

Azad R. Kareem 

Control and Systems
Engineering Department
University of Technology
Baghdad, Iraq
drazadnarj@hotmail.com

Marwa F. Jassim

Control and Systems
Engineering Department
University of Technology
Baghdad, Iraq
marwaalkubaissy@gmail.com

Mustafa Q. Ali 

College of Islamic Sciences
University of Baghdad, Iraq
mustafaquassay85@gmail.com

Received on: 14/06/2017
Accepted on: 05/04/2018
Published online: 25/8/2018

A Proposed Method for the Sound Recognition Process

Abstract-One of the most important issues in the signal-processing world is the issue of sounds differentiations. It has many applications in identifying the sources of the sounds, and many researches nowadays serves in this field and all of them looking for the best way to have a high accuracy implementation of the discrimination process. Sounds can be recognized under suitable recording conditions by converting the sound signal from time to frequency domain (because Sound signal of a source differs from other sources by the frequency contents. This property serves to differentiate between sounds, and differences become visually apparent when the spectrograms of the signals are compared). All the classical methods based on the amplitude comparison of the spectrum. The main problem faces the recognition process is the 100% system accuracy cannot be achieved. A proposed strategy suggested solving the problem. It is based on the comparison of slops between spectrogram sections instead of the magnitude comparison and taking the minimum differences between the pattern and references stored in a database. The tested examples with the help of Matlab program proved that the proposed method is more accurate than the conventional methods.

Keywords: sound recognition, frequency analysis, spectrogram, slop vector.

How to cite this article: A.R. Kareem, M.F. Jassim and M.Q. Ali, "A Proposed Method for the Sound Recognition Process," Engineering and Technology Journal, Vol. 36, No. 8, pp. 939-945, 2018.

1. Introduction

The recognition of sounds is the ability of a machine or program to convert the sound signals to coding patterns and saving them as references, comparing a new sound pattern enters the system with the reference patterns, selecting the nearest pattern and giving the decision.

Sound signal that is perceptible by humans has frequencies from about 20 Hz-20 kHz. This signal is containing the physical characteristics of the sources. Any sound signal can be represented graphically or mathematically either in time or frequency domains. The time domain describes the physical signals respecting to time. The signal or function can be converted between the time domain and frequency domain using mathematical operation called Fourier Transform or Fast Fourier Transformation (FFT) and then the waveform is called spectrogram [1]. The spectrogram has two sections; magnitude and phase spectrums. The magnitude spectrum has data about the comparative magnitude of the frequency formed for the signal. The phase spectrum consists of data about the comparative phase or time relations for the frequency formed. The process of recognition in the classical methods is accomplished by the magnitude comparison for the spectrograms and giving the

decision according to the degree of the similarity of the test sound with reference sounds. Sometimes the presence of the sound level difference between test and reference patterns makes the need for normalization necessary with the recognition processes.

In 2011, Manikandan et al. [2] proposed a real-time speech recognition system using two approaches; the first uses (MFCC) Mel Frequency Cestrum Coefficients with a recognition accuracy of 93.3% and the second approach uses Cochlear filter banks with (ZC) Zero Crossings as feature input for recognition with an accuracy 98.6%. Also in 2011, Guo et al. [3] designed a system for environmental sound recognition using time-frequency intersection patterns with a recognition rate for about 92%. An Arabic speech recognition was designed by Zaid and Abdulsattar in 2013 [4] with a recognition rate between 88.4% - 90.8%. While in 2015 [5] Li and Binwu propose an animal sound recognition based on double feature of spectrogram in real environment with a recognition rate about 80%.

This work proposes a new strategy by using slops comparison of the spectrogram sections in order to maximize the recognition accuracy, and

<https://doi.org/10.30684/etj.36.8A.15>

2412-0758/University of Technology-Iraq, Baghdad, Iraq

This is an open access article under the CC BY 4.0 license <http://creativecommons.org/licenses/by/4.0>

dispense with the normalization process. The operations are done by using the Matlab program.

