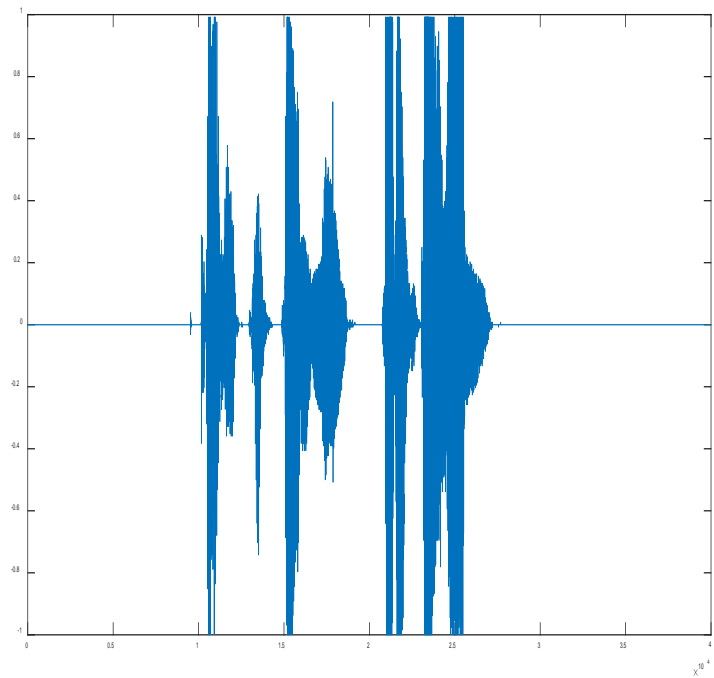
X-axis → time, Y-axis → amplitude

Figure 8. First experimental speech signal of third person.



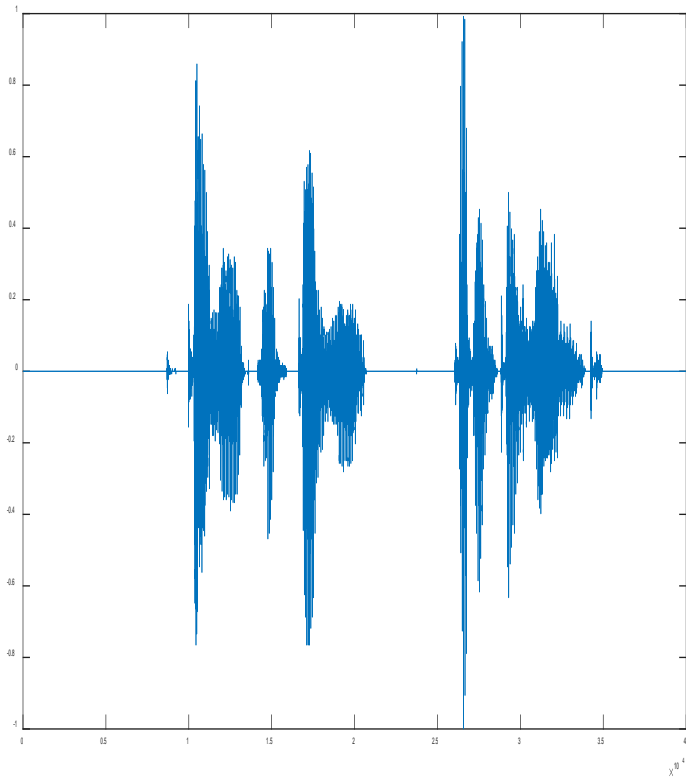
X-axis → time, Y-axis → amplitude

Figure 9. Second experimental speech signal of third person.



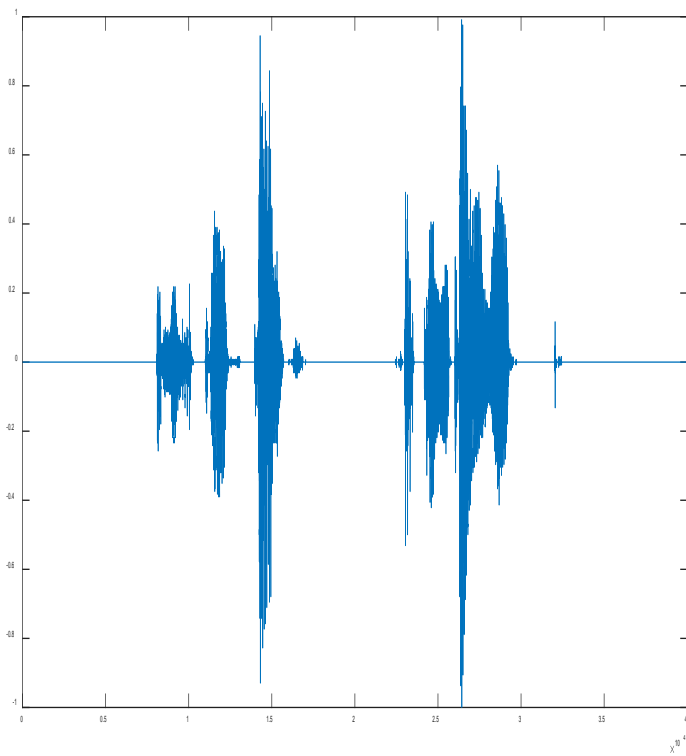
X-axis → time, Y-axis → amplitude

Figure 10. Third experimental speech signal of third person.



