A new method for voice signal features creation

Majed O. Al-Dwairi[1], Amjad Y. Hendi[2], Mohamed S. Soliman[3], Ziad A.A. Alqadi[4]
[1,2]Department of Communication Engineering Technology, Faculty of Engineering Technology,
Al-Balqa Applied University, Jordan
[3]Department of Electrical Engineering, Faculty of Engineering, Taif University, Taif, Kingdom of Saudi Arabia
[3]Department of Electrical Engineering, Faculty of Energy Engineering, Aswan University, Aswan, Egypt
[4]Department of Computer and Network Engineering, Faculty of Engineering Technology,
Al-Balqa Applied University, Jordan

Article Info

Article history:

Received Aug 7, 2018
Revised Apr 18, 2019
Accepted Apr 25, 2019

Keywords:

Crest factor
Dynamic range
Features
Wave file
Window

ABSTRACT

Digital audio is one of the most important types of data at present. It is used in several applications, such as human knowledge and many security and banking applications. A digital voice signal is usually of a large size where the acoustic signal consists of a set of values distributed in one column (one channel) (mono signal) or distributed in two columns (two channels) (stereo signal), these values usually are the results of sampling and quantization of the original analogue voice signal. In this paper we will introduce a method which can be used to create a signature or key, which can be used later to identify or recognize the wave file. The proposed method will be implemented and tested to show the accuracy and flexibility of this method.

Corresponding Author:

Majed O. Al-Dwairi,
Department of Communication Engineering Technology,
Faculty of Engineering Technology,
Al-Balqa Applied University,
P.O. Box: 15008, Amman 11134, Jordan.
Email: majeddw@bau.edu.jo

1. INTRODUCTION AND RELATED WORK

Digital audio is one of the most important types of data at present. It is used in several applications, such as human knowledge and many security and banking applications. A digital voice signal is usually of a large size where the acoustic signal consists of a set of values distributed in one column (one channel) (mono signal) or distributed in two columns (two channels) (stereo signal), these values usually are the results of sampling and quantization of the original analogue voice signal [1, 2].

Since the volume of the audio file is large, [3, 4] it is difficult to conduct the matching of two voices using all the values, where the process of matching will require a large amount of time, which in turn leads to delay in the process of sound recognition [5-7]. Table 1 shows the results of voice matching with itself, and here we can see that the bigger wave file size will increase the matching time, and the process of matching requires a big amount of time [8, 9].

To decrease the recognition time [10], we have to seek a method based on features extraction, this method will generate a set of features for any wave file, this set must be a unique and can be used as a key or a voice signature to retrieve or recognize the wave file. Any normalized wave file can be represented by a sinusoidal signal as shown in Figure 1. [1, 3], this signal can characterize by the following parameters: amplitude, frequency and phase shifting. If the features are based on these parameters, to any changes on these parameters must not affect the extracted voice features.

Table 1. Matching time for different wave files

Wave file	File size (elements)	Number of values	Matching time (seconds)
W1	36787×2	73574	0.006000
W2	39730×2	79460	0.008000
W3	33844×2	67688	0.007600
W4	17658×2	35316	0.005000
W5	41202×2	82404	0.007900
W6	36787×2	73574	0.006000
W7	63274×2	126548	0.014000
W8	48049×2	96098	0.010000
W9	55916×2	111832	0.013000
W10	89760×2	179520	0.019000
Average		92601	0.0097
Cost of 1 value		9700/92601=0.1048 microseconds	

