Achieving a Secure Information System by Creating Person-Word-Sound Signature

Mohamad Al-Laham

Amman university college -Al-Balqa Applied University, Jordan

Abstract
Different wave files will be analyzed in order to create a wave file signature. This signature can be used later by any recognition tool to recognize the person and the word. The results of sound analysis can be used to create a sound password which can be used in any information system project. The purpose of this paper is to investigate sounds and select appropriate parameters to create a sound signature for each person- word. The selected parameters values for each person-word will use a set of parameters values and this set will be used as a signature and we will prove the set used to create different signature are different in values. In this research a new model based on a local binary pattern will be introduced. This model will be implemented using various spoken words and phrases for different persons to insure that the extracted features are unique for each person-word.

Key words:
Wave file, sound signature, sound parameters, signature set, LBP, ANLBP.

1. Introduction

In the field of securing applications, using voice recognition technology has become an important issue. The voice recognition (some time called speaker recognition) can be defined as the mission of determining a human ID through utilizing the characteristics of his voice [1, 2]. Voice recognition includes both physiological and behavioral features, such as tone and accent. Researchers over the years have introduced and developed several methods for identifying humans using their vice effectively. In addition, the voice recognition is used in security applications and virtual personal assistants like Google Assistant, so that they can recognize the voice of the phone owner and distinguish it from others [3].

The voice recognition can be classified into voice identification and voice verification. The speaker identification is the procedure of identifying a person from a set of sounds recorded using a particular pronunciation [4, 5, 6]. On the other hand, the speaker verification is the procedure of accepting or rejecting the proposed identity of the speaker [4]. These two process always use the same evaluation method under commonly used metrics. The terms are sometimes used interchangeably in literature [1, 7].

Human voice is by nature analogue waves that can be transformed into digital waves through applying sampling and quantization techniques. Voice recognition is also related to voice diarization, while the input audio stream is divided into homogeneous segments according to the voice identity [8, 9, 10]. Human voice signals include significant digital data type because of the urgent need for this data type by many applications. These applications, including security systems application [11, 12, 13, 14], need high speed execution. Unfortunately, voice signals have a hug size where this size will slow down the implementation speed, thus affects the system efficiency negatively.

Figure 1: mu waveforms

In this paper, a new model will be developed to analyze the voice signal and retrieve its parameters (i.e., Sigma, Mu as shown in Figure 1, Peak factor, and Dynamic range). Where Sigma is an estimate of the standard deviation of the voice signal (wave) normal distribution, Mu is an estimate of the mean of the voice signal normal distribution, Peak Factor can be defined as the ratio of maximum value to the Root Mean Square RMS value of the wave, and Dynamic Range is the ratio of the loudest undistorted sound to the quietest discernible sound, that the system is able to produce. The size of Voice signal depends mainly on the recording time and the sampling rate [15, 16, 17, 18, 22, 23]. The sampling rate (fs) (sometimes called sampling frequency) is defined as the average number of samples obtained in one second - sample per second, and calculated as present in Eq. 1.

$$fs = \frac{1}{T} \qquad (1)$$

where T is the time in seconds.

Manuscript received October 5, 2021
Manuscript revised October 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.10.30

2. Proposed Model

The proposed model for analyzing the voice signal consists of a number of steps that are run several times in order to define the parameter values. These parameters create the sound signal that will be utilized as the signature for each person. The procedure steps necessary to analyze the voice signal file and to estimate the parameters needed to create the person voice signature as seen in Figure 2. Where the most common properties used to analyze wave files as in [19] are prediction of the population mu, the predicted sigma, the dynamic range, and the peak factor (crest factor).

