# The Method and Algorithmic Support of the Determining the Presence of Information Message in a Priori an Uncertain Sound Signal

**Vasiliy E. Gai**

Institute of Radioelectronics and Information Technologies
Nizhny Novgorod State Technical University n.a. R.E.Alekseev
K. Minina, 24, Nizhny Novgorod, Russian Federation, 603950
vasiliy.gai@gmail.com

**ABSTRACT**

*The problem of voice activity detection (VAD) is one of the main in the system of signal's processing. In this work VAD is considered in terms of finding a priori indefinite useful signal in the observed signal. Suggested solution is based on the theory of active perception. Using this theory allows to significantly (with probability 1) identify the system elements and the structure of the signal and structure of connections between them with a given $\varepsilon$-fidelity by Kolmogorov. Assumed solution is without lacks detectors signals using in the basis a correlation receiver. In the work also performed analysis of existing algorithms of voice activity detection. Common lack of these algorithms is using rather complicated algorithms estimate the parameters of the noise and signal, and using Fourier and wavelet transforms to calculating the spectra. In the offered algorithm in the construction of the signal spectrum are used only addition and substraction operations, wherewith achieving a high processing speed. These results of speech signals segmentation demonstrate the effectiveness of offered algorithm.*

**Keywords:** digital signal processing, theory of active perception, voice activity detection.

**2010 Mathematics Subject Classification:** 60G35, 93E10, 94A13.
**1998 Computing Classification System:** C.3, H.5.5, I.5.4.

## 1 Statement of the Problem

From the point of view of an observer, sound signal has a useful signal (informational massage) and a noisy signal. The useful signal is an information, which needed for the observer to decision making in the context of the problem, and the noise is all the other information. In this work, the useful signal is considered as a systemic formation. In this case it must contain the structural elements and connections. Consequently, the problem of useful signal detection can be reduced to the problem of structural elements dedication of signal and connections between them.

Confirmation of this approach is the statement of the author's work (Haykin, 2001; Tikhonov, 1984), in which declare that without presence information about the signal is impossible to distinguish the useful signal from the noise.

The problem of informational message dedication from the received signal in statistical radio engineering corresponds to the problem of signal detection.

Let be on a final time interval $[0;T]$ is accepted the sound signal $f(t)$, which is a function from the useful signal $s(t,\lambda)$ and the noise $n(t)$:

$$f(t) = F(s(t,\lambda), n(t)), 0 \leq t \leq T, \tag{1.1}$$

where $\lambda = \lambda_1, \ldots, \lambda_m$ is vector of signal parameters. It is assumed, that the direct observation $f(t)$ of the received signal is only available.

Let be unknown the fact presence or absence of informational massage $s(t,\lambda)$ in the received signal $f(t)$. We record the received signal $f(t)$ (1.1) in the following way:

$$f(t) = \theta \cdot s(t,\lambda) + n(t), 0 \leq t \leq T. \tag{1.2}$$

Here $\theta$ is random quantity, which can take two value: $\theta = 0$ (message is absent) and $\theta = 1$ (message is present). It is needed by adopted implementation $f(t)$ on the interval $T$ to make a decision: present or absent the useful signal $s(t,\lambda)$.

The described problem is a typical formulation of the problem of signal detection on the background noise.

Decision making about the presence or absence of the useful signal is always accompanied by two types of errors:

1. take a wrong decision about presence of a signal (type I error);
2. take wrong decision about absence of signal (type II error).

Thereby, the aim of the work is reliable (with probability 1) detection of system elements of the signal and the structure connections between them with specified $\varepsilon$-fidelity by Kolmogorov.

## 2 Correlation Receiver

The classical approach to the useful signal detection from the position of statistical radio engineering, built on the basis of correlation receiver (see fig.1), assumes the decision of the Fredgolm equation of the first kind:

$$g(\overline{T}) = \int_{x \in \overline{T}} s(t,x) f(x) dx, \tag{2.1}$$

where $g(t)$ is the observed signal, $f(x)$ is a priori unknown change in the time of amplitude of registration signal at the device input, $s(t,x)$ is the useful signal.

Result is obtained after filtration, in the context using the Fredgolm's equation, is incorrect, as minor changes in the recorded signal $f(t)$ are able to lead to acceptably large changes in the solution.

By incorrect problem is understood the problem which isn't done at least one of the conditions, characterized correctly formulated problem (Parker, 1989). The concept of "correct problem" was introduced by Hadamard in 1923.

On the other hand, the disadvantage of useful signal detection on the basis of the correlation receiver consists in the necessity of knowing the signal $s(t,x)$. In consideration of the problem of detection is solved in the conditions of a priori uncertainty, this method cannot work.

