A Study of the Stability of Sound Signal Description[1,2]

V. E. Gai

Nizhni Novgorod State Technical University n. a. R. E. Alekseev, K. Minina street, 24, 603950 Nizhni Novgorod
e-mail: iamuser@inbox.ru

Abstract—The paper describes the method of allocation of the most distortion-resistant sound signal sections. Preliminary signal description is created by means of U-transformation. Results of computing experiments are adduced.

DOI: 10.1134/S1054661814040051

1. INTRODUCTION

The solution of any problem of digital processing of a sound signal begins with the creation of a signal description. There are various methods to create a sound signal description. These are Fourier transforms, wavelet analysis, and *U*-transformation. The question of stability of a sound signal description arises while solving a signal-processing problem in conditions of noise. One such task is media data analysis, which consists in locating a set sound fragment in a sound stream or a file.

The present work is devoted to the study of the stability of description distortion created by means of *U*-transformation [1]. *U*-transformation is the basic transformation used in the theory of active perception (TAP). Within TAP a sound signal is considered as system unit. For detection of system elements integral transformation is used, and for identification of communications between elements, spatial derivation. A spectral signal description is the result of differential structure identification.

Suppose $f(t)$ is an analyzed sound signal observed over a finite amount of time. The result of the application of *U*-transformation to the signal f is a multilevel (coarse-fine) spectral representation of $D = \{d_{ij}\}$, $i = \overline{1, K}$, $j \in \overline{1, M_i}$, where K is a number of decomposition levels, M_i is the number of signal segments on i decomposition level, d_{ij} is the spectrum including L spectral coefficients (number of filters used), $d_{ij}\{k\}$ is k's spectral coefficient ($k = \overline{1, L}$), f_{ij} is a waveform segment of f on which the range d_{ij} is calculated. *U*-transformation is realized by means of a set of filters $\{F_i\}$. Figure 1 shows an example set consisting of four filters.

Let us assume for filter sets $\{F_i\}$ that $(+1 \longrightarrow 1)$ and $(-1 \longrightarrow 0)$. We obtain the binary operator sets $\{V_i\}$. The set-theory operations of integration (addition) and intersection (multiplication) are admissible for these operators. As an algebraic result we have $A_V = \langle\{V_i\}: +, \times q\rangle$ [1].

In the function A_V there are the following functional groups:

1—P_{ni}—groups of three elements (called complete) are formed with the three operators (V_i, V_j, V_k), which correlate as follows: $V_i + V_j + V_k = e_1$ (e_1 is the unit). The complete group admits two descriptions: intersection operation $V_iV_jV_k$ (multiplication operation, number of inversions is even) and integration operation $\bar{V}_i + V_j + V_k$ (addition operations, number of inversions is odd). The number of possible patterns of complete groups, taking into account the inversions of operators, is four. The complete group pattern is a compact of four connected elements.

2—P_{si}—groups of four elements (called closed) are formed with the four operators (V_i, V_j, V_n, V_m), where $(V_i, V_j, V_k) \in P_{ni}$, $(V_n, V_m, V_k) \in P_{nj}$, with the description $V_iV_j + V_nV_m$ (where the desired number of operator inversions is odd) and the unit is $V_i + V_j + V_n + V_m = e_1$. The number of possible closed group patterns, taking into account operator inversions, is eight. The closed group pattern is a compact of eight connected elements.

Sets $\{P_{ni}\}$, $\{P_{si}\}$ are finite and have a capacity of 35 and 105, respectively (these sets are formed by 16 operators). The sets $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$ are the sets of standards used to solve the problem of recognition of the object of study in standard space. The notation of the complete group in addition operation is P_{nia}, in multiplication operation is P_{nim}.

The formation of a sound signal description within a *U*-transformation is based on the use of the algebra of groups. Spectral and correlation analyses are carried out using closed and complete groups. Complete groups make it possible to explore the correlation of communications between operators. Closed groups

[1] The article was translated by the authors.