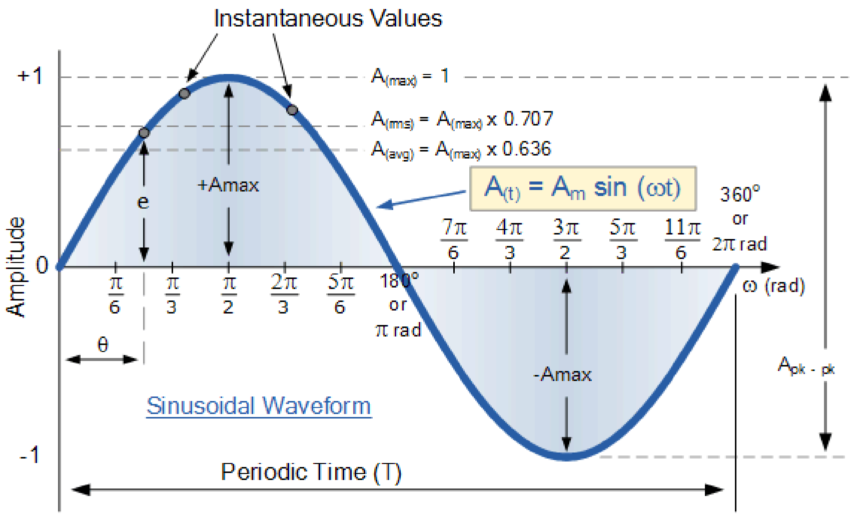

Figure 1. Sinusoidal signal

Many authors proposed some techniques of voice features extraction based on calculation: Crest factor, dynamic range, sigma (mean of the normalized data), and Mu (standard deviation of the normalized data). [11, 12]. The crest factor [4] is the ratio of peak value to RMS value of waveform as shown in Figure 2. This ratio is also called to peak-to-RMS ratio.

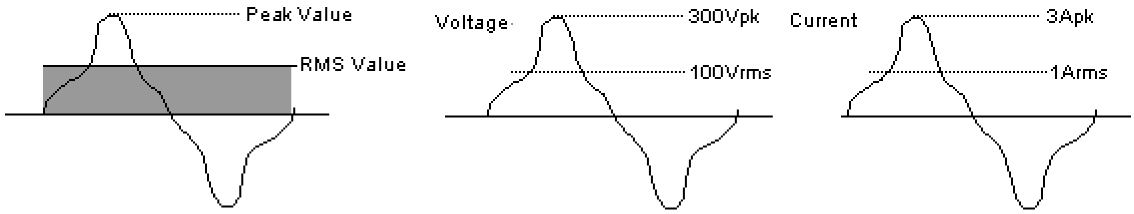

Figure 2. Calculating crest factor

Dynamic range [4-6] is the ratio between the largest and smallest intensity values of a changeable sound that can be reliably transmitted or reproduced by a particular sound system, measured in decibels. It's the measurement between the noise floor and the maximum sound pressure level and what a microphone can capture. In [9, 12] a method was proposed to generate voice signal features base on the above-mentioned parameters, any changes in amplitude, frequency, and phase shift will be reflected as some changes in voice signal features, thus will lead to more difficulties in the voice recognition process. Here we must notice that any change in the voice parameters must not change the voice features. For example, let us take the sinusoidal signals listed in Table 2 as shown in Figure 3.

Table 2. Original signal Y and some versions with changes in amplitude, frequency and phase shifting

Signal	Features(x=-4×π:0.001:4×π)			
	Crest factor (dB)	Dynamic range (dB)	Sigma	Mu
$Y = 5\sin(10x + 10)$	3.0103	82.2975	3.5356	-2.7846e-005
$Y1 = 15\sin(10x + 10)$	3.0103	82.2975	10.6068	-8.3539e-005
$Y2 = 5\sin(20x + 10)$	3.0103	90.4159	3.5356	-2.7686e-005
$Y3 = 5\sin(10x + 40)$	3.0103	82.2975	3.5356	-3.0453e-006

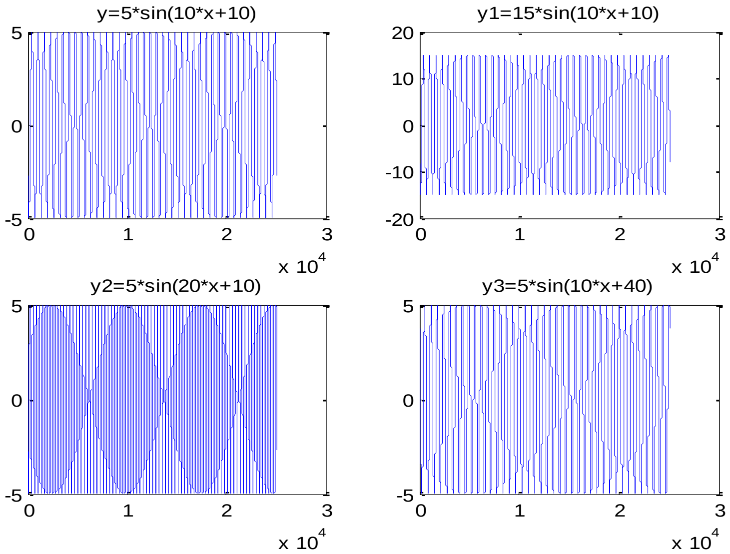

Figure 3. Changing the signal parameters

We can see from the results shown in Table 2, that some parameters remain the same, and the others changed, thus the features consisting of a set of these parameters will also change. If we record the previous sinusoidal signals and create a wave file using various sampling frequencies, then applying the proposed method in [13] we receive the results shown in Tables 3 thru 6.