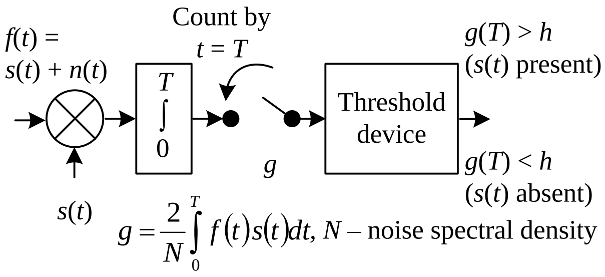

Figure 1: Correlation Receiver

## 3 System Approach

The assumed solution of the problem of signal detection is based on the use of a system approach to signal processing. From the point of view of system analysis, the problem of detection of the useful signal can be divided into three stages (see fig.2):

1. preparation of the data;
2. analysis of data;
3. decision.

Systems approach to signal analysis realizes the theory of active perception (Utrobin, 2004). In the context of this theory the signal is considered as a system formation.

From the theory of active perception these steps of data processing can be formulated as follows (see fig.2): the formation of the initial description, the formation of the system of features, classification.

The theory of active perception describes operations that allow to distinguish the signal structural elements and their connections. Integral transformation is used for detection system elements in the theory of active perception, and to identify links between elements is used spatial differentiation. The result of the identification of the differential structure of the signal is spectral description.

Integration and differentiation transformations together are a composition which is called $U$-transform: $U = d \circ \int$.

It should be noted that in the theory of active perception, compared to a Fredgolm equation, the order of operations applied to the signal is reverse: at first the integration is performed, and then is differentiation. This allows to reduce the solution to the incorrect problem to correct.

Transformations of integration and differentiation for one-dimensional signals realized with the help of four-base dimensional filter-coatings ($F_0$, $F_1$, $F_2$, $F_3$, see fig.3).

## 4 The proposed method for the detection of the useful signal

Let us consider the proposed scheme of information transformations (see fig.4), based on the theory of active perception without shortcomings of correlation receiver and adapted to the specifics of one-dimensional signals.

