DOI 10.1007/s11018-023-02150-8
Measurement Techniques, Vol. 65, No. 10, January, 2023

IMPROVING THE METHOD FOR MEASURING THE ACCURACY INDICATOR OF A SPEECH SIGNAL AUTOREGRESSION MODEL

V. V. Savchenko

UDC 53.082.4; 004.934

The problem of determining the accuracy of an autoregressive model of a speech signal is considered, and a method for measuring the accuracy index in the sliding observation window mode is proposed. As an indicator of accuracy, we used a modified value of the COSH-distance (hyperbolic cosine) of the autoregressive model relative to the eponymous (single phoneme) Schuster periodogram as a reference spectral sample. To study the possibilities of the proposed method, a full-scale experiment was set up and carried out, in which the object of study was a set of autoregressive models of different orders. These models were obtained by Berg's method for the vowel speech sounds of a test speaker. According to the results of the performed measurements for each vowel, the optimal values of the autoregressive order and the corresponding optimal autoregressive model were found. It is shown that this optimization made it possible to increase the accuracy of the autoregressive model of the speech signal by more than 60%, depending on the sound of the test speaker's speech and the characteristics of his vocal tract. The results obtained are intended for use in automatic processing and digital speech transmission systems with radical data compression based on linear prediction coefficients.

Keywords: *acoustic measurements, speech acoustics, vocal tract, speech signal, autoregressive model, small sample problem.*

Introduction. The autoregression model of the speech signal is widely used in systems of automatic processing and transmission of speech over digital communication channels to encode speech information with radical data compression [1, 2]. In conditions of a priori uncertainty of the thin structure of the speech signal, the autoregression model is adapted to it in the mode of a sliding observation window [3, 4]. At the same time, the length τ of the window is strictly limited from above by the length of the intervals of stationarity of the speech signal [5]. As a result, in practice, the problem of small samples often arises [6, 7]. It manifests itself in the insufficient accuracy of the autoregressive model used, associated with an unreasonable choice of order [8]. The severity of this problem becomes especially obvious, given that in the theory there is no strict criterion of the accuracy of the autoregressive model, and the existing criteria, for example, Akaike, Schwartz, etc. [9], are not very suitable for application to the speech signal, since they are designed for homogeneous ergodic processes. Therefore, the task of measuring the accuracy index of the autoregression model on a sample of finite volume $N = \tau F$, where F is the sampling rate of observations.

The purpose of the study is to develop methodological foundations for optimizing autoregression models of the speech signal, taking into account its variability and features of the speaker's vocal tract [10, 11]. The article is written in the development of the results of previous works [2, 3].

Problem statement. Since the autoregressive model is directly related to the linear power spectrum envelope of the speech signal [12, 13], we will define this spectrum within the speech frame $x(n)$, where $n = 0, 1, \dots, N-1$. When choosing the method of spectral analysis, we will follow the logic of the work [7, 14] and give preference to the discrete Fourier transform as a kind of linear signal processing that is not related to the known effects of parametric estimates of the power spectrum: displacements and splitting of spectral lines in conditions of small sample observations. Based on the discrete Fourier transform, Schuster's periodogram [15]:

$$G_x\left(f_m\right) = (FM)^{-1}\left|\sum_{n=0}^{M-1} x(n)\exp\left(-\,\mathrm{j}\,2\pi nmM^{-1}\right)\right|^2, \quad m = 0, 1, \dots, M-1 \tag{1}$$

National Research University Higher School of Economics, N. Novgorod, Russia; e-mail: vvsavchenko@yandex.ru. Translated from Izmeritel'naya Tekhnika, No. 10, pp. 58–63, October, 2022. Original article submitted June 30, 2022. Accepted August 9, 2022.

defines an instantaneous power spectrum estimate (PSE) signal on a selected set of discrete frequencies $f_m = m\Delta f = mM^{-1}F$. If $M = 2^k > N$, where k is some integer, the periodogram is designed for the use of fast Fourier transform algorithms [16]. In this case, only the first N of M samples of the time series $\{x(n)\}$ on the right side of (1) are nonzero. For example, with the duration of the speech frame τ = 20–30 ms and the sampling rate F = 8 kHz (coordinated with the bandwidth of a standard telephone communication channel) where k = 10, the dimension will be $M = 1024 \gg N = 160–240$.