2. The Proposed System

The algorithm of the proposed sound recognition system is as follows:

- A. Reading the sound signals.
- B. Signal analysis.
- C. Vectors calculation.
- D. Comparison with stored references.
- E. The differences measurement .
- F. Take the sound of vector with minimum difference.

The main process architecture is illustrated in Figure 1.

I. Read the sound signals

The microphone is used to convert the sound signal into electrical signal, this signal represented in time domain as shown in Figure 2. The waveform consists of a collection of sin waves, and each sine wave represents a tone or point on the spectrum of the signal.

II. Signal analysis

The main component for the system are shown in Figure 3 [6].

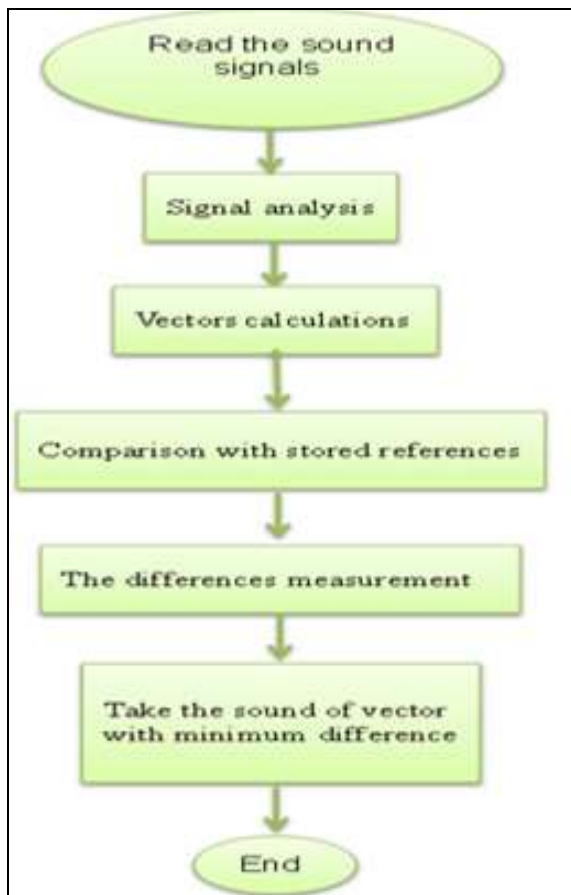


Figure 1: Flow graph of proposed system algorithm

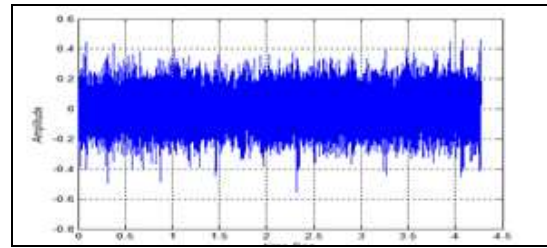


Figure2: Time domain representation of a sound signal

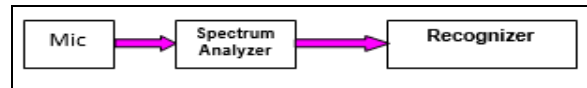


Figure 3: Sound recognition main system

- Spectrum analyzer: converts the sound signals to spectrograms which represent graphically the variation of intensities level versus the frequency for the sound signals.
- The recognizer: is a computerized program used to compare between spectrogram of the sound signals and deciding which the nearest one is, then recognize it.

Signal analysis is performed by converting it from time to frequency domain using the Fast Fourier Transform (FFT)

$$Y = FFT(X) \tag{1}$$

Where X represents the time domain of the sound signal, and Y represents spectrogram of the signal in frequency domain as shown in Figure 4.

