

Correlation analysis of voice communications in Russian language in the airspace of the Russian Federation

Nikita Andriyanov^{1,2,*}, Anatoliy Buldakovsky², and Kirill Risukhin²

¹Ulyanovsk State Technical University, 432027 Severny Venets 32, Ulyanovsk, Russia

²Ulyanovsk Civil Aviation Institute, 432071 Mozhaiskogo 8/8, Ulyanovsk, Russia

Abstract. The spectral characteristics of audio recordings of typical phrases of radio exchange of air traffic controllers and civil aviation pilots are investigated. Based on the obtained spectral characteristics, experimental covariance functions are constructed that show statistical relationships within a single voice message in the time domain. The main attention is paid to the possibility of describing speech messages using mathematical models of random sequences.

1 Introduction

The rapid development of transport systems, including air, land and water vehicles, is currently accompanied by the active use of automated or automatic systems of various kinds. However, such systems are more focused on the development of autonomous vehicles. Today, such devices are used only for solving highly specialized tasks and usually are performed completely offline, i.e. in the absence of a person inside. In this regard, such vehicles a priori have lower safety requirements.

A completely different situation is in the field of passenger transportation. An important element that ensures flight safety remains radio communication today. In accordance with the Federal Aviation Rules in Russia [1], the use of the Russian language is allowed on the territory of the Russian Federation. This, on the one hand, facilitates mutual understanding between pilots and controllers, but on the other hand, requires both of these skills to quickly switch from one language to another. In noisy environments, voice recognition accuracy can be enhanced by special digital signal processing systems. In this paper, we propose methods for describing the voice messages of some typical phrases using mathematical models of random processes [2-4], which can also be used in different processing algorithms.

2 Modern mathematical models for representing time sequences

The properties of heterogeneity and multidimensionality inherent in real big data remain outside the scope of most mathematical models that either describe spatially homogeneous signals or are too complex for the subsequent formation of processing and prediction

* Corresponding author: nikita-and-nov@mail.ru

algorithms. The most relevant for the description and processing of multidimensional inhomogeneous random fields and their time sequences is the use of a mixture of deep Gaussian processes. A feature of these processes is a characteristic nested structure, when the model parameters themselves are realizations of a lower level process. In [5–7], a good approximation quality was shown, which is achieved when describing real processes by deep Gaussian models. In [5], the possibility of describing unsteady random processes using this approach was shown. Unfortunately, the published work did not resolve the issue of describing the correlation and dynamics of changes in this correlation within the same model layer. This means that the behavior of the reference of a deep Gaussian process depends only on hidden embedded random variables at this point and auxiliary random additions, but not on neighboring samples. This significantly reduces the scope of such models. Therefore, at the initial stage of the study of mathematical models suitable for describing speech messages, we will use autoregressive models or autoregressive models with multiple roots of characteristic equations [8,9]. These models are well studied, possess properties important from the point of view of implementation (for example, quasi-isotropy, the minimum number of necessary parameters) and allow one to determine not only random processes, but also fields of arbitrary dimension, as well as to form subsequent processing algorithms that are qualitatively superior to algorithms for most random fields based on conventional autoregressions. In addition, more complex doubly stochastic models, which are described in [10–12], can be used to describe inhomogeneous signals.

3 Spectral characteristics of typical radio messages

The papers results obtain by using Audacity editor to record 20 typical phrases of radio exchange in Russian. Table 1 presents the phrases and the distribution of the number of records belonging to various male announcers.

Figure 1 shows the spectra of the phrase "Preflight communication check" (sounds like "Predpoletnaya proverka svyazi"), which were also formed in Audacity. The signal level in [dB] is plotted along the Y axis, and the frequency in [Hz] is plotted along the X axis.

