A STATISTICAL PATTERN RECOGNITION APPROACH TO ROBUST RECURSIVE IDENTIFICATION OF NON-STATIONARY AR MODEL OF SPEECH PRODUCTION SYSTEM

Milan Ž. Marković[1], Branko D. Kovačević[2] and Milan M. Milosavljević[1]

[1]*Institute of Applied Mathematics and Electronics, Kneza Miloša 37, 11000 Belgrade, Yugoslavia*
[2]*University of Belgrade, Faculty of Electrical Engineering, Bulevar Revolucije 73, 11000 Belgrade, Yugoslavia, Fax: (381-11) 324-8681*

ABSTRACT

In this work[1], we propose a robust recursive procedure based on WRLS algorithm with VFF and frame-based quadratic classifier for identification of non-stationary AR model of speech production system. Also, two versions of the frame-based quadratic classifier design procedure, iterative quadratic classifications procedure (CIQC) and its real-time modification (RTQC), are considered. A comparative experimental analysis is done according to the results obtained in analyzing speech signal with voiced and mixed excitation segments. Experimental results justify that two main problems of LPC speech analysis, non-stationarity of LPC parameters and non-appropriateness of AR modeling of speech (particularly on the voiced frames), can be solved by application of the proposed robust procedure. As for the comparison of CIQC and RTQC algorithm, it has been observed that superior results are obtained by using the proposed method with RTQC algorithm and it is recommended for use in the non-stationary AR speech model identification.

1. INTRODUCTION

The linear prediction analysis (LPC) of speech signal [1,2] is based on a linear model of a speech production system [3]. In a discrete time-domain, this model is given by:

$$s(k) + \sum_{i=1}^{p} a_i s(k-i) = e(k) \qquad (1)$$

where $s(k)$ is a speech sample, $\{a_i\}$ $(i=1,...,p)$ are the parameters of AR model (LPC parameters) of order p and $e(k)$ is a sample of speech excitation signal. In the conventional LPC analysis, LPC parameters are estimated by either autocorrelation or covariance method [2]. Both algorithms minimize the sum of squared residuals (a difference between a speech sample and its linear prediction) representing the least squares (LS) type algorithms. These algorithms are optimal if the excitation signal can be represented as innovation random process of white noise type.

However, there are two main problems in application of the conventional LPC methods. First problem consists in an inherent non-stationarity of AR model of the speech production system while second problem is in fact that speech excitation signal does not match the assumption of the white-noise type signal, particularly on the voiced speech frames. In the other words, the voiced speech production system is not adequately modeled by AR model.

In order to solve both of the above mentioned problems, we propose a robust recursive procedure with statistical pattern recognition approach for efficient identification of non-stationary AR speech model. The algorithm is based on weighted recursive least squares (WRLS) algorithm with variable forgetting factor (VFF) and a frame-based quadratic classifier of non-stationary signals. In this paper, the stress is posed on the aspects of the applied statistical pattern recognition method. Namely, the iterative quadratic classifications method [6] for design of the frame-based quadratic classifier of non-stationary signals [7] and its modified procedure for real-time applications [8] are considered.

Comparative experimental results, presented in the paper, justify that the proposed robust recursive method is superior to the non-robust WRLS algorithm with VFF in analyzing the real speech signal with voiced and mixed excitation frames. Also, a comparative analysis of two proposed procedures for the frame-based classifier design is presented and discussed.

The paper is organized as follows. Description of the proposed method is given in Section 2. Experimental analysis is presented in Section 3. Conclusions and summary are provided in Section 4.

2. DESCRIPTION OF THE ALGORITHM

The equation (1) can be rewriten in the linear regression form:

This work was supported by Ministry of Sciences and Technology of Republic Serbia, project No. 0403 through Institute of Mathematics SANU, Belgrade, and project No. 1007 through University of Belgrade, Faculty of Electrical Engineering.

0-7803-2431-5/95 $4.00 © 1995 IEEE

$$s(k) = Z^T(k)\theta + e(k) \quad (2)$$

where $\theta^T=\{a_1 \ \dots \ a_p\}$ is the vector of LPC parameters, and $Z^T(k)=\{s(k-1) \ \dots \ s(k-p)\}$ is the observations vector. As the non-recursive sliding window methods [4,5], the application of the WRLS algorithm with VFF represents a way for solving the problem of identification of non-stationary AR model of speech production system. Based on equation (2), the WRLS algorithm with VFF is given by:

$$\Gamma(k) = \frac{1}{\rho}\left[\Gamma(k-1) - \frac{\Gamma(k-1)\cdot Z(k)\cdot Z^T(k)\cdot\Gamma(k-1)}{\rho + Z^T(k)\cdot\Gamma(k-1)\cdot Z(k)}\right] \quad (3)$$

