A Novel Methodology to Extract Voice Signal Features

Saleh Khawatreh
AL-Ahliyya AMMAN
University

Belal Ayyoub
Balqa Applied University

Ashraf Abu-Ein
Balqa Applied University

Ziad Alqadi
Balqa Applied University

ABSTRACT

A novel methodology to manipulate wave file and create a feature array for each wave file will be introduced, this array can be used later on to recognize the voice file. A set of experiments will be performed in order to prove the uniqueness of the calculated feature array, and that the created feature array for a certain wave file does not match any other feature array for other wave files. The proposed methodology will minimize the efforts of voice recognition by mean of minimizing the time of feature array creation and minimizing the size of the calculated array.

General Terms

Voice recognitions, artificial intelligence

Keywords

Wave file, feature array, histogram.

1. INTRODUCTION

Voice recognition systems have become the important applications for speech recognition technology. In this paper, an animal identification (ID) detection system based on animal voice pattern recognition algorithm has been developed. Voice waveforms (stereo with 2 channels (two column array) or mono with one channel (one column array) are classified according to their amplitude, frequency, and shape, as well as the sites on the scalp at which they are recorded. The most familiar classification uses EEG waveform frequency [1], [2], and [3]. In [4] author used Time Frequency Transforms for signal decomposition in setting frame theory and provides new insights in application to data analysis (see Figure 1), which shows the amplitude of the two channels of the voice wave file bird.wav.
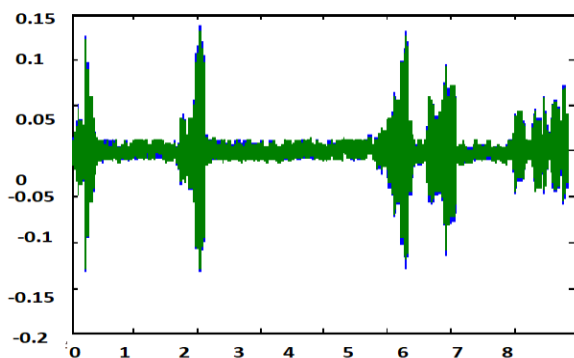

Fig 1: Amplitude of the bird.wav signal

Wave files can be represented histogram, which includes the repetition of each value in wave form of voice signal [5], [6],[7] and [8]. Figure 2 shows the histogram of the bird.wav voice signals with amplitude range value from -0.1319 to 0.1377.
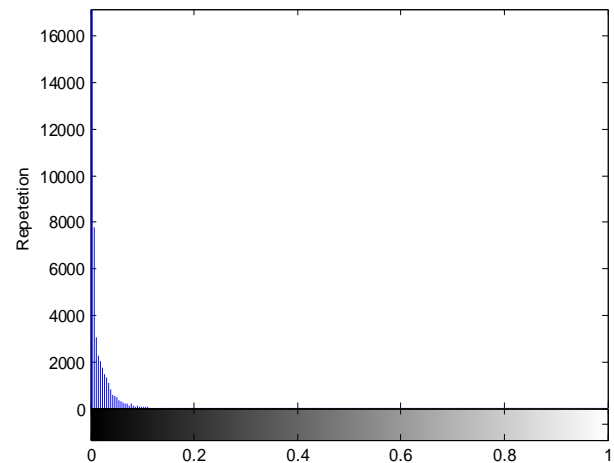

Fig 2: Histogram of the bird.wav

In Figure 2 which illustrating the histogram of the bird.wav signal the main objective of voice recognition system is to understand the spoken information and it must be completely accurate giving a 100% of recognition ratio. Many developments have been done in the field of voice recognition [9] and mostly all the developed proposed methods are based on: Voice analysis, feature extraction, modeling and matching [10]. These methods [11], [12],[13] and [14] are mostly suffering from the efforts needed to perform the process of voice recognition and the lowest recognition ratio. A method proposed in [15] is a good example of one of the current development which is based on extracting voice features based on voice analysis to calculate some parameters such as estimated population (mu), dynamic range, peak factor, Power spectral density, and zero crossing rate. These parameters require a deep processing and gathering them in one array to create a voice signature leads some time to decrease the recognition system efficiency by reducing the recognition ratio.

2. THE PROPOSED METHODOLOGY

The proposed methodology can be implemented applying the following steps:

2.1 Array features creation and initialization

Array features (AF) is a 4-element array: the first element reflects the repletion of zeros, the second reflects the repetition of ones, the third reflects the repetition of twos, while the fourth element reflects the repetition of threes. Initially array have to be reset to zeros

2.2 Prepare the wave file (mono or stereo wave file)

This step can by implemented by getting the wave file (WF) and converting it from one column (mono file) or from 2 columns to one row file.

