QUADRATIC CLASSIFIER WITH SLIDING TRAINING DATA SET IN ROBUST RECURSIVE IDENTIFICATION OF NON-STATIONARY AR MODEL OF SPEECH

Milan Marković
Institute of Applied Mathematics and Electronics
Kneza Miloša 37, 11000 Belgrade, Yugoslavia
e-mail: emarkovm@ubbg.etf.bg.ac.yu

ABSTRACT

In this work, a robust recursive procedure based on WRLS algorithm with VFF and a quadratic classifier with sliding training data set for identification of non-stationary AR model of speech production system is proposed. Experimental analysis is done according to the results obtained in analyzing speech signal with voiced and mixed excitation segments. Presented 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 using the proposed robust procedure.

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). 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 conventional LPC methods. First problem consists of an inherent non-stationarity of AR model of speech production system while second problem is in fact that speech excitation signal does not match the assumption of 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, robust recursive procedures with a frame-based statistical pattern recognition approach for efficient identification of non-stationary AR speech model are proposed in [4,5,6]. These algorithms are based on weighted recursive least squares (WRLS) algorithm with variable forgetting factor (VFF) and a frame-based quadratic classifier of non-stationary signals. Namely, the iterative quadratic classifications method [7] for design of the frame-based quadratic classifier of non-stationary signals, as proposed in [4], and its modified procedure for real-time applications, described in [5,6], are considered.

In order to simultaneously solve both of the main problems in conventional LPC speech analysis more efficiently, a robust recursive procedure based on WRLS algorithm with VFF and a quadratic classifier with sliding training data set is proposed. Presented experimental results 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 using quadratic classifier with sliding training data set and the real-time frame-based quadratic classifier of non-stationary signals, as the preceding procedure proposed in [5,6], 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:

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

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 [8,9], 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:

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

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

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 [10], 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, a procedure for robustification the WRLS algorithm with VFF based on the statistical non-stationary pattern recognition approach is proposed. This procedure consists of the application of quadratic classifier with sliding training data set in a combined non-robust/robust recursive AR speech analysis procedure [5]. 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 case, the value of ρ is heuristically determined by the expression: $\rho=1/(1-|r_{norm}|/2)$, where r_{norm} is the normalized value obtained by dividing the current residual with the maximal residual on the current frame. In addition, the maximal residual value is updated on the frame-by-frame manner. In this way, the algorithm assigns less weights to the large residuals, so that it is robust in the sense of its insensitivity to the spiky excitations on the voiced frames. In this heuristic procedure, the quadratic classifier with sliding training data set 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 [8]. 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 by applying the iterative quadratic classifications procedure [7] 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. 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. As a difference from the frame-based quadratic classifier of non-stationary signals which is adapted on a frame by frame basis [5,6], the parameter adaptation of quadratic classifier with sliding training data set is done after each speech sample. Namely, for given: N - sliding training data set length, ω_i, $i=1,...,c$; c classes of training data set described by corresponding parameter estimates, $M^{(i)}$ - mean vector of class i, $\Sigma^{(i)}$ - covariance matrix and $P^{(i)}$ - a priori probability of class i, a classified residual sample (*input sample*) in ($k+1$)-time instance is included to corresponding class (based on the classification result) while (k-N+1) residual sample (*output sample*) is excluded from belonging class. In this procedure, we differ the following four cases:

(1) input sample is classified into class *i*,

(2) output sample is belong to class *i*,

(3) input and output sample belong to class *i*, and

(4) neither of considered samples belong to class *i*.

