ISSN 0001-4338, Izvestiya, Atmospheric and Oceanic Physics, 2012, Vol. 48, No. 4, pp. 384–390. © Pleiades Publishing, Ltd., 2012.
Original Russian Text © N.D. Tsybul'skaya, S.N. Kulichkov, A.I. Chulichkov, 2012, published in Izvestiya AN. Fizika Atmosfery i Okeana, 2012, Vol. 48, No. 4, pp. 434–441.

Studying Possibilities for the Classification of Infrasonic Signals from Different Sources

N. D. Tsybul'skaya, S. N. Kulichkov, and A. I. Chulichkov

Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Pyzhevskii per. 3, Moscow, 119017 Russia

e-mail: snk@ifaran.ru

Received July 19, 2011

Abstract—Atmospheric infrasonic signals were classified on the basis of data obtained in the United States (University of Alaska, Fairbanks) and in the Antarctic region (Windless Bight) from 1980 to 1983. The data archive included five classes of signals from different sources: explosions, mountain associated waves, microbaroms, volcanic infrasound, and auroral infrasonic waves. This classification was based on the theory of testing statistical hypotheses. The possibilities of separating these classes were studied. It was shown that the signals (from the archive used) that are characteristic of explosions and volcanic activity can be rather easily separated from those characteristic of mountain associated waves, microbaroms, and auroral infrasonic waves.

Keywords: signal classification, acoustics, statistical hypotheses.

DOI: 10.1134/S0001433812040147

1. INTRODUCTION

Infrasonic signals from more than a million different events (explosions, fires, bolides, storms in the ocean, volcanic eruptions, tsunami, thunderstorms, other meteorological sources, etc.) have been recorded at the international monitoring network [1].

Figure 1 shows an infrasonic signal recorded at a distance of 304 km from a surface explosion equivalent to 20–70 t of TNT [2]. Figure 2 shows an infrasonic signal recorded at a distance of 980 km from the place of an equivalent explosion. It follows from Figs. 1 and 2 that the recorded signals significantly differ in form. In Fig. 1, the waveform corresponds to a classical infrasonic signal at a distance of one cycle of ray paths from its source. This signal is characterized by the presence of several types of infrasonic arrivals (surface, stratospheric, mesospheric, and thermospheric) which significantly differ in amplitude and waveform. Such signals are not characteristic of any other infrasonic sources (fires, bolides, storms in the ocean, volcanic eruptions, tsunami, thunderstorms, other meteorological sources, etc.). Therefore, identifying a signal such as that corresponding to a pulsed source presents no difficulty. The situation is quite different in identifying signals recorded at a distance of a few cycles of ray paths from an explosion (see Fig. 2). Such signals are oscillating and very long in time. Some arrivals can be pronounced in their waveforms (this is easily seen in Fig. 2). However, it is impossible to visually identify these arrivals as those corresponding to a pulsed source, because similar waveforms are also observed in infrasonic signals from storm waves in the ocean (microbaroms), auroras, mountain waves, etc. (see below).

Therefore, it is necessary to develop special methods to classify different types of infrasonic signals. The proposed approach of such a classification is based on the methods of testing statistical hypotheses [4].

The SigLib data archive, which includes infrasonic signals recorded in the United States (University of Alaska, Fairbanks) and in the Antarctic region (Windless Bight) from 1980 to 1983 [5], was used in this work.

At the University of Alaska, a team of specialists studying infrasonic signals analyzed library data [5]. They singled out 172 signals from natural and anthropogenic sources.

All the signals from the data archive [5] were divided into five classes: signals from explosions (class 1, Explosion Test), mountain associated waves (class 2, MAW), microbaroms (class 3, Microbarom), volcanic infrasound (class 4, VOL), and auroral infrasonic waves (class 5, AIW). The signals were recorded with a few (3–4) sensors located at some distance from one another, which made it possible to record one and the same signal in a few channels and, thus, single out a useful signal against the background of noise.

Figure 3 shows the wave forms that are characteristic of each of the classes of infrasonic signals.

The characteristic features of signals should be found for each of the classes, and a method of classifying signals should be proposed.

