# AR Speech Model Parameter Estimation Using a Robust Non-Recursive Estimator

Ana Lazović, Mihailo Bjekić and Aleksandra Marjanović

*Abstract*—**In this paper a robust block linear prediction (RBLP) method for autoregression (AR) model parameter estimation of speech signal is considered. The considered method consists of two separate iterative steps which are described in detail. The method is tested on both synthesized and natural human speech in the presence of outliers and additive measurement noise. A comparative analysis of the RBLP method and the conventional linear prediction method shows that the robust method gives results which are less biased and have a smaller variance.**

*Index Terms*—**AR model, outliers, RBLP, estimation, speech.**

## I. Introduction

Statistical signal processing is frequently based on the assumption of a priori known probability data distribution, stationarity, linearity as well as independence of stochastic processes [1]. In most engineering problems, real system model parameter estimation methods assume that stochastic processes have Gaussian distribution, which is, in most cases, justified. The assumption of Gaussian distribution allows for a straightforward implementation of optimal estimators. On the other hand, in the case of large realizations of stochastic perturbations, as well as non-Gaussian additive measurement noise, the assumption is not justified. Optimal estimation methods based on the assumption of Gaussian distribution are extremely sensitive to deviations from the assumed Gaussian distribution, where, in some cases, even a small deviation from the assumed distribution results in significant deterioration in estimator performance.

Even though in many cases, the Gaussian distribution is justified, there are many situations where it has been determined that the actual data distribution is far from Gaussian. One of the main reasons for the distribution deviations are the impulse perturbations, i.e. outliers.

As it can be seen in [2], outliers can be found in various areas, such as image processing, speech signal processing, sensor networks, medicine, industry, etc. Because of the prevalence of outliers in different areas, outlier detection and elimination are of great importance.

Based on the fact that the conventional linear prediction (CLP) methods assume the Gaussian distribution of the residuals, CLP methods assign equal weights to all residuals [3]. If the residual distribution varies from the Gaussian distribution, the results of the CLP methods may be inaccurate with both a large bias and variance. Due to the unsatisfactory results, robustification must be performed.

There are two different approaches to estimator robustification, the diagnostic approach and statistically robust approach [4]. Diagnostic approach detects outliers and replaces them using classical parameter estimation, as discussed in [5]. The statistically robust approach uses the entire dataset and bounds the influence of outliers using influence functions.

In this paper, robust estimation of autoregression (AR) model of speech signal is discussed. Due to the impulse-type quasiperiodic excitation of the vocal tract, distribution of the residuals varies from the Gaussian distribution. This variation from the Gaussian distribution is considered to be a result of the presence of outliers, so the performance of CLP methods decreases.

Robust parameter estimation techniques are efficient even without the a priori known statistical characteristics of the perturbations in the system [6]. Robust estimation techniques are less sensitive to the presence of impulse perturbations in the system, that is, robust estimation techniques enable outlier detection in the signal and its elimination.

Specific application of outlier detection in AR parameter estimation of speech signal can be found in [7], where a voice-based Parkinson's disease detection is considered.

In this paper, robust block linear prediction (RBLP) method for an AR parameter estimation of speech signal, described in detail in [8], is analyzed. The algorithm in [8] is improved, especially the second step of the algorithm. The outliers considered in [8] originate only from the impulse-type quasiperiodic excitation, whereas, in this paper, outliers originating from the measurement errors, also referred to as additive outliers (AO) [4], will be analyzed. The effects of measurement noise, as well as the effects of different types of nonlinearities are considered. The RBLP method will be tested on both synthesized and natural voiced signals in the presence of additive measurement noise.

## II. Estimation problem

The AR model of the speech signal [9], considered in this paper, shown in the Fig. 1, is given in the form:

$$E(z)G(z)V(z)L(z) = \frac{E(z)}{A(z)} = S(z), \qquad (1)$$

Ana Lazović is a PhD student at the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: analazovic1995@gmail.com).

Mihailo Bjekić is a PhD student at the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: mihailobjekic@gmail.com).