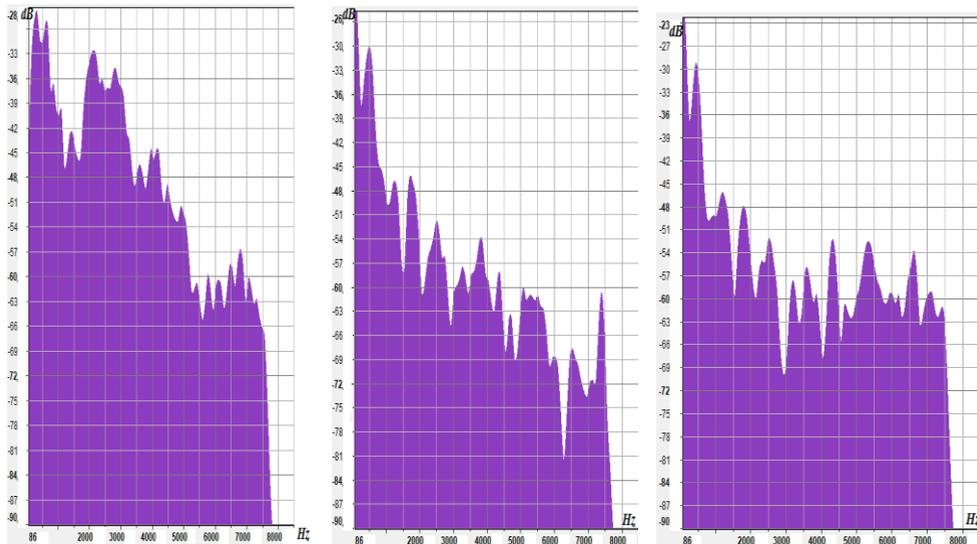


Fig. 1. Spectral characteristics for the phrase "Preflight communication check" recorded by different speakers.

Analysis of Figure 1 shows that the same phrase recorded by different speakers generally has different characteristics of voice messages in the frequency domain. However, the presented dependences also have pronounced similarities associated with the presence of harmonic oscillations in the number of words and the general attenuation of the spectral components with increasing frequency, as well as with the close arrangement of the peaks and valleys of the spectrum.

Moreover, the spectra for all identical records are quite similar, therefore, to simplify the analysis, we used single spectra obtained by simple mathematical averaging of signal level values over a specific frequency range in all the obtained spectra.

Table 1. Investigated sample phrases

№	Phrase 1	Speaker 1-2-3
1	Preflight communication check	4-3-3
2	I accept you at five	3-4-3
3	Information Alpha on board	3-3-4
4	Allow Parking Launch	4-3-3
5	Launch allowed	3-4-3
6	Allow preliminary	3-3-4
7	Ready to write down exit conditions	4-3-3
8	Exit scheme is Golim 07	3-4-4
9	Wind near the earth one zero zero degrees five meters per second	3-3-4
10	Ready to take off	4-3-3
11	I take off	3-4-3
12	We collect level three six zero	3-3-4
13	Next work with	4-3-3
14	Secondary control	3-4-3
15	Drop level seven zero	3-3-4
16	Runway ILS two seven Approach	4-3-3
17	Course capture	3-4-3
18	Landing allowed	3-3-4
19	Parking lot one one on RD three	4-3-3
20	Aeroflot one five four	3-4-3

It should be noted that all phrases presented in Table 1 was recorded in Russian language. Therefore, it was not the variant that is presented as a translation in the table that was recorded, processed, and studied, but the one that was originally received in Russian. This is due to the fact that in the airspace of the Russian Federation on domestic flights radio communication in Russian can often be used.

Figure 2 shows spectra for all 20 phrases. The countdown is from top to bottom and from left to right, i.e. the first line corresponds to the averaged spectra (over all records) of Phrases No. 1-4, and the fifth is for Phrases No. 17-20. The X-axis represents the frequency in [Hz], on a logarithmic scale, the Y-axis represents the level in [dB].