$$\theta(k) = \theta(k-1) + \Gamma(k)\cdot Z(k)\cdot\left[s(k) - Z^T(k)\cdot\theta(k-1)\right] \quad (4)$$

where $\Gamma(k)$ is the gain matrix and ρ is the variable forgetting factor VFF. The value of VFF less than one makes the WRLS algorithm adaptive to the non-stationarity of the estimated LPC parameters. In order to obtain the reliable estimates of the non-stationary LPC parameters, the value of VFF is determined at each time instances by using a modified generalized ratio (MGLR) algorithm [9], which enables fully automatic detection of the instants of abrupt changes in stationarity of a speech signal. In the other words, the value of VFF changes at each time instances according to the ammount of the estimated LPC parameters variability.

In order to solve the problem of non-appropriateness of AR modeling of speech production system, particularly on the voiced frames, we propose a procedure for robustification the WRLS algorithm with VFF based on the statistical pattern recognition approach. This procedure consists of the application of the frame-based quadratic classifier in a combined non-robust/robust recursive AR speech analysis procedure [8]. The non-robust procedure represents the WRLS algorithm with VFF given by equations (3) and (4) while the robust procedure is the same algorithm with variable factor, $\rho > 1$, which value changes according to the value of corresponding residual sample. In this heuristic procedure, the frame-based quadratic classifier of non-stationary signals is used to classify the residual speech samples into the two classes. The first class consists of "small" residual samples and the second one consists of "big" residual samples. The classification of the k-th residual sample selects either the non-robust (first class) or the robust (second class) recursive AR procedure for LPC parameters estimation at the k-time instance. This method is based on the well-known assumption of the excitation for voiced speech as innovative process from mixture distribution, such that a large portion of the excitations are from a normal distribution with a very small variance while a small portion of the glottal excitations are from an unknown distribution with a much bigger variance [4]. In this case, the classifier is a very simple, one-dimensional, and mean vectors and covariance matrices are means and variances, respectively. The classification consists of two steps: *initialization* and *adaptation*.

Initialization: On the initial frame of signal, the initial quadratic classifier is obtained applying the iterative quadratic classifications procedure [6] based on an initial partition that is heuristically chosen.

Adaptation: The initial classifier is then applied in the classification of the residual speech samples obtained by the proposed recursive AR speech analysis procedure on the next frame of signal with N samples size. The result of k-th residual sample classification selects either the non-robust recursive procedure or robust recursive procedure to estimate vector of AR parameters in the k-time instance. The obtained vector is used to determine the (k+1) residual sample and the procedure is continuing. The classification result of entire frame represents the initial partition of that frame and is used to start the proposed iterative clustering procedure (algorithm CIQC) or its real-time modification (algorithm RTQC) to produce the initial quadratic classifier for the next frame of N residual samples, and so on.

3. EXPERIMENTAL ANALYSIS

The efficiency of the proposed algorithm is elaborated according to the results in AR modeling of real speech signal. The signal consists of five isolately spoken vowels ("A", "E", "I","O","U") and ten isolately spoken digits ("1", "2", ... ,"0") from one speaker. The signal is sampled with f_S=10kHz and preemphasized with q=1. All experimental results are obtained by using AR model of 10th order.

As the objective quality measure, we used MAR (Mean Absolute Residual) criterion:

$$J = 1/M \cdot \sum_{i=1}^{M} |s(i) - \hat{s}(i)| \quad (5)$$

where $s(i)$ is the speech sample at the i-th instance, $\hat{s}(i)$ is its linear prediction, and M is total number of speech samples.

Tables 1 and 2 show means (E) and standard deviations (σ) of the MAR criterion values obtained through the analysis of the five vowels and ten digits by using the proposed robust recursive AR speech analysis procedure with CIQC and RTQC algorithm for the frame-based quadratic classifier design.

The values presented in Table 1 are calculated according to ten following values of N (length of speech segment): 50, 70, 90, 100, 150, 200, 250, 300, 350, and

400 speech samples, while, in the case of digits analysis, Table 2, the values are calculated according to five following values of N: 50, 100, 150, 200, and 300 speech samples.