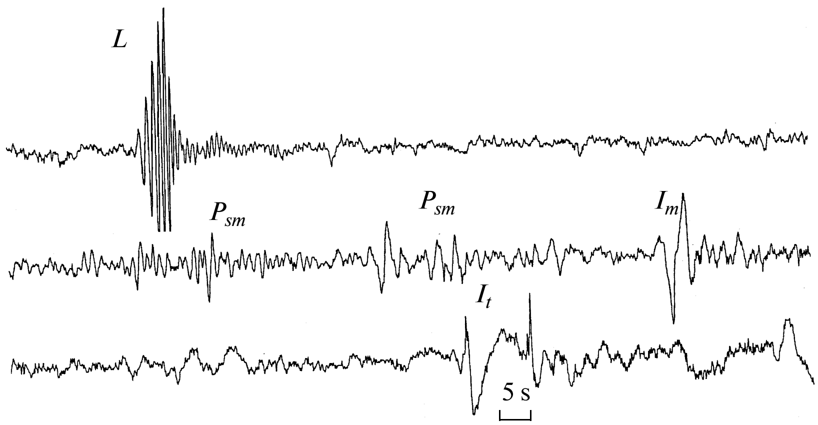

Fig. 1. Infrasonic signal recorded at a distance of 304 km from the place of a surface explosion equivalent to 20–70 t of TNT [2]. L denotes a surface infrasonic arrival; P_{sm} denotes stratomesospheric infrasonic arrivals; I_m denotes a mesospheric infrasonic arrival; and I_t denotes a thermospheric infrasonic arrival.

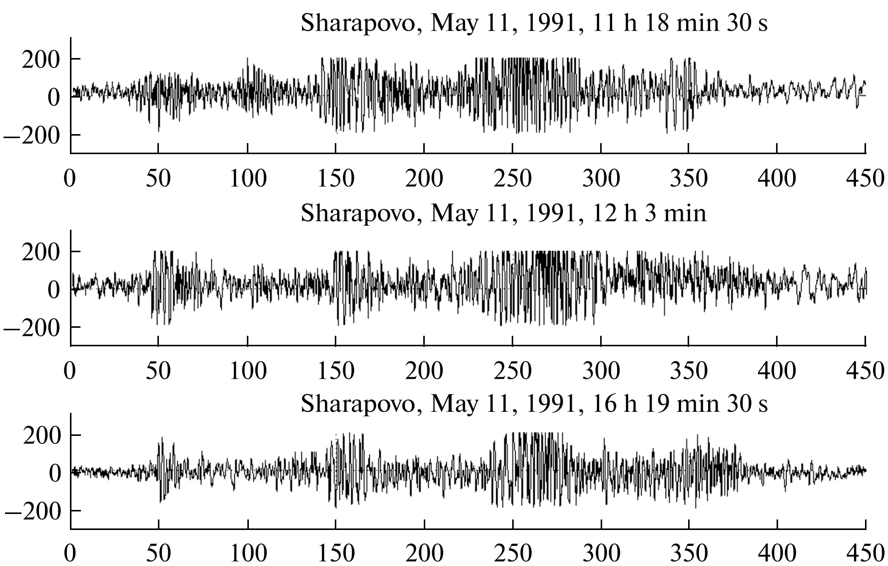

Fig. 2. Infrasonic signal recorded at a distance of 980 km from the place of a surface explosion equivalent to 20–70 t of TNT [2]. Seconds are plotted on the abscissa.

In this work, the possibilities for signal classification are studied on the basis of an analysis of their form [6, 7]. In the methods of a morphological analysis, the form of signal implies information that is general for the elements of a given class and independent of the recording conditions. In this case, the coefficient of signal enhancement is considered unknown; therefore, the methods of analyzing the form of signal must be invariant under variations in its amplitude. The features of signal forms that are characteristic of time intervals of 5 to 10 min and reflected in the correlations of signal values are analyzed in this work. The separability of signals belonging to each of the classes from all the rest of the signals is studied. It is shown that it is possible to distinguish between signals from bomb explosions and volcanic infrasound and those from the rest of sources.