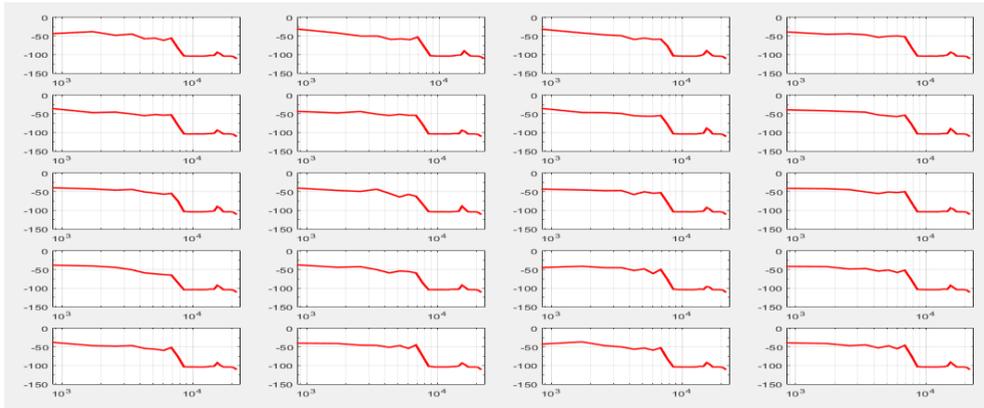


Fig. 2 Averaged spectrum of typical speech messages of phraseological radio exchange in [dB].

According to the graphs of Figure 2, a comparison was made with the source data for each implementation. As a result of this study, we can conclude that, on the whole, it did not lead to severe distortions of the shape of the spectrum of signals, which is well explained by the same features characteristic of each sound, and that in the future may serve as the basis for the development of machine learning algorithms for recognizing speech message data.

4 Correlation representation of typical phrases and identification of parameters

However, all previously considered values characterize the signal level in [dB], and the change in the signal over time is more convenient to consider in amplitude values, for example, [V]. Given the level of the reference voltage $U_0 = 1$ mV, we use the expression

$$S[\text{dB}] = 10 \log(U/U_0), \quad (1)$$

where S is spectrum in [dB], U is spectrum in [V], U_0 is reference voltage.

Using the formula (1), the desired signal level can be expressed in [V]

$$U[\text{V}] = 10^{S/10-3}. \quad (2)$$

Next, we use the following relationship of the amplitude spectrum of a signal with its energy spectrum or power spectral density

$$G(F) = |S(F)|^2, \quad (3)$$

where $G(F)$ is signal energy spectrum, $|\cdot|$ is module calculation operation, $S(F)$ is signal spectrum.

Finally, we can use the inverse Fourier transform to find the covariance function (CF) in accordance with the expression

$$R(t) = \frac{\int_{-F_{\max}}^{F_{\max}} G(F) e^{j2\pi Ft} dF}{\sigma^2}, \quad (4)$$

where σ^2 is maximum transformation value.

Figure 3 presents calculated CFs for all 20 phrases. The countdown is from top to bottom and from left to right, i.e. the first line corresponds to the phrase of phrase No. 1-4, and the fifth is for the phrase No. 17-20. The interval k is plotted along the X-axis, and the CF values are plotted along the Y-axis.

Analysis of the curves in Figure 3 shows that most of the obtained CFs have a decreasing character. The presence of many extremes is associated with quasiperiodicity, which may be present in a speech signal. In addition, it should be noted the possibility of forming autoregressive sequences with similar CFs.

