EdiTracker

Dedicated software module

STC-S338

User manual
## INTRODUCTION

General

Manpower requirements

Typography conventions

Copyright

### 1 GENERAL INFORMATION

1.1 About the system and the producer
1.2 Program allocation

### 2 MODULE INSTALLATION AND LAUNCH

### 3 “RECORDER PARAMETERS” TAB

3.1 Guidelines for the analysis of recorder characteristics
3.2 Amplitude-Frequency Response (AFR)
3.3 Noise and interference
3.4 Total harmonic distortion (THD)
3.5 Detonation and Speed
3.6 Parasitic amplitude modulation (PAM)
  3.6.1 In-phaseness of recorded signals
  3.6.2 Separation between the adjacent related tracks

### 4 “INDICATORS OF DIGITAL PROCESSING” TAB

### 5 “SPECTRUM ESTIMATION” TAB

5.1 Analysis objects and their properties
5.2 Spectrum analysis instruction

### 6 “PHASE SCANNING” TAB

6.1 Physical meaning of the continuity indicator
6.2 Phase estimation
6.3 Additional options

### 7 “BACKGROUND SCANNING” TAB

7.1 General information
7.2 Techniques of background spectrum analysis
7.3 Method of signal mean value controlling (“Mean value” method)

### 8 “AUDITORY ANALYSIS” TAB

8.1 Guidelines for the detection of audio recording authenticity breaches by means of auditory-linguistic analysis
8.2 Description of the work with the SIS II software when performing auditory linguistic analysis of audio recordings

8.2.1 Opening sound files

8.2.2 Visualization of data to be listened

8.2.3 Audio recordings playback

8.2.4 Operations with the marks

8.3 Step-by-step user instructions for performing auditory-linguistic analysis of an audio recording for the detection of audio recording authenticity breaches

8.4 Comparison of the results of auditory and instrumental analysis

9 “CONCLUSIONS” TAB

10 MODULE SHUTDOWN

APPENDICES

Appendix A: List of audio recording authenticity breaches

Appendix B: Description of indicators used in auditory-linguistic analysis of audio recordings

Appendix C: Glossary
INTRODUCTION

General

This User’s Guide contains information how to install and operate the dedicated software module EdiTracker. The document is primarily intended for the users having special skills in speech records assessment and describes users’ actions for successful sound editing.

This paper does not replace academic, reference books and manuals from the manufacturers of the operating system and common software.

Manpower requirements

Staff, producing the installation of the specialized EdiTracker, should have professional skills to install general and special software.

Staff, working with the dedicated EdiTracker, should have basic skills to operate with applications in the operating systems Microsoft Windows and should know how to expertise speech audio/sound records.

Typography conventions

The following typographic conventions are used in the manual:

<table>
<thead>
<tr>
<th>Font</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Body text of the manual</td>
</tr>
<tr>
<td>Italic</td>
<td>The first appearance of a term. Meaning of the term is explained here or in the appendix. Also it is used to attract attention or to make up notes.</td>
</tr>
<tr>
<td>Bold</td>
<td>Names of software components and interface elements (headings, buttons, etc.).</td>
</tr>
<tr>
<td>BoldItalic</td>
<td>Names of files and paths to them.</td>
</tr>
</tbody>
</table>

Menu selection is marked with an arrow →, i.e. the combination Menu→Command should be understood as following: select Menu and then find the item Command.
INTRODUCTION

To indicate the importance of any information, the following comments and notes are used in the manual:

- **Note:** Useful information.

- **Warning:** Important information.

- **Caution:** Essential instructions which are obligatory to be fulfilled to prevent any fatal error in the system functioning.

---

**Copyright**

*EdiTracker* is trademark of Speech Technology Center Ltd. All rights reserved.

All other companies and products mentioned in the manual are property of their respective owners.

The software includes modules of cross-platform application framework **Qt 4.7.0** (http://qt.nokia.com) distributed under the terms of the GNU LGPL 2.1 license http://www.gnu.org/licenses/lgpl-2.1.html.

---

Any part of this publication may not be reproduced, transmitted, stored in a retrieval system or translated into any language in any form or by any means, without the written permission of Speech Technology Center, Ltd.
1 GENERAL INFORMATION

1.1 About the system and the producer

Name | Dedicated software module **EdiTracker**
Conditional name | STC-S338
Producer | Speech Technology Center, Ltd.
Postal address | Russia, 196084, St. Petersburg, 4 Krasutskogo str.
Telephone | +7 (812) 325-88-48
Fax | +7 (812) 327-92-97
1.2 Program allocation

The EdiTracker dedicated software module for the specialized Sound Editor SIS II is the 32-bit Windows application, designed for detecting signs of tape tampering in audio recordings. The software can run as a plug-in from the SIS II sound editor (for details see the respective user manual).