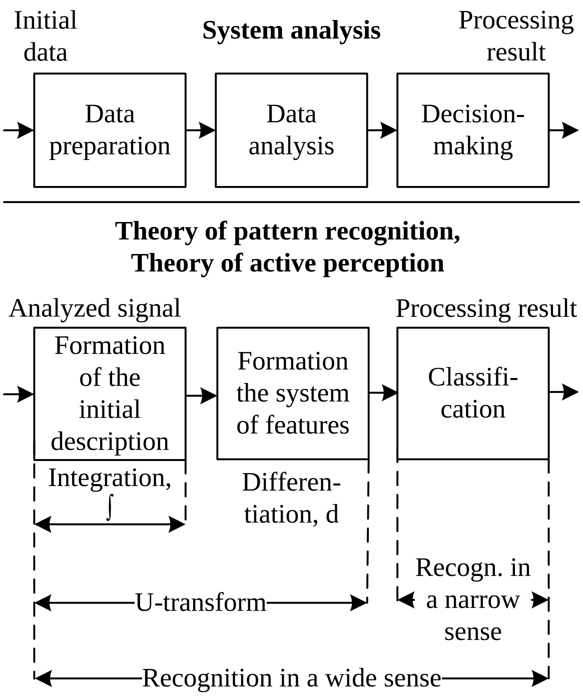

Figure 2: Stages of Information processing

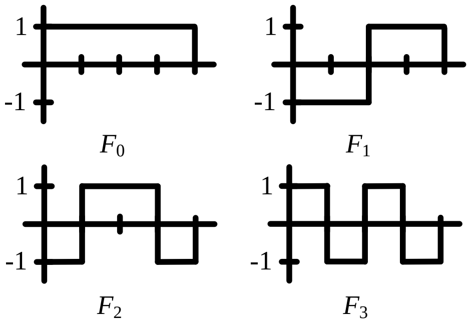

Figure 3: Basis functions

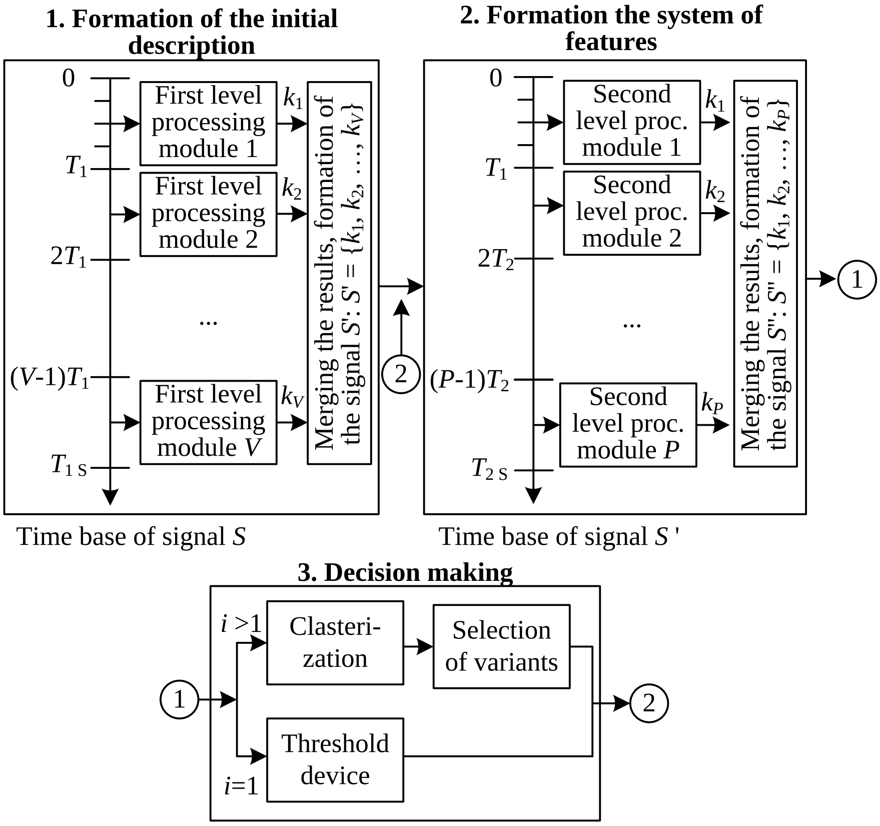
1. Formation of the initial description
0
$T_1$
$2T_1$
$(V$-1$)T_1$
$T_{1\,S}$
First level processing module 1
First level processing module 2
...
First level processing module $V$
$k_1$
$k_2$
$k_V$
Merging the results, formation of the signal $S'$: $S' = \{k_1, k_2, \ldots, k_V\}$
Time base of signal $S$
2
2. Formation the system of features
0
$T_1$
$2T_2$
$(P$-1$)T_2$
$T_{2\,S}$
Second level proc. module 1
Second level proc. module 2
...
Second level proc. module $P$
$k_1$
$k_2$
$k_P$
Merging the results, formation of the signal $S''$: $S'' = \{k_1, k_2, \ldots, k_P\}$
1
Time base of signal $S$ '
3. Decision making
1
$i$ >1
Clasteri-zation
Selection of variants
$i$=1
Threshold device
2

Figure 4: The proposed scheme

This device differs from the correlation receiver that correlation receiver outputs a binary decision: yes or no useful signal, and proposed is determines the location of the useful signal in the received signal (performs segmentation of the received signal).
The stages of transformation of the original signal are shown in fig.4 and included:

1. formation of the initial description: calculation $U$-transform of $S$ signal and construction of its envelope (to create a signal $S'$) by by mean square deviation the first second and third coefficients $U$-transform;

2. formation of features' system: calculation $U$-transform envelope (to create a signal $S''$) is used only zero-coefficient of generated spectral representation;

3. decision-making: clustering of signal $S''$ and selecting the best option for segmentation.

Steps 2 and 3 are repeated iteratively $M$ times, thus generated $M$ segmentation results of the received signal with different accuracy of useful signal localization.
Processor modules, which used on formation steps of the initial description and features' system, are shown on the fig.5 (modules operate in parallel portions of the signal). Each section of the signal, which is sent to the module, is divided into four segments.
Table of symbols used in schemes:

- $i$ - the level of analysis, $i = 1 : M$.