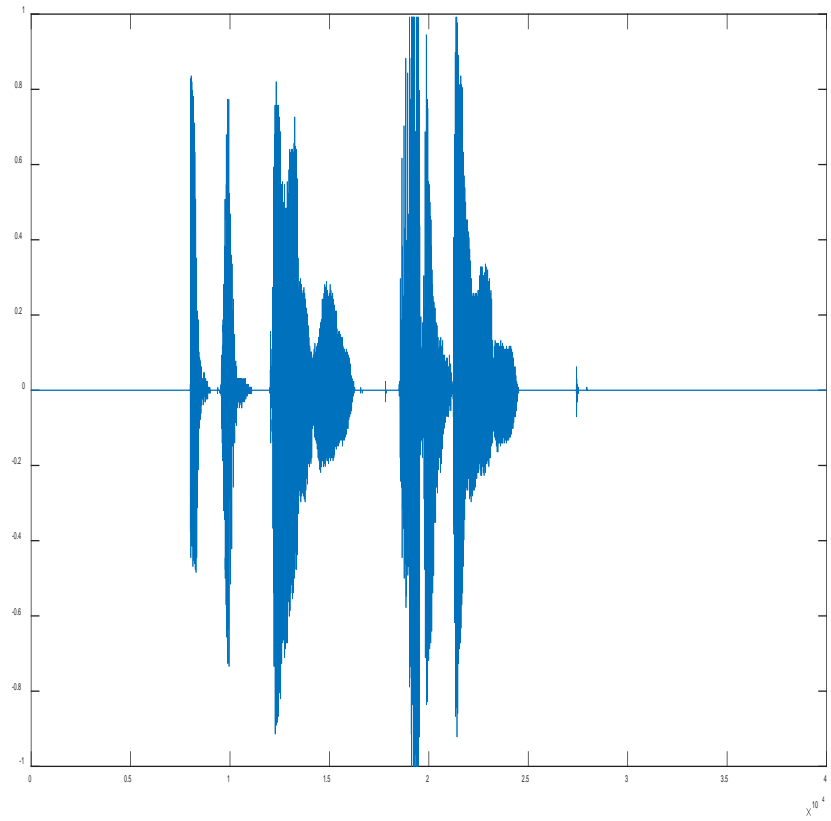
X-axis → time, Y-axis → amplitude

Figure 11. First experimental speech signal of fourth person.



X-axis → time, Y-axis → amplitude

Figure 12. Second experimental speech signal of fourth person.



X-axis → time, Y-axis → amplitude

Figure 13. Third experimental speech signal of forth person.