For the cases (1), (2), and (3), the parameter adaptation of class *i* is necessary in order to update classifier for classification of residual sample in (*k*+2)-time instance. The adaptation is done by using recursive formulae (5)-(10) in which the parameter estimates with index *i* correspond to the case (1), the estimates with index *o* to the case (2), and the estimates with index *oi* to the case (3). These formulae are derived from the recursive formula for determination of the "leave-one-out" classifier parameter estimates, defined in [7].

$$M_i^{(i)}=\frac{1}{N_i+1}\left[\sum_{j=1}^{N_i}x_j^{(i)}+x_i^{(i)}\right]=M^{(i)}+\frac{1}{N_i+1}A \quad (5)$$

$$\Sigma_i^{(i)}=\frac{N_i-1}{N_i}\left[\Sigma^{(i)}-\frac{N_i}{\left(N_i+1\right)^2}AA^T\right] \quad (6)$$

$$M_o^{(i)}=\frac{1}{N_i-1}\left[\sum_{j=1}^{N_i}x_j^{(i)}-x_o^{(i)}\right]=M^{(i)}-\frac{1}{N_i-1}B \quad (7)$$

$$\Sigma_o^{(i)}=\frac{N_i-1}{N_i-2}\left[\Sigma^{(i)}-\frac{N_i}{\left(N_i-1\right)^2}BB^T\right] \quad (8)$$

$$M_{oi}^{(i)}=\frac{1}{N_i}\left[\sum_{j=1}^{N_i}x_j^{(i)}+x_i^{(i)}-x_o^{(i)}\right]=M^{(i)}+\frac{1}{N_i}\left(A-B\right) \quad (9)$$

$$\begin{aligned}\Sigma_{oi}^{(i)}&=\Sigma^{(i)}+\left(A-B\right)\left(A-B\right)^T\Big/\left(N_i\left(N_i-1\right)\right)+\\&\left(\left(N_i-1\right)A+B\right)\left(\left(N_i-1\right)A+B\right)^T\Big/\left(N_i^2\left(N_i-1\right)\right) \quad (10)\\&-\left(A-\left(N_i+1\right)B\right)\left(A-\left(N_i+1\right)B\right)^T\Big/\left(N_i^2\left(N_i-1\right)\right)\end{aligned}$$

where: x_i - input sample, x_o - output sample, $A=x_i^{(i)}-M^{(i)}$, $B=x_o^{(i)}-M^{(i)}$ and N_i is a number of residual samples in class *i*. Also, the a priori class probability is adapted by using the following formulae: $P_i^{(i)}=P^{(i)}+1/N_i$; $P_o^{(i)}=P^{(i)}-1/N_i$; $P_{oi}^{(i)}=P^{(i)}$.

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, the MAR (Mean Absolute Residual) criterion is used [5]:

$$J = 1/\mathrm{M} \cdot \sum_{i=1}^{\mathrm{M}} |s(i) - \hat{s}(i)| \qquad (11)$$

where $s(i)$ is the speech sample at the i-th time 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 five vowels and ten digits by using the proposed robust recursive AR speech analysis procedure with two versions of quadratic classifier design procedure (with the quadratic classifier with sliding training data set (QCSTS) and the real-time frame-based quadratic classifier (RTQC) [5,6]).

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

		RTQC		QCSTS	
V	Length	E	σ	E	σ
A	3690	50.04	0.619	48.51	0.564
E	3690	72.89	0.454	72.74	0.445
I	3690	39.45	0.563	38.84	0.246
O	3690	27.14	0.166	27.02	0.080
U	3690	10.37	0.044	10.34	0.049

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

		RTQC		QCSTS	
D	Length	E	σ	E	σ
1	6690	34.33	0.236	35.28	0.394
2	6690	27.17	0.122	27.32	0.372
3	5690	27.36	1.743	26.67	1.105
4	6690	24.61	0.606	24.52	0.294
5	6690	17.32	0.244	17.57	0.726
6	7690	21.24	0.269	21.18	0.209
7	6690	33.25	0.626	33.89	0.695
8	7690	17.52	0.737	17.34	0.228
9	5690	37.80	1.189	37.19	0.179
0	5690	21.29	0.040	21.27	0.104

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.

The other quality criteria that are considered in this paper are: bias, variance, and sensitivity to the pitch impulses of AR parameter estimates obtained by using the proposed algorithms. Figures 1, and 2 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, QCSTS and RTQC, of the proposed robust recursive procedure in analyzing the vowel “A” and the digit “1”, 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.