According to the methodology of parametric spectral analysis, we define the autoregressive model in the frequency domain by the general expression [17]:

$$G_p\left(f_m\right) = \sigma_p^2 T \left|1 - \sum_{i=1}^{p} a_p(i) \exp\left(-\mathrm{j}2\pi imM^{-1}\right)\right|^{-2}, \tag{2}$$

where $T = F^{-1}$; σ_p^2 is a scale factor; $a_p(i)$ is the ith element of the vector coefficient of the pth order linear autoregression. The parameters σ_p^2 and $a_p(i)$ are determined from the sample $\{x(n)\}$ according to the spectral analysis method used. For example, in the Berg method, for these purposes, a recursive computational procedure of the form [15]:

$$\begin{gathered}\forall q = \overline{1,\ p}: \quad a_q\left(i\right) = a_{q-1}\left(i\right) + c_q a_{q-1}\left(r - i\right),\ i = \overline{1,\ q}; \quad c_q = \frac{2\sum\limits_{n=q}^{N-1} \eta_{q-1}\left(n\right) \nu_{q-1}\left(n-1\right)}{\sum\limits_{n=q}^{N-1}\left[\eta_{q-1}^2\left(n\right) + \nu_{q-1}^2\left(n-1\right)\right]}; \\ \eta_q\left(n\right) = \eta_{q-1}\left(n\right) + c_q \nu_{q-1}\left(n-1\right); \quad \nu_q\left(n\right) = \nu_{q-1}\left(n-1\right) + c_q \eta_{q-1}\left(n\right),\ n = 0, 1, \ldots, N-1\end{gathered} \tag{3}$$

is used when it is initialized by the system of equations

$$\forall n < N: \nu_0\left(n\right) = \eta_0\left(n-1\right) = x\left(n\right).$$

The final values of the recursion (3) together with (2) determine the autoregression model of the spectral envelope $\overline{G}_x(f_m)$ of the speech signal. The accuracy of this model depends on its order. Here, the scale multiplier σ_p^2 does not play any role because it has nothing to do with the shape of the spectral envelope. Therefore, the task is to determine the accuracy index of the autoregressive model (2) depending on the order p and its subsequent measurement regime using a sliding window of observations $\{x(n)\}$.

Accuracy index. We rewrite expression (2) in the form of an inverse transformation

$$G_p(f_m) = \sigma_p^2 T \left|A_p(\mathrm{j}f_m)\right|^{-2}$$

of the square of the modulus of the complex transmission coefficient $A_p(\mathrm{j}f_m)$ of the pth order transverse filter

$$A_p(\mathrm{j}f_m) = 1 - \sum_{i=1}^{p} a_p(i) \exp\left(-\mathrm{j}2\pi imM^{-1}\right), \ m = \overline{0;\ M-1}. \tag{4}$$

This filter is used to equalize in the frequency domain the envelope $\overline{G}_x(f_m)$ of the Schuster periodogram (1), which in this case acts as a reference spectral sample [3]. The signal at the output of the leveling filter

$$G_y\left(f_m;\ p\right) = \left|A_p\left(\mathrm{j}f_m\right)\right|^2 G_x\left(f_m\right) = \sigma_p^2 T \frac{G_x\left(f_m\right)}{G_p\left(f_m\right)}, \quad m = \overline{0; M-1} \tag{5}$$

determines the PSE of the sequence $\{y(n)\}$ of impulses that excite the vocal tract of a test speaker within the framework of the inverse task of speech formation [18]: from the speech signal to its voice source.

Ideally, PSE (5) is a rectangular sequence $K = 0.5FF_0^{-1}$ of harmonic components with a shift between themselves by the frequency of the fundamental pitch F_0 [19]. However, in practice, the shape of the spectral envelope $\overline{G}_y(f_m;\ p)$ can differ significantly from the rectangular one, primarily due to the imperfection of the autoregression model used (2). From this point of view, the order of the model is of fundamental importance. We optimize the value of p within the finite set (R-set) of alternatives $p_1 < p_2 < \ldots < p_R$.