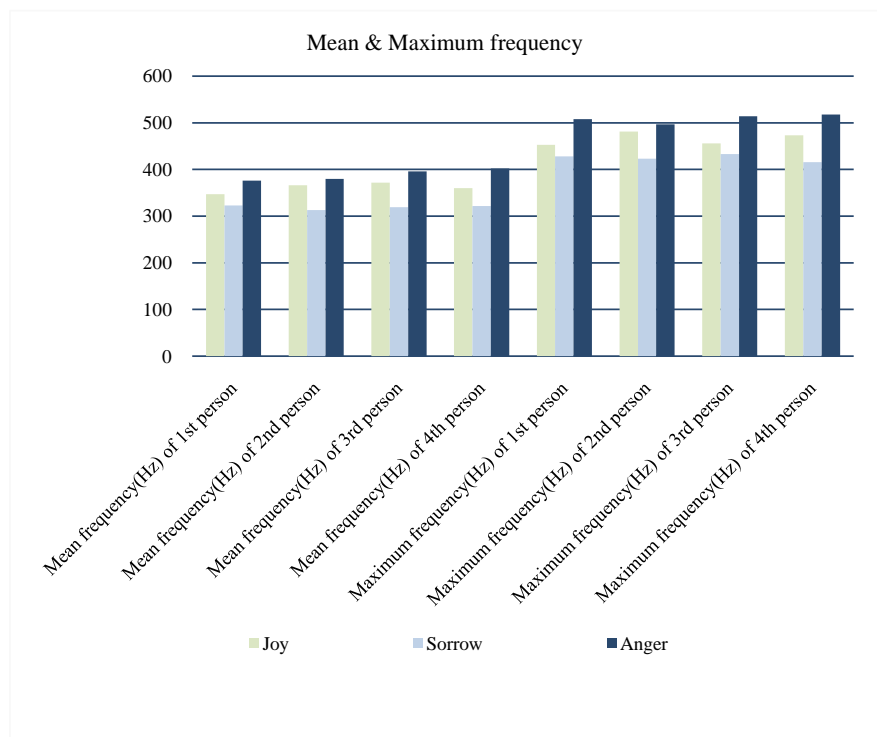


Figure 14. Comparison chart for frequency.

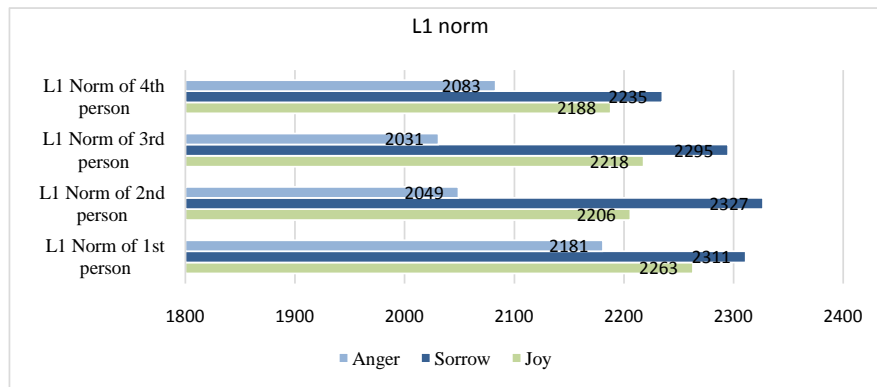


Figure 15. Comparison chart for L_1 norm.

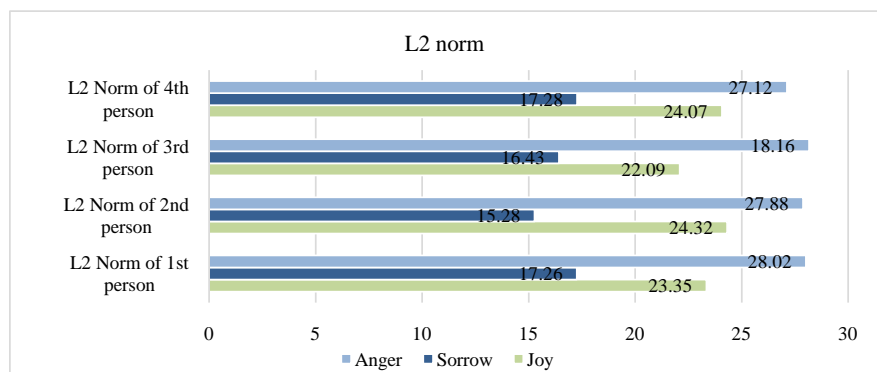


Figure 16. Comparison chart for L_2 norm.

Table 1. Data chart for first person.

1 st person	Mean frequency (Hz)	Maximum frequency (Hz)	L_1 Norm	L_2 Norm
Joy	347	453	2263	23.35
Sorrow	323	428	2311	17.26
Anger	376	508	2181	28.02

Table 2. Data chart for second person.

2 nd person	Mean frequency (Hz)	Maximum frequency (Hz)	L_1 Norm	L_2 Norm
Joy	366	481	2206	24.32
Sorrow	313	423	2327	15.28
Anger	380	497	2049	27.88

Table 3. Data chart for third person.

3 rd person	Mean frequency (Hz)	Maximum frequency (Hz)	L_1 Norm	L_2 Norm
Joy	372	456	2218	22.09
Sorrow	319	433	2295	16.43
Anger	396	514	2031	28.16

Table 4. Data chart for fourth person.

4 th person	Mean frequency (Hz)	Maximum frequency (Hz)	L_1 Norm	L_2 Norm
Joy	360	473	2188	24.07
Sorrow	322	416	2235	17.28
Anger	402	518	2083	27.12

Table 5. Data table for all frequency.

Frequency ↓	Different types of emotions of people		
	Joy	Sorrow	Anger
Mean frequency (Hz)	347 (1 st person)	323 (1 st person)	376 (1 st person)
	366 (2 nd person)	313 (2 nd person)	380 (2 nd person)
	372 (3 rd person)	319 (3 rd person)	396 (3 rd person)
	360 (4 th person)	322 (4 th person)	402 (4 th person)
Maximum frequency (Hz)	453 (1 st person)	428 (1 st person)	508 (1 st person)
	481 (2 nd person)	423 (2 nd person)	497 (2 nd person)
	456 (3 rd person)	433 (3 rd person)	514 (3 rd person)
	473 (4 th person)	416 (4 th person)	518 (4 th person)

Table 6. Data table for all L_p norm.