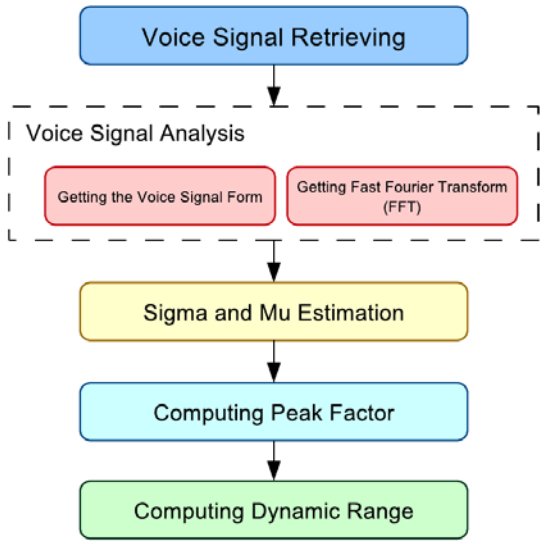

Figure 2: Proposed Model

2.1 Voice Signal Retrieving Figures

In this step, the voice signal is uploaded to the system in order to undergo the analysis process.

2.2. Voice Signal Analysis

In this step, the voice signal analysis process begins by plotting the signal and Getting Fast Fourier Transform (FFT). Where the FFT is a significant measurement technique in the field of audio and phonetics measurement. The signal is converted into individual spectral components and thus provides frequency information about the signal.

2.3. Sigma and Mu Estimation

In this step, the voice signal normal distribution means (Mu) is calculated in addition to calculate the voice signal normal distribution standard deviation (Sigma).

2.4. Computing the Peak Factor

Here the peak factor is calculated through finding the RMS value of the signal as in Eq. 2.

$$RMS = \sqrt{\frac{\sum x^2}{N}} \qquad (2)$$

where x represents the voice signal.
Next, finding the peak value of the signal which is the absolute maximum value (P). Then the Peak Factor can be calculated by using Eq. 3.

$$Peak\ Factor\ = 20 * log_{10}(\frac{P}{RMS}) \qquad (3)$$

2.5. Computing the Dynamic Range

Before calculating the Dynamic Range, the minimum value (minval) of the voice signal is defined, and then the Dynamic Range can be calculated by using Eq. 4.

$$Dynamic\ Range\ = 20 * log_{10}(\frac{P}{minval}) \qquad (4)$$

The voice signal file represented by a wave file (either it was stereo or mono) is to be reshaped into one row matrix as shown in Figure 3, that based on local binary pattern model (LBP) [11]. The features will be extracted applying the following steps:

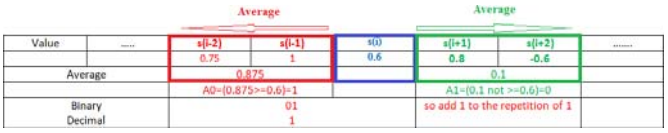

Figure 3: ANLBP Process

Algorithm 1 ANLBP implementation code

```
clear all
[a fs]=waveread('f1.wav');
f=zeros(4,1);
[n1,n2]=size(a);
tic
for i=3:n1-2 do
        a0=((a(i-1,1)+a(i-2))/2)>=a(i,1);
        a1=((a(i+1,1)+a(i+2))/2)>=a(i,1);
        ind=a0+2*a1;
        f(ind+1,1)=f(ind+1,1)+1;
end for
toc
```

- Get the speech.
- Reshape the speech (mono or stereo) to one row array.
- Initialize the features vector (array) to zeros (4 elements array because the number of extracted features equals 4).

- For each sample in the speech row apply average neighbor LBP (ANLBP) as shown in Figure 3 and Algorithm 1.
- Save the extracted features to be used later on in a recognition system.

3. Experiment

Table 1 illustrates the phrases about voice signals that will be investigated in this article [20, 21]. The proposed model is run several times using different voice signal for different persons and different words. For each voice signal file the parameters are defined and the calculated results for random 10 persons are illustrated in Table 2 for (No) word and Table 3 for (Yes) word.

Table 1: Used Voice Signal Phrases

Phrase	***Total Samples***	Time (*sec*)	Sample Rate (*fs*)	bits per Sample
yes	16000	0.0500	8000	8
no	16000	0.0500	8000	8

Table 2: Parameters for the word No

Person #	***Sigma***	Mu	Peak Factor	Dynamic Range
1	0.10797	+0.0055149	19.3225	42.0761
2	0.10501	-0.0044922	19.5682	42.1442
3	0.10058	+0.0108270	19.9902	42.1442
4	0.16781	-0.0033291	15.5021	42.1442
5	0.10474	-0.0068405	19.579	42.0761
6	0.12345	+0.0018433	18.1695	42.0761
7	0.14133	-0.0111220	16.9687	39.8245
8	0.12569	+0.0150620	17.9519	38.7904

In addition, the proposed model can be utilized to clarify the characteristics of any voice signal files. Figures 4, 5, 6, and 7 show a sample of the characteristics of one of the investigated voice signal files.