2. A MATHEMATICAL MODEL OF SIGNALS BELONGING TO EACH OF THE CLASSES

The signals under study were recorded at sufficiently long distances from their sources and their duration was 5–10 min. Therefore, the whole signal

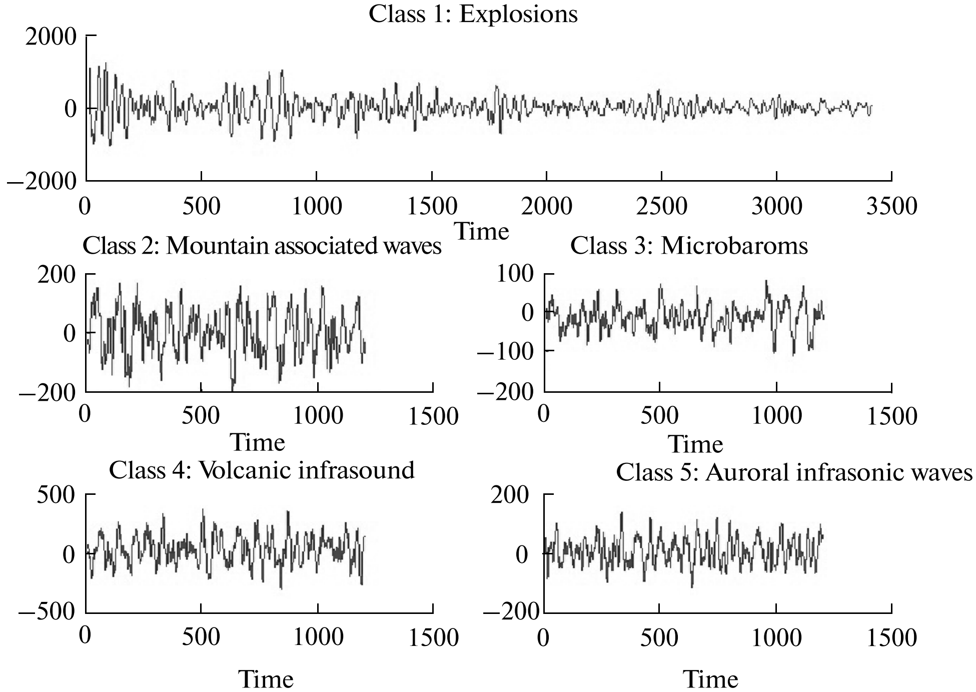

Fig. 3. Signals characteristic of each of the five classes.

was divided into fragments containing $n = 600$ readings in the same manner as was done in the "caterpillar" method [8]; i.e., the first fragment started with the first reading of the signal and ended with reading n, the second fragment started with the second reading of signal and ended with reading $n + 1$, etc. (Fig. 4). All these fragments were treated as the random-vector realizations $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$ with the dimensionality n.

Based on the requirement for invariance under variations in the coefficient of signal enhancement, the normalization was performed by dividing each coordinate of vector by its norm.

It was assumed that the correlation properties of normalized vectors are the same inside each of the classes, but they differ from class to class.

The results of an analysis of these signals showed that their mathematical expectations calculated as an arithmetic mean of the coordinates of each signal's vector $\left(E\xi = \frac{1}{n}\sum_{j=1}^{n} \xi_j\right)$, are close to zero, and the matrices of covariations are close to the Toeplitz matrices (i.e., their matrix elements along each diagonal are constant). This suggests that the random vectors of a given class can be treated as the realizations of stationary random processes. The covariation matrices are shown in Figs. 5a and 5b by the example of signals belonging to different classes (Explosion Test and AIW).

Then, it was assumed that each of the five signal classes is a set of random vectors with zero mathematical expectation and specified correlation matrices; sample covariance matrices were used in the classification algorithm.

3. STATEMENT OF THE PROBLEM OF CLASSIFICATION

Now, let us consider the problem of dividing vectors into two classes. It is assumed that the elements of each class are random vectors from the Euclidean space R_n with the zero mathematical expectation and the correlation matrices V (for the first class) and W (for the second class).