[2] This paper uses the materials of the report submitted at the 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies," Samara, The Russian Federation, September, 23–28, 2013.

Received April 10, 2013

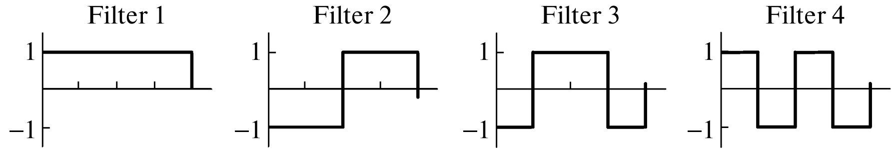

Fig. 1. Basis functions.

make possible to explore the correlation of communications between complete groups. Drawing an analogy between the algebra of groups and language, it is possible to apply the following comparison: an operator is a language alphabet, a complete group is a word, and a closed group is a phrase.

A closed group is formed on the basis of two full groups, and of three operators (see Fig. 2).

In [1] it is noted that the greatest resistance to distortions belongs to descriptions created on the basis of closed groups, and the least is to those created on the basis of operators.

2. SIGNAL RESEARCH ALGORITHM

Let us define indicators that can be used for allocation of stable segments of a signal *A*. For this purpose we will analyze a signal *A*:

1—a single-level *U*-transformation of a signal A is calculated, the signal is divided into *M* segments. For each signal segment the set of complete (P_{ni}) and closed (P_{si}) groups is calculated. The signal description represents the set of values of operators and the groups calculated on each signal segment:

$$D_A = \bigcup_{i=\overline{1,M}} \langle F_i, P_{ni,i}, P_{si,si} \rangle;$$

2—under signal *A* by means of an operator of distortion *B* and *C* signals are formed (distortion level of signal *A* has to be constant, however, considering that if signal *A* is superimposed with random noise, signals *B* and *C* won't be equivalent):

$$B = \mathrm{NOISE}[A],\ C = \mathrm{NOISE}[A],$$

where NOISE[•] is an operator of superimposed noise on a signal;

3—descriptions of signals *B* and *C* are calculated: D_B, D_C. The structure of descriptions of D_B and D_C is equivalent to the structure of D_A;

4—the descriptions are compared, as follows: D_A–D_B and D_A–D_C, resulting in arrays generating $C_{B,V}$, $C_{B,ni}$, $C_{B,pi}$, $C_{C,V}$, $C_{C,ni}$, $C_{C,pi}$ ($C_{B,V}[i]$ is the number of operators which coincide for *i* segment of descriptions D_A–D_B; $C_{B,ni}[i]$ is the number of complete groups which coincide for *i* segment of descriptions D_A–D_B; $C_{B,pi}[i]$ is the number of closed groups which coincide for *i* segment of descriptions D_A–D_B).

As a result of the above mentioned arrays it is possible to draw a conclusion about the stability of the segments in accordance with the following rule: If $C_B[i] > T$ and $C_C[i] > T$, then *i* segment of a signal is distortion-resistant (*T*—threshold value).

Let us use this method to define the characteristics of distortion-resistant signal segments.

After a number of experiments, it has been established that for allocation of stable signal segments only closed groups can be used. Analysis of the descriptions of distorted and undistorted signals generated by means of closed groups shows how to allocate indicators in order to divide signal segments into stable and unstable:

1—"P1" indicator: the greater the difference between maximum and minimum mass of groups in a set describing a signal segment *S*, the more likely it is that segment description of a signal will not change while being distorted:

$$D_1 = \mathrm{MAX}(\mathrm{MASS}(P, A)) - \mathrm{MIN}(\mathrm{MASS}(P, A)),$$

where *P* is a set of groups, calculated using a signal *A*, MAX [•] is an operator calculating the maximum value of numbers in the array, MIN [•] is an operator calculating the minimum value of numbers in the array, and MASS [•] is the operator calculating the masses of groups including those in the array *P*;