- $M$ - number of the levels of analysis;

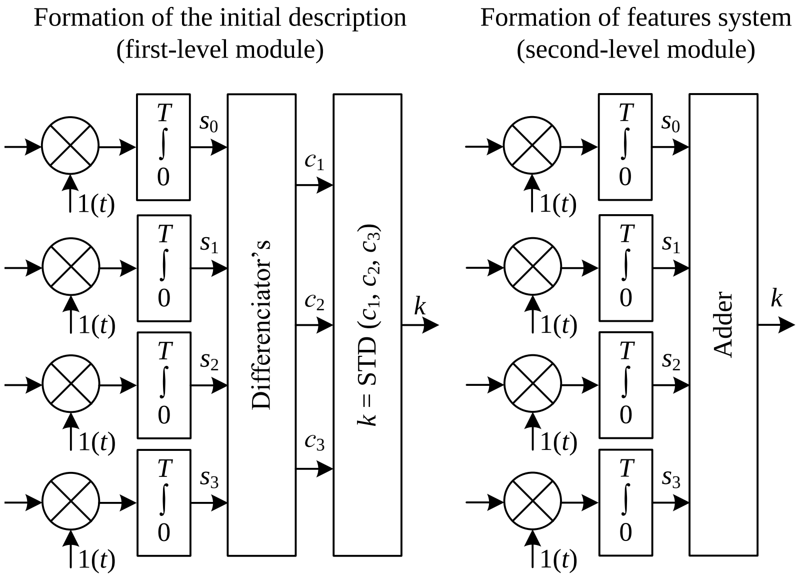

Figure 5: Processor modules

- $V$ - number of the $S'$ signal's segments, $T_{1S} = V \cdot T_1$;
- $P$ - number of the $S''$ signal's segments, $P = 2^i$, $T_{2S} = P \cdot T_2$;
- $s_0, s_1, s_2, s_3$ - results of integration;
- $c_0, c_1, c_2, c_3$ - spectral coefficients;
- $k$ - standard deviation of $c_1, c_2, c_3$ (first-level module);
- $k = s_0 + s_1 + s_2 + s_3$ (second-level module).

In the first-level processing module, in the block "Differentiator's", are used filters $F_1$, $F_2$, $F_3$, which used to segments integration results and in the second-level processing module in the block "Adder" is used filter $F_0$ (see fig.5).
Threshold device (see fig.4) with $i = 1$ check the presence of activity in the analyzable signal:

- if $S_0^{''} < T$ ($T$- threshold), so the useful signal in the observed signal is absent and subsequent analysis of this signal becomes pointless;
- if $S_0^{''} \geq T$, so analysis of the hierarchical development of the signal begins from the third-level, as on the second-level are not so much information (only two coefficients) for localization of the useful signal.

Hereby, this scheme consists of several elements, like correlation receiver, however, essentially differ from it.
Clustering of the signal $S''$ is satisfied by an K-means algorithm: $[C, N]$ = KMEANS$[S'', i]$, where $C = \{c_{ij}\}$ is a set of centroids, generated by the signal $s_i$, $N = \{n_{ij}\}$ is a set of signal's samples which relevant to centroids, $i$ - the number of segments on the $i$-th level of the signal analysis. Then, it is satisfied the processing of the clustering results, which consists from 3 steps:

```
Data: clusters $C$
Result: norm. clusters $C$
for i = 3:M do
    for j = 1:i do
        $c_{ij} = c_{ij} - min(c_i)$ $c_{ij} = c_{ij}/max(c_i)$
    end
end
```

**Algorithm 1:** Normalization of clusters' centroids on the $i$-th level

```
Result: initialized maps $P$
for i = 3:M do
    for j = 1:(i-1) do
        for k = 1:$2^{(i-1)}$ do
            $p_{ij}(k) = 0$
        end
    end
end
```

**Algorithm 2:** Initialization of maps of signal segmentation

Table of symbols used in algorithms:

- $i$ - number of level;
- $j$ - number of map on the level;
- $c_i$ - set of centroids on the $i$-th level (it is assumed that on the $i$-th level centroids are sorted by amplitude);
- $c_{ij}$ - $j$-th centroid on the $i$-th level;
- $n_{ij}$ - set of samples of the clustered signal put to $j$-centroid (number of maps on the unit less than the decomposition level);
- $|n_{ij}|$ - number of samples of signal $S''$ which refer to $j$-centroid;
- $M$ - number of levels for analysis.

Note: centroid with value of $c_{ij} = 0$ doesn't include in further analysis, as suppose that it applies to a part of the signal, on which the useful signal is absent.
After generation of set maps by each map $p_{ij}$ calculates the coefficient:

$$K_{ij} = |max(S_i^{''}(M_{NS})) - min(S_i^{''}(M_S))|, \tag{4.1}$$

where $M_{NS}(k) = \begin{cases} 1, p_{ij}(k) = 0, \\ 0, p_{ij}(k) = 1, \end{cases}$ $M_S(k) = \begin{cases} 1, p_{ij}(k) = 1, \\ 0, p_{ij}(k) = 0, \end{cases}$ $i = \overline{3 : M}, j = \overline{1 : (i-1)}, k = \overline{1 : 2^{(i-1)}}$.
Described coefficient is calculated as follows: calculate the absolute difference between maximum value of the signal part, pertaining to pause and minimum value pertaining to the part

```
Data: initial segment maps $P$
Result: modified segment maps $P$
$P = \{p_{ij}\}, j\overline{1, i-1}$
for i = 3:M do
    for j = 1:(i-1) do
        if j == 1 then
            for k = 1:$|n_{i1}|$ do
                $p_{i1}(n_{i1}(k)) = 1$
            end
        else
            $p_{ij} = p_{ij-1}$;
            for k = 1:$|n_{ij}|$ do
                $p_{ij}(n_{ij}(k)) = 1$;
            end
        end
    end
end
```