To solve this problem, we use an approach based on the classical theory of solving problems of testing statistical hypotheses [4]. The assumption that the correlation matrix of the vector produced is V is called a hypothesis, and the assumption that the correlation matrix is W ($W \neq V$) is called an alternative.

In order to solve the problem of classification, we divide the space R_n into two regions S and its complement $\bar{S}$. If the realization of a random vector falls within the region S, it belongs to the first class, which corresponds to the accepted hypothesis; otherwise, it belongs to the second class. Let us explain from what considerations we plot the acceptance region S.

Let us assume that the hypothesis is true. We consider the Karhunen–Loéve basis $\{\mathbf{e}_j, j = 1, ..., n\}$ composed of the eigenvectors of the matrix V and corre-

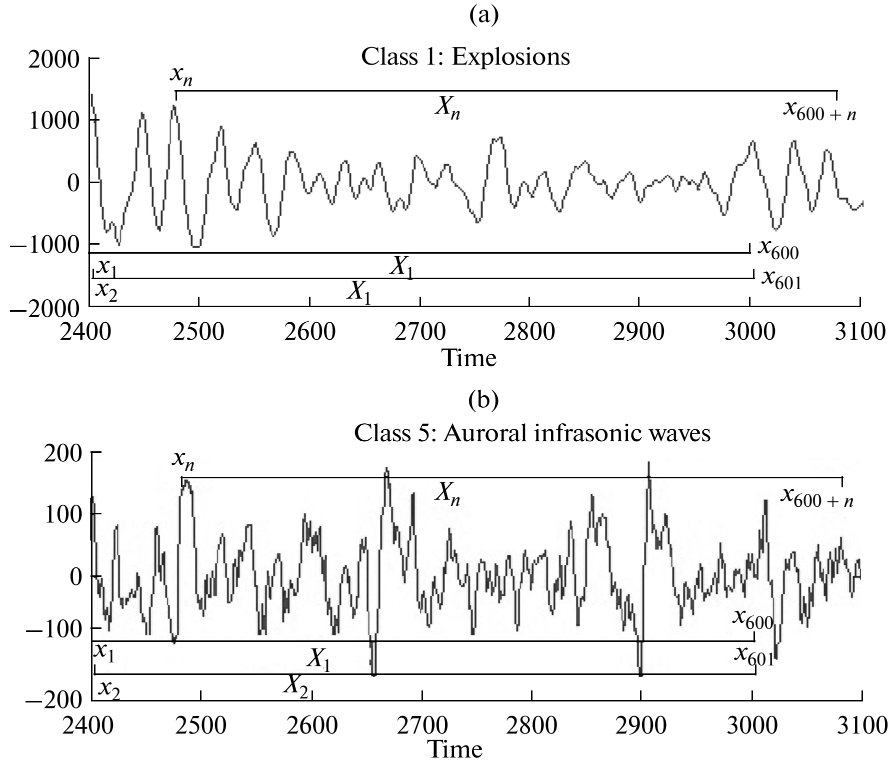

Fig. 4. Graphical description of the covariation matrix for signals belonging to (a) explosions (Explosion Test) and (b) auroral infrasonic waves (AIW).