2—"P2" indicator: the greater the sum of group masses including into a set of groups *P*, describing a signal segment *A*, the more likely it is that segment description of a signal will not change while being distorted:

$$D_2 = \mathrm{SUM}(\mathrm{MASS}(P, A)),$$

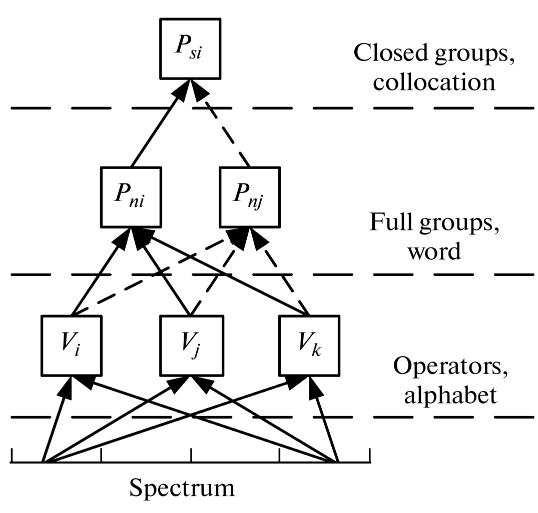

Fig. 2. Hierarchy of operators and groups.

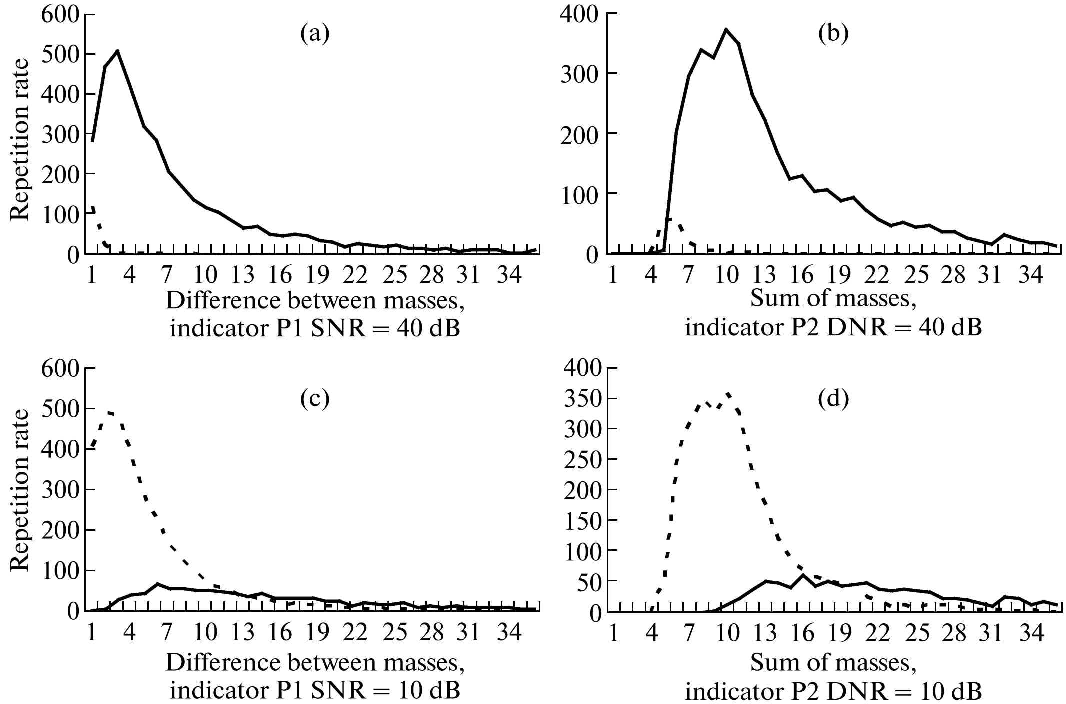

Fig. 3. Threshold identification.

where SUM[•] is the operator of the calculation of the sum of the array elements.