Table 1 Means (E) and standard deviations (σ) of the MAR criterion values obtained in the vowels analysis

Vow.	Length	CIQC		RTQC	
		E	σ	E	σ
A	3690	49.58	0.314	50.04	0.619
E	3690	72.08	0.389	72.89	0.454
I	3690	39.78	0.341	39.45	0.563
O	3690	27.88	0.793	27.14	0.166
U	3690	10.81	0.248	10.37	0.044

Table 2 Means (E) and standard deviations (σ) of the MAR criterion values obtained in the digits analysis

Digits	Length	CIQC		RTQC	
		E	s	E	s
1	6690	35.42	1.177	34.33	0.236
2	6690	27.63	0.377	27.17	0.122
3	5690	32.70	5.759	27.36	1.743
4	6690	27.64	3.887	24.61	0.606
5	6690	20.35	2.189	17.32	0.244
6	7690	29.63	7.604	21.24	0.269
7	6690	37.70	3.045	33.25	0.626
8	7690	20.55	0.644	17.52	0.737
9	5690	39.24	0.672	37.80	1.189
0	5690	22.15	0.818	21.29	0.040

The other quality criteria that are considered in this paper are: a bias, a variance, and a sensitivity to the pitch impulses of AR parameter estimates obtained by using the proposed algorithms. Figures 1, 2, and 3 show the examples of the estimated trajectories of the first LPC parameter (AR_1) obtained by using the non-robust WRLS algorithm with VFF and two versions of the proposed robust recursive procedure (with CIQC and RTQC algorithms) in analyzing the vowel " A" and the digits " 1", and " 6", respectively. The estimated trajectories are compared to the reference parameter trajectory obtained by standard LPC sliding window method with the window length smaller than pitch period estimate. In case of the vowels analysis the sliding window length is NL=40 samples. In the case of digits analysis the sliding window length is NL=50 samples. The tops of this trajectory present the best parameter estimates due to LPC analysis window in these moments occupies the speech samples from closed-glottis period [5].

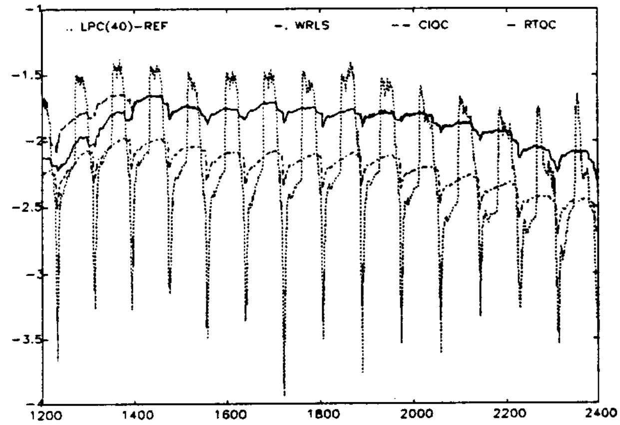

Figure 1: Trajectories of AR_1 parameter estimates obtained by using: LPC(40)-REF, WRLS, CIQC, and RTQC algorithms, in analyzing of vowel: "A".

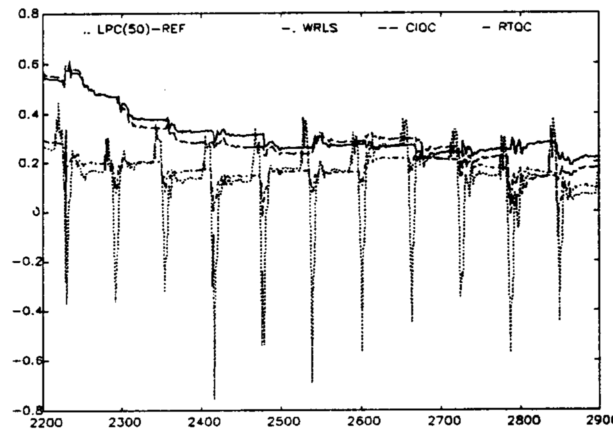

Figure 2: Trajectories of AR_1 parameter estimates obtained by using: LPC(50)-REF, WRLS, CIQC, and RTQC algorithms, in analyzing of digit: "1".

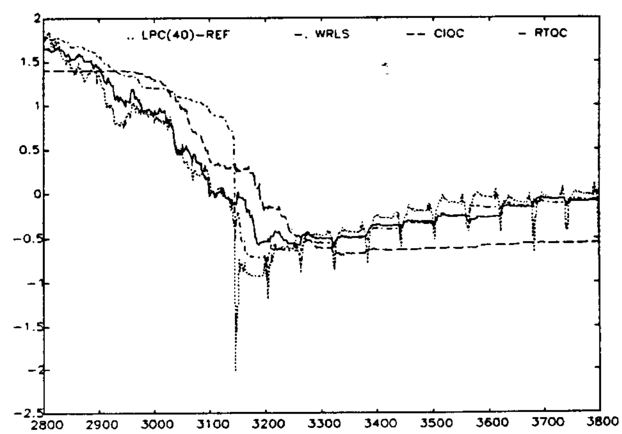

Figure 3: Trajectories of AR_1 parameter estimates obtained by using: LPC(50)-REF, WRLS, CIQC, and RTQC algorithms, in analyzing of digit: "6".

