ON ROBUST AR SPEECH MODEL IDENTIFICATION BASED ON QUADRATIC CLASSIFIER WITH HEURISTICALLY DECISION THRESHOLD

Milan Marković

Institute of Applied Mathematics and Electronics
Kneza Miloša 37, 11000 Belgrade, Yugoslavia
e-mail: markovic@kiklop.etf.bg.ac.yu

Abstract

In this work, a robust recursive procedure for identification of nonstationary AR speech model based on a quadratic classifier with heuristically decision threshold is considered. Two versions of the robust procedure based on frame-based quadratic classifier and quadratic classifier with sliding training data set and heuristically decision threshold are evaluated and compared through analyzing natural speech signal with voiced and mixed excitation segments. Obtained results show that the considered robust procedure with the quadratic classifier with sliding training data set and heuristically decision threshold achieves more accurate AR speech parameter estimation, provides improved tracking performance, and achieves better discrimination capabilities for possible application in some vowel recognition systems.

1. Introduction

Linear prediction coding (LPC) of speech signal [1] is based on a linear model of speech production system:

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

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, the LPC parameters are estimated by either autocorrelation or covariance method [1]. These algorithms are optimal if the excitation signal is an innovation random process of white Gaussian noise type. However, there are two main problems in application of the conventional LPC methods. The first problem consists of an inherent nonstationarity of speech production system, while the second problem is, in fact, that the speech excitation does not match the assumption of white Gaussian noise, particularly on the voiced speech frames. In the other words, the AR model cannot adequately represent the voiced speech.

In order to solve both of the above mentioned problems, robust recursive procedures with a statistical pattern recognition approach for efficient identification of nonstationary AR speech model are proposed in [2,3,4,5]. These algorithms are based on the weighted recursive least squares (WRLS) algorithm with variable forgetting factor (VFF) and a frame-based quadratic classifier of nonstationary signal [2,3,4], as well as a quadratic classifier with sliding training data set [5]. The basis of the mentioned procedures is the assumption of two-class nature of the voiced speech excitation, such that a large part of the excitation is from a normal distribution with a very small variance while a small part is also from the normal distribution but with a much bigger variance. Since the quadratic classifier is the optimal classifier in the case of Gaussian distributed data, this approach is emphasized compared to other robust nonrecursive [6,7] and recursive [8] speech analysis approaches based on Huber's M-estimation theory, as well as robust recursive procedure based on Kalman filtering [9].

In order to improve tracking performance of the considered robust recursive algorithm, procedures based on the frame-based quadratic classifier with a heuristically decision thresholds are proposed in [10,11,12], instead of the optimal decision threshold used in [2,3,4]. Also, a procedure based on the quadratic classifier with sliding training data set and heuristically decision threshold is proposed in [13], instead of the optimal decision threshold used in [5]. Comparative experimental analysis, presented in [10,11,12,13], justifies the use of the heuristically decision thresholds in the proposed robust procedure.

In this paper, a comparative analysis of the two considered robust procedures with heuristically decision threshold, based on the frame-based quadratic classifier, proposed in [12], and the quadratic classifier with sliding training data set, proposed in [13], is presented. The parameter estimation accuracy of the considered robust procedures is evaluated in order to determine the most appropriate solution regarding the criterion of simultaneous solving both of the main problems of the AR speech analysis. The considered procedures are evaluated through analyzing natural speech signal with voiced and mixed excitation frames. Besides, we consider a possibility of applying the Bayes error estimates for evaluating the proposed parameter estimation procedures of nonstationary AR speech model in vowel recognition systems. The Bayes error estimates obtained by a very reliable k-NN procedure [14] are used for such purpose. Presented experimental results have shown that the considered robust recursive procedure based on quadratic classifier with sliding training data set and heuristically decision threshold achieves more accurate AR speech parameter estimation, provides improved tracking performance, and achieves better discrimination capabilities for possible application in some vowel recognition systems.

2. Robust estimation procedure

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

$$s(k) = Z^T(k)\theta + e(k) \tag{2}$$

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. Similarly as the nonrecursive sliding window methods [6,7], the application of WRLS algorithm with VFF represents a way for solving the problem of identification of nonstationary AR model of speech production system. Based on the equation (2), the WRLS algorithm with VFF is given by:

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

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

where $\Gamma(k)$ is the gain matrix and ρ is VFF. The value of VFF less than one makes the WRLS algorithm adaptive to the nonstationarity of LPC parameters. In order to obtain the reliable estimates of nonstationary LPC parameters, the value of VFF is determined at each time instances by the modified generalized likelihood ratio (MGLR) algorithm [15,16], which enables fully automatic detection of the instants of abrupt changes in stationarity of speech signal. In the other words, the value of VFF changes at each time instances according to the amount of LPC parameter variability, which is expressed through the value of the MGLR discrimination function [8,16].