Table 3: Parameters for the word Yes

Person #	***Sigma***	Mu	Peak Factor	Dynamic Range
1	0.112720	-0.0050181	18.1745	33.2552
2	0.073035	+0.0055439	22.7045	42.0074
3	0.120180	-0.0107570	18.3689	36.6502
4	0.118810	-0.0053940	18.4945	42.1442
5	0.109960	+0.0204810	19.0218	42.1442
6	0.101550	-0.0073560	19.8435	42.1442
7	0.101480	-0.0063700	19.8552	42.1442
8	0.063259	+0.0018003	23.9742	38.7904

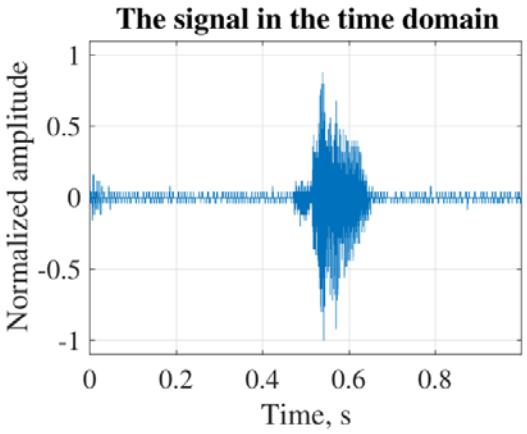

Figure 4: Amplitude of the wave file (No)

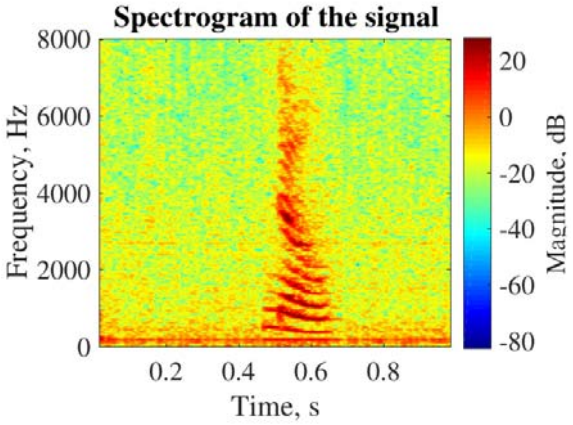

Figure 5: Spectrogram of the wave file (No)

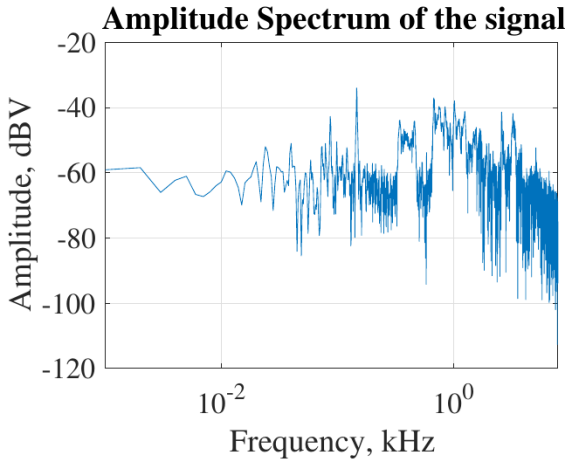

Figure 6: Amplitude spectrum of the wave file (No)