2.3 For each value in WF p(i) do the following

2.3.1. Find a logical value k0 using the following equation

$$K0 = (P(i+1) + P(i-1) \geq 2\,P(i))$$

2.3.2. Find a logical value k0 using the following equation

$$K1 = (P(i+2) + P(i-2) \geq 2\,P(i))$$

2.3.3. Find the index of IAF

$$IAF = (k1(2) + K0 + 1)$$

2.3.4. Add one to AF (IAF)

Below is a calculated example:

Let us do the calculation for the value pointed by the red:

0.0153	0.0115	0.0115	0.0134	0.0077

$$K0 = (P(i+1) + P(i-1) \geq 2\,P(i))$$

$$= 0.0134 + 0.0115 \geq 2\ (0.0115)$$

$$= 0.0249 >= 0.0230 = 1$$

$$K1 = (P(i+2) + P(i-2) \geq 2\,P(i))$$

$$= 0.0077 + 0.0153 \geq 2(0.0115)$$

$$= 0.0230 \geq 0.0230 = 1$$

IAF $= k1(2) + k0 = 3$

So, add 1 to the repetition of 3 (AF (4) =AF (4) +1)

Table 1 shows how the calculated parameters k0 and k1 are used to increment the repetition of values from 0 to 3:

Table 1: Calculated parameters k1 and k0

Parameter k1	Parameter k0	Repetition to be incremented
0	0	0
0	1	1
1	0	2
1	1	3

3. IMPLEMNTATION AND RESULTS DISCUSSION

The mentioned methodology was implemented using Matlab environment, deferent wave file was used. Figure3 and Figure 4 shows the wave of the two channels of the bird stereo file:
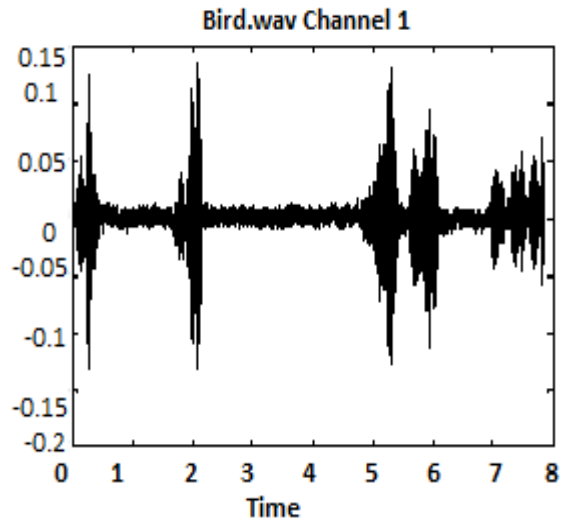

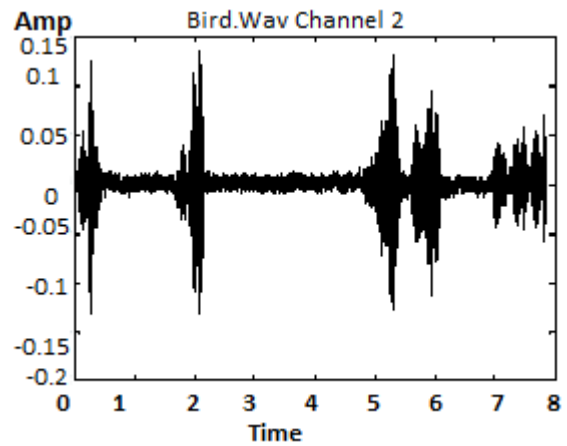**Fig 3: The Bird wave of the first channel**

Fig 4: The Bird wave of the second channel

In the table which is illustrated in the section below (see Table 2), all features which extracted for some stereo voices Table 2: Samples of the extracted features for some stereo voices
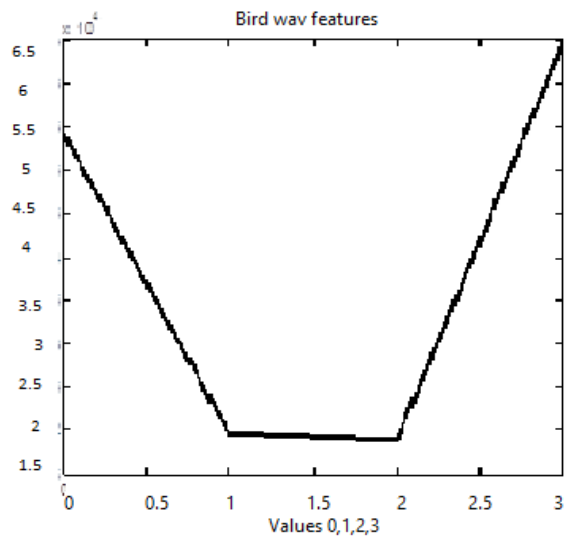

Fig 5: Values 0 to 3 distribution for file bird.wav

In the Figure 5 which illustrated the distribution of the values 0 to 3 which are to be used as extracted features for the file bird.wav.