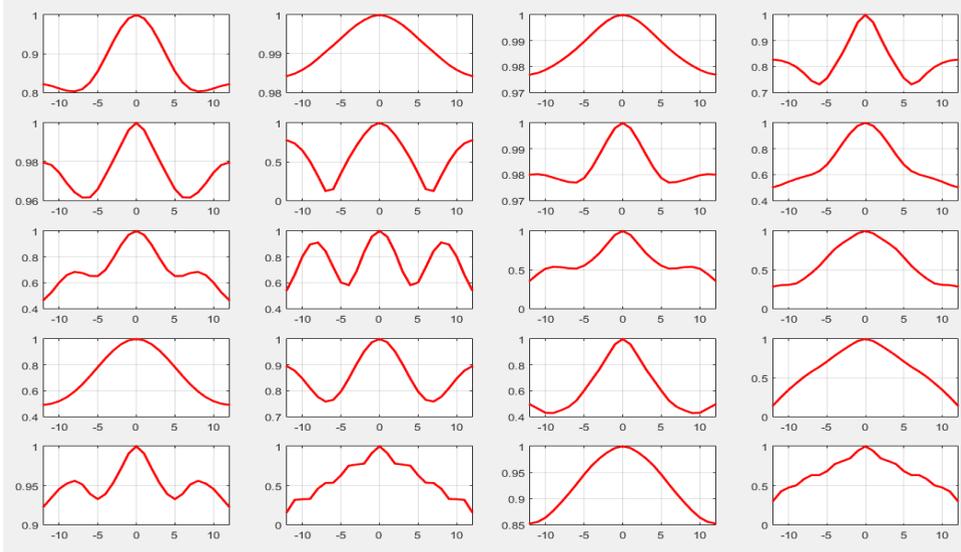


Fig. 3 CF of typical speech messages of phraseological radio exchange.

To identify the parameters of the model, we write an arbitrary order autoregressive model

$$x_i = \rho_1 x_{i-1} + \rho_2 x_{i-2} + \dots + \rho_m x_{i-m} + \xi_i, \quad i = 1, 2, \dots, M, \quad (5)$$

where m is order of autoregression.

By selecting the coefficients $\rho_1, \rho_2, \dots, \rho_m$ one's can get a Gaussian random processes $\{x_i\}$, $i = 1, 2, \dots, M$, with a variety of correlation properties. Then for the CF values, we can use the relation

$$R_x(k) = \rho_1 R_x(k-1) + \rho_2 R_x(k-2) + \dots + \rho_m R_x(k-m), \quad k > 0. \quad (6)$$

The CF of sequence (5) also obeys the recurrence expression (6).

Substitution values $k = 1, 2, \dots, m$ in (6) leads to a well-known system of Yule-Walker equations [13], which, for example, for second-order systems takes the following form

$$\begin{aligned} \rho_1 + \rho_2 R(1) &= R(1), \\ \rho_1 R(1) + \rho_2 &= R(2). \end{aligned} \quad (7)$$

The solution of this system allows one's to find the coefficients $\rho_1, \rho_2, \dots, \rho_m$ of autoregressive equations (5) according to the values set in advance or estimated on the basis of the experiment values of CF $R_x(1), R_x(2), \dots, R_x(m)$.

We will use the previously found CF values, then we can compose a system of equations with 12 unknowns

$$\left\{ \begin{array}{l} \rho_1 + \rho_2 R(1) + \rho_3 R(2) + \rho_4 R(4) + \dots + \rho_{12} R(11) = R(1), \\ \rho_1 R(1) + \rho_2 + \rho_3 R(1) + \rho_4 R(2) + \dots + \rho_{12} R(10) = R(2), \\ \dots \\ \rho_1 R(10) + \rho_2 R(9) + \rho_3 R(8) + \rho_4 R(7) + \dots + \rho_{12} R(1) = R(11), \\ \rho_1 R(11) + \rho_2 R(10) + \rho_3 R(9) + \rho_4 R(8) + \dots + \rho_{12} = R(12). \end{array} \right. \quad (8)$$

Figure 4 shows graphs of approximation of CFs of typical phrases No. 1 (a), No. 4 (b) and No. 10 (c) with the 3rd-order (dashed line), 6th-order (dash-dotted line) and 12th-order (solid line) models. The circles indicate the obtained experimental values.