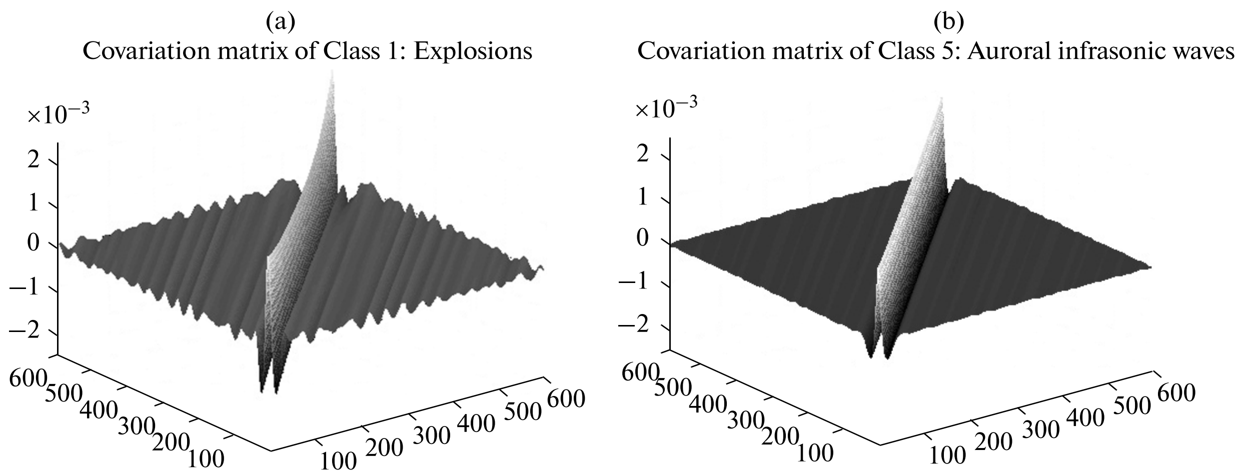

Fig. 5. Portions of signals from (a) explosions and (b) auroral infrasonic waves which are chosen to array the covariation matrices using the caterpillar method.

sponding to the eigenvalues of $\sigma_j^2, j = 1, \ldots, n$ that are ordered so that $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_n^2$; then, the random vector $\xi \in R_n$ with the zero mathematical expectation and the covariance matrix V are written in the form $\xi = \sum_{j=1}^{n} \alpha_j \mathbf{e}_j$, where the expansion coefficients α_j are the uncorrelated random quantities with the zero mathematical expectation and the variance equal to σ_j^2, $j = 1, \ldots, n$ [7]. After a transformation with the aid of the matrix $V^{-1/2}$, we obtain the vector $V^{-1/2}\xi = \sum_{j=1}^{n} \frac{\alpha_j}{\sigma_j} \mathbf{e}_j$, whose expansion coefficients have a unit variance, and its squared norm $t(\xi) = \left\| V^{-1/2}\xi \right\|^2 \equiv \left(\xi, V^{-1}\xi\right)$ has a mathematical expectation equal to the dimensionality n of the space R_n. Then, on the basis of the Chebyshev inequality, for any $\varepsilon > 0$, one can write

$$P(t(\xi) \geq \varepsilon) \leq \frac{n}{\varepsilon}. \qquad (1)$$

This relationship is used to characterize the agreement between the realization $\mathbf{x}$ of the random vector $\xi \in R_n$ and the hypothesis. Substituting $\varepsilon = t(\mathbf{x})$ in (1), we obtain $P(t(\xi) \geq t(\mathbf{x})) \leq \frac{n}{t(\mathbf{x})}$, which can be interpreted in the following way: the higher the value of $t(\mathbf{x})$ obtained for the realization $\mathbf{x}$ is, the lower the probability is that the value of $t(\xi)$ exceeding that of $t(\mathbf{x})$ will appear if the hypothesis is true. The value of $\alpha_V(\mathbf{x}) = \frac{n}{t(\mathbf{x})}$ is the upper boundary of the probability of obtaining the ξ realization that is in no better agreement (than $\mathbf{x}$) with the hypothesis. The random quantity $\alpha_V(\mathbf{x})$ is called the hypothesis reliability and is used as a characteristic of agreement between the $\mathbf{x}$ realization and the hypothesis [7].

Similar reasonings show that an agreement between the $\mathbf{x}$ realization of the random vector $\xi \in R_n$ and the alternative is due to $\alpha_W(\mathbf{x}) = \frac{n}{(\mathbf{x}, W^{-1}\mathbf{x})}$.