**Algorithm 3:** Generation of maps of signal segmentation

activity. The map, for which the coefficient is equal to $1$ or less $0.1$ from the set of the map is excluded.

Next from set of the maps on each level is selected one map which has the highest signal coverage. This is prevented loss of data in the voice activity signal, however, leads to an increase in errors of the first kind. As a result, on each level is selected one map of the signal segmentation. Therefore, the result of the work of algorithm is a set of maps $P = \{p_i\}, i = \overline{3, M}$.

## 5 Detection of the speech signal

### 5.1 Description of the task

In speech processing the problem of the detection is segmentation of signal speech activity (useful signal) and pause (noise). Results of the segmentation of audio signal are used by identification/verification of the speaker, speech recognition (B. Putra, 2013; Ye, 2013; Gopu and Neelaveni, 2013; Sinha and Sanyal, 2009). In the standard G.729b, describing the narrowband speech codec, the method of voice activity detection is used to compress the signal.

The spacing interval in the speech signal, usually, is defined the cause of its formation. In the colloquial speech the pause is occurred when organs of articulation are plugged condition that is associated with the pronunciation of the occlusive consonants. The duration of these pauses is 0.1 seconds (Zellner, 1994). While reading the duration of pauses on linguistic boundaries syntagmas (sequence of words or morphemes) not exceeding 0.75 seconds and between sentences vary from 0.5 to 1.5 seconds.

One of the problems that arises when a pause is noisiness of the initial signal. Therefore, algorithm of signal segmentation must be resistant to the presence of various kinds of noise in the signal. It should be noted that the human auditory system also provides mechanisms of noise perception (Altman, Vartanian, Andreeva, Vaitulevich and Malinina, 2005):

1. efferent feedback reduces the sensitivity of the cochlea by the noise presence and reduces the risk of overload protects the cochlea from the damage by loud sound (Nicholls, Martin, Fuchs, Moore and Stuart, 2011);

2. binaural interaction of the right and left channels of the auditory system allow to improve the speech intelligibility;

3. stapedius muscle of the middle ear allows to increase the stiffness the chain of the auditory ossicles, which leads to a reduction of their energy and to compensate for high-intensity noise.

## 5.2 Review of existing algorithms

Most of the algorithms fulfilling segmentation of audio signal, work as follows:

1. source signal is divided into segments;

2. evaluates the parameters of background noise; at the same time is used the assumption that characteristics of background noise practically remain constant for a long period of time;

3. for determining the presence or absence of signal in the $i$-th segment, its characteristics in this segment are compared with calculated characteristics of noise (used by a critical rule).

Let us consider some of the algorithms which is used to extract the active voice.

Described segmentation method of audio signal in (Sohn and Sung, 1998) to noise parameters estimation used autoregressive method. The calculated values noise parameters, the signal processing are updated with data on the probability of the presence of the voice signal in this segment. To improve the accuracy of the speech allocation with small amplitude using hidden Markov model of the first order.

In (Lee and Hasegawa-Johnson, 2009) is proposed improving the method of assessment of noise described in (Sohn and Sung, 1998), namely, eliminate defects, connected with impossibility to accurate predict the presence of the speech signal at the analyzable signal.

Algorithm for detection voice activity (Saeedi, Ahadi and Faez, 2013) as a decision rule is used the method of finite vectors and to estimate the parameters of the signal and noise - wavelet transform.

In (Tan, Borgstrom and Alwan, 2010) as a decision rule is used method, based on the likelihood test. The likelihood ratio for signal parts, related to speech and noise is calculated separately. Description of the signal segments is performed using a Fourier Transform.

It should be noted that the existing methods of allocation pauses usually try to optimize one of the followings parameters: accuracy of the allocation pauses, temporary delay, computational complexity.

Consider the existing methods can mark their common lack: to build models of signal and noise used by Fourier or wavelet transform, which computing require considerable resources. The proposed method is free of this lack.