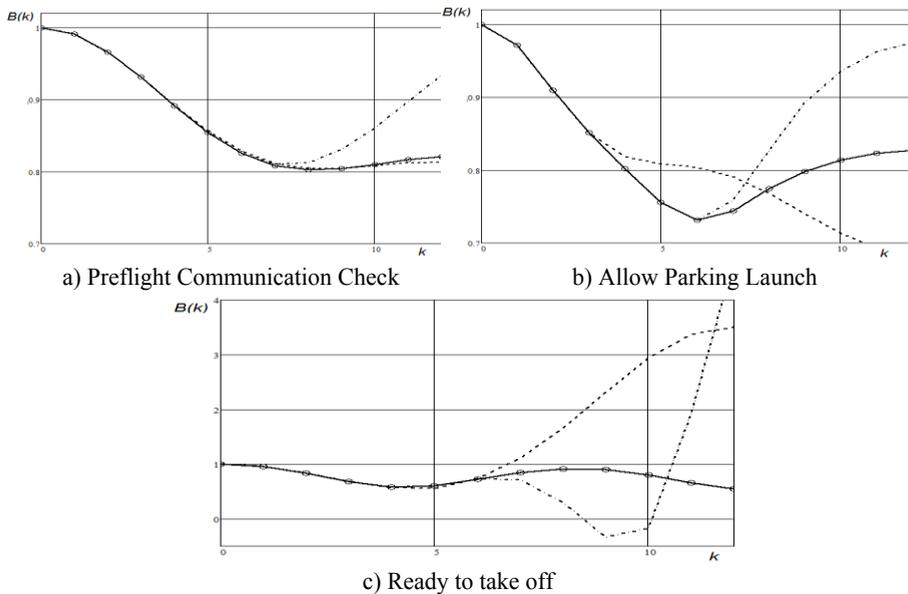


Fig. 4 Approximation of CF using autoregressive models.

An analysis of the obtained dependences allows us to conclude that the higher the order of the model, the more accurately it is able to convey the correlation properties of speech messages. However, in some cases (Figure 4 a), lower-order models also provide CFs which are close to real, while significantly reducing computational costs.

5 Simulation of voice messages using autoregressive models

Let us implement the autoregressive models of different orders according to the identified parameters for typical message No. 1 in accordance with the obtained solutions of the system (8). Figure 5 shows the obtained random processes based on the 12th-order (solid

line), 6th-order (dash-dotted line) and 3rd-order (dashed line) models. The X-axis represents the numbers of samples, the Y-axis represents the values of the random process.

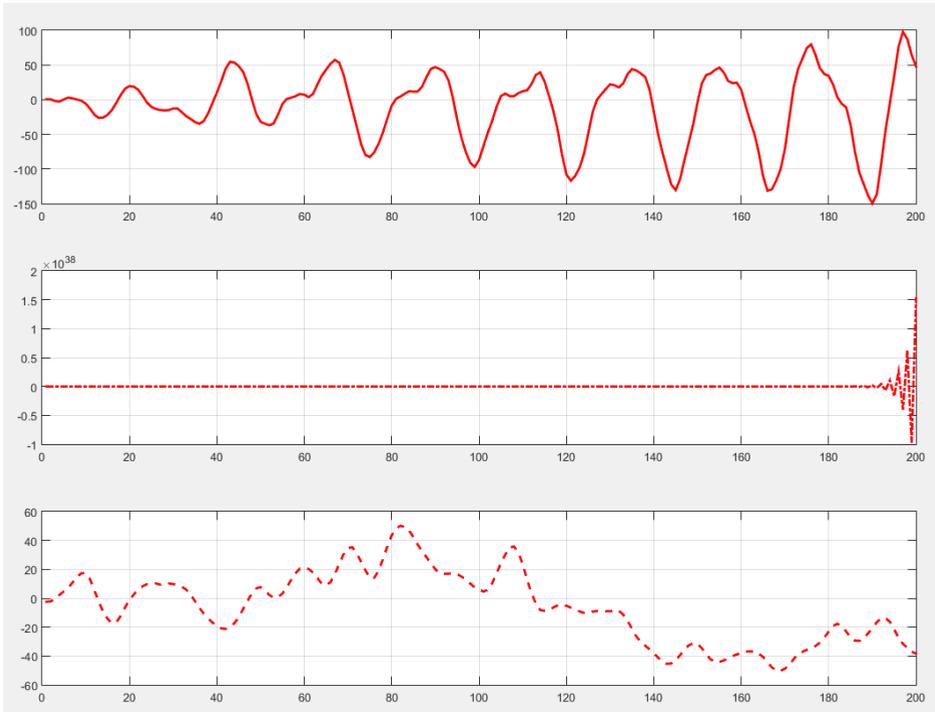


Fig. 5 Time domain presentation of signals from CF based on phrase No. 1.