In order to solve the problem of inappropriateness of AR modeling of speech production system, particularly on the voiced frames, a procedure for robustifying the WRLS algorithm with VFF by applying nonstationary pattern recognition method is proposed [2,3,4]. This procedure consists of the application of frame-based quadratic classifier in a combined nonrobust/robust recursive AR speech analysis scheme. The nonrobust procedure represents the WRLS algorithm with VFF ($\rho<1$) given by the equations (3) and (4). In this case, as mentioned before, ρ changes at each time steps according to the value of MGLR discrimination function [8,16]. On the other hand, the robust procedure is the WRLS algorithm with variable factor ρ which 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 this way, the algorithm assigns less weight 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 heuristically procedure, the frame-based quadratic classifier of nonstationary 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 "large" residual samples. The classification of the *k*-th residual sample selects either the nonrobust (first class) or the robust (second class) recursive AR procedure for LPC parameter estimation at the *k*-th time instance. In this case, the classifier is very simple, i.e. it is one-dimensional, and the mean vectors and covariance matrices are reduced to the means and variances, respectively.

Instead of the quadratic classifier with optimal decision threshold, as described in [2,3,4], procedures based on heuristically decision thresholds are considered. Namely, as it is shown in [14], the use of the optimal decision threshold in Bayes error estimation procedure when the limited training data set is available is highly inefficient, and a heuristically determination of the decision threshold is suggested. The problem of the limited training data set is also presented in the application of the quadratic classifier in AR speech parameter estimation and, thus, similar procedures for determination of the decision thresholds are proposed. The following values of heuristically decision thresholds are proposed: *Thresh1*=*FHP*$*\sigma_1$ [10], *Thresh2*=$(\sigma_1+\sigma_2)/2$ [11], and *Thresh3*=σ_2 [12], where σ_1 and σ_2 represent frame-based adaptive standard deviations of the first and second class, respectively, and *FHP* is a heuristically determined adaptive factor (*FHP*=3 is used in [10]). Comparative experimental results, presented in [10,11,12], justify the use of the frame-based quadratic classifier with heuristically decision threshold in the considered robust recursive AR speech identification procedure. Based on the evaluation analysis from [12], the most accurate results are obtained by using the robust procedure with the heuristically decision threshold, *Thresh3*=σ_2 [12].

As a difference from the frame-based quadratic classifier of nonstationary signals, which is adapted on a frame by frame basis, the parameter adaptation of the quadratic classifier with sliding training data set is done after each speech sample [5,13]. The adaptation is done by using recursive formulae, given in [5,13], which are derived from the recursive formula for determination of the "leave-one-out" classifier parameter estimates, defined in [14]. In order to further enhance the robustness and adaptivness of the proposed robust procedure, instead of using the optimal threshold [5], the use of the quadratic classifier with sliding training data set and heuristically decision threshold is proposed in [13], with the value of *Thresh*= *FHP*$*\sigma_1$, where *FHP*=1.5. Experimental analysis, presented in [13], justifies the use of the quadratic classifier with sliding training data set and proposed heuristically decision threshold in the robust recursive AR speech identification procedure.

3. Experimental analysis

In this Section, the comparative experimental analysis of using the robust procedures based on the frame-based quadratic classifier with heuristically decision threshold, proposed in [12] (denoted as RTQCH procedure), and the quadratic classifier with sliding training data set and heuristically decision threshold, proposed in [13] (QCSTSH procedure), is presented. The efficiency of the considered algorithms is tested on natural speech signal. The signal consists of five isolately spoken Serbian 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 [7]. All experimental results are obtained by using the 10th order AR model. As the objective quality measure, we used MAR (Mean Absolute Residual) criterion [3]: $J = 1/M \sum_{i=1}^{M} |s(i) - \hat{s}(i)|$, 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. In Table 1, MAR criterion values obtained by the proposed robust procedures, as well as the nonrobust WRLS algorithm with VFF, in analyzing vowels and digits are presented. In this case, the N=100 samples is used as the length of the training data set both for the frame-based quadratic classifier, and the quadratic classifier with sliding training data set. According to the results presented in Table 1, it could be concluded that the best results are obtained by using QCSTSH procedure.

Table 1: MAR criterion values: vowels and digits.