## 5.3 Computing Experiment

Perform testing described algorithm by segmentation commands and connected speech. Compare the results with the results of segmentation on the basis of the existing algorithms.
Testing of segmentation algorithms commands were performed on the audio signals recorded by five speakers (signal duration - 120 seconds). The following commands were announced: ”stop”, ”forward”, ”back”, ”left”, ”right”, ”turn on”, ”turn off”, ”signal”, ”motion”, ”evacuation”. For testing algorithms under segmentation of connected speech was used database, which includes 12 records duration of 60 seconds. Sampling rate of recorded signals is 16kHz, depth coding is 16 bit.
Estimation of stability segmentation algorithms to signal distortion was carried when exposed to signal following noise:

1. theoretical (model):
   - normal;
   - uniform;
2. real:
   - highway noise;
   - motor sound of a military truck;
   - sound of the helicopter engine;

An example of the segmentation of a noisy signal is shown in fig. 6 (signal is distorted by additive Gaussian noise, signal / noise ratio = -8 dB).

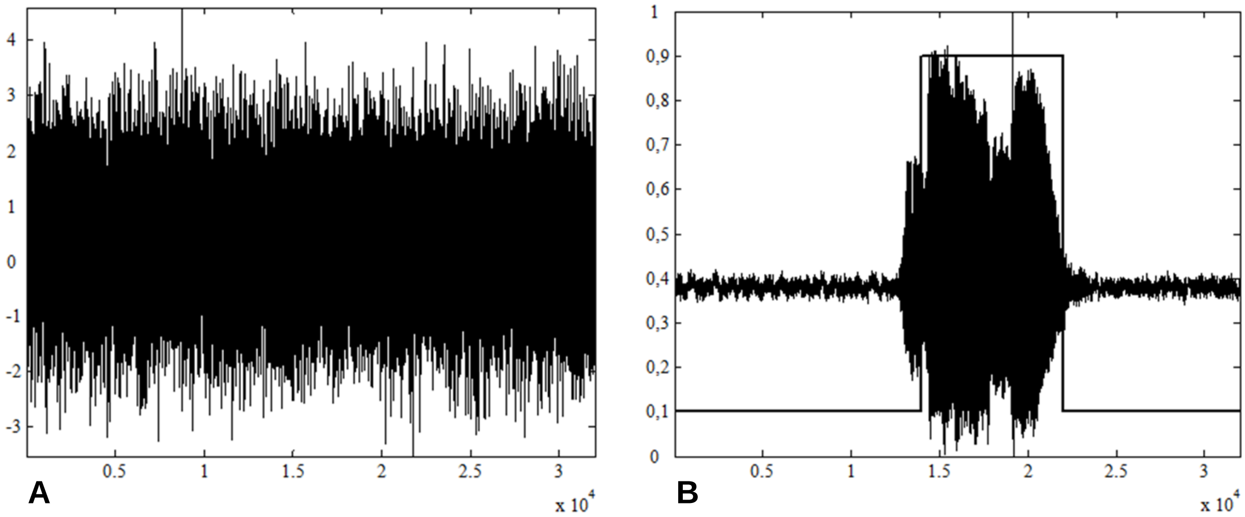

Figure 6: Segmentation of signal: a) noisy signal, b) the initial + segmentation result

In table 1 and table 2 are shown values of segmentation faults of the first kind (first column) and second kind (second column) by segmentation commands and connected speech. Dash in the table means that the algorithm is not able to detect a person's voice.