3. COMPUTING EXPERIMENT

Let us consider pictures that prove the possibility of using the herein-mentioned indicators for dividing signal segments into stable and unstable. In Fig. 3 are given the statistics of existence of stable and unstable segments at low noise level (signal/noise ratio is equal to 40 dB) in a signal. These statistics are calculated by means of C_B and C_C arrays. When forming C_B and C_C arrays the first N by weight of closed groups was removed. In Fig. 3a the unbroken line represents the frequency change of the appearance of stable segments, depending on the difference between the mass of groups calculated on the signal segment i, and the dotted line shows the same thing, only for unstable segments. In Fig. 3b the unbroken line represents the change in frequency of the appearance of stable segments, depending on the sum of mass of groups calculated on signal segment i, the dotted line shows the same thing, only for unstable segments. The range of values of mass differences and also the sums of masses are standardized on a segment [0; 100].

Table 1. Statistics of signal segment stability (uniform noise)

Noise 40 dB	Noise 20 dB	Noise 10 dB	Noise 0 dB
Segment length: 64			
9640/3610	7851/5399	6385/6865	1570/11680
9212/4038	6606/6644	4747/8503	613/12637
Segment length: 128			
2588/1404	2543/1449	1728/2264	728/3264
2305/1687	2355/1637	1316/2676	653/3339
Segment length: 256			
1363/485	1363/485	1042/806	497/1350
1255/593	1255/593	843/1005	566/1281
Segment length: 512			
959/401	959/401	820/540	251/1109
908/452	908/452	908/452	133/1227
Segment length: 1024			
462/245	364/343	364/343	195/512
425/282	425/282	547/160	85/622

The experiments conducted show that the peak position in the figure (it is marked with a triangle) can be used as a threshold to dividing distortion-resistant and distortion-labile segments. In Figs. 3c and 3d the same diagrams are shown, only for a signal/noise ratio equal to 10 dB. Comparing Figs. 3a and 3c we can see that an increase of noise level causes a decrease in the quantity of stable segments; meanwhile the peak position is practically unchanged (it moves slightly to the left).

This fact can be used to divide stable and unstable segments. It should be noted that the structure of diagrams for different noise levels is similar.

In Tables 1 and 2 are presented statistics on the segments of steady and unsteady signals under different con-

Table 2. Statistics of signal segment stability (normal noise)

Noise 40 dB	Noise 20 dB	Noise 10 dB	Noise 0 dB
Segment length: 64			
9640/3610	6385/6865	2880/10370	107/13143
9212/4038	5584/7666	1849/11401	179/13071
Segment length: 128			
2588/1404	2161/1831	916/3076	536/3456
2305/1687	1954/2038	1042/2950	53/3939
Segment length: 256			
1363/485	1042/806	622/1226	171/1677
1255/593	843/1005	566/1282	57/1791
Segment length: 512			
959/401	685/675	470/890	251/1109
908/452	908/452	537/823	133/1227
Segment length: 1024			
462/245	364/343	234/473	37/670
425/282	323/384	170/537	11/696

ditions (for different noise levels and for different segment length). To the left of the slash are presented data for first indicator (P1) and to the right, the second (P2).

CONCLUSION

This paper presents the method that would make it possible to calculate the sections of a sound signal resistant to distortions. For allocation of such sections two indicators are introduced. The method offered can find a practical application in the analysis of media data that will make it possible to increase resistance of the description created to different distortions.

REFERENCES

1. V. A. Utrobin, "Physical interpretation of image algebra elements," Usp. Fiz. Nauk **174** (10), 1089–1104 (2004).

Vasilii E. Gai was born in 1984. Year of graduation and the name of institution of higher education: 2006, Vladimir State University. The year of defend a thesis (PhD): 2009. Associate professor of Nizhniy Novgorod State Technical University n. a. R. E. Alekseev. Research interests: digital signal processing. Number of publications (monographs and articles): 100.