Heuristic Method of Arabic Speech Recognition

Khalid Saeed and Mohammad Nammous

Department of Real-Time Systems
Faculty of Computer Science
Bialystok University of Technology, Poland[*]

http://aragorn.pb.bialystok.pl/~zspinfo/

ABSTRACT – This paper deals with the application of Toeplitz matrices and their minimal eigenvalues together with a number of different types of Neural Networks on Speech Recognition. The speech signal is looked at as an image and it is treated graphically. The object of our research in this work is some spoken Arabic words – the digits from zero to ten, recorded from a set of twenty persons. The approach has shown promising results and in most cases it gave very high success rate through the high recognition rate it presented.

1. INTRODUCTION

There have been proved [1, 2, 3] that the characteristics of the non-increasing sequence of minimal eigenvalues extracted from Toeplitz matrices establish an interesting way of image description. The series of these eigenvalues converges to a limit of specific value depending on the characteristics of the input image, to which the matrices belong. The theory had been applied on recognition of scripts of cursive character and with different designation of input data [3]. The following, however, shows the general new way of describing images of voices and their classification.

2. PRE-PROCESSING SIGNALS

Each recorded file can contain only one voice with a silence region before and after the right signal. The authors used the standard format PCM, with frequency 22050 Hz, 16-bits mono, as a standard in this work. Figure 1 shows a simple graph of speech preparation for feature extraction and classification.

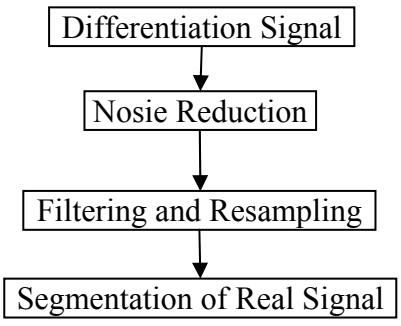

Fig. 1 Speech signal preparation for further process.

More information and details about how to prepare the signals to recognition can be found in [4, 5, 6]. And now, the new signal is ready to the next step of feature extracting.

[*] *This work is supported by the Rector of Technical University of Bialystok (grant no.W/WI/03/04)*

3. VOICE BASE

Our base consists of recorded voices for twenty people from six different countries, not only Arabic. The total number of recorded samples is 5472 divided into two groups. For each person and voice we choose five samples to be a test set (totally 1100 samples), while the remaining samples (4372 samples) are chosen to be the teaching set.

4. EXAMPLES

Experiments showed that power spectrum estimation of Burg's model is one of the best approaches for the purpose of presenting waveform graphically; in our tests it was used alone or as input method for other algorithms, so its parameters always had a deciding effect on the last results. In our experiments we tested all possible combinations for the following values of Burg's parameters:

The length of FFT (*NFFT*) : 32, 64, 128, 256, 512, 1024
The prediction order (*P*) : 8, 10, 12, 16, 20, 24, 28, 32, 40

That means we had about 54 individual situations for each method, but simply we will present only a few cases for each value of the first parameter (*NFFT*).
Firstly we will show the results when using Burg method on these samples without using any additional algorithms. Next, we test the Eigenvalues algorithm with two of its modifications, the first modification depends on reversing the coordinates of Burg's curve feature points before using the Toeplitz system, the second modification shows the dependence of using the difference between the next points of Burg curve as coordinates put directly in Toeplitz model, the last method to be applied is the Projection one. It depends on the curve obtained by Burg method, projecting every selected characteristic point onto the nearest of twelve axes selected according to certain rules mentioned and discussed in details in some previous work [7].
The following simple table (Table 1) shows the best recognition rate for each of the mentioned algorithms and the adjacent method of classification. We are not presenting the detailed parameters of the obtained results as they are given in details in some previous work [8] .