Table 3. Wave files features using sampling frequency=1000

Signal	Features			
	Crest factor (dB)	Dynamic range (dB)	Sigma	Mu
Y	0.38699	68.0301	0.95644	-2.3515e-005
Y1	0.12484	58.7133	0.98575	-2.5364e-005
Y2	0.38699	76.3296	0.95644	-2.3524e-005
Y3	0.38699	68.0301	0.95644	-3.0453e-006

Table 4. Wave files features using sampling frequency=2000

Signal	Features			
	Crest factor(dB)	Dynamic range(dB)	Sigma	Mu
Y	0.38699	68.0301	0.95644	-2.3515e-005
Y1	0.12484	58.7133	0.98575	-2.5364e-005
Y2	0.38699	76.3296	0.95644	-2.3524e-005
Y3	0.38699	68.0301	0.95644	-3.0453e-006

Table 5. Wave files features using sampling frequency=3000

Signal	Features			
	Crest factor (dB)	Dynamic range (dB)	Sigma	Mu
Y	0.38699	68.0301	0.95644	-2.3515e-005
Y1	0.12484	58.7133	0.98575	-2.5364e-005
Y2	0.38699	76.3296	0.95644	-2.3524e-005
Y3	0.38699	68.0301	0.95644	-3.0453e-006

Table 6. Wave files features using sampling frequency=4000

Signal	Features			
	Crest factor(dB)	Dynamic range(dB)	Sigma	Mu
Y	0.38699	68.0301	0.95644	-2.3515e-005
Y1	0.12484	58.7133	0.98575	-2.5364e-005
Y2	0.38699	76.3296	0.95644	-2.3524e-005
Y3	0.38699	68.0301	0.95644	-3.0453e-006

From the results shown in these tables we can raise the following facts:

- For the same file the extracted features using various sampling frequencies remain the same.
- The features of the wave file are changed when adjusting the sinusoidal signal parameters (amplitude, frequency and phase shifting). Table 7 shows the features of different wave files with different parameters, each of them is a unique set and it can be used to identify the specific wave file with specific parameters.

Table 7. Wave files features

Wave file	Sampling frequency	Features			
		Crest factor	Dynamic range	Sigma	Mu
W1	44100	14.0422	83.9346	0.095352	-1.2441e-005
W2	44100	15.9654	82.9461	0.06817	-5.8692e-006
W3	44100	15.9017	84.5294	0.082403	-5.7592e-006
W4	44100	18.0547	76.6782	0.026053	-3.9937e-005
W5	44100	18.782	82.7243	0.048047	-1.8901e-005
W6	44100	16.8202	75.464	0.026106	-9.4489e-006
W7	44100	16.8423	79.2805	0.04041	-1.0895e-005
W8	48000	17.5073	79.392	0.037913	-1.4748e-005
W9	44100	16.7265	79.1359	0.040274	-1.9031e-005
W10	44100	19.0837	81.8915	0.042164	-3.304e-005

In [14] a method for voice feature extraction was proposed, this method is based local binary pattern to find the repetition of the values 0, 1, 2 and 3. This method is very effective in creating a wave file signature, but each repletion value is a big number and it will be increased when the wave file size increased. The obtained voice features can be used later to recognize the voice, and the voice features can be passed to a recognizer tool capable to process any application related to voice processing [15, 16].

Our paper will focus in building an effeciet algorithm to create a unique features array for each wave file, this features array can be used as a signature to recognize or retrieve a wave file. The created signature will remain the same for a wave file with deferent sampling frequency, this will reduce the memory space rquires for storing wave files.

2. THE PROPOSED METHOD

The proposed in this paper method uses an algorithm which is based on dividing the wave file into windows with fixed number of values, this algorithm can be implemented applying the following sequence of steps:

- Get the digital wave file.
- Reshape the wave file (mono or stereo voice) into one row.
- Initialize a 4 element features victor to zeros (First element points to the repetition of zeros, second element points to the repletion of ones, third element points to the repletion of twos, and the fourth element points to the repletion of threes).
- Divide voice row of values into windows with fixed size of voice values.
- Select the window size and number of windows
- While not of end windows do
- For each window find the average of the first half av0, and the average of the second half av1.
- Compare av0 with the value in the center of the first half, if av0 is greater or equal to the center value make b0=1, else make b0=0.
- Compare av1 with the value in the center of the second half, if av1 is greater or equal to the center value make a1=1, else make a1=0.
- Convert the binary number a1a0 to decimal d.
- Add 1 to features array element with index=d
- endwhile

Figure 4 shows an example of how to calculate the repetition for the first window with size equal ten values.