The optimization is based on the obvious understanding: the higher the accuracy of the autoregressive model, the more pronounced in PSE (5) are the harmonic components and the closer the shape of their envelope $\{\overline{G}_y(f_m;\ p)\}$ to the rectangular G_0 = const will be for all $m < M$. When taking this into account, we will use as a criterion function a modified measure of COSH distance (COSH, hyperbolic cosine) ρ(p) with the property of scale invariance in the frequency domain [20]:

$$\begin{aligned}\rho\,(p) &= \sqrt{\left[M^{-1}\sum_{m=0}^{M-1} G_0\overline{G}_y^{-1}\,(f_m;\ p)\right]\left[M^{-1}\sum_{m=0}^{M-1}\overline{G}_y\,(f_m;\ p)\,G_0^{-1}\right]} - 1 \\ &= M^{-1}\sqrt{\left[\sum_{m=0}^{M-1}\overline{G}_y^{-1}\,(f_m;\ p)\right]\left[\sum_{m=0}^{M-1}\overline{G}_y\,(f_m;\ p)\right]} - 1 \geq 0\,.\end{aligned} \tag{6}$$

Measure (6) is identically zero if the spectral envelope is equal to an arbitrary constant: $\overline{G}_y(f_m;\ p) = G_0\ \forall m < M$. As follows from the results of the work [3], it is in this version of the PSE (5) that the accuracy of the autoregression model (2) will be maximum. With an increase in ρ(p), the accuracy decreases. Measure (6) can then be considered an objective indicator of the accuracy of the generated autoregressive model. Hence follows this decision-making criterion in the problem of optimizing the model order on the set R of the alternatives under consideration:

$$p^* = \operatorname{Arg}\min_{\forall r \leq R} \rho\,(p)\Big|_{p\ =\ p_r}\,. \tag{7}$$

The optimal model satisfies expression (2) when you select its order $p \leq p_R$ according to rule (7). Practical implementation is reduced to the measurement of indicator (6) in the mode of the sliding window of observations.

Method for measuring the accuracy index. The structural diagram of the device for measuring the accuracy index (6) and optimizing the order of the autoregressive model (2) is presented in Fig. 1. The device contains series-connected elements: analog-to-digital ADC transformer; FB frame builder; the first unit of digital spectral analysis DSA1; blocks for aligning SEF and detecting ED the spectral envelope; blocks for calculating the trajectory of partial results BCPI and selecting the minimum MSS of the partial accuracy indicators; determination and reading of the results RR; two clock pulse generators CPG1, CPG2; the second block of DSA2. Red lines indicate many dimensional (dimension M) functional connections. Adjustable parameters of the device are the sampling period T of the speech signal and the length τ of the observation window.

The operation of the device begins with the launch of generators that are necessary to synchronize the operation of individual elements of the computational process (3)–(6). With the help of the generators periodically (with a period of τ)

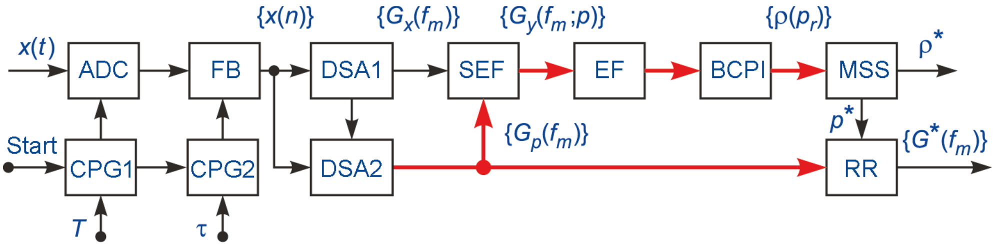

Fig. 1. Block diagram of the device for measuring the accuracy indicator and the optimal order of the autoregressive model of a speech signal: ADC, analog-to-digital converter; FB, frame builder; DSA1, DSA2, the digital spectral analysis blocks; SEF, spectral envelope flattening block; ED, envelope detector; BCPI, block for calculating partial indicators; MSS, minimum selection scheme; RR, result reading block; CPG1, CPG2, the clock pulses generators.