Aleksandra Marjanović is with the School of Electrical Engineering, University of Belgrade, 73 Bulevar kralja Aleksandra, 11020 Belgrade, Serbia (e-mail: amarjanovic@etf.bg.ac.rs).

where $E(z)$ is the $z$ transform of the excitation of the glottal tract, $G(z)$ is the transfer function of the glottal tract, $V(z)$ is the transfer function of the vocal tract and $L(z)$ is the transfer function of the radiation on the lips. Combined transfer function of $G(z)V(z)L(z)$ can be represented as an inverse filter given by:

$$A(z) = \frac{1}{G(z)V(z)L(z)}. \qquad (2)$$

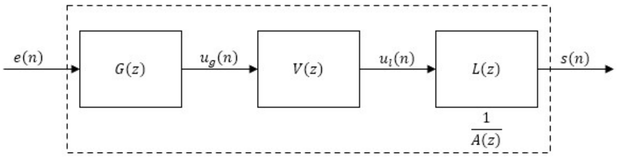


Fig. 1. A linear model of the speech production system.

In this paper, two impulse-type quasiperiodic excitations of the AR model, shown in the Fig. 2, are considered. First type of excitation is the train of Dirac pulses, while the second type of excitation is the twice differentiated Strube's glottal wave [10], which is a more accurate representation of the speech signal excitation. Both excitations are normalized and periodic with the period equal to the fundamental period of speech.

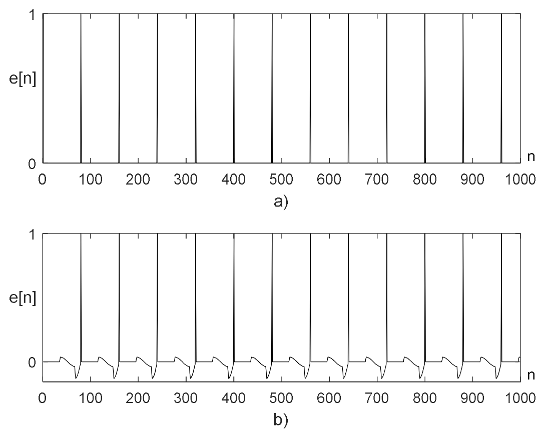


Fig. 2. Impulse-type quasiperiodic excitation of the AR speech model with the fundamental period of 8 ms with the sampling frequency of 10 kHz: a) Dirac pulse train; b) twice differentiated Strube's glottal wave.

AR model of the speech signal of the $p$-th order, in the time domain, is given by (3).

$$s(n) + \sum_{i=1}^{p} a_i s(n-i) = e(n) \qquad (3)$$

In the case of the $N + p$ known speech signal samples $s(n)$, (3) can be rewritten in the matrix form:

$$S = H\theta + E, \qquad (4)$$

where $S$ is a vector of the speech signal samples with the length of $N$, $\theta$ is a vector of coefficients $a_i$ of the AR model with the length of $p$, $E$ is a vector of excitation samples $e(k)$ with the length of $N$ and $H$ is a matrix of speech signal observations with the dimensions $N \times p$.

$$S^{\mathrm{T}} = [s(k+1)\, s(k+2) \ldots s(k+N)] \qquad (5)$$

$$\theta^{\mathrm{T}} = [a_1\, a_2 \ldots a_p] \qquad (6)$$

$$E^{\mathrm{T}} = [e(k+1)\, e(k+2) \ldots e(k+N)] \qquad (7)$$

$$H = \begin{bmatrix} -s(k) & \cdots & -s(k+1-p) \\ -s(k+1) & \cdots & -s(k+2-p) \\ \vdots & \ddots & \vdots \\ -s(k+N-1) & \cdots & -s(k+N-p) \end{bmatrix} \qquad (8)$$

One of the CLP methods for speech parameter estimation is the least-squares (LSQ) method [8]. In the presence of outliers, LSQ method does not give satisfactory results. The reason for decreased performance of the LSQ method lies in the fact that the LSQ method is based on the assumption of the Gaussian data distribution and assigns equal weights to all samples including outliers. For the purpose of overcoming the problems that the LSQ method encounters, robustification is proposed.