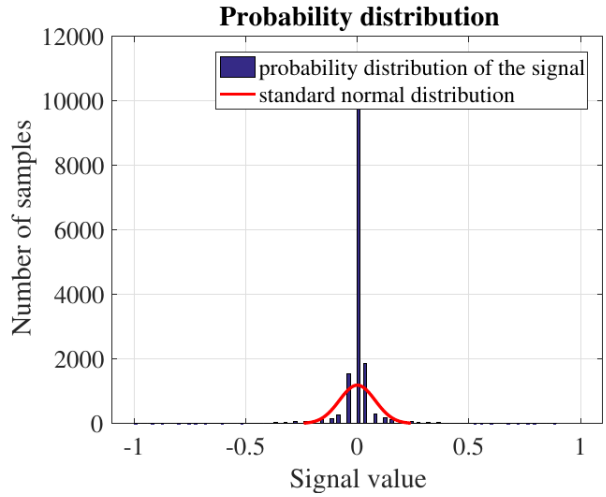

Figure 7: Probability distribution of the wave file (No)

Ten spoken words-phrases were selected for various persons; the features for each person-word were extracted using ANLBP, Tables 4 and 5 show the obtained results for two persons.

Table 4: Results for Person 1

Speech #	***Spoken Words***	Features			
1	Good morning	831	7513	7072	16012
2	Open the door	1320	7917	6961	16488
3	How are you	1396	6743	6044	16542
4	My name is Laham	3180	12955	10012	18443
5	Please shut down the computer	5371	14953	12589	20782
6	Good by	1090	4744	4932	16287
7	I am fine	1863	7513	6391	17074
8	This is my password	4865	10274	9252	20318
9	Please let me in	2153	9598	5667	17237
10	This is my voice	4793	7996	7558	20000

One hundred persons were chosen, and for each person a wave files were recorded for the words: no, yes, left, right, up, down, stop, proceed, reject, thanks. For each person-word a signature with a set of parameters value was created such as shown in Table 2 and 3. The calculation result shows that each signature is a unique, and does not match any other set. So each signature can be used as a password to recognize a person and a word. The proposed ANLBP model provided an excellent efficiency by decreasing the extraction time to an average of 0.001000 seconds. The extracted features were unique for each person-word, thus these features can be easily used to identify the person and the spoken word-phrase.

Table 5: Results for Person 2

1.1.1.1 Speech #	***1.1.1.2 Spoken Words***	1.1.1.2.1 Features			
1	Good morning	0652	3531	3277	8252
2	Open the door	1152	3490	3018	8681
3	How are you	1163	2926	2561	8710
4	My name is Lahham	1993	6227	4516	9557
5	Please shut down the computer	3515	6639	5505	11186
6	Good by	0896	2052	2117	8459
7	I am fine	1419	3292	2684	9023
8	This is my password	2828	4799	4277	10448
9	Please let me in	1162	4713	2741	8709
10	This is my voice	2299	4144	3837	9891

4. Conclusion

A suitable wave file parameters were chosen, and a unique signature for person- word was created. This signature can be passed to any recognition tool to recognize the person and the word. This signature can be applicable in any information system projects which require a highly secure key.

Acknowledgments

This work has been carried out during sabbatical leave granted to the author Mohamad M. AL-Laham from Al-Balqa Applied University (BAU) during the academic year 2020/2021.

References

[1] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, D. Zhang, Biometric recog- nition using deep learning: A survey, arXiv preprint arXiv:1912.00271 (2019).

[2] M. M. Al-Laham, Reducing security concerns when using cloud computing in on- line exams case study: General associate degree examination (shamel) in jordan, Int. J. Comput. Sci. Inf. Technol 7 (6) (2015) 131–144.

[3] D. B. Yoffie, L. Wu, J. Sweitzer, D. Eden, K. Ahuja, Voice war: Hey google vs. alexa vs. siri, Harvard Business School (2018) 25.

[4] S. Furui, Speaker recognition, Scholarpedia 3 (4) (2008) 3715.

[5] R. Ranjan, Speaker recognition and performance comparison based on machine learning, Turkish Journal of Computer and Mathematics Education (TURCO- MAT) 12 (14) (2021) 2297–2306.

[6] Q. Zhong, R. Dai, H. Zhang, Y. Zhu, G. Zhou, Text-independent speaker recogni- tion based on adaptive course learning loss and deep residual network, EURASIP Journal on Advances in Signal Processing 2021 (1) (2021) 1–16.

