

# A Note on Biometrics and Voice Print: Voice-Signal Feature Selection and Extraction – A Burg-Töeplitz Approach

Khalid Saeed<sup>#</sup>

<sup>#</sup> Faculty of Computer Science,  
Bialystok Technical University,  
Wiejska 45A, 15-351 Bialystok, Poland.

[aida@ii.pb.bialystok.pl](mailto:aida@ii.pb.bialystok.pl)

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

**Abstract:** *This work presents new applications of Töeplitz matrix eigenvalues approach in image description, feature extraction and recognition. It discusses the possibility of treating the speech signal graphically in order to extract the essential image features as a basic step in successful data mining applications in the biometric techniques. The considered object here is the human-voice signal. The suggested frequency spectral estimation and Töeplitz-based approach, built on linear predictive coding principle, has proved the possibility of selecting signal features from the power spectral plot and entering Töeplitz matrix in a manner similar to its application on images of written texts, signature, palm-print, face geometry or fingerprints, the topics that have shown a success rate of about 98% in many cases. The extracted feature-carrying image comprises the elements of Töeplitz matrices to consecutively compute their minimal eigenvalues and introduce a set of feature vectors within a class of voices. The required computations were performed in MATLAB proving speech-signal image recognition in a simple and easy-to-use way. This stems from the fact that the presented problem solution and its Matlab implementation do not require to implement any special hardware and can be used in tandem with other biometric technologies in hybrid systems for multi-factor verification.*

## 1 Introduction

Voice identification in terms of both speech and speaker authentication has its unique significance and role in almost all known biometric techniques. The reason is simply that people seek for an easy-to-use, reliable and safe system to show and prove their authenticity whenever needed. Each biometric type of technology demands a user to play their active part in the process of being identified and authenticated. They involve user's writing on paper, stamping their fingerprints, showing their open eye to a camera, pressing their hand to show its geometry, and the most common way of identifying

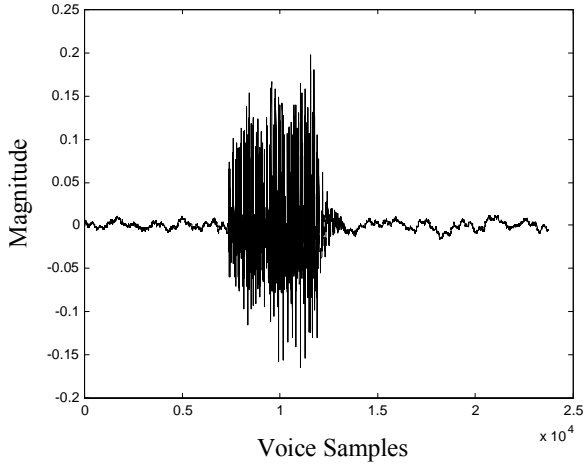
the user by their recalling a code or a password to enter the identifying machine with. Voice biometric methods of human identification, however, need nothing but the human utterance to obtain their voiceprint. The term "voiceprint" was coined as everybody has their own unique and characteristic voice parallel to the so called fingerprint left by fingers of an individual. This term is used in most voice biometric solutions as a template of our unique voice features manifested while entering when entering the identifying system.

In this paper, the author introduces some examples and experiments to show how this voiceprint looks like from the graphical point of view and how it is identified for recognition. The basic idea is derived from applying Töeplitz matrix minimal eigenvalues algorithm [1] to Burg's model [2]. This implies a graphical approach for feature extraction, selection and hence signal-image description confronting the conventional and traditional methods. Töeplitz matrix approach is employed to verify a variety of biometrics, including the recognition of hand and machine written texts [3], off [4] and on-line [5] signature, face [6], and voice [7]. In all, it has proved a promising success rate.