## III. RBLP METHOD

One of the robust estimators used for estimating AR model parameter coefficients is the approximate maximum likelihood estimator, shortly M-estimator [10]. M-estimators are based on the approximation of the unknown probability density function, and its corresponding nonlinear score function, which reduces the effects of outliers.

The minimization problem whose solution is a robust M-estimation of the parameter vector $\theta$ is defined by:

$$J_N(\hat{\theta}) = \sum_{i=1}^{N} \rho \left[\frac{s_i - h_i^T \hat{\theta}}{d}\right], \qquad (9)$$

where $s_i$ is the $i$-th element of $S$, $h_i$ is the $i$-th row of $H$, $N$ is the dimension of $S$, $d$ is the scaling factor and $\rho$ is the nonlinear score function [8].

The minimization problem can be reformulated into:

$$\sum_{i=1}^{N} \frac{1}{d} h_{ij}\, \psi \left[\frac{s_i - h_i^T \hat{\theta}}{d}\right] = 0, \qquad j = 1, \ldots, p, \qquad (10)$$

where $h_{ij}$ is the element in the $i$-th row and the $j$-th column of $H$, and $\psi$ is the influence function $\psi(\cdot) = \rho'(\cdot)$.

In this paper, a non-recursive RBLP method for AR parameter estimation is considered [8]. The proposed RBLP method starts from the estimates obtained using one of the CLP methods, such as the LSQ method, and further consists of two separate iterative steps.

### A. Dutter algorithm

The first step of the non-recursive RBLP is based on M-estimation [10]. The used nonlinear score function is Huber's score function:

$$\rho(x) = \begin{cases} \frac{x^2}{2}, & |x| \le k \\ k|x| - \frac{k^2}{2}, & |x| > k \end{cases}, \qquad (11)$$

where $k$ ensures the desired efficiency in the case of nominal Gaussian distribution. The appropriate influence function is given by:

$$\psi(\mathrm{x}) = \begin{cases} x, & |x| \le k \\ k \ \mathrm{sign}(x), & |x| > k \end{cases}. \tag{12}$$

Because the solution of the system (10) cannot be obtained in the closed form, Dutter iterative procedure is used [10]. For the initial guess $\theta_0$ of the unknown vector $\theta$, result of the LSQ algorithm is used, while the initial scaling factor $d_0$ is obtained by (13).

$$d_0 = \mathrm{median}\left(\frac{|s_i - \mathrm{median}(s_i)|}{0.6745}\right) \tag{13}$$

Dutter iterative method consists of several steps:

Step 1: Calculation of the nonnormalized residual:

$$n_i(\hat{\theta}_0) = s(i) - h_i^T \hat{\theta}_0. \tag{14}$$

Step 2: Calculation of a new estimation of the scaling factor $d_1$:

$$d_1^2 = \frac{1}{(N-p)E\{\psi^2(z)\}} \sum_{i=1}^{N} d_0^2 \psi^2[\varepsilon_i(\hat{\theta}_0)], \tag{15}$$

where the normalized residual is given by:

$$\varepsilon_i(\hat{\theta}_0) = \frac{n_i(\hat{\theta}_0)}{d_0}, \tag{16}$$

and $E\{\psi^2(z)\}$ is the mathematical expectation for the standard normal random variable:

$$E\{\psi^2(z)\} = \int_{-\infty}^{\infty} \psi^2(z) p(z) dz, \tag{17}$$

where $p(z)$ is the probability density function of the standard Gaussian random variable $z$ ($N(0,1)$). For the adopted parameter $k = 1.5$, (17) results in 0.7785.