III. Vectors calculation

The spectrograms converted into digital form named as the M-vector (magnitude vector). These Spectrograms can be divided into several parts (It is partitioned into 20 part), and the bands are not uniformly distributed in order to increase the efficiency of the process , then by averaging the magnitude of each consecutive part, the M-vector will be obtained. The output figure will be as shown in Figure 5. Here it is obvious that the M-vector is $M = [1.5 \ 1.96 \ 1.15 \ 0.6 \ 0.7 \ 0.98 \ 0.5 \ 0.41 \ 0.72 \ 0.41 \ 0.39 \ 0.2 \ 0.1 \ 0.08 \ 0.06 \ 0.023 \ 0.01 \ 0.01 \ 0 \ 0 \ 0] * 10^{-3}$

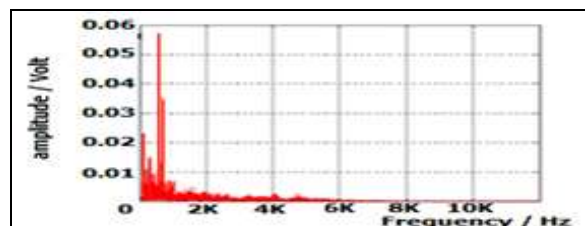


Figure 4: The spectrogram of the signal

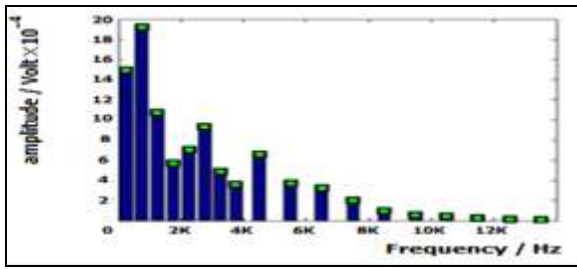


Figure 5: Amplitude vector of the sound signal

The classical methods use the normalization process to control the overall sound level or spectrogram magnitude.

This work proposed a new method for the recognition process based on the slop vectors comparison to eliminate the need for normalization step as well as increasing the method efficiency. The calculation of the slop vector is done by taking the slop between each consecutive two points in the (M-vector), this slop is determined as given in the relation [7]:

$$\text{Slop} = (y_2 - y_1) / (x_2 - x_1) \tag{2}$$

where x_1, y_1 are the coordinates of the first point in the (m-vector) and x_2, y_2 are the coordinates of the second point, the first item of the slop vector is obtained by taking the slop between these two points, in the same way the slop vector will be obtained as shown in Figure 6, which represents the slop diagram of the signal and then the slop vector (S-vector) is given as $S = [0.8 \ -1.7 \ -1 \ 0.3 \ 0.5 \ -0.9 \ -0.3 \ 0.48 \ -0.3 \ -0.08 \ -0.019 \ -0.081 \ 0.02 \ -0.2 \ -0.01 \ -0.01 \ 0 \ 0 \ 0 \ 0] * 10^{-5}$.

IV. Comparison with a stored reference

It is necessary to generate a database of references to be used in the recognition process by comparing the vector of unknown pattern with the database of known references.

Table 1 represents the references using the classical method (normalized magnitude vectors), it contained 16 vectors named from (A to M) each one represents the m-vector of a sound signal. Table 2 represents the patterns using the classical method, while Tables 3 & 4 represent the references & patterns respectively using the proposed method (slop vectors), each row represent the slop vector of a specific signal.

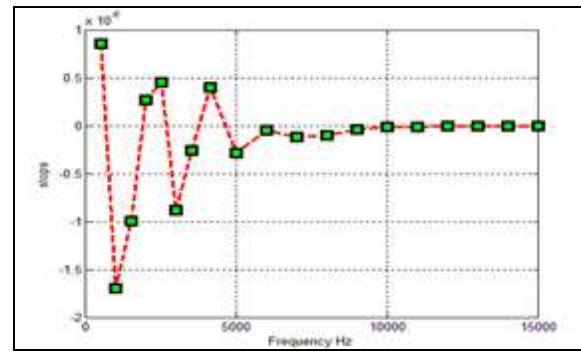


Figure 6: The slops of the sound signal diagram

V. The differences measurement

The pattern recognition problem solution is depending on the minimum difference method by comparing the unknown pattern vector of a test signal with the database of the references vectors. Mean Absolute Error (MAE) is used as a comparison approach to compare between vectors and is defined as a quantum used to measure how forecasts are close. The mean absolute error is given by [8]:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \tag{3}$$

When n = elements no.
 f = unknown vector's value.
 y = reference vector's value.