Table 1 *Recognition results with different methods of processing and classification*

Recognition Results / *Method of Processing*	**with Classical Classification Method**		**with Radial NN**		**with Probabilistic NN**	
	Wrongly-Recognized Samples	Recognition Rate	Wrongly-Recognized Samples	Recognition Rate	Wrongly-Recognized Samples	Recognition Rate
Burg Method	70	93.64%	**28**	**97.45%**	**19**	**98.72%**
Eigenvalues Algorithm	149	86.45%	**46**	**95.82%**	673	38.82%
I Modification of Eigenvalues Alg.	603	45.18%	271	75.36%	866	21.27%
II Modification of Eigenvalues Alg.	478	56.55%	261	76.27%	953	13.36%
Projection Method	436	60.36%	655	40.45%	479	56.45%

5. CONCLUSIONS

Minimal eigenvalues algorithm used in data reduction and similarity measurements [3] has shown very interesting results in speech processing and recognition when working with neural networks [8, 9]. In this work, however, the heuristic research of eleven different experiments made on 5472 samples is summarized and given here with a comparison between the considered methods of classification and recognition of Arabic spoken-digits zero to ten. This hybrid way of testing and identification has shown as good results as in written text recognition. Table 1 shows how the minimal eigenvalues algorithm has not given more than 86.45% in best cases when used in its original version and without neural networks. However, when used with radial neural networks [10] the rate of recognition was improved to 95.82% which is very close to the efficiency of

Burg method, the approach which has furnished the highest recognition rate (98.72%) when applied with the probabilistic neural networks. The latter in turn proved to be impractical to use with the minimal eigenvalues or any of its modifications as it gave a much lower recognition rate than the classical classification methods, in many other cases radial NN working with Eigenvalues algorithm or its modifications increased the efficiency by 30% percent more than the classical methods of classification.

The recognition rate for the projection method was low. And it seems that this method does not fit speech recognition in its graphical-based approach, at least so far. This and other aspects of interest in this work are being worked on for better results and improvement to get similarly good results as in the case of handwritten and printed script recognition [7, 11, 12].

6. REFERENCES

[1] K. Saeed, "On the Realization of Digital Filters, " Proc. 1st International Conference on Digital Signal Processing and its Applications – *IEEE-DSPA'98*, Vol.1, pp. 141-143, Moscow 1998.

[2] K. Saeed, "A New Approach in Image Classification," Proc. 5th International Conference on Digital Signal Processing and its Applications – *IEEE-DSPA'03*, Vol.1, pp. 49-52, Moscow 2003.

[3] K. Saeed, "Image Analysis for Object Recognition," *Bialystok Technical University Press,* Poland 2004.

[4] Lyons Richard G, "Wprowadzenie do cyfrowego przetwarzania sygnałów," *WKiŁ*, Warszawa 1999 (in Polish).

[5] K. Saeed, Nammous M., "An Image-Based Algorithm to Identify Spoken Arabic Digits," Computer Information Systems and Applications, Vol.1, pp.55-66, *WSFiZ Press*, Bialystok, Poland 2004.

[6] Sadaoki Furui: "Digital Speech Processing, synthesis and recognition, second edition", z serii "Digital Speech Processing and Communication", wydawca serii: K.J. Ray Liu.

[7] K. Saeed, "A Projection Approach for Arabic Handwritten Characters Recognition," P. Sincak and J. Vascak (Editors), Quo Vadis Computational Intelligence? New Trends and App. in Comp. Intelligence, pp. 106-111, *Physica-Verlag*: Berlin 2000.

[8] K. Saeed, M. Nammous, "A New Step in Arabic Speech Identification: Spoken Digit Recognition," *Springer-Verlag*, under publication New York 2005.

[9] Shigeru Katagiri: "Handbook of Neural Networks for Speech Processing," *Artech House*, Boston 2000.

[10] M.W. Mak, W.G. Allen and G.G.Sexton, "Speaker identification using radial basis functions", The Third International Conference on Artifical Neural networks, *University of Northumbria at Newcastle*, U.K 1998.

[11] Saeed K., Tabędzki M., "Cursive-Character Script Recognition using Toeplitz Model and Neural Networks," 7th ICAISC - International Conference on Artificial Intelligence and Soft Computing, June 7-11, Zakopane 2004. Proceedings published by *Springer-Verlag Berlin Heidelberg*: Lecture Notes in Computer Science, L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L. Zadeh (Editors), *LNCS* 3070, pp.658-663, Berlin 2004.

[12] K. Saeed, M. Tabędzki, "A New Hybrid System for Recognition of Handwritten-Script," *International Scientific Journal of Computing*, *Institute of Computer Information Technologies*, Volume 3, Issue 1, pp. 50-57, 2004, Ternopil, Ukraine 2004.