Step 3: Residual Windsorization:

$$\Delta_i = \begin{cases} n_i(\hat{\theta}_0), & |\varepsilon_i(\hat{\theta}_0)| \le k \\ kd_1, & \varepsilon_i(\hat{\theta}_0) > k \\ -kd_1, & \varepsilon_i(\hat{\theta}_0) < k \end{cases}. \tag{18}$$

Step 4: Calculation of the regression coefficients $\theta$ using LSQ estimation:

$$\Delta\theta = [H^{\mathrm{T}} H]^{-1} H^{\mathrm{T}} \psi_v; \ \psi_v = \{\Delta_1, \dots, \Delta_N\}. \tag{19}$$

Step 5: Update of the estimation of the parameter $\theta$:

$$\hat{\theta}_1 = \hat{\theta}_0 + q\Delta\theta, \tag{20}$$

where $q$ is the correction factor defined by:

$$q = \min\left[\frac{1}{2\,erf(k)}, 1.9\right]. \tag{21}$$

Step 6: Repetition of the steps 1-5 with the new estimations $\hat{\theta}_1$ i $d_1$ as the initial guesses until the fulfillment of the termination conditions:

$$\left|\hat{\theta}_1^{(m)} - \hat{\theta}_0^{(m)}\right| < \eta \left|\hat{\theta}_0^{(m)}\right|, \tag{22}$$

$$|d_1 - d_0| < \eta \left|\hat{\theta}_0^{(m)}\right|, \tag{23}$$

where $\eta > 0$ is a conveniently chosen small number and $\hat{\theta}^{(m)}$ is the $m$-th element of the vector $\hat{\theta}$.

### B. *Weighted Least Squares Algorithm*

The second step of the RBLP method, considered in [8], is the Weighted Least Squares Algorithm (WLSQ). The estimations obtained by the Dutter algorithm are used as the initial guesses for the iterative WLSQ algorithm, which further reduces the effects of outliers.

In this paper an improvement of the WLSQ algorithm, given in [8], is presented.

The minimization problem (10), in the case of the WLSQ algorithm, is given in the form:

$$\sum_{i=1}^{N} h_{ij} w_{i0} \left(s(i) - h_i^{\mathrm{T}} \hat{\theta}\right) \approx 0; j = 1,2, \dots, p, \tag{24}$$

where coefficients $w_{i0}$ are given by:

$$w_{i0} = \begin{cases} \dfrac{\psi\left(\frac{s(i) - h_i^{\mathrm{T}}\hat{\theta}_0}{d_0}\right)}{\frac{s(i) - h_i^{\mathrm{T}}\hat{\theta}_0}{d_0}}, & s(i) \ne h_i^{\mathrm{T}}\hat{\theta}_0 \\ 1, & s(i) = h_i^{\mathrm{T}}\hat{\theta}_0 \end{cases}, \tag{25}$$

where $\hat{\theta}_0$ and $d_0$ are the initial estimations obtained using the Dutter algorithm.

Matrix notation of the system is given by:

$$W_0 = diag\{w_{10}, \dots, w_{N0}\}; S^{\mathrm{T}} = \{s(1), \dots, s(N)\}. \tag{26}$$

The new estimation of the vector of parameters $\hat{\theta}$ is calculated using:

$$\hat{\theta} = (H^{\mathrm{T}} W_0 H)^{-1} H^{\mathrm{T}} W_0 S. \tag{27}$$

The WLSQ algorithm is repeated using the new estimation $\hat{\theta}$ as the initial guess for the vector parameter $\theta$. The scaling factor $d_0$ changes with each iteration, and its new value is calculated using (15) with an appropriate $E\{\psi^2(z)\}$, which is the main improvement of the algorithm given in [8] The algorithm is repeated until the terminal condition (22) is satisfied.

The influence functions that can be used in this step of the RBLP algorithm are either Andrews (28) or Tukey (29) nonlinearity [10]. The influence functions are shown in Fig. 3.

$$\psi(x) = \begin{cases} \sin(x/a), & |x| < a\pi \\ 0, & |x| > a\pi \end{cases}, \ a \in [0.45\ 0.65] \tag{28}$$

$$\psi(x) = \begin{cases} x(1-(x/a)^2), & |x| < a \\ 0, & |x| > a \end{cases}, \ a \in [1.5\ 2] \tag{29}$$

The influence functions (28) and (29) can lead to divergence of the algorithm because of their non-convexity. By limiting the number of the iterations of the WLSQ algorithm to only a few, the problem of the non-convex influence functions can be overcome, while the quality of the estimation does not deteriorate.