VI. Take the sound of the vector with Minimum difference

After comparing each pattern with all the references, the test signal is recognized according to the minimum number as shown in Table 5, which represents the average difference of magnitude vectors using classical method, and Table 6 shows the average difference of slop vectors using proposed method. The results of the recognition process proved the efficiency of the proposed method by increasing the accuracy of the system; it has up to 100% accuracy for the tested data while in the classical method the accuracy was 75% for the same data.

The test is done in ideal acoustically isolated environment, note that the efficiency of the system changed and affected by circumstances, if it's not ideal it will relatively decrease the efficiency of the system this is what can happened when using the system in a specific application and needing to some processing according to its environment.

3. Conclusions

The proposed system has been implemented using the slop method, and the experimental results of the recognition process proved the

efficiency of the proposed method over the classical method (the normalized magnitude method) by increasing the accuracy of the system; the slop method has up to 100% accuracy for the tested data in ideal environment while in the classical method the accuracy was 75% for the same data. In addition to this method does not add calculations or complexities to the design of the system and requires the same period of time that needed in classical methods, also there is no need for the normalization step with the new algorithm.

References

[1] M. Al-Akaidi, "Fractal Speech Processing", Cambridge University, 2004.
 [2] J. Manikandan, B. Venkataramani, K. Girish, H.Karthic and V. Siddharth," Hardware Implementation of Real-Time Speech Recognition System using TMS320C6713 DSP," Department of Electronics and Communication Engineering (ECE), National Institute of Technology, 24th Annual Conference on VLSI Design, INDIA, 2011.

[3] X. Guo ,Y.Toyoda and H.Li, " Environmental sound recognition using time-frequency intersection patterns",Applied Computational Intelligence and Soft Computing Volume 2012, Article ID 650818, 6 pages.
 [4] Z.Y. Mohammed and A.M. Khidhir, "Real-Time Arabic Speech Recognition", International Journal of Computer Applications, Volume 81 – No.4, November 2013.
 [5] Y. Li and Z. Binwu, "Animal Sound Recognition based on double feature of spectrogram in real environment," International Conference on Wireless Communications & Signal Processing (WCSP), Pages 1-5, IEEE Conference Publications, 2015.
 [6] B. Schuller, "Intelligent Audio Analysis, Springer, Germany, 2013.
 [7] B.G. Thomas et al., "Thomas' Calculus," 13th edition, Pearson, 2014.
 [8] S. Deviant, "Practically Cheating Statistics Handbook," Printed in the United States of America, 2010.