L_p ↓	Different types of emotions of people		
	Joy	Sorrow	Anger
L_1 Norm	2263 (1 st person)	2311 (1 st person)	2181 (1 st person)
	2206 (2 nd person)	2327 (2 nd person)	2049 (2 nd person)
	2218 (3 rd person)	2295 (3 rd person)	2031 (3 rd person)
	2188 (4 th person)	2235 (4 th person)	2083 (4 th person)
L_2 Norm	23.35 (1 st person)	17.26 (1 st person)	28.02 (1 st person)
	24.32 (2 nd person)	15.28 (2 nd person)	27.88 (2 nd person)
	22.09 (3 rd person)	16.43 (3 rd person)	28.16 (3 rd person)
	24.07 (4 th person)	17.28 (4 th person)	27.12 (4 th person)

5. Conclusion

In the above discussion, we have discussed many issues of voice signal. Our aim was to give a complete aspect of how voice signals behave in different ways of emotion of people. By processing the voice signal through wavelet analysis, we have seen that the signal changes its structure in case of three different moods (joy, sorrow and anger) and give characteristic values of mean frequency, maximum frequency, L_1 norm and L_2 norm. In this way by making a relationship among these characteristic calculations, we can easily determine the emotion of voice signal. These results not only help to determine the emotion but also they

give important information about speakers. Emotion detection through voice signal analysis plays an important role in machine learning, robotics, artificial intelligence, data mining, voice biometric, intelligence division, forensic department etc. With the help of voice biometric, forensic and voice identification process, our intelligence department will be able to identify many criminals, as a result our cybercrime will greatly reduce. Now a day's voice automation is very popular and widely used medium. Already, many things are being done through voice commands and in the near future, this sector will be spread. Voice security is very popular among the many trusted security media. Therefore, in the present world, emotion detection, voice identification, specification and speaker recognition are a very important task. In this paper, we have generated codes with the help of MATLAB Programming and recorded our sound by regular headphone so some noise may be mixed with the original voice. We will try to improve this issue in the future.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Gabor, D. (1946) Theory of Communication. *Journal of Institute of Electrical and Electronics Engineers*, **93**, 429-457. <https://doi.org/10.1049/ji-3-2.1946.0076>
- [2] Littlewood, J.E. and Paley, R.E.A.C. (1931) Theorems on Fourier Series and Power Series. *Journal of the London Mathematical Society*, **6**, 230-233. <https://doi.org/10.1112/jlms/s1-6.3.230>
- [3] Rioul, O. and Vetterli, M. (1991) Wavelets and Signal Processing. *Institute of Electrical and Electronics Engineers of Signal Processing Magazine*, **8**, 14-38. <https://doi.org/10.1109/79.91217>
- [4] Badsha, M.F., Islam, M.R. and Bulbul, M.F. (2018) Object Detection by Point Feature Matching Using Matlab. *Advances in Image and Video Processing*, **6**, 22-29.
- [5] Favero, R.F. (1994) Compound Wavelets: Wavelets for Speech Recognition. *Institute of Electrical and Electronics Engineers Communications Letters*, **17**, 600-603. <https://doi.org/10.1109/TFSA.1994.467280>
- [6] Meyer, Y. (1993) Wavelets, Algorithms and Applications. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [7] Agbinya, J.I. (1996) Discrete Wavelet Transform Techniques in Speech Processing. *International Technical Conference of Institute of Electrical and Electronics Engineers*, **2**, 514-519.
- [8] Pasti, L., Walczak, B., Massart, D.L. and Reschiglian, P. (1999) Optimization of Signal Denoising in Discrete Wavelet Transform. *Chemometrics and Intelligent Laboratory Systems*, **48**, 21-34. [https://doi.org/10.1016/S0169-7439\(99\)00002-7](https://doi.org/10.1016/S0169-7439(99)00002-7)
- [9] Boll, S.F. (1979) Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Proceedings*, **27**, 56-72. <https://doi.org/10.1109/TASSP.1979.1163209>
- [10] Chen, J.-F. and Ser, W. (2000) Speech Detection Using Microphone Array. *Electronics Letters*, **36**, 100-102.

tronic Letters, **36**, 181-182. <https://doi.org/10.1049/el:20000140>

- [11] Rabiner, L. and Juang, B. (1993) *Fundamentals of Speech Recognition*. Prentice Hall PTR, Englewood Cliffs, NJ.