Table 1: Results of testing algorithms on commands

| Noise | SNR | Alg. 1 [9] | | Alg. 2 [10] | | Alg. 3 [11] | | Alg. 4 [12] | | Proposed | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Uniform | 20 | 0.3 | 7.6 | 2.4 | 1.7 | 9.0 | 0.6 | 0.2 | 4.6 | 2.4 | 1.6 |
| | 10 | 0.9 | 22.6 | 5.8 | 1.2 | 11.1 | 0.1 | 1.6 | 14.5 | 2.6 | 1.56 |
| | 0 | - | - | 16.3 | 0.6 | 16.3 | 0.0 | 8.8 | 5.1 | 3.7 | 2.2 |
| | -10 | - | - | 24.8 | 0.5 | 24.8 | 0.0 | 9.9 | 25.2 | 3.1 | 11.3 |
| Normal | 20 | 0.1 | 4.4 | 4.8 | 1.6 | 10.2 | 0.3 | 1.0 | 12.5 | 2.8 | 1.4 |
| | 10 | - | - | 9.9 | 0.5 | 13.4 | 0.0 | 6.8 | 9.7 | 2.8 | 1.4 |
| | 0 | - | - | 24.8 | 0.5 | 23.3 | 0.1 | 12.1 | 24.3 | 4.6 | 5.5 |
| | -10 | - | - | 24.8 | 0.0 | 24.8 | 0.0 | 15.8 | 9.5 | 6.8 | 15.8 |
| HighWay | 20 | 0,4 | 3.9 | 2.1 | 2.1 | 8.3 | 3.2 | 0.4 | 24.2 | 2.4 | 1.6 |
| | 10 | 2.7 | 7.5 | 3.4 | 1.5 | 9.0 | 2.9 | 0.5 | 35.7 | 2.8 | 1.4 |
| | 0 | - | - | 8.0 | 0.7 | 11.5 | 6.7 | 1.8 | 36.4 | 7.2 | 1.1 |
| | -10 | - | - | 23.0 | 0.3 | 20.4 | 10.7 | 3.6 | 37.8 | 2.7 | 19.3 |
| Truck | 20 | 0.4 | 4.1 | 1.7 | 3.3 | 7.7 | 4.7 | 0.2 | 28.7 | 2.4 | 1.6 |
| | 10 | 0.4 | 5.7 | 1.7 | 2.8 | 7.6 | 7.8 | 0.2 | 38.0 | 2.4 | 1.6 |
| | 0 | 0.4 | 25.6 | 1.6 | 4.6 | 6.2 | 19.8 | 0.1 | 53.2 | 2.8 | 1.4 |
| | -10 | 1.0 | 37.1 | 2.4 | 14.4 | 5.6 | 35.7 | 0.8 | 44.9 | 6.7 | 2.8 |
| Helic. | 20 | 0.6 | 4.3 | 1.9 | 2.2 | 8.1 | 3.3 | 0.4 | 19.6 | 2.2 | 1.7 |
| | 10 | 2.2 | 5.7 | 4.2 | 1.7 | 9.2 | 2.4 | 1.1 | 28.1 | 2.6 | 1.5 |
| | 0 | 1.1 | 37.8 | 12.7 | 0.1 | 12.4 | 1.8 | 3.2 | 18.2 | 3.9 | 1.8 |
| | -10 | - | - | 23.4 | 0.0 | 20.9 | 1.7 | 11.9 | 13.5 | 8.2 | 9.2 |

## 6 Conclusion

An algorithm for detecting a priori undetermined signal in the observed signal was developed. The offered algorithm is based on the theory of active perception. Using this theory allows to reveal structural elements of the signal and the connection between them, and, consequently, to detect the useful signal.

The offered algorithm by the parameter ”segmentation accuracy” doesn't inferior to the existing algorithms by speech signals segmentation, and in some cases - shows the best results. Segmental signal must contain parts of pauses, the location of these areas does not matter, if not as the pause will be selected segment with a minimum amplitude Algorithm allows to control the signal segmentation accuracy by selecting the level of decomposition, on which the segmentation is done.