T	WRLS with VFF	RTQCH	QCSTSH
A	52.34	50.43	49.29
E	74.56	73.05	71.98
I	39.72	39.81	39.88
O	27.87	27.04	26.97
U	10.40	10.32	10.66
1	36.94	34.07	34.54
2	28.56	27.46	27.26
3	28.78	26.72	26.37
4	27.07	24.46	24.09
5	19.39	17.25	17.28
6	22.28	21.14	21.17
7	34.99	34.22	32.91
8	17.89	17.38	17.28
9	38.72	37.45	37.19
0	21.96	21.39	21.04

The other quality criteria that are considered in this paper are bias, variance, and sensitivity to the intensity of pitch impulses of the estimated AR speech parameters obtained by using the considered robust procedures. Fig. 1, and 2 show the typical examples of estimated trajectories of the first LPC parameter (AR_1), obtained by using RTQCH, and QCSTSH algorithms in analyzing digits "7" and "8", respectively. The estimated trajectories are compared with the reference parameter trajectory obtained by standard LPC sliding window method with the window length smaller than the pitch period estimate. The local tops of this trajectory present the best parameter estimates, due to the fact that the LPC analysis window in these moments occupies the speech signal samples from the closed-glottis period [7].

Based on the experimental results, presented in Fig. 1 and 2, it could be concluded that the trajectories of AR_1 parameter estimates, obtained by the QCSTSH procedure, possess improved tracking characteristics, lower bias, lower variance, lower sensitivity to the intensity of pitch impulses, and more adaptivity to the abrupt changes of the LPC parameters than the RTQCH algorithm. This is also justified in analyzing the rest of the test examples.

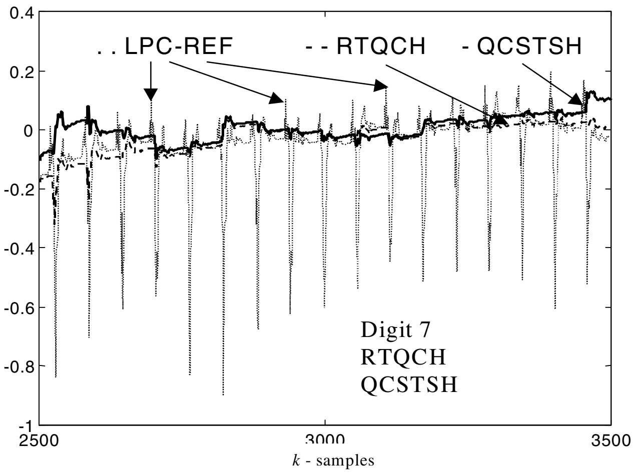

Figure 1: Trajectories of AR_1 parameter estimates obtained by using: LPC-REF, RTQCH, and QCSTSH algorithms in analyzing of digit: “7”.

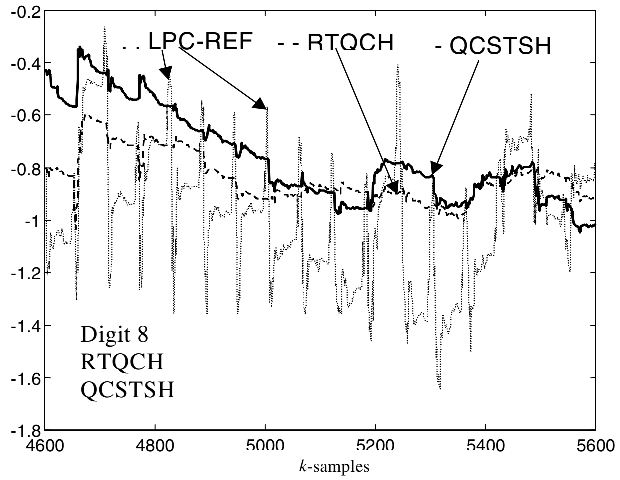

Figure 2: Trajectories of AR_1 parameter estimates obtained by using: LPC-REF, RTQCH, and QCSTSH algorithms in analyzing of digit: “8”.

As an another criterion for evaluating the proposed methods, we consider the estimated discrimination capabilities of the RTQCH and QCSTSH procedures related to the isolated spoken vowel recognition problem. In this sense, we estimate 1000 LP vectors from each of the five isolated spoken Serbian vowels, “A”, “E”, “I”, “O”, and “U”, by using the proposed estimation procedures. Then, we estimate Bayes error on the given training data set by using a very reliable estimation procedures based on the *k*-NN approach, combining the resubstitution and leave-one-out estimation methods [14]. The obtained Bayes error estimates are presented in Table 2. Actually, we have used only the second (II) and the third (III) option of the Bayes error estimation procedure, proposed in [14].