[7] M. M. AL-Laham, Encryption-decryption rgb color image using matrix multi- plication, International Journal of Computer Science & Information Technology 7 (5) (2015) 109–119.

[8] D. Garcia-Romero, D. Snyder, G. Sell, D. Povey, A. McCree, Speaker diarization using deep neural network embeddings, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 4930– 4934.

[9] X. Xiao, N. Kanda, Z. Chen, T. Zhou, T. Yoshioka, S. Chen, Y. Zhao, G. Liu, Y. Wu, J. Wu, et al., Microsoft speaker diarization system for the voxceleb speaker recognition challenge 2020, in: ICASSP 2021-2021 IEEE International Confer- ence on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 5824–5828.

[10] F. Landini, J. Profant, M. Diez, L. Burget, Bayesian hmm clustering of x-vector sequences (vbx) in speaker diarization: theory, implementation and analysis on standard tasks, Computer Speech & Language 71 (2022) 101254.

[11] M. J. Aqel, Z. A. Alqadi, I. M. El Emary, Analysis of stream cipher security algorithm, Journal of information and computing science 2 (4) (2007) 288–298.

[12] J. Al-Azzeh, Z. Alqadi, M. Abuzalata, Performance analysis of artificial neural networks used for color image recognition and retrieving, international Journal of Computer Science and Mobile computing 8 (2) (2019) 20–33.

[13] S. B. Sadkhan, S. F. Jawad, Security evaluation of cryptosystems based on orthog- onal transformation, in: 2020 6th International Engineering Conference “Sustain- able Technology and Development”(IEC), IEEE, 2020, pp. 222–226.

[14] H. G. Zaini, Z. AlQadi, Analysis of ffann used for pattern recognition, Interna- tional Journal of Computer Science and Mobile Computing 10 (3) (2021) 55–65.

[15] Z. Alqadi, B. Zahran, Q. Jaber, B. Ayyoub, J. Al-Azzeh, Enhancing the capacity of lsb method by introducing lsb2z method, International Journal of Computer Science and Mobile Computing 8 (3) (2019) 76–90.

[16] J. A.-a. A. Sharadqh, B. Ayyoub, Z. Alqadi, J. Al-azzeh, Experimental investi- gation of method used to remove salt and pepper noise from digital color image, International Journal of Research in Advanced Engineering and Technology 5 (1) (2019) 23–31.

[17] H. G. Zaini, Image segmentation to secure lsb2 data steganography, Engineering, Technology & Applied Science Research 11 (1) (2021) 6632–6636.

[18] S. Yadav, S. Taterh, M. A. Saxena, A literature review of various techniques avail- able on image denoising (2021).

[19] M. Nosrati, R. Karimi, H. Nosrati, A. Nosrati, Taking a brief look at steganogra- phy: Methods and approaches, Journal of American Science 7 (6) (2011).

[20] G. Qaryouti, S. Khawatreh, Z. Alqadi, M. Abu Zalata, Optimal color image recog- nition system (ocirs), International Journal of Advanced Computer Science and Technology 7 (1) (2017) 91–99.

[21] A. Al-Qaisi, A. Manasreh, A. Sharadqeh, Z. Alqadi, Digital color image classifi- cation based on modified local binary pattern using neural network, International Journal on Communications Antenna and Propagation (I. Re. CAP) 9 (6) (2019) 403–408.

[22] Shayeb, I., Asad, N., Alqadi, Z., &Jaber, Q. (2020). Evaluation of speech signal features extraction methods. Journal of Applied Science, Engineering, Technology, and Education, 2(1), 69-78.

[23] Rasmi, M., Al-salameen, F., Al-Laham, M. M., & Al-Fayomi, A. (2016). Enhancing RGB Color Image Encryption-Decryption Using One-Dimensional Matrix. In The 17th International Arab Conference on Information Technology (ACIT'2016), Beni-Mellal, Morocco.

Mohamad Al-Laham Associate Professor of Computer Information Systems specializing in Software Engineering in the field of Human Computer Interaction. He has published various research articles in the field of Information Security. Works at Al-Balqa Applied University.