From the results shown in Table 2 important Facts can be seen as illustrated below:

- Each wave file has unique features array.

- The features can be used as a signature to recognize the wav file.

- The extracted features are very sensitive to any changes in the wave file, as shown in Table 3, thus signature can be used to discover any distortion in the wave file.

- The proposed methodology is very simple to be implemented thus it reduces the deep processing of voice analysis eliminating the required time needed for voice processing.

- The proposed methodology simplifies the voice histogram by creating a new one consisting of only four values.

- The created feature are which can be used as a voice signature contain only four values, thus the array size is very

Table 2: Samples of extracted features for some stereo voices

Voice file	Feat-ure 1	Feat-ure 2	Feat-ure 3	Feat-ure 4
Bird.wav	54075	19533	18947	64616
Cow.wav	27260	3209	3342	28112
Cow2.wav	14438	2504	1208	16833
Dog.wav	31655	11615	11091	37772
Dolphin.wav	27757	9397	8953	32476
Donkey.wav	82767	30644	29122	95450
Duck.wav	144450	50842	48556	167219
Elephant.wav	15908	5965	5245	17911
Horse.wav	12625	2319	1392	14753

Small (32 bytes), so using this array for different voice application will lead to efficient voice processing

Table 3: The methodology sincerity to the changes in the wave file

Changes in wave file(only one element)	Features				
Bird.wav **WF(100)=0**	Before changes	54075	19533	18947	64616
	After changes	54075	19534	18947	64615
Cow.wav **WF(100)=0.5**	Before changes	27260	3209	3342	28112
	After changes	27257	3210	3344	28112
Dog.wav **WF(220)=0.5**	Before changes	31655	11615	11091	37772
	After changes	31655	11614	11091	37773

4. CONCLUSION

A novel Methodology for voice feature extraction was proposed, it was proved that the methodology very simple and effective and it can be used to create a voice signature which can be used for different application such voice recognition. There is a need to develop new hybrid Methods that will give better performance in robust speech recognition area.

5. REFERENCES

[1] Blume WT, Kaibara M. *Atlas of Pediatric Electroencephalography*. 2nd ed. Philadelphia: Lippincott-Raven; 1999.

[2] Fisch B, Spehlmann R. *Fisch and Spehlmann's EEG Primer*. 3rd ed. Amsterdam: Elsevier; 1999.

[3] Niedermeyer E, Lopes da Silva F. *Electroencephalography :Basic Principles, Clinical Applications, and Related Fields*. 5th ed. Baltimore: Williams & Wilkins; 1993.

[4] Mijail Guillemard, Gitta Kutyniok, Holger Boche, Friedrich Philipp,"Signal Analysis with Frame Theory and Persistent Homology", June 2013.

[5] Ioannides AA, Poghosyan V, Dammers J, Streit M. Real-time neural activity and connectivity in healthy individuals and schizophrenia patients. Neuroimage. 2004 Oct. 23(2):473-82.

[6] Stern JM, Engel J. *An Atlas of EEG Patterns*. Philadelphia: Lippincott Williams & Wilkins; 2004.

[7] Barker J., Marxer R., Vincent E., and S. Watanabe, "The third 'chime' speech separation and recognition challenge: Dataset, task and baselines," in 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU). IEEE, 2015, pp. 504–511.

[8] Johns Hopkins. Pacemakers for the brain. *Johns Hopkins Med Lett Health After 50*. 2004 Sep. 17(7):1-2

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

[10] Shreya Narang1, Ms. Divya Gupta2, Speech Feature Extraction Techniques: A Review, IJCSMC, Vol. 4, Issue. 3, March 2015, pg.107 – 114.

[11] Santosh K.Gaikwad and Pravin Yannawar, A Review, International Journal of Computer Applications A

Review on Speech Recognition Technique Volume 10–No.3, November 2010

[12] Rybach, D.; C. Gollan; G. Heigold; B. Hoffmeister; J. Lööf; R. Schlüter; H., Ney September 2009. "The RWTH Aachen University Open Source Speech Recognition

[13] Jaslene, Dawood Dilber, Feature "Selection and Extraction of Audio", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 5, Issue 3, March 2016

[14] Sanjivani S. Bhabad Gajanan K. Kharate International Journal of Advanced Research in Computer Science and Software Engineering, An Overview of Technical an Overview of Technical Progress in Speech Recognition Volume 3, Issue 3, March 2013

[15] Khaled Matrouk, Abdullah Al-Hasanat, Haitham Alasha'ary, Prof Ziad Al-Qadi, Prof Hasan Al-Shalabi Speech Fingerprint to Identify Isolated Word-Person World Applied Sciences Journal 31 (10): 1767-1771, 2014