When classifying the vector ξ, it would be natural to consider that the hypothesis is true; i.e., the correlation matrix of vector ξ corresponds to V if, at the $\mathbf{x}$ realization of the vector ξ, its reliability $\alpha_V(\mathbf{x})$ is not less than the reliability $\alpha_W(\mathbf{x})$ of the alternative. In view of the fact that the errors of both the first and second kinds result in different losses, we consider that the vector ξ (according to the $\mathbf{x}$ realization) belongs to the hypothesis if $\alpha_V(\mathbf{x}) - \alpha_W(\mathbf{x}) \geq c$, where the threshold value is the problem parameter regulating the relation between the errors of the first and second kinds. After the corresponding transformations, the region S of the accepted hypothesis is determined by the following relation:

$$S = \{x \in R^n : (x, V^{-1}x) - (x, W^{-1}x) \leq c_\alpha\}. \qquad (2)$$

4. EMPIRICAL CONSTRUCTION OF THE MODEL OF SIGNAL CLASSES

In this problem, the sample covariation matrices arrayed according to a sample obtained from the archive with the caterpillar method were used for classification; in this case, signals from all sensors were used. The mathematical expectations of random vectors were assumed to be zero, and the sample vectors were normalized.

One of the five signal classes under analysis was chosen, and the sample covariation matrix V was arrayed on the basis of the obtained sample of vectors belonging to this class.

If the archive contains L signals of a chosen class, each signal was recorded with k sensors and the number of counts is N; then, as a result, the number of sample vectors is $(N - n)kL$. Figure 4 shows the characteristic graphs divisible by the period of signal fragments (by the example of explosions and auroras).

The vectors similarly obtained from the archive for the other four classes were treated as sample values of a random vector distributed according to the alternative, and the sample covariance matrix W was arrayed on their basis.

5. TESTING SIGNAL SEPARABILITY AND DETERMINING CRITICAL LEVELS

The signal separability was tested in the following way:

(1) The value of criterion (2) was calculated for each pair of classes. If x belongs to S, then it is considered to correspond to the first-class signal. If x does not belong to S, then it corresponds to the second-class signal.

(2) For each x of the first class, $(x, V^{-1}x) - (x, W^{-1}x) = s_1(x)$ was calculated. For each x of the second class, $(x, V^{-1}x) - (x, W^{-1}x) = s_2(x)$ was calculated.

(3) For different c_α, the values of the function $d_i(c_\alpha)$ (the number of the x vectors of the ith class, for which $s_i(x)$ is smaller than or equal to c_α) were calculated.

(4) The probability of the correct solution (i.e., the correctness of the accepted hypothesis) $P_1(c_\alpha) = d_1(c_\alpha)/N_1$ (N_1 is the number of the first-class vectors) was estimated, as was the probability of a wrong decision (the incorrectness of the accepted alternative) $P_2(c_\alpha) = d_2(c_\alpha)/N_2$ (N_2 is the number of second-class vectors).

(5) The values of the c_α thresholds were determined using the obtained graphs.

The c_α thresholds were determined for the following two cases: for division into five classes and for division into two sets. The first set included signals of the first (explosions) and fourth (volcanic infrasound) classes, and the second set included signals of the second (mountain associated waves), third (microbaroms), and fifth (auroral infrasonic waves) classes.

The $P_1(c_\alpha)$ and $P_2(c_\alpha)$ graphs obtained for the case of division into two sets are given in Fig. 6. The c_α level is chosen to be 1500.

For some sensors, the readings of signals (from the archive) had insufficient separability. The sample vectors including the readings of such sensors were not taken into account in constructing the empirical models of classes.

6. THE RESULTS OF SIGNAL CLASSIFICATION

At the thresholds chosen, signals were finally classified according to the following algorithm:

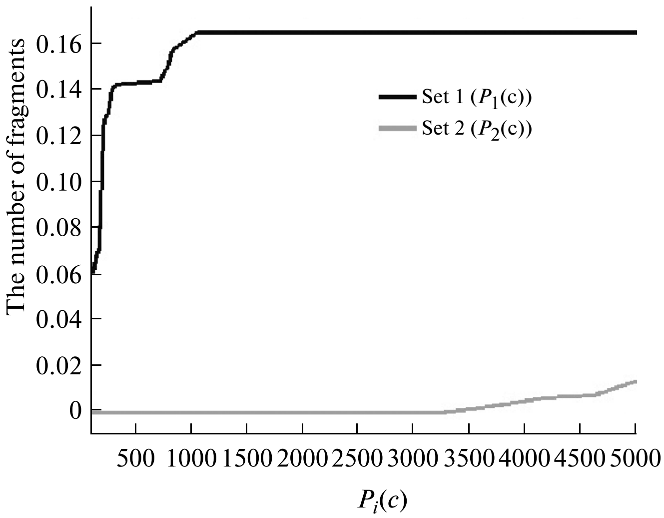

Fig. 6. Graphs of $P_1(c_\alpha)$ and $P_2(c_\alpha)$ for the case of dividing into two sets.