the frames of the voice signal $\{x(n)\}$ are formed and updated at the inputs of the blocks DSA1, DSA2. Accordingly, PSE (1), (2) will be periodically updated at their outputs. Moreover, according to (2), the model is updated simultaneously in R different variants $\{G_p(f_m)\}$ depending on the order $p = p_r$ for all $r \leq R$. i.e. in accordance with the expression (5) there will be R variants of the spectral envelope $\overline{G}_y(f_m;\ p)$. To isolate them in the device one uses a spectral envelope detector. Its practical implementation works according to a recirculator scheme [16]:

$$\overline{G}_y\ (f_m;\ p) = b\overline{G}_y\ (f_m - \Delta f;\ p) + G_y\ (f_m;\ p)\ ,\ \ m = \overline{1, M}\ , \tag{8}$$

based on the procedure of accumulating or smoothing the speech signal in the frequency domain. The constant b on the right side of (8) affects the inertia of such accumulation. The quantitative index of inertia $\theta = -\Delta f / \ln\ b$ must be consistent with the frequency F_0 of the fundamental pitch of the speech signal. Hence in the first approximation we will write

$$b\ = \exp\ \left(-\ \Delta f / F_0\right)\ = \exp\ \left(-\ F / \left(M F_0\right)\right)$$

and, for example, in the operating range of values F_0 = 100–200 Hz [3], for F = 8 kHz, M = 1024 we obtain b = 0.93–0.97.

The main result of the operation of the device is that the indicator of the potentially achievable accuracy of the autoregression model (2) is superimposed in the R-channel scheme of the minimum selection block MSS (cf. Fig. 1) according to criterion (7) for selecting the optimal order. In accordance with the order p^* the optimal autoregression model is read using the RR block $\{G^*(f_m)\}$.

The program and the results of the experiment. In the experiment, the pilot version of the device was used (cf. Fig. 1), implemented on the basis of the author's software module Phoneme Training.[1] The program is publicly available, its interface is described in detail in the work [11].

The objects of the experimental study were the signals of the six vowel sounds of Russian speech of the test speaker — the author of this article. Sufficiently long (2–3 s) duration of each signal was initially acquired for automatic sequential segmentation into a set of frames of duration τ = 30 ms with a time shift duration of 10 ms. At the sampling rate F = 8 kHz, the dimension of frame N was 240 samples. The shape of the observation window is rectangular. As a result, a representative database of L = 200–300 eponymous samples was formed for each vowel sound of speech. Next, for each sample the Schuster periodogram (1) with dimension M = 1024 and frequency increment Δf = 8000:1024 = = 7.8125 Hz is calculated. Also the autoregressive model of the speech signal determines the spectral Berg estimates (2) for R = 40 alternative versions of order p = 10–50. In accordance with Levinson's recursion (3), the total complexity of the computational procedure is of the order of $3Np_R = 3 \cdot 240 \cdot 50$ = 36,000 elementary multiplication-addition operations on the interval of one speech frame. This volume with a margin corresponds to the performance of modern speech systems and technologies [21].

In the course of the experiment, for each of the 40 alternatives for b = 0.96, according to (8), estimates of the spectral envelope $\overline{G}_y(f_m;\ p)$ were obtained according to which, in accordance with criterion (7), we established the optimal value of the order p^*, the corresponding autoregression model (2) and the indicator (6) of its accuracy. Then the results were averaged according to the results of L tests for each vowel phoneme of the test speaker. As a result, the relative error of experimental estimates [3] was $\varepsilon = 1.65L^{-1/2} = 1.65(200–300)^{-1/2}$ with a confidence probability of 0.9 no more than 10–12%.

Figure 2 shows a graph of the operating characteristic $\rho(p)$ of the proposed method (1)–(7) in relation to the vowel sound of speech "a". Vertical segments in the control points of this graph determine the boundaries of the confidence interval of the values of the indicator according to the results of multiple measurements. The curve $\rho(p)$ has a single minimum p^* = 32 that indicates the effectiveness of criterion (7) and, in general, of the acoustic measurement method. The achieved effect can be characterized by the value of the gain B in terms of the accuracy of the autoregressive model (2) due to the optimization of its order p^* with respect to its standard value p = 10 [2]:

$$B\ = \frac{\rho\ (10) - \rho^*}{\rho\ (10)}\ 100\ . \tag{9}$$

[1] Phoneme Training. Information system for phonetic analysis and speech training [site]. URL: https://sites.google.com/site/frompldcreators/ produkty-1/phonemetraining (accessed 20.09.2022).