Table1: The references using the classical method

File Name	500	1000	1500	2000	2500	3000	3500	4000	5000	6000	7000	8000	9000	10000	11000	12000	13000	14000	15000	16000
A	0.77818	1	0.55821	0.26700	0.36682	0.48477	0.25576	0.1876	0.34442	0.1949	0.16826	0.10536	0.05101	0.03011	0.02219	0.01382	0.0104	0.00782	0.00636	0.00516
B	0.71894	1	0.84385	0.32732	0.34921	0.37078	0.37584	0.23216	0.21402	0.15073	0.10862	0.09007	0.03275	0.02067	0.01259	0.00915	0.00577	0.00578	0.00392	0.00349
C	0.86106	1	0.50217	0.32089	0.35743	0.34284	0.26033	0.19375	0.33098	0.20578	0.14157	0.09347	0.04645	0.02829	0.01666	0.01162	0.00909	0.00796	0.0056	0.00416
D	0.91259	1	0.53754	0.32598	0.25282	0.29234	0.23188	0.19938	0.21866	0.14663	0.09755	0.06381	0.03328	0.01915	0.01287	0.00940	0.00667	0.00698	0.00394	0.00359
E	0.88354	1	0.50618	0.34869	0.44399	0.40327	0.34496	0.3452	0.33407	0.24944	0.18233	0.10935	0.05711	0.02672	0.0211	0.01724	0.01285	0.0103	0.00779	0.0065
F	0.72917	1	0.59687	0.25691	0.43947	0.39542	0.37488	0.24991	0.19455	0.15288	0.10344	0.05982	0.03341	0.01901	0.01105	0.009	0.00533	0.00598	0.00393	0.00347
G	1	0.83339	0.41497	0.19620	0.30567	0.24554	0.21825	0.3183	0.21385	0.19145	0.12361	0.07114	0.04151	0.01575	0.0111	0.00853	0.00619	0.00522	0.00433	0.00348
H	0.87738	1	0.45105	0.34174	0.28008	0.19527	0.17154	0.17382	0.17298	0.11616	0.08582	0.04785	0.02978	0.01583	0.01205	0.00822	0.00574	0.00491	0.00386	0.00306
I	0.57513	1	0.48918	0.38853	0.41463	0.58128	0.34891	0.31811	0.43883	0.25741	0.24629	0.19718	0.11317	0.08732	0.05515	0.03029	0.02175	0.01795	0.0143	0.01256
J	0.60237	1	0.549	0.37516	0.35841	0.25582	0.27988	0.20205	0.20663	0.18114	0.09951	0.07967	0.04149	0.02078	0.01581	0.01003	0.0065	0.00542	0.00393	0.00318
K	1	0.91381	0.79551	0.48863	0.46355	0.45883	0.41978	0.29124	0.36883	0.23693	0.17594	0.10343	0.07015	0.03682	0.02922	0.01715	0.01117	0.00924	0.00702	0.0057
L	0.55378	1	0.45014	0.32285	0.32715	0.29077	0.22775	0.1522	0.21538	0.13807	0.10023	0.07009	0.04478	0.0242	0.01947	0.01257	0.00802	0.0062	0.00422	0.00349
P	0.74786	1	0.47832	0.35872	0.36533	0.27238	0.26147	0.18441	0.25844	0.19109	0.13425	0.0671	0.04135	0.02542	0.01512	0.01215	0.00904	0.00816	0.00535	0.00415
O	0.84133	1	0.54071	0.35155	0.34915	0.49764	0.35059	0.28343	0.31846	0.20293	0.14722	0.08931	0.08186	0.03635	0.02615	0.01788	0.01264	0.01116	0.00852	0.00522
N	0.51801	1	0.5405	0.50602	0.40215	0.3991	0.33547	0.22883	0.26237	0.19002	0.13782	0.09153	0.04823	0.02492	0.01602	0.00605	0.00609	0.00634	0.0037	0.00301
M	0.84395	1	0.45777	0.34903	0.34302	0.48129	0.42828	0.27369	0.27941	0.22129	0.1648	0.12028	0.07473	0.03425	0.03037	0.01581	0.00987	0.01132	0.00682	0.00551

Table 6: The average difference of magnitude vectors using proposed method.