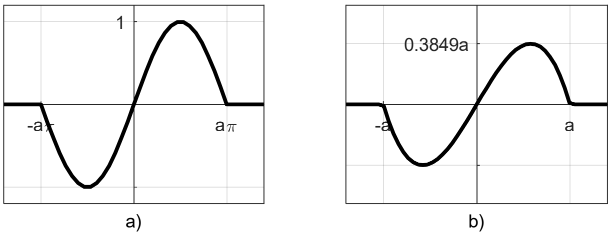


Fig. 3. Nonlinear influence functions $\psi(\cdot)$: a) Andrews nonlinearity; b) Tukey nonlinearity.

## IV. Results and Discussion

In this paper, the RBLP method is tested on both synthesized and natural voiced signal "a" in the presence of additive measurement noise. For the purpose of testing the RBLP method, estimations of the AR model parameters are obtained by performing both a CLP method and the RBLP method at the sliding rectangular window with the length of 256 samples.

In the second step of the RBLP method, Andrews nonlinearity, with the parameter $a = 0.5$ and $E\{\psi^2(z)\} = 0.5617$, is used, while the number of iterations is limited to four iterations.

### A. *Synthesized vowel "a"*

The considered AR model of the order $p = 8$ is the same as in [8], with the parameters: $a_1 = -2.22$, $a_2 = 2.89$, $a_3 = -3.08$, $a_4 = 3.27$, $a_5 = -2.77$, $a_6 = 2.35$, $a_7 = -1.7$ and $a_8 = 0.75$. The excitation of the AR model is described in detail in the second section. The output of the AR model is noised by additive Gaussian noise $N(0{,}0.0002)$ and contaminated with outliers. One of the most common types of outliers in time series, such as the speech signal, are the AO [11], which are the type of outliers used in this model. The outliers are generated as a random variable with the Gaussian distribution $N(0{,}0.02)$ with the probability of occurrence of 5% [4].

In the Fig. 4, results of the RBLP method for parameter $a_1$, performed on noised and non-noised synthesized vowel "a", in the case of train of Dirac pulses excitation (Fig. 2(a)), are shown.

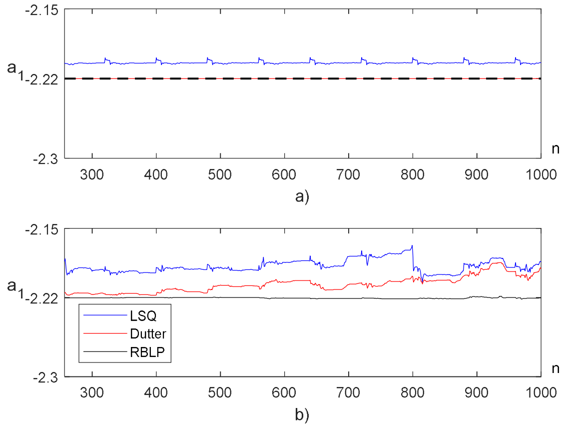


Fig. 4. Comparative analysis of the CLP and the RBLP methods performed on synthesized vowel "a", parameter $a_1$ estimates for: a) non-noised b) noised train of Dirac pulses excitation.

In the Fig. 4(a) the results of the RBLP method without both additive Gaussian noise and outliers, are shown. For signals as simple as this, it can be seen that the first step of the RBLP method gives satisfactory results that are both unbiased and insensitive to the position of the sliding window, unlike the results of the LSQ algorithm. In the Fig. 4(b) the same type of excitation is used, but in the presence of both additive Gaussian noise and outliers. In the presence of noise and outliers, the importance of the second step of the RBLP method can be seen. The WLSQ algorithm ensures an unbiased estimation with a smaller variance than the first step of the RBLP method.

In the Fig. 5 the results of the RBLP method for the parameter $a_1$, in the case of an excitation that is twice differentiated Strube's glottal wave (Fig. 2(b)), are shown.