(1) A fixed signal was chosen from the archive, and $(N-n)kl$ sample vectors were plotted using the caterpillar method for this signal.

(2) Each of the sample vectors was classified on the basis of criterion (2) with some chosen threshold of c_α. Five criteria (each with its threshold) were used in dividing signals into five classes. One criterion was used in dividing signals into two sets.

(3) The number of sample vectors belonging to each of the classes was calculated.

(4) When dividing signals into five classes, it was assumed that the signal can reliably be referred to the class with number k_0 if the sum of signal fragments belonging to this class exceeds 20000 for the first, third, and fifth classes, 10000 for the second class, and 30000 for the fourth class. When classifying into two sets, the given quantity divided by the number of class vectors was bound to exceed 0.8.

All in all, 57 signals were analyzed:

(i) one signal from the class of explosions (recorded in three channels);

(ii) 15 signals from the class of mountain associated waves (each of these signals was recorded in four channels);

(iii) 15 signals from the class of microbaroms (each of these signals was recorded in four channels);

(iv) 5 signals from the class of volcanic infrasound (each of these signals was recorded in four channels);

(v) 21 signal from the class of auroral infrasonic waves (each of these signals was recorded in four channels).

Analyzing the results of dividing signals into five classes, one can conclude that the proposed algorithm rather clearly separates the signals belonging to the first and fourth classes (set no. 1) from those belonging to the second, third, and fifth classes (set no. 2).

In fact, all signals of the first and fourth classes are correctly classified; however, 13 (of 51) signals belonging to the second, third, and fifth classes were also erroneously referred to the first and fourth classes.

The results of classifying signals into two sets are given in the histogram shown in Fig. 7. The sample matrix V based on the fragments of signals belonging to the first and fourth classes and the sample matrix W based on the fragments of signals belonging to the second, third, and fifth classes were used in the classifica-

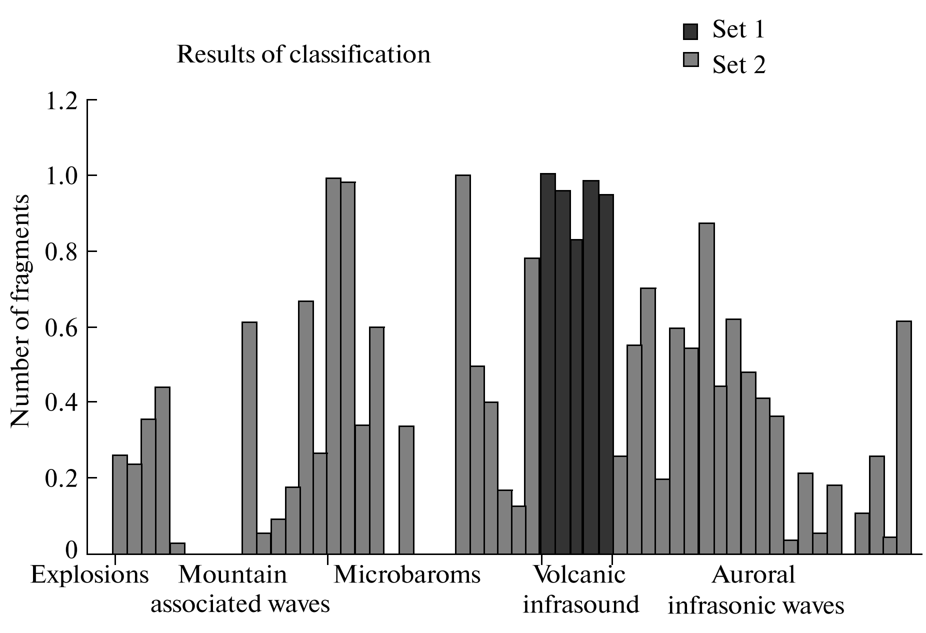

Fig. 7. Histogram of dividing signals into two sets at $c_\alpha = 1500$.