Ref Pattern	A	B	C	D	E	F	G	H	I	J	K	L	P	O	N	M
AA	6.292E-08	3.87E-07	1.84E-07	2.7E-07	4.4E-07	5.14E-07	5.58E-07	4.43E-07	3.48E-07	3.94E-07	2.38E-07	3.58E-07	3.3E-07	1.09E-07	4.05E-07	1.88E-07
BB	3.489E-07	9.74E-08	2.58E-07	3.27E-07	4.44E-07	1.58E-07	5.22E-07	5.42E-07	4.48E-07	2.74E-07	3.78E-07	2.58E-07	2.58E-07	3.57E-07	3.12E-07	3.48E-07
CC	1.463E-07	3.03E-07	5.33E-08	1.99E-07	4.07E-07	3.93E-07	4.5E-07	3.4E-07	3.41E-07	2.75E-07	2.63E-07	2.89E-07	2.15E-07	1.77E-07	2.85E-07	2.09E-07
DD	2.31E-07	2.72E-07	1.8E-07	6.23E-08	4.55E-07	3.98E-07	4E-07	2.88E-07	4.15E-07	3.45E-07	3.3E-07	2.52E-07	2.45E-07	2.17E-07	3.53E-07	2.28E-07
EE	1.664E-07	3.98E-07	2.38E-07	2.82E-07	5.77E-08	4.22E-07	3.95E-07	4.11E-07	4.78E-07	3.7E-07	2.04E-07	3.82E-07	2.93E-07	2.14E-07	3.91E-07	1.83E-07
FF	5.038E-07	2.38E-07	3.7E-07	4.29E-07	5.38E-07	6.73E-08	4.71E-07	5.87E-07	5.27E-07	2.97E-07	5.1E-07	3.49E-07	3.08E-07	5.12E-07	3.64E-07	4.98E-07
GG	4.912E-07	4.95E-07	4E-07	3.37E-07	7.17E-07	4.39E-07	8.58E-08	5.29E-07	6.48E-07	4.98E-07	4.34E-07	5.32E-07	4.48E-07	5.03E-07	5.51E-07	4.89E-07
HH	4.78E-07	4.58E-07	3.44E-07	3.43E-07	5.89E-07	5.28E-07	5.16E-07	8.07E-08	5.82E-07	3.93E-07	5.18E-07	3.73E-07	3.27E-07	4.47E-07	4.16E-07	4.49E-07
I	3.963E-07	3.78E-07	3.32E-07	4.09E-07	2.38E-07	5.01E-07	7.06E-07	5.88E-07	9.37E-08	3.42E-07	5.88E-07	2.8E-07	4.04E-07	3.37E-07	3.31E-07	4.02E-07
JJ	3.428E-07	1.93E-07	2.12E-07	3.19E-07	4.04E-07	2.83E-07	5.12E-07	4.18E-07	3.78E-07	7.2E-08	3.51E-07	1.68E-07	1.68E-07	3.38E-07	1.45E-07	3.1E-07
KK	2.014E-07	4.18E-07	2.8E-07	2.88E-07	6.51E-07	5.41E-07	4.88E-07	5.38E-07	5.99E-07	4.4E-07	5.85E-08	4.06E-07	3.69E-07	2.88E-07	4.65E-07	2.68E-07
LL	2.898E-07	2.36E-07	1.84E-07	2.83E-07	3.19E-07	2.91E-07	5.29E-07	3.91E-07	2.74E-07	1.28E-07	3.68E-07	1.03E-07	1.85E-07	2.79E-07	1.47E-07	2.85E-07
PP	2.754E-07	2.51E-07	1.88E-07	2.74E-07	4.28E-07	3.18E-07	4.57E-07	3.47E-07	3.84E-07	1.81E-07	2.97E-07	1.67E-07	7.4E-08	2.5E-07	1.83E-07	2.41E-07
OO	9.834E-08	3.18E-07	1.78E-07	2.21E-07	4.34E-07	4.9E-07	5.42E-07	4.05E-07	3.58E-07	3.52E-07	2.3E-07	3.03E-07	2.68E-07	8.72E-08	3.43E-07	8.89E-08
NN	3.817E-07	2.85E-07	2.73E-07	3.87E-07	3.83E-07	3.69E-07	6.11E-07	4.68E-07	4.03E-07	1.28E-07	4.18E-07	1.88E-07	2.17E-07	3.13E-07	5.06E-08	3.23E-07
MM	1.254E-07	3.25E-07	2.31E-07	2.58E-07	5.17E-07	4.8E-07	5.35E-07	4.51E-07	4.52E-07	3.58E-07	2.03E-07	2.82E-07	2.82E-07	1.07E-07	3.18E-07	5.54E-08