0.3	0.2	0.5	0.6	0	0.7	0.8	0.1	0.4	0.6	-	-	-	-	-

Sum =1.6000 sum=2.6000
Av=0.3200 Av=0.5200
Av<0.5 so a0=0 Av>0.1 so a1=1
Binary=10 feature index=2, add 1 to repetition 2

Figure 4. Calculation example

3. IMPLEMENTATION AND EXPERIMENTAL RESULTS

3.1. First experiment

A sinusoidal signal was taken, the proposed method was implemented to get the features of this signal, and we change one of the signal parameters (amplitude, frequency and phase shifting), for each changed signal we calculate the signal features, the results of this experiment are listed in Table 8. From Table 8 we can see that all the four signals have the same feature array, this means that changing the signal parameter does not affect the voice features.

Table 8. Experiment 1 results

Signal	Features(x=-4×π:0.001:4×π)			
$Y = 5\sin(10x + 10)$	4	1255	1254	0
$Y1 = 15\sin(10x + 10)$	4	1255	1254	0
$Y2 = 5\sin(20x + 10)$	4	1255	1254	0
$Y3 = 5\sin(10x + 40)$	4	1255	1254	0

3.2. Second experiment

In this experiment we recorded the previous four signals as a wave file using different sampling frequencies; the results of this experiment are listed in Tables 9 thru 12.

Table 9. Features with sampling frequency=1000

Wave file	Features			
Y	2136	172	192	13
Y1	2350	63	87	13
Y2	2094	177	218	24
Y3	2137	171	191	14

Table 10. Features with sampling frequency=2000

Wave file	Features			
Y	2136	172	192	13
Y1	2350	63	87	13
Y2	2094	177	218	24
Y3	2137	171	191	14

Table 11. Features with sampling frequency=3000

Wave file	Features			
Y	2136	172	192	13
Y1	2350	63	87	13
Y2	2094	177	218	24
Y3	2137	171	191	14

Table 12. Features for Wave file: Y with different sampling frequencies

Sampling frequency	Features			
1000	2136	172	192	13
2000	2136	172	192	13
3000	2136	172	192	13
4000	2136	172	192	13

From the results of experiment 2 we can raise the following facts:

- Fixing the parameters of digital signal and recording it with various sampling frequencies will keep the features of the wave signal without any change.
- Changing any voice parameter and record it with a new sampling frequency will change the voice features.

3.3. Thered Experiment

Different wave files were taken and treated using the proposed method, Table 13 shows some sample results of this experiment. From the results shown in Table 13, we can see that the set of each feature values is a unique set, thus we can use this set as a key or signature to retrieve or recognize the desired wave file. From the obtained experimental results, we can rase the following facts:

- For a wave file the created features array is a unique array, thus we can use this array as a signature to recognize the file.
- Changing the sampling frequency does not affect the features array values, thus there is no need to store extra copies (with deferent sampling frequencies) of a file, and this will minimize the storage size required to store the wave file data base.
- Features array is a simple data structure which contains only four values, this will simplify the architecture of the recognition tool such as artifiacial neural network.

Table 13. Experiment 3 results

Wave file	Size	Features			
W1	36787×2	585	2989	3250	533
W2	39730×2	1170	5978	6500	1066
W3	33844×2	1755	8967	9750	1599
W4	17658×2	2340	11956	13000	2132
W5	41202×2	2925	14945	16250	2665
W6	36787×2	3510	17934	19500	3198
W7	63274×2	4095	20923	22750	3731
W8	48049×2	4680	23912	26000	4264
W9	55916×2	5265	26901	29250	4797
W10	89760×2	5850	29890	32500	5330

4. CONCLUSION

A simple and highly accurate method was proposed to create a wave file features, which can be used as a wave file key or signature. The proposed method was implemented and tested using various wave file and it was shown from the obtained experimental results that: a) The created array of features is a unique for each wave file, thus it can be used as key to identify the wave file; b) For a recorded wave file with different sampling frequencies, the feature array does not change, remains the save using various sampling frequencies, thus make the proposed method more flexible, and efficient by reducing the required processor time and memory space needed for the process of voice recognition

REFERENCES

[1] M. Guillemard, et al., "Signal Analysis with Frame Theory and Persistent Homology," Jun 2013.