tion algorithm. The abscissa is the classes of signals; the shades of gray correspond to the sets of signals. The height of the columns is proportional to the portion of signal fragments that satisfy condition (2) (i.e., the number of fragments (for which (2) is fulfilled) related to the total number of fragments in the signal). If the column height exceeded 0.8, the signal was considered to belong to set 1 (classes 1 and 4); otherwise, the signal was considered to belong to set 2 (classes 2, 3, and 5). Analyzing the graph, one can conclude that all signals from the first set were quite correctly classified. Only three signals from the second set were erroneously referred to the first set. The results suggest that the quality of the classification algorithm is good.

7. CONCLUSIONS

An approach related to the theory of statistical testing hypotheses was used to solve the classification problem [4]. An empirical class model was constructed using the caterpillar method; in this case, the signal was divided into fragments of the same duration (600 readings) [8].

At the first stage, a set of signal fragments 600 readings in length was studied, the possibilities of separating signal classes were determined, signal fragments were tested for their self-descriptiveness, and the levels of accepted hypotheses were determined according to informative signal fragments.

At the second stage, on the whole, signals were classified according to the chosen thresholds of classifying signal fragments.

As a result, a rather well distinction was found between the two sets of signals. The first set included the signals of the first (explosions) and fourth (volcanic infrasound) classes; the second set included the signals of the second (mountain associated waves) third (microbaroms), and fifth (auroral infrasonic waves) classes.

For the control sample, when referring signals to these two sets, all signals from the first set (explosions and volcanic infrasound) were correctly classified. As for the second set (microbaroms, mountain associated waves, and auroral infrasonic waves), 3 out of 51 signals were erroneously referred to signals of the first set.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 11-07-00338-a and 11-05-00890) and state contract nos. 70/GF/N-11 and 14.740.11.0203.

REFERENCES

1. P. Grenard, "Status of the IMS Infrasound Network," in *Proceedings of the Infrasound Technology Workshop, October 18–22, 2010, Tunis, Tunisia.*
2. S. N. Kulichkov, G. A. Bush, A. I. Svertilov, "New Type of Infrasonic Arrivals in the Geometric Shadow Region at Long Distances from Explosions," Izv., Atmos. Ocean. Phys. **38** (4), 397–402 (2002).
3. N. D. Tsybulskaya, S. N. Kulichkov, and A. I. Chulichkov, "Analysis of Classification Possibility Infrasound Signals from Different Sources Based on Correlation Ability," in *Proceedings of the International Conference "Comprehensive Nuclear-Test Ban Treaty: Science and Technology, June 8–10, 2011, Vienna, Austria.*
4. E. L. Lehman, *Testing Statistical Hypotheses* (Wiley, New York, 1959; Nauka, Moscow, 1979).
5. Library of Typical Infrasonic Signals, Matlab Files of Infrasonic Sensor Data with Online Analysis Parameters for Each 2.5 min of Data (Geophysical Institute, University of Alaska, 2007).
6. Yu. P. Pyt'ev and A. I. Chulichkov, *Methods of Morphological Image Analysis* (Fizmatlit, Moscow, 2010) [in Russian].
7. S. N. Kulichkov, A. I. Chulichkov, and D. S. Demin, *Morphological Analysis Infrasonic Signals in Acoustics* (Novyi Akropol', Moscow, 2010) [in Russian].
8. N. E. Golyandina, *The Caterpillar-SSA Method: Time Series Analysis* (St. Petersburg State University, St. Petersburg, 2004) [in Russian].
9. Yu. P. Pyt'ev, *Matematical Modeling Methods for Measuring and Computing Systems* (Fizmatlit, Moscow, 2004) [in Russian].