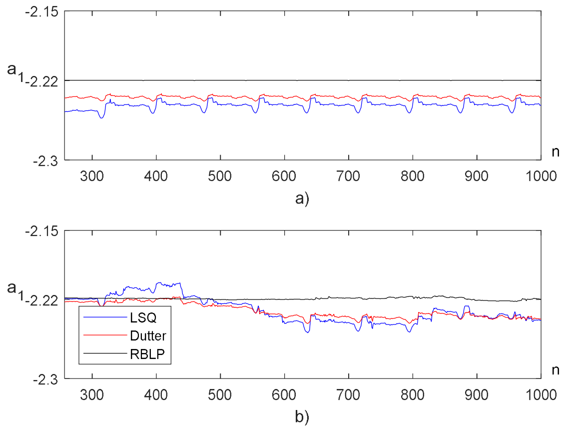


Fig. 5. Comparative analysis of the CLP and the RBLP methods performed on synthesized vowel "a", parameter $a_1$ estimates for: a) non-noised b) noised twice differentiated Strube's glottal wave excitation.

When there is no noise nor outliers (Fig. 5(a)), in the contrast to the case when the excitation is the train of Dirac pulses (Fig. 4(a)), the first step of the RBLP method does not give good enough results. The unsatisfactory results of the Dutter method, which reflect in the presence of a large bias in the parameter estimation, are improved using the WLSQ algorithm, whose results are the same as in (Fig. 4(a)). In the presence of noise and outliers (Fig. 5(b)) the variance of the estimation of the parameters is bigger than in the previous case, while the estimation is still unbiased.

In the Fig. 6 the results of the RBLP method, for the twice differentiated Strube's glottal wave excitation, with noise, for both constant and variable scaling factor $d$ in the WLSQ algorithm, are shown. It can be seen that the result of the RBLP algorithm with variable scaling factor has both smaller variance and bias, which is the main improvement of the algorithm in [8].

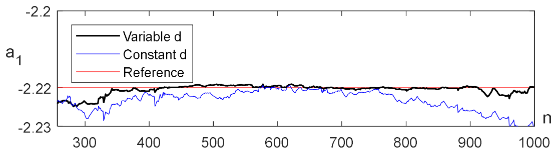


Fig. 6. Comparative analysis of the RBLP with constant and variable scaling factor in the second step of the algorithm.

### B. *Natural human spoken vowel "a"*

The algorithm was tested on the natural human spoken vowel "a" which was recorded with the sampling frequency of 10 kHz and filtered with a low-pass filter with the cut-off frequency of 4 kHz. Because the exact values of the parameters of the AR model are not a priori known, for the reference values an estimation provided by the CLP method, performed on the sliding window with the length shorter than the fundamental period, is adopted [8]. When the sliding window for the CLP method does not contain the part of the signal originating from the excitation, the estimation is considered unbiased. For the length of the sliding window of the CLP method, 37 samples are adopted, as the fundamental period of the test signal is 47 samples. The adopted order of the AR model is $p = 8$.

The filtered signal is given in Fig. 7(a), while the estimates for the parameter $a_1$ are given in Fig. 7(b). For the reference values of the parameter $a_1$, results of the CLP method, with a shorter window length, when there is no effect of the impulse type excitation, are adopted. In the Fig. 7(b), there is no effect of the impulse type excitation when the CLP estimates are at their higher values.

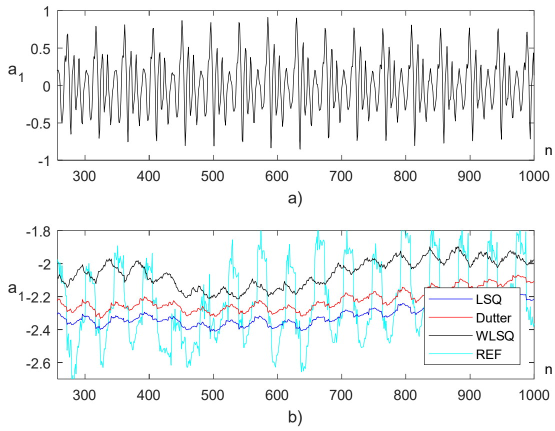


Fig. 7. Comparative analysis of the CLP and the RBLP methods performed on natural human spoken vowel "a": a) Filtered spoken vowel "a"; b) $a_1$ parameter estimates for the CLP and the RBLP methods.