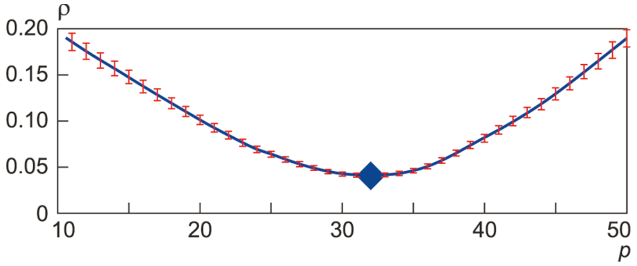

Fig. 2. Performance characteristic (with confidence limits according to the results of multiple measurements of the accuracy index) of the proposed method for measuring the accuracy of the autoregression model on the example of the vowel sound "a" of the speech of test speaker.

According to the data of Fig. 2 the value calculated by formula (9) was $B = 79.48\%$.

To illustrate the obtained result, Fig. 3 presents the family of Berg spectral estimates for autoregressive models of order $p = 10$ (curves 1–3) versus Schuster's periodogram (curve *4*) on the example of the same (typical) frame $\{x(n)\}$ of the vowel sound "a". All curves are normalized by the average power to unity. The model of order $p = 32$ is optimal in terms of accuracy. Figure 4 shows the graph corresponding to order $\overline{G_y}(f_m;\ p)$ of the PSE (5) at the output of the leveling filter (4), as well as the envelope $p = 32$ (shown by the dashed line). Its shape is close to rectangular. The indicator (6) in this case is characterized by the minimum $\rho^* = \rho(32) \approx 0.04$ (cf. Fig. 2).

The conclusions drawn remain true for the rest of the vowel sounds of the test speaker's speech. At the same time, according to criterion (7), their optimal model order values can vary greatly from each other, for example, for the "o" sound we have $p^* = 22$. At the same time, one can recognize in each case the winning values obtained from the score (9). A histogram of their distribution for vowel phonemes is shown in Fig. 5. Optimization made it possible to increase the accuracy of the autoregression model of the speech signal by more than 60%, depending on the speech sound of the test speaker and the specifics of his vocal tract. The maximum gain was achieved for the (Russian) upper vowel sounds "и", "у", "ы", which can be attributed to the features of a particular speaker.

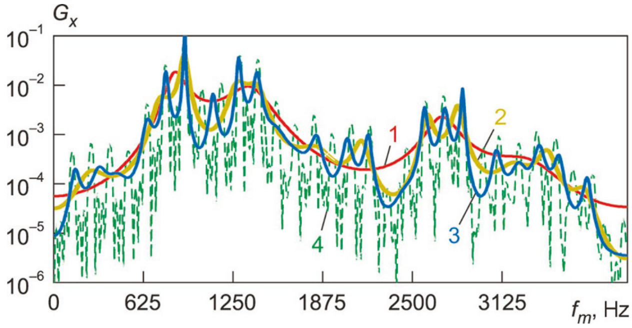

Fig. 3. Berg spectral estimates for the speech sound frame "a" of orders 32; 50 (curves 1–3) against the background of the Schuster periodogram (curve 4).

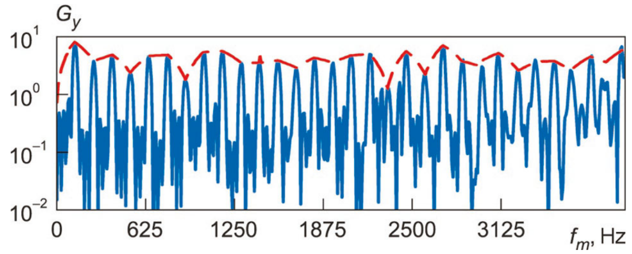

Fig. 4. Graphs of the normalized power spectrum estimation (5) of the speech sound signal "a" at the output of the flattening filter (4) with order of the autoregression model.

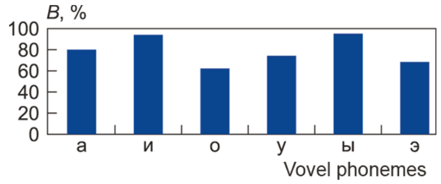

Fig. 5. Histogram of distribution of the score *B* (9) by vowel phoneme signals.