The same algorithm has also shown its possible application in hybrid systems [3] where multiple forms of classifying and identifying tools are fused in one system. The image of a voice signal in any of its classical forms is rather complicated and usually does not convey exactly similar images of the same signals, even when spoken by the same person. However, Burg's model [2] inspired the author and his team to undertake a promising area of study, and they have managed to discover some new facts. These facts concern the possibility of looking at the voice-signal image in a manner similar to any other object image. This enabled extending Töeplitz matrix applications to cover speech signal description, as well. The experiments and their results in this work have either been published [8] or are under publishing [7]. All of the experiments were performed in Matlab.

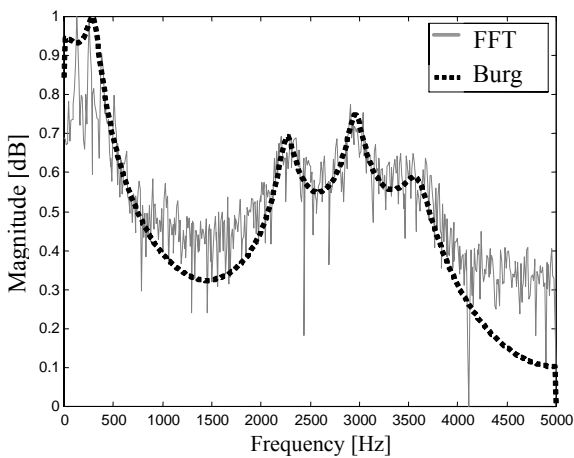
## 2 Theoretical considerations – A review

In this section, a brief discussion is given to show how the speech signal is seen as an object image. Consider the original waveform of a recorded voice in Fig. 1.



**Fig. 1.** The speech waveform of a recorded voice.

Fig. 2, however, depicts the performance of obtaining the speech-signal image, which can be suitable for further image analysis. In its shown shape (the solid line in Fig. 2), the spectrum resulting from applying standard FFT (Fast Fourier Transform) is not practical to analyze as a speech image. It is difficult to select and extract the image features in a manner similar to the analysis of common object images [7]. Experiments have shown that the power spectrum estimation of Burg's model is the best for this purpose [9]. For most of the phonemes, the spectral estimation furnishes a smooth envelope, while local extremes can be seen clearly (Fig. 2).



**Rys. 2.** The standard FFT spectrum and its estimation by *Linear Predictive Coding*.

## Feature selection by LPC estimation

In order to reach Burg's approximating model given in Fig. 2, let us consider the idea of LPC (Linear Predictive Coding) estimation. The theory of linear predictive coding is given in [10]. A brief explanation is given here.

The signal sample  $u(n)$  can be approximated as the linear combination of the  $P$  previous samples, with  $n > 0$ .

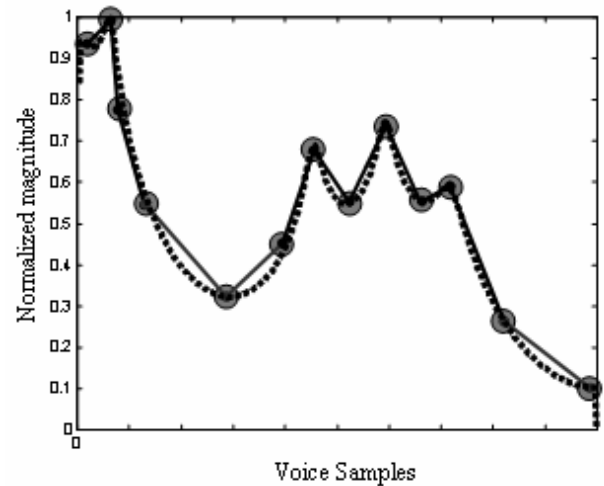
The  $n^{\text{th}}$  sample estimator  $\tilde{u}(n)$  is defined as:

$$\tilde{u}(n) = -\sum_{p=1}^P a_p u(n-p) \quad (1)$$

where  $a_p$  – prediction coefficients,  $p = 1, 2, \dots, P$ , and  $P$  – prediction order. The difference between  $u(n)$  and  $\tilde{u}(n)$  is called the prediction error  $e(n)$ . Hence,

$$e(n) = u(n) - \tilde{u}(n) = u(n) + \sum_{p=1}^P a_p u(n-p) \quad (2)$$

In this way, the obtained spectrum will then be very clear to extract the acoustic image from it and to easily fix the feature and characteristic points on it (Fig. 3). Figure 3 shows the places of maxima, minima and points of inflection appearing on the voiceprint, which results from Burg's power spectrum estimation.