It can be seen that the estimates provided by the Dutter algorithm are less biased than the starting LSQ estimates, while the WLSQ algorithm gives significantly less biased estimates.

## V. Conclusion

In this paper, an improvement to the RBLP algorithm, described in detail in [8], has been proposed. The main improvement of the algorithm is in the second step, the WLSQ algorithm, where it is proposed that the scaling factor changes with each iteration. The experimental results given in [8] have been confirmed. The algorithm has also been tested to the presence of the additive measurement noise as well as outliers, and it has been shown that the RBLP method is much more robust than the CLP methods. In the case of the natural human speech, the results of the RBLP method have a significantly smaller bias than the results of the CLP methods. Overall, the RBLP method proves to be much better for estimating the AR model parameters of the speech signal for both synthetized and natural human speech in the presence of outliers.

## Acknowledgment


This research was partially supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Contracts No. TR32038.


## References


[1] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh and M. Muma, "Robust Estimation in Signal Processing: A tutorial style treatment of fundamental concepts", *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, July, 2012.

[2] K. Singh and S. Upadhyaya, "Outlier detection: applications and techniques", *International Journal of Computer Science Issues*, vol. 9(1), no. 3, pp.307–323, January, 2012.

[3] K.Y. Lee, B. G. Lee, I. Song and S. Ann, "Robust Estimation of AR Parameters and Its Application for Speech Enhancement," in IEEE International Conference on Acoustics, Speech and Signal Processing, San Francisco, USA, vol. 1, pp. 309–312, March, 1992.

[4] M. Muma and A. M. Zoubir, "Bounded Influence Propagation $\tau$-Estimation: A New Robust Method for ARMA Model Estimation", *IEEE Transactions on Signal Processing*, vol. 65, no. 7, pp. 1712–1727, April, 2017.

[5] A. Azad and L. Mili, "Robust Speech Filter And Voice Encoder Parameter Estimation using the Phase-Phase Correlator", *IEEE Transactions/ACM on Audio, Speech, and Language Processing*, vol. 28, pp. 592–604, 2020.

[6] B. Kovačević, Z. Banjac i Ž. Đurović, *Filtracija stohastičkih signala – Optimalni, adaptivni i robusni estimatori parametara i stanja*, Beograd, Srbija: Akademska misao, 2017.

[7] A. H. Poorjam, M. S. Kavalekalam, L. Shi, Y. P. Raykov, J. R. Jensen, M. A. Little and M. G. Christensen, "Automatic Quality Control and Enhancement for Voice-Based Remote Parkinson's Disease Detection", IEEE Access, 2020.

[8] M. Veinović, B. Kovačević and M. Milosavljević, Robust Non-recursive AR Speech Analysis, *Signal Processing*, vol. 37, pp. 189–201. 1994.

[9] D. Bajić, M. Veinović, B. Kovačević i M. Milosavljević, „Praćenje formantnih trajektorija primenom robusnih metoda linearne predikcije", u ETRAN, Zlatibor, Srbija, str. 493–496, Jun, 1995.

[10] B. Kovačević, M. Milosavljević, M. Veinović and M. Marković, *Robust Digital Processing of Speech Signals*, Berlin, Germany: Springer, 2017.

[11] A. S. Ahmar, S. Guritno, Abdurakhman, A. Rahman, Awi, Alimuddin, I. Minggi, M. A. Tiro, M. K. Aidid, S. Annas, D. U. Sutiksno, D. S. Ahmar, K. H. Ahmar, A. A. Ahmar, A. Zaki, D. Abdullah, R. Rahim, H. Nurdiyanto, R. Hidayat, D. Napitupulu, J. Simarmata, N. Kurniasih, L. A. Abdillah, A. Pranolo, Haviluddin, W. Albra and A. N. M. Arifin, "Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO)", *Journal of Physics: Conference Series*, vol. 954, March, 2018.