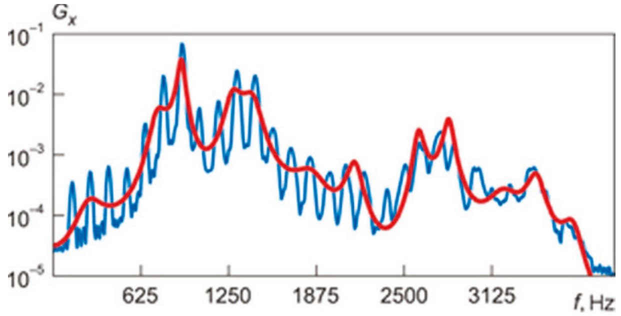

Fig. 6. The average Schuster periodogram (blue line) compared with the Berg estimate of the order $p = 32$ (red line) at the input of the flattening filter (4).

Discussion of the results obtained. Unlike Berg's method, Schuster's periodogram currently finds almost no independent application in the problems of spectral analysis of time series due to its inappropriate statistical characteristics [15]. Instead of Schuster's periodograms, for these purposes one uses the methods of Bartlett, Welch, and others, based on the effect of averaging periodogram PSE on samples of observations of large volume $N \gg M$. However, according to the conditions of the task considered in this article, the exact opposite relation holds, which excludes statistical averaging and thereby limits the possibility of Shuster's periodogram as a PSE of the speech signal. Therefore, a natural question arises regarding the validity of expression (1) as a reference spectral sample. The answer to this is given in the results of the second, final, stage of experimental studies, within the framework of which the comparative efficiency of the method proposed above was studied in two options for specifying a spectral sample: in a practically implemented version of the periodogram (1), and also in a speculative (hypothetical) version of Welch

$$G_x\left(f_m\right) = I^{-1}\sum_{i=1}^{I} G_x^{(i)}\left(f_m\right), \quad \forall m \le M-1\,, \tag{10}$$

with *I*-fold averaging of Schuster's partial periodograms on the observation interval with a total duration of τ*I*. Here the symbol $G_x^{(i)}(f_m)$ denotes the partial periodogram (1) within the boundaries of the *i*th speech frame $\{x(n)\}$ of finite volume *N*.

The results obtained are presented in Fig. 6 as graphs of two spectral estimates of the phoneme "a" signal of the test speaker: Welch (10) with $I = 240$ (blue line) and Berg (red line) with the optimal order $p^* = 32$ of the model (2). Berg's estimate was not averaged, i.e., one of the frames $\{x(n)\}$ was used, however, it agrees well with the envelope average periodogram estimate.

The PSE envelope at the output of the leveling filter in this case is no longer more uniform in comparison with the similar graph in Fig. 4. Moreover, its accuracy indicator (6) $\rho(32) = 0.051$ did not go beyond the corresponding confidence interval [0.02; 0.06] at the optimum point of the working characteristic (cf. Fig. 2). Hence we can conclude that the averaging of the Schuster periodogram according to the Welch method (10) gives almost nothing in terms of improving the accuracy of the formed autoregression model (2). Schuster's periodogram in this case plays the role of a non-final PSE, and a source of useful information regarding the fine structure of the speech signal. In this capacity, it is flawless due to the linearity of the processing of the speech signal during the discrete Fourier transform [7].

Conclusion. As a result of the study, an indicator of the accuracy of the autoregression model of the speech signal and a method for measuring it in the sliding window of observations mode of finite duration τ are proposed. Within the framework undertaken to optimize the autoregression model, the Berg spectral estimate and the Schuster periodogram mutually complement each other, although in practice they usually compete with each other [15, 17]. In this case, the optimization criterion was for the first time rigorously formulated in terms of the inverse problem of speech production. In contrast to criteria such as AIC, BIC, and the like [8, 9], the proposed criterion does not require synchronization of the observation window with the pitch period of the speech signal, which greatly weakens the precision of the problem due to a priori uncertainty [3].

The results obtained are intended for use in automatic processing and digital speech transmission systems with radical data compression based on linear prediction coefficients [1, 2].