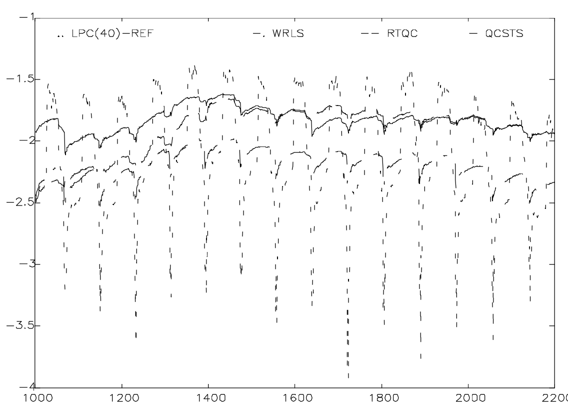

Figure 1: Trajectories of AR_1 parameter estimates obtained by using: LPC(40)-REF, WRLS, RTQC, and QCSTS algorithms, in analyzing of vowel: “A”.

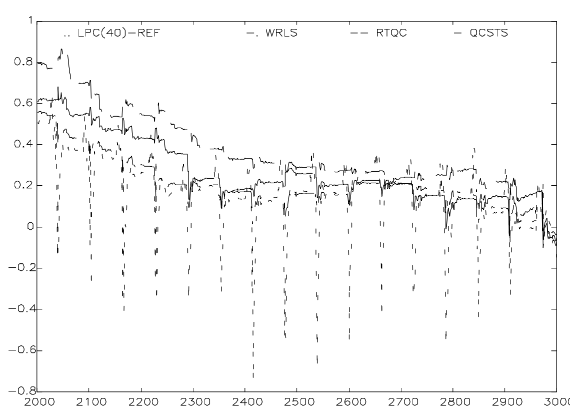

Figure 2: Trajectories of AR_1 parameter estimates obtained by using: LPC(50)-REF, WRLS, RTQC, and QCSTS algorithms, in analyzing of digit: “1”.

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 [9].

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 (with QCSTS 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 with VFF). As for comparison between RTQC and QCSTS algorithms, Tables 1 and 2 and Figures 1 and 2, it can be concluded that better results are obtained by using QCSTS version of the proposed robust recursive procedure, both for vowels and for digits analysis. Namely, the trajectories of AR_1 parameter estimate, obtained by using the proposed robust procedure with QCSTS algorithm, have lower bias than trajectories obtained by the same procedure with RTQC algorithm. Also, the results obtained by using the proposed robust procedure with QCSTS algorithm are better in sense of both MAR objective criterion (values of E in Tables 1 and 2) and sensitivity to the changing of sliding training data set length N (values of σ in Tables 1 and 2).

Based on the entire analysis, the proposed robust recursive procedure with QCSTS algorithm for the quadratic classifier design and adaptation is recommended as a efficient solution of the problem of AR modeling of speech production system.

4. CONCLUSION

In the paper, a robust recursive procedure for identification of non-stationary AR model of speech production system based on WRLS algorithm with VFF and a quadratic classifier with sliding training data set is introduced. The 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, RTQC and QCSTS, of the proposed robust procedure is considered. It has been observed that better results are obtained by using QCSTS algorithm. Based on the entire analysis, the proposed robust recursive procedure with QCSTS algorithm for quadratic classifier design and adaptation is recommended as a efficient solution of the problem of AR modeling of speech production system.

ACKNOWLEDGEMENTS

This work was supported by Ministry of Sciences and Technology of Republic Serbia, project No. 401, through Institute of Mathematics SANU, Belgrade.

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] 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.

[5] 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.

[6] M.Marković, B.Kovačević, M.Milosavljević, "A Statistical Pattern Recognition Approach to Robust Recursive Identification of Non-Stationary AR Model of Speech Production System," *In Proc. of ICASSP-95*, Detroit, USA, May 8-12, 1995.

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

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

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

[10] 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.