Figures 6 and 7 show similar implementations for phrases No. 4 and No. 10 respectively.

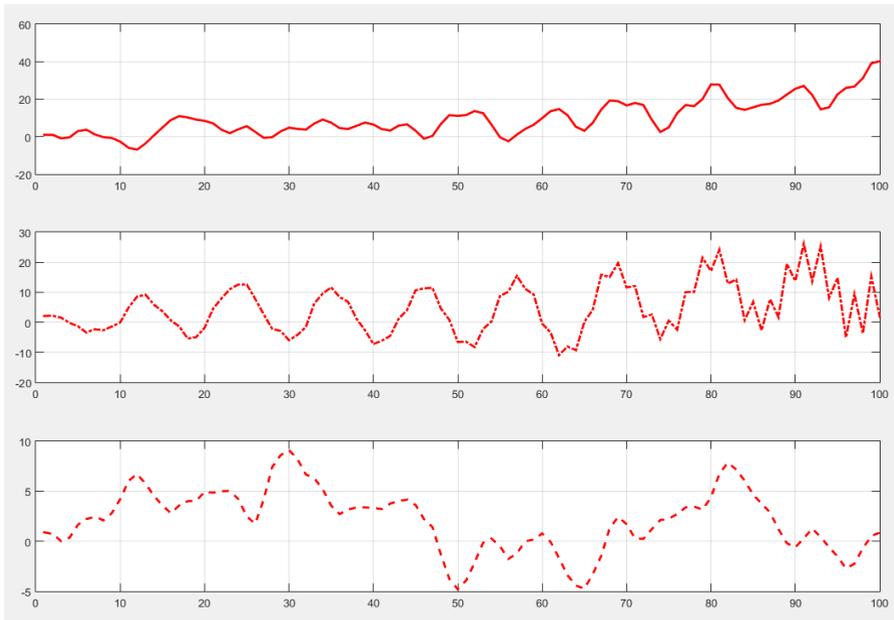


Fig. 6 Time domain presentation of signals from CF based on phrase No. 4.

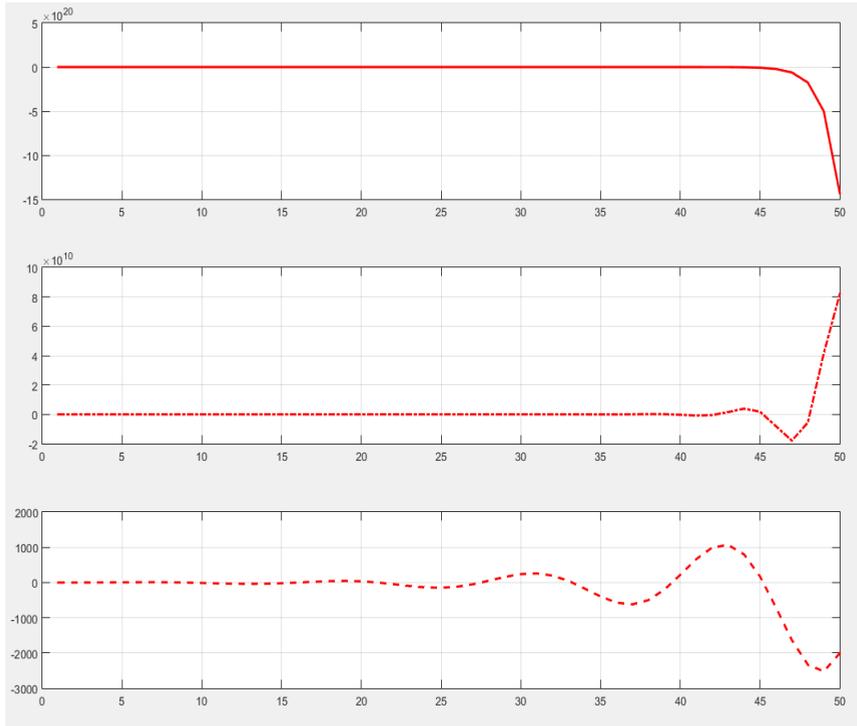


Fig. 7 Time domain presentation of signals from CF based on phrase No. 10.