Table 2: Results of testing algorithms on commands

| Noise | SNR | Alg. 1 [9] | | Alg. 2 [10] | | Alg. 3 [11] | | Alg. 4 [12] | | Proposed | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Uniform | 20 | 5.4 | 0.2 | 2.2 | 1.9 | 8.2 | 0.1 | 16.3 | 0.7 | 7.6 | 2.6 |
| | 10 | 4.5 | 9.0 | 6.5 | 1.3 | 17.6 | 0.0 | 37.7 | 0.9 | 9.0 | 3.0 |
| | 0 | 0.0 | 17.5 | 41.0 | 0.1 | 37.1 | 0.0 | - | - | 9.3 | 2.5 |
| | -10 | 0.0 | 17.5 | - | - | - | - | - | - | 11.8 | 5.9 |
| Normal | 20 | 8.9 | 0.2 | 4.5 | 1.5 | 13.0 | 0.1 | 26.1 | 0.7 | 8.1 | 2.8 |
| | 10 | 14.8 | 1.8 | 22.3 | 0.4 | 26.2 | 0.0 | 68.8 | 1.4 | 10.5 | 2.2 |
| | 0 | 0.0 | 17.5 | - | - | - | - | - | - | 9.7 | 3.5 |
| | -10 | 0.0 | 17.5 | - | - | - | - | - | - | 11.4 | 12.2 |
| HighWay | 20 | 7.6 | 0.2 | 0.7 | 2.5 | 4.8 | 0.3 | 7.3 | 1.9 | 7.7 | 2.6 |
| | 10 | 12.6 | 0.1 | 1.3 | 2.0 | 10.6 | 0.2 | 12.4 | 1.8 | 8.4 | 2.4 |
| | 0 | 13.6 | 6.0 | 3.8 | 1.2 | 22.0 | 0.2 | 26.0 | 1.7 | 8.1 | 3.1 |
| | -10 | 22.6 | 10.1 | 41.5 | 0.3 | - | - | - | - | 12.3 | 4.6 |
| Truck | 20 | 5.1 | 0.2 | 0.6 | 2.8 | 0.1 | 14.7 | 5.1 | 2.5 | 7.5 | 2.7 |
| | 10 | 2.0 | 6.8 | 0.7 | 2.5 | 0.0 | 15.3 | 7.9 | 2.4 | 8.5 | 2.5 |
| | 0 | 0.4 | 14.5 | 1.3 | 2.0 | 0.1 | 15.3 | 15.1 | 2.3 | 9.2 | 2.2 |
| | -10 | 0.5 | 14.8 | 4.7 | 4.3 | 0.2 | 15.3 | 32.8 | 2.7 | 9.2 | 4.5 |
| Helic. | 20 | 6.5 | 2.0 | 0.7 | 2.5 | 1.0 | 10.7 | 9.2 | 1.5 | 8.2 | 2.5 |
| | 10 | 5.0 | 8.4 | 1.6 | 1.9 | 2.5 | 10.7 | 17.1 | 1.3 | 9.4 | 2.6 |
| | 0 | 3.3 | 12.9 | 9.6 | 1.6 | 6.2 | 10.7 | 37.8 | 1.5 | 10.1 | 3.4 |
| | -10 | 3.7 | 15.4 | 35.0 | 4.6 | 14.6 | 10.7 | - | - | 12.6 | 10.5 |

It is shown that the algorithm can be used in solving the problem of segmentation commands and connected speech. The advantages of offered algorithm is low computational complexity, ease of implementation and the absence of adjustable parameter.

## Acknowledgment

This work was supported by the Ministry of Education and Science of Russian Federation under contract No 02.G25.31.0061 12/02/2013 (Government Regulation No 218 from 09/04/2010).

## References

Altman, J. A., Vartanian, I. A., Andreeva, I. G., Vaitulevich, S. F. and Malinina, E. S. 2005. Tendencies in auditory physiology, *Uspekhi fiziologicheskikh nauk* **36**: 3–23.

B. Putra, S. 2013. Secure speaker verification at web login page using cepstral features, *International Journal of Applied Mathematics and Statistics* **35**: 92–107.

Gopu, G. and Neelaveni, R. 2013. A proposed novel approach of voice controlled prosthetic arm for differently abled, *International Journal of Imaging and Robotics* **9**: 37–47.

Haykin, S. 2001. *Communication Systems*, Wiley, New York.

Lee, B. and Hasegawa-Johnson, M. 2009. Estimation of high-variance vehicular noise, *In-Vehicle Corpus and Signal Processing for Driver Behavior* pp. 221–232.

Nicholls, J. G., Martin, R. A., Fuchs, P. A., Moore, J. W. and Stuart, A. E. 2011. *From neuron to brain*, Sinauer associates, Sunderland, MA.

Parker, S. B. 1989. *McGraw-Hill Dictionary of Scientific and Technical Terms*, McGraw-Hill, New York.

Saeedi, J., Ahadi, S. M. and Faez, K. 2013. Robust voice activity detection directed by noise classification, *Signal, Image and Video Processing* **2**: 1–12.

Sinha, T. S. and Sanyal, G. 2009. Understanding of speech and speaker model for recognition of a language, *International Journal of Artificial Intelligence* **9**: 107–125.

Sohn, J. and Sung, W. 1998. Voice activity detector employing soft decision based noise spectrum adaptation, *Proceedings of IEEE International Conference Acoustics, Speech and Signal Processing*, IEEE, pp. 365–368.

Tan, L. N., Borgstrom, B. J. and Alwan, A. 2010. Voice activity detection using harmonic frequency components in likelihood ratio test, *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, IEEE, pp. 4466–4469.

Tikhonov, V. 1984. *Optimal signal reception*, Radio and Communication, Moscow.

Utrobin, V. A. 2004. Physical interpretation of the elements of image algebra, *Advances in Physical Sciences* **47**: 1017–1032.

Ye, S. 2013. Speech emotion analysis and recognition based on the pca feature extraction model, *International Journal of Applied Mathematics and Statistics* **51**: 127–135.

Zellner, B. 1994. *Pauses and the temporal structure of speech in E. Keller (Ed.) Fundamentals of speech synthesis and speech recognition*, Wiley, New York.