**Fig. 3.** The idea of a voiceprint. The acoustic image is extracted from Burg's model of estimation with the  $n$  feature points on it. The computations are performed in MATLAB.

These easily selected features help us understand the basic signal data, which are essential to classify the tested speech-signal. Therefore, this is the basic step to treat the speech-signal as an object-image and then apply the known image analysis methods for object

recognition [11]. Burg's model, together with the whole software implementation, is given in [2]. The method used for analyzing Burg's model and building the feature vector is Töeplitz matrices and their minimal eigenvalues.

### Töeplitz Matrices and their minimal eigenvalues

Therefore, after obtaining a better selection of the essential characteristic points of the signal-image by Burg's model through its estimation, the selected points enter the stage of description. This is achieved by the proposed algorithm of Töeplitz forms and their minimal eigenvalues as given in the following discussion.

In one of its practical applications, the coordinates  $(x_i, y_i)$  of the features in Fig. 3 are put into the rational function (3) to define its numerator and denominator:

$$f(s) = \frac{x_0 + x_1s + x_2s^2 + \dots + x_ns^n + \dots}{y_0 + y_1s + y_2s^2 + \dots + y_ns^n + \dots} \quad (3)$$

The number of the points is  $n$  and they are the points marked with circles in Fig. 3. From (3) Taylor series is evaluated in a simple way described in [12], [13]:

$$T(s) = \alpha_0 + \alpha_1s + \alpha_2s^2 + \dots + \alpha_ns^n + \dots \quad (4)$$

Then, Töeplitz matrices are formed from these coefficients:

$$[A_0] = \alpha_0 = \frac{x_0}{y_0},$$

$$[A_i] = \begin{bmatrix} \alpha_0 & \alpha_1 \\ \alpha_1 & \alpha_0 \end{bmatrix}, \dots$$

$$[A_i] = \begin{bmatrix} \alpha_0 & \alpha_1 & \dots & \alpha_i \\ \alpha_1 & \alpha_0 & \dots & \alpha_{i-1} \\ \dots & \dots & \dots & \dots \\ \alpha_i & \alpha_{i-1} & \dots & \alpha_0 \end{bmatrix} \quad (5)$$

for  $i = 1, \dots, n$

Then the minimal eigenvalues  $\lambda_{\min}^{(i)}$  of these Töeplitz matrices are evaluated for  $i = 1, 2, \dots, n$  in such a way that for each submatrix, the  $i^{\text{th}}$  minimal eigenvalue is computed. The resulting series of  $\lambda_{\min}^{(i)}$  is monotonically nonincreasing giving a very stable tool for image description (signal-image, here).

The stable feature vector (6) is hence formed from these Eigenvalues

$$\Phi_i = (\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_n) \quad (6)$$

It has experimentally been shown [8], [7] that in a class of voices an individual voice feature vector is quite unique and has its independent series of minimal eigenvalues in (6) among other signals within the tested class. This is practical and very efficient in small classes of voices. An example for that is to have a spoken password of 1-3 uttered digits [7]. Also, the emergency telephones usually have three digits that might be uttered to the phone set. Hence, and as a result, the voice is recorded in the destination office.

To summarize the discussion in this theoretical section, we would reach a conclusion that all the main and essential stages of the suggested system are shown in the block diagram of Fig. 4. It shows the procedure of the whole signal recognition system starting with the recorded signal acquisition and its preprocessing, ending with voiceprint extraction, description and hence classification for identification and recognition.