Thus, it is possible to simulate speech messages for which CFs will be close enough to CFs based on real data. However, in some cases, models may diverge. In this case, you must either change the order of the model, or introduce additional restrictive conditions.

6 Conclusion

A statistical analysis of typical phrases of voice messages of radio exchange is carried out. The use of autoregressive models of random processes to describe such messages is proposed, and the capabilities of simulating signals with close correlation properties are established. The main results of the work are as follows:

1) 20 typical phrases of radio exchange were recorded, in different voices so that each phrase would have 10 entries. The necessary sample was collected for research.

2) On the basis of a representative sample, spectral characteristics were constructed for each studied voice message, the data were averaged, and the CF values of the studied voice messages were obtained.

3) Using the Yule-Walker equations, the parameters of autoregressive models are identified by the calculated values of the CF of the studied voice messages. It is shown that this identification method makes it possible to obtain CFs arbitrarily close to real ones when the order of the model is increased.

4) Signals were generated using autoregression models of different orders, the parameters of which were identified on the basis of experimental values of CF.

Thus, the study allows us to conclude that it is possible to increase the efficiency of processing voice messages during radio communications in civil aviation based on mathematical modeling of typical phrases.

References

1. *The procedure for the implementation of radio communications in the airspace of the Russian Federation: Federal Aviation Rules* **362**, 66 (2012)
2. V. R. Krashennnikov, E. Yu. Lebedeva, V. K. Kapyrin, *Variation of the boundaries of speech commands to improve recognition of speech commands by their cross-correlation portraits*, Bulletin of the Samara Scientific Center of the Russian Academy of Sciences **15**, 4(4) (2013)
3. K. K. Vasiliev, V. E. Dementiev, N. A. Andriyanov, *Comparative analysis of point forecasting of doubly stochastic and autoregressive models*, DSPA: Issues of the use of digital signal processing **6**, 2 (2016)
4. N. A. Andriyanov, A. I. Buldakovsky, K. V. Risukhin *Investigation of the probabilistic characteristics of typical civil aviation radio communications*, DSPA: Issues of using digital signal processing. **1** (2019)
5. A. Damianou and N. Lawrence, *Deep Gaussian processes*, Artificial Intelligence and Statistics **29** (2013)
6. D. Zhenwen, A. Damianou, J. González and N. Lawrence, *Variational auto-encoded deep Gaussian processes*, CoRR (2015)
7. C. Viroli, G. J. McLachlan *Deep Gaussian Mixture Models*, arXiv:1711.06929v1 (2017)
8. N. A. Andriyanov, Yu. N. Gavrilina, *Image Models and Segmentation Algorithms Based on Discrete Doubly Stochastic Autoregressions with Multiple Roots of Characteristic Equations*, CEUR Workshop Proceedings **2076** (2018)
9. N. A. Andriyanov, M. N. Sluzhivyi, *Solution for the problem of the parameters identification for autoregressions with multiple roots of characteristic equations*, CEUR Workshop Proceedings **2391** (2019)
10. N. A. Andriyanov, V. E. Dement'ev *Application of mixed models of random fields for the segmentation of satellite images*, CEUR Workshop Proceedings **2210** (2018)
11. K. K. Vasiliev, N. A. Andriyanov, *Synthesis and analysis of doubly stochastic models of images*, CEUR Workshop Proceedings **2005** (2017)
12. K. Vasiliev, V. Dementiev and N. Andriyanov, *Representation and processing of multispectral satellite images and sequences*, Procedia Computer Science **126** (2018), doi.org/10.1016/j.procs.2018.07.208 eid: 2-s2.0-85056672834
13. K. K. Vasiliev *Methods of signal processing: a training manual*, Ulyanovsk (2001)