EdiTracker allows:

- Analysis of the technical characteristics of the recorder, used by the expert for audio data input into a PC;
- Examination of the sound signal for traces of its previous digital processing, using spectrum analysis techniques and detection of anti-alias filtering signs;
- Audio recording analysis using a high-precision spectrum analyzer for detecting stationary harmonics in the signal (power source noise etc.).
- Examination of the sound signal for tampering signs, using methods involving high-precision analysis of stationary frequencies phase values;
- Examination of the sound signal for tampering signs, using spectrum analysis techniques and detection of changes in background noise dynamics;
- Auditory-linguistic analysis of an audio recording, with designation of suspicious areas by special marks;
- Visualisation of audio data analysis findings in the sound windows;
- Compilation of instrumental and auditory analysis data in the form of conclusions, which can be stored as a text file.
Module is installed on a PC, on which the specialized sound editor SIS II has been already installed.

Software installation must be performed by OS administrator.

The EdiTracker software module is implemented as a set of dynamic-link libraries: sis_edtr.dll, editrck.dll, antm_r.dll, antmd.dll, copy.dll, el.dll, f50.dll, fon.dll, magn.dll.

The EdiTracker add-on (plugin) is connected to the specialized Sound Editor SIS II as a part of the additional signal processing modules.

To install the software, run the EdiTracker.exe file located in the root directory on the distribution disk. Below the basic steps of module installation are presented.

In welcome window (Figure 1) click Next> and follow instructions of the Installation Wizard appearing on the screen.

![Welcome window](image)

Figure 1 – “Welcome” window
On completion of installation, click **Finish** to exit the Setup Wizard program (Figure 2).

![Figure 2 - Window of successful module installation](image)

After module installation and sound editor **SIS II** launch, the **EdiTtrack** item will be added to the **Modules** menu (Figure 3).

Before running **EdiTtrack** you should activate the data window with the sound signal you wish to process or input the audio signal via the I/O device (STC-H246), intended for measuring characteristics and forming electrical signals in the sound frequency range.

Other types of sound cards can also be used for signal input, but in this case the recording conditions can fail to meet all technical requirements that can lead to the loss of data.

⚠️ **The current version of EdiTtrack** accepts for analysis only audio recordings inputted at 44 100 Hz sampling frequency or higher.

If you have several windows with audio data opened simultaneously, feature analysis will be performed only for the signal in the currently active window.

💡 **Some analysis procedures (e.g. Spectrum Estimation, Phase Scanning)** are rather lengthy. To speed up the process, we recommend you to cut out the fragment of interest from the signal and place it into a separate window for further processing (for more information about processing of audio data see the respective user manual “SIS II. Sound editor. User’s Guide”).
The *EdiTracker* main window looks as a dialog window, shown in Figure 4.

The main window contains several tabs (modules) providing access to the respective analysis procedures used for the examination of audio recordings and recorder characteristics. They are the following:

- **Recorder Parameters**;
- **Indicators of Digital Processing**;
- **Spectrum Estimation**;
- **Phase Scanning**;
- **Background Scanning**;
- **Auditory analysis**;
- **Conclusions**.
View of the Recorder Parameters tab is shown in Figure 4. This module is used to measure the following technical characteristics of the recorder:

- Amplitude-Frequency Response (AFR);
- Noise and interference spectrum;
- Signal-to-Noise Ratio (SNR);
- Total Harmonic Distortion (THD);
- Parasitic Amplitude Modulation index (PAM index);
- Speed deviations;
- Weighted detonation;
- Separation between the adjacent related tracks;
- AFR mismatch between the channels;
- Delay between the channels;
- Effective frequency ranges.

Technical characteristics are computed separately for the left and right channels of the recorder. Left channel data is graphed in green, while right channel data is shown in yellow.

To receive methodical assistance when using this tab, click Procedure.

### 3.1 Guidelines for the analysis of recorder characteristics

Two test signals are provided for analysis on the distribution CD.

The test1s.wav test signal (stereo, 48 kHz) (it is in the following folder by default: C:\Program Files (x86)\Speech Technology Center\SIS II\ET\TestSignals) contains:

- 100 seconds of a 3150 Hz harmonic signal with an amplitude of 32 000 (nominal signal level) for measuring recorder speed deviations, detonations, parasitic amplitude modulation;
- 100 seconds of silence for measuring noise spectral components;
- 11 7-frequency signals of 6 seconds duration each, with a 1 second pause between them, for measuring Amplitude-Frequency Response (AFR). Each harmonic has an amplitude of 3200 (-20 dB of the nominal level);
- 100 seconds of a 1000 Hz harmonic signal with an amplitude falling from 32 000 to 0 according to the cube law for measuring this characteristic;
- 6 seconds of a 315 Hz harmonic signal with an amplitude of 32 000 for measuring Total Harmonic Distortion (THD);
“RECORDER PARAMETERS” TAB

- 6 seconds of a 