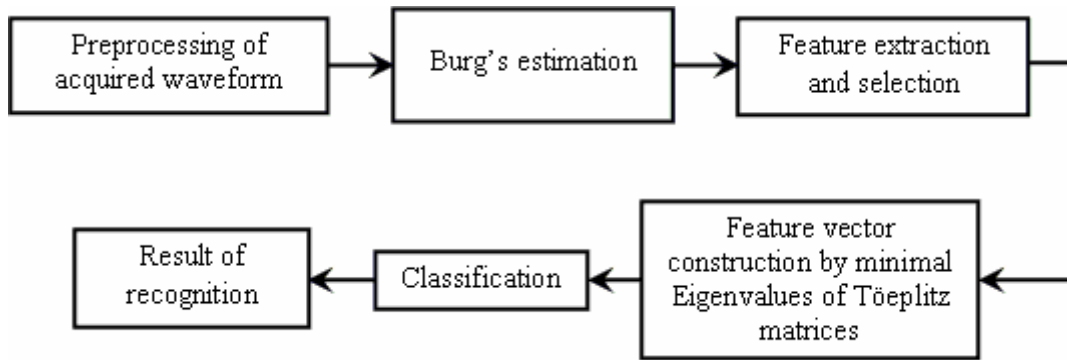


Fig. 4. Block diagram for the proposed voiceprint system

### 3 Feature vectors and NN classification – Examples and results

Although the minimal eigenvalues algorithm is itself a good classifier [11], [13], its basic role is when it acts as a descriptor and a feature extractor [3-9]. Therefore, when the features  $(x_i, y_i)$  are transformed into the feature vector in (6), the processing enters the stage of classification. Töeplitz matrices have proved very good performance with neural networks NN as classifiers. The success rate reached about 98% in *speaker identification* when classifying Töeplitz feature vectors by Radial Function neural networks [7]. That was why Töeplitz forms indicated their best performance in hybrid systems [14], [15].

To compare the results of classification of the presented system in *speech recognition*, we will consider two examples. Then, as a result of Example 1, TABLE I will show a comparison between Töeplitz-based processing and the conventional methods like projection [16]. The classification in all cases was by neural networks. Two kinds of neural networks are applied, namely, the radial and probabilistic ones.

#### Example 1 *Speech recognition* [8]

In a base of twenty recorded voices for twenty people from six different countries, the total number of recorded samples was 5472 divided into two groups. For each person and each voice, five samples were taken to be the test set (totally 1100 samples), while the remaining samples (4372 samples) were the teaching set. All possible combinations for the following values of Burg's parameters were considered in [2]:

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

Therefore, we experimented on 54 individual cases for each method of classification to study the *speech recognition*. TABLE I shows the approximate results giving the average success rate obtained for almost all the 54 combinations. Projection method is not mentioned in this work but has many applications in image recognition and also works very well with NN [16] [17], particularly in hybrid approaches [3]. In TABLE I, the projection approach is shown as a processing method, exactly as an image descriptor.

**TABLE I** SPEECH RECOGNITION RESULTS FOR EXAMPLE 1 WITH DIFFERENT CLASSIFICATION APPROACHES APPLYING THREE MAIN METHODS OF PROCESSING. BURG'S METHOD MEANS DIRECT CLASSIFICATION FROM IT TO THE CLASSIFIERS.