Based on the experimental results, Figures 1 and 2 are the examples, it can be concluded that the trajectories of AR_1 parameter estimates obtained by the two versions ot the proposed robust recursive AR speech analysis procedure (CIQC and RTQC algorithms) have lower bias, lower variance, and lower sensitivity to the pitch impulses than the non-robust recursive least squares procedure with variable forgetting factor (WRLS). As for comparison between CIQC and RTQC algorithms, it is concluded that the two algorithms have similar behaviour in case of the vowels analysis, Table 1 and Figure 1. But, in case of the digits analysis, Table 2 and Figures 2 and 3, better results are obtained by using RTQC version of the proposed robust recursive procedure. Even more, in some cases of digits analysis, Figure 3 is an example, the results obtained by application of CIQC algorithm are worse than ones obtained by the application of non-robust WRLS algorithm with VFF. An explanation of some difference in the CIQC algorithm behaviour in the analyzing vowels and digits could be described by the following. The vowels are example of speech signal with voiced frames. The digits are speech signal with voiced and mixed excitation frames. In the latter case, the two-class normal distribution model of speech excitation is not completely appropriate and this might be a reason why the results obtained by CIQC version of the proposed algorithm sometimes are not the best. In that case, the inherent robustness of RTQC procedure to the non-validity of the assumed model is clearly expressed.

4. CONCLUSION

In the paper, we introduce a robust recursive procedure for identification of non-stationary AR model of speech production system based on WRLS algorithm with VFF and frame-based quadratic classifier of non-stationary signals. The comparative experimental analysis is performed on the real speech signal: isolately spoken vowels and digits. Experimental results justify that two main problems of LPC speech analysis, non-stationarity of LPC parameters and limited validity of AR model of speech (particularly on the voiced frames), can be solved by using the method proposed in the paper. Namely, it has been observed that lower bias, lower variance, and lower pitch sensitivity of estimated parameter trajectories are obtained by the proposed method compared to the non-robust WRLS algorithm with VFF. Also, the comparative analysis of two versions, CIQC and RTQC, of the frame-based quadratic classifier design procedure is considered. It has been observed that both algorithms show similar results in the vowels analysis. But, in the case of the digits analysis, it has been observed that better results are obtained by using RTQC algorithm. Namely, RTQC algorithm shows more robustness to the non-validity of assumed two-class normal distribution classification model of speech excitation in case of the digits analysis. Based on the entire analysis, the proposed robust recursive procedure with RTQC algorithm for frame-based quadratic classifier design is recommended as a efficient solution of the problem of AR modeling of speech production system.

REFERENCES

[1] B.S.Atal, S.L.Hanauer: "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Amer.*, Vol. 50, pp. 637-655, 1971.

[2] J.D.Markel, A.H.Gray, Jr.: *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.

[3] G.C.Fant: *Acoustic Theory of Speech Prediction*, The Hague, The Netherlands: Monton, 1970.

[4] Chin-Hui Lee: "On robust Linear Prediction of Speech," *IEEE Trans. on ASSP*, Vol. 36, No. 5, May 1988., pp 642-650.

[5] M.Veinović, B.Kovačević, M.Milosavljević: "Robust Non-recursive AR Speech Analysis", *Signal Processing*, Vol. 37, No. 2, May 1994., pp. 189-201.

[6] K.Fukunaga: *Introduction to Statistical Pattern Recognition*, ACADEMIC PRESS INC., Harcourt Brace Jovanovich, Publishers, Boston-San Diego-New York-London-Sydney-Tokyo-Toronto, 1990.

[7] M.Marković, M.Milosavljević, B.Kovačević: "Clustering in Non-Stationary Pattern Recognition Systems," *SIGNAL PROCESSING VII: Theories and Applications, M. Holt, C. Cowan, P. Grant, W. Sandham (Eds.), 1994. European Association for Signal Processing (Proc. of EUSIPCO-94*, Edinburgh, Scotland, 13-16 September, 1994.) pp. 1803-1806.

[8] M.Marković: *The application of statistical pattern recognition methods in adaptive parameter estimation of AR speech model,* (in Serbian) Master Thesis, University of Belgrade, Faculty of Electrical Engineering, 1992.

[9] M.Milosavljević, I.Konvalinka: "The modified generalized likelihood ratio algorithm (MGLR) for automatic detection of abrupt changes in stationarity of signals," *Proc. of the Twenty-Second Annual Conference on Information Science and Systems*, Princeton, NY, 1988.