Table 2: Bayes error estimates obtained in vowel recognition problem

Procedure	Bayes error estimates (%)	
	II option	III option
WRLS with VFF	0.52	0.57
RTQCH	0.14	0.16
QCSTSH	0	0

Based on the results presented in Table 2, we can conclude that better results are obtained by using the QCSTSH algorithm. In the other words, better discrimination capabilities of the AR speech parameter estimates for possible applications in the vowel recognition systems are obtained by using the QCSTSH procedure.

4. Conclusion

In this paper, robust recursive procedures for parameter estimation of nonstationary AR model of speech production system based on the WRLS algorithm with VFF and quadratic classifier with heuristically decision threshold are elaborated. The comparative experimental analysis of the two procedures with heuristically decision thresholds, based on the frame-based quadratic classifier and the quadratic classifier with sliding training data set, is performed on the natural speech signal, isolately spoken vowels and digits. Experimental results show that the robust procedure based on the quadratic classifier with sliding training data set gives better results regarding the criterion of simultaneous solving of the two main problems of LPC speech analysis, the nonstationarity of LPC parameters and limited validity of AR model of speech, particularly on the voiced frames. Namely, it has been observed that lower bias, lower variance, lower sensitivity to the pitch impulses and better tracking characteristics of the estimated AR parameter trajectories, as well as better discrimination capabilities in vowel recognition problems, are obtained by the quadratic classifier with sliding training data set and the heuristically decision threshold.

5. References

[1] J.Markel, A.Jr.Gray, *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.
[2] M.Marković, M.Milosavljević, B.Kovačević, “Clustering in Non-Stationary Pattern Recognition Systems,” in *Proc. of EUSIPCO-94*, pp. 1803-1806.
[3] 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.
[4] M.Marković, B.Kovačević, M.Milosavljević, “A Statistical Pattern Recognition Approach to Robust Recursive Identification of Non-Stationary AR Model of Speech Production System,” *IEEE Trans. on Speech and Audio Proc.*, Vol. 4, No. 6, Nov. 1996, pp. 456-460.
[5] M.Marković, “Quadratic Classifier with Sliding Training Data Set in Robust Recursive Identification of Nonstationary AR Model of Speech,” in *Proc. of EUSIPCO-96*, Sept. 13-16, 1996, pp. 491-494.
[6] C.-H.Lee, “On robust Linear Prediction of Speech,” *IEEE Trans. on ASSP*, Vol. 36, No. 5, May 1988., pp 642-650.
[7] M.Veinović, B.Kovačević, M.Milosavljević, “Robust Non-recursive AR Speech Analysis,” *Signal Processing*, Vol. 37, No. 2, May 1994., pp. 189-201.
[8] B.Kovačević, M.Milosavljević, M.Veinović, “Robust recursive AR speech analysis,” *Signal Processing*, Vol. 44, 1995, pp. 125-138.
[9] T.Yang, J.-H.Lee, K.-Y.Lee, K.-M.Sung, “On robust Kalman filtering with forgetting factor for sequential speech analysis,” *Signal Processing*, Vol. 63, 1997, pp. 151-156.
[10] M.Marković, “Quadratic Classifier with Heuristically Decision Threshold in Robust Recursive AR Speech Analysis,” *Proc. of ECMCS’99*, 1999, pp. 28-41.
[11] M.Marković, “On Possible Applications of Quadratic Classifier with Heuristically Decision Threshold in Robust AR Speech Analysis,” *Proc. of WISP’99*, Budapest, 4-7 September, 1999, pp. 236-241.
[12] M.Marković, “On determining heuristically decision threshold in robust AR speech model identification procedure based on quadratic classifier,” in *Proc. of Fifth Intern. Symp. on Signal Proc. and its Applications, ISSPA'99*, Brisbane, 1999, pp. 131-134.
[13] M.Marković, “Robust Recursive AR Speech Analysis Based on Quadratic Classifier with Sliding Training Data Set and a Heuristic Decision Threshold,” in *Proc. of EUSIPCO-2000*, Sept. 4-8, 2000, Finland.
[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, ACADEMIC PRESS INC., 1990.
[15] M.Milosavljević, I.Konvalinka, “The modified generalized likelihood ratio algorithm (MGLR) for automatic detection of abrupt changes in stationarity of signals,” in *Proc. of the Twenty-Second Annual Conference on Information Science and Systems*, Princeton, NY, 1988.
[16] M.Marković, M.Veinović, M.Milosavljević, B.Kovačević, “On parameter estimation of nonstationary AR model of speech,” in *Proc. of DSPA’98*, Moscow, 1998, Vol. I, pp. I-190 – I-197.