REFERENCES

1. J. Gibson, *Entropy*, **20**, No. 10, 7502018 (2018), https://doi.org/10.3390/e20100750.
2. V. V. Savchenko, *Radioelectron. Commun. Syst.*, **64**, No. 11, 592–603 (2021), https://doi.org/10.3103/S0735272721110030.
3. A. V. Savchenko and V. V. Savchenko, An adaptive method for measuring the pitch frequency using a two-level autoregressive model of a speech signal, *Izmerit. Tekh.*, No. 6, 60–66 (2022).
4. E. Jaramillo, J. K. Nielsen, anf M. G. Christensen, *27th Eur. Signal Processing Conf.* (EUSIPCO), 2019, pp. 1–5, https://doi.org/10.23919/EUSIPCO.2019.8902763.
5. V. V. Savchenko, *Radiophys. Quantum Electron.*, **60**, No. 1, 89–96 (2017), https://doi.org/10.1007/s11141-017-9778-y.
6. S. Cui, E. Li, and X. Kang, *IEEE Int. Conf. on Multimedia and Expo* (*ICME)*, London, United Kingdom, 2020, pp. 1–6, https://doi.org/10.1109/ICME46284.2020.9102765.
7. Sh. Ando, *J. Acoust. Soc. Am.*, **146**, No. 11, 2846 (2019), https://doi.org/10.1121/1.5136873.
8. V. V. Savchenko and A. V. Savchenko, *Radioelectron. Commun. Syst.*, **62**, No. 5, 276–286 (2019), https://doi.org/10.3103/S0735272719050042.
9. J. Ding, V. Tarokh, and Y. Yang, *IEEE Trans. Inform. Theory*, **64**, No. 6, 4024–4043 (2018), https://doi.org/10.1109/TIT.2017.2717599.
10. V. V. Savchenko, *Radioelectron. Commun. Syst.*, **63**, 532–542 (2020), https://doi.org/10.3103/S0735272720100039.
11. V. V. Savchenko and L. V. Savchenko, Method for Measuring the Intelligibility of Speech Signals in the Kullback–Leibler Information Metric, *Meas. Tech.*, 62, No. 9, 832–839 (2019), https://doi.org/10.1007/s11018-019-01702-1.
12. M. Tohyama, Spectral envelope and source signature analysis, in: *Acoustic Signals and Hearing*, Academic Press (2020), pp. 89–110, https://doi.org/10.1016/B978-0-12-816391-7.00013-9.
13. P. Sun, A. Mahdi, J. Xu, and J. Qin J., *Speech Commun.*, **101**, 57–69 (2018), https://doi.org/10.1016/j.specom.2018.05.006.
14. V. V. Savchenko and L. V. Savchenko, The method of autoregresion modeling of the speech signal on the basis of its discrete Fourier transform and scale-invariant measure of information mismatch, *Radiotech. Electron.*, **66**, No. 11, 1100–1108 (2021), https://doi.org/10.31857/S0033849421110085.
15. S. L. Marple, *Digital Spectral Analysis with Applications*, 2nd edn., Dover Publications, Mineola, New York (2019), 432 p.
16. *Radioelektronnye sistemy. Osnovy postroeniya i teoriya: Spravochnik*, Ed. Ya. D. Shirman, 2nd edn., Radiotekhnika Publ., Moscow (2007), p. 657.
17. J. Benesty, J. Chen, and Y. Huang, Linear prediction, in: *Springer Handbook of Speech Processing, Part B*, Springer, New York (2008), pp. 111–124, https://doi.org/10.1007/978-3-540-49127-9_7.
18. A. Palaparthi, and I. R. Titze, *Speech Commun.*, **123**, 98–108 (2020), https://doi.org/10.1016/j.specom.2020.07.003.
19. L. R. Rabiner and R. W. Shafer, *Theory and Applications of Digital Speech Processing*, Pearson, Boston (2010), 1060 p.
20. A. V. Savchenko, V. V. Savchenko, and L. V. Savchenko, *Optim. Lett.*, No. 7 (2021), https://doi.org/10.1007/s11590-021-01790-5.
21. S. Kumar, S. K. Singh, and S. Bhattacharya, *Int. J. Speech Technol.*, **18**, 521–527 (2015), https://doi.org/10.1007/s10772-015-9296-2.