[2] Barker J., et al., "The third „chime‟ speech separation and recognition challenge: Dataset, task and baselines," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE*, pp. 504-511, 2015.

[3] S. K. Dastoor, "Comparative Analysis of Steganographic Algorithms impacting the information in the Speech Signal for enhancing the Message Security in next Generation Mobile devices," *(WICT2011), 2011 World Congress on Information and Communication Technologies. IEEE*, pp. 279-284, 2011.

[4] Bin, et al., "Dynamic range estimation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1618-1636, 2006.

[5] B. Wu, et al., "An analytical approach for dynamic range estimation," in *ACM/IEEE 41 Design Automation Conference (DAC-04), San Diego, Calif.*, Jun 2004.

[6] S. Millers and D. Childers, "Probability and random processes," Academic Press, pp. 370-5, 2012.

[7] F. Gouyon, et al., "Classifying percussive sounds: a matter ofzero-crossing rate in Proceedings of the COST G-6," *Conference on Digital Audio Effects (DAFX-00), Verona, Italy*, Dec 2000.

[8] D. D. Jaslene, "Feature Selection and Extraction of Audio," *International Journal of Innovative Research in Science, Engineering and Technology*, vol/issue: 5(3), Mar 2016.

[9] Sanjivani S., et al., "An Overview of Technical Progress in Speech Recognition," *International Journal of Advanced Research* in *Computer Science and Software Engineering*, vol/issue: 3(3), Mar 2013.

[10] M. I. Ouloul, et al., "An Efficient Face Recognition Using SIFT Descriptor in RGB-D Images," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 5(6), pp. 1227-1233, Dec 2015.

[11] S. Narang and D. Gupta, "Speech Feature Extraction Techniques: A Review," *IJCSMC*, vol/issue: 4(3), pp. 107-114, Mar 2015.

[12] Rybach, et al., "The RWTH Aachen University Open Source Speech Recognition system," Sep 2009.

[13] K. Matrouk, et al., "Speech Fingerprint to Identify Isolated Word-Person," *World Applied Sciences Journal*, vol/issue: 31(10), pp. 1767-1771, 2014.

[14] S. Khawatreh, et al., "A Novel Methodology to Extract Voice Signal Features," *International Journal of Computer Applications*, vol/issue: 179(9), pp. 0975-8887, Jan 2018.

[15] E. Benmalek, et al., "Voice Assessments for Detecting Patients with Parkinson's Diseases in Different Stages," *International Journal of Electrical and Computer Engineering*, vol/issue: 8(6), pp. 4265-4271, Dec 2018.

[16] F. P. George, et al., "Recognition of emotional states using EEG signals based on time-frequency analysis and SVM classifier," *International Journal of Electrical and Computer Engineering*, vol/issue: 9(2), pp. 1012-1020, Apr 2019.

BIOGRAPHIES OF AUTHORS

Majed O. Dwairi, in communication systems, was born on 10 of dicember 1968 in Jordan. He received his diploma degree in 1994 and PhD degree from Ukraine state Academy in 1998in the field of multichannel communication. An associate professor in the department of communication Engineering Technology, faculty of Engineering Technology/Al-Balqa Applied University Amman - Jordan. His research interests include optical communication networks, digital communications, signal and image prossecing, Antenna design, and microstrip patch antennas.

Amjad Y. Hendi, in Radio and TVsystems. He received his diploma degree in 1994 and PhD degree from Ukraine state Academy in 1998in the field of Radio &Tv systems. His research interests include digital communications, signal and image prossecing, Antenna design, optimization techniques in antenna design and antenna measurement techniques and microstrip patch antennas.

Mohamed S. Soliman an assistant professor in the department of Electrical Engineering, Faculty of Energy Engineering, Aswan University, Egypt. Currently, he is with the department of Electrical Engineering, Faculty of Engineering Taif University, Saudi Arabia. His research interests include wireless communications, phased and timed array signal processing, UWB microstrip patch antennas, dielectric Resonant antennas, numerical methods in electromagnetics, optimization techniques in antenna design and antenna measurement techniques. Dr. Soliman is a member of the IEEE-AP Society, KAUST chapter, Saudi Arabia.

Ziad AL-Qadi, in computer engineering, was born on 09 March 1955 in Jordan. He received his diploma degree in 1980 and PhD degree from Ukraine in 1986 in the field of Computer Engineering. Currently he is a Professor at the Computer Engineering Department, Faculty of Engineering Technology, Al-Balqa Applied University, Jordan. His main interest include signal processing, pattern recognition, algorithms, modelling and simulations.