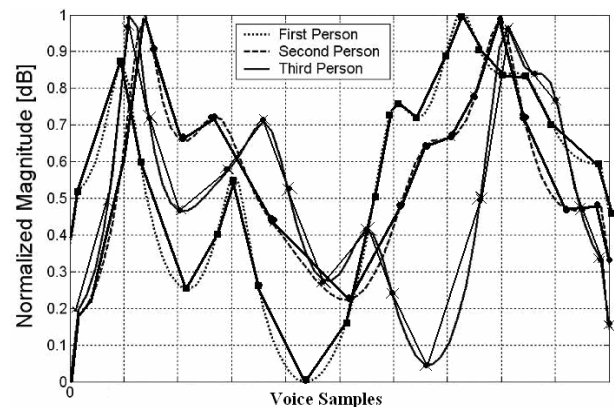
<i>Method of Processing</i>	<b>Recognition rate %</b>		
	<b>Conventional classification</b>	<b>Radial NN classification</b>	<b>Probabilistic NN classification</b>
<b>Burg's estimation and Töeplitz algorithm</b>	87	<b>95</b>	39
<b>Burg's estimation and Projection method</b>	60	41	57

#### Example 2 *Speaker recognition* [7]

Consider Fig. 5 showing the voiceprints of three people (three speakers) uttering the same word.

From Fig. 5, it can easily be noticed that the general shapes represent the same origin (voice), but they also show some small but sufficient differences proving they belong to different speakers.

Therefore, we have one word uttered by three different people and the question is how to identify the speakers from their voiceprints of Fig. 5 applying the approach considered in this paper. According to the proposed idea, the same graph is used to identify both the speaker and his speech by only knowing his voiceprint from his spoken word. The task may seem very difficult to implement on computer. The three prints of Fig. 5 are similar enough for the computer to decide.



**Fig. 5.** Three different estimated signals ready for extracting the voiceprint feature vectors of their speakers. The plots furnish the output of Burg's model. Performance was in MATLAB.

This was proved in [7] and was shown that they represent the same voice. On the other hand, as can be noticed from the figure itself, they seem to show sharp differences that allow distinguishing one from another. It is these visible differences that make it possible to differentiate one speaker from another. This system was accomplished and published in [7] as a new graphical treatment of acoustic signal identification for the purpose of speech and speaker recognition.

#### 4 Conclusions

This tutorial has mainly shown a general review of applying the algorithm of Töeplitz matrices and their minimal eigenvalues [11] as a feature extractor and descriptor of voiceprints created by linear predictive estimation in a method proposed by Burg [2]. The results of the experiments have proved the possibility of treating the signal-image as an object-image for processing and classification. This initial step in Image Analysis and Processing *may* open a new subject in Digital Voice-Signal Processing for the purpose of voiceprint identification and recognition. The "may" is because despite the presented in this and other works promising results, speech signal is still intricate to understand as an object-image and needs more more comprehensive study of the voice-print of every single phoneme to identify leading to perfect recognition. What is certain is that Töeplitz approach is proved to work successfully in hybrid systems of both processing and classification.

Although for the time being the aim is not commercial and has rather a research character, the author and his team have been working on a multi-factor verification system to use in a number of life aspects. Some of the examples are the mobile telephones in which the owner is verified by both his voice and fingerprints. Another interesting application is in the security systems where it is safer to have a password, which requires the recognition of a spoken word (or words) and its speaker. Many other aspects and applications are to be elaborated as a part of the future work of the author's team.

#### Acknowledgement

I would like to express my gratitude to professor Adam Dąbrowski for his kind invitation to give this tutorial. I would also like to thank my assistant and coauthor of many papers on voice analysis M. K. Nammous, for the varieties of experiments he conducted to verify my image analysis approaches on voice recognition.

This work was supported by the Rector of Białystok Technical University (grant no. W/WI/3/04).

#### References

- [1] Gray R. M., "Töeplitz and Circulant Matrices: A Review," Technical Report, *Stanford University Press*, 2000.
- [2] Ingle V.K., Proakis J.G., "Digital Signal Processing Using MATLAB," *Brooks Cole*, July 1999.
- [3] Saeed K., Tabędzki M. "Intelligent Feature Extract System for Cursive-Script Recognition," *4<sup>th</sup> IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology – IEEE-WSTST'05*, Muroran, Japan, 2005. In: Abraham A., Dote Y., Furuhashi T., Köppen, M., Ohuchi A. and Ohsawa Y. (Editors.), *Soft Computing as Transdisciplinary Science and Technology* (series: *Advances in Soft Computing*), Springer-Verlag Berlin Heidelberg, Germany, 2005, pp. 192-201.
- [4] Saeed K., Adamski M., "Extraction of Global Features for Off-line Signature Recognition," Intern. Conf. on *Advanced Computer Systems, Computer Information Systems and Industrial Management Applications – ACS-CISIM'05*, Ełk, Poland, 2005, pp. 429-436.
- [5] Saeed K.: Efficient Method for On-Line Signature Verification. Proc. Intern. Conf. on *Computer Vision and Graphics - ICCVG'02*, vol. 2, Zakopane, Poland, 2002, pp. 25-29.
- [6] Saeed K.: Minimal-Eigenvalue-Based Face Feature Descriptor. In: Dрамиński M., Grzegorzewski P., Trojanowski K., Zadrożny S. (Eds), *Issues in Intelligent Systems Models and Techniques. Institute of System Research*, Polish Academy of Sciences, Akademicka Oficyna Wydawnicza EXIT, Warsaw, Poland, 2005, pp. 185-196.
- [7] Saeed K., Nammous M., "A Simple Speech-and-Speaker Identification System," Accepted for publication in *IEEE Trans. on Industrial Electronics*, IEEE Computer Society, 2006.
- [8] Saeed K., Nammous M., "Heuristic Method of Arabic Speech Recognition," *Proc. 7<sup>th</sup> Intern. Conf. on Digital Signal Processing and its Applications – IEEE-DSPA'05*, Moscow, Russia, 2005, pp. 528-530
- [9] Saeed K., Kozłowski M., "An Image-Based System for Spoken-Letter Recognition. In: Petkov N. and Westenberg M. (Eds), *Lecture Notes in Computer Science – LNCS 2756*, Springer-Verlag Heidelberg, Germany, 2003, pp. 494-50.
- [10] Tadeusiewicz R., "Sygnał mowy," *WKiŁ*, Warsaw, 1988 (in Polish).
- [11] Saeed K., "Image Analysis for Object Recognition," *Publications of Białystok Technical University*, Poland, 2004.
- [12] Guillemin E. A., "A Summary of Modern Methods of Network Synthesis - Advances in Electronics,"

Vol.III, *Academic Press*, pp. 261-303, New York, 1951.

- [13] Saeed K., "Computer Graphics Analysis: A Criterion for Image Feature Extraction and Recognition," Vol. 10, Issue 2, 2001, pp. 185-194, *MGV - International Journal on Machine Graphics and Vision*, Institute of Computer Science, Polish Academy of Sciences, Warsaw.
- [14] Saeed K., AlBakoor M., "TMNN-Based Method for Arabic Character Recognition," Accepted for publication in *Computing and Informatics*, Slovak Academy of Sciences, 2006.
- [15] Saeed K., Tabędzki M., "A New Hybrid System for Recognition of Handwritten-Script," *Computing - Intern. Scientific Journal*, vol. 3, no. 1, Ternopil, Ukraine, 2004, pp. 50-57.
- [16] Burr D. J., "Experiments on Neural Net Recognition of Spoken and Written Text," *IEEE Transactions on Acoustic, Speech, and Signal Processing*, pp. 1162-1168, vol. 36, July 1988.
- [17] Saeed K., "A Projection Approach for Arabic Handwritten Characters Recognition," In: Sincak P. and Vascak J. (Eds), *Quo Vadis Computational Intelligence? New Trends and Applications in Computer Intelligence*, Physica-Verlag, Berlin, Germany, 2000, pp. 106-111.

---

This paper sample was published in the *Proceedings of the IEEE Scientific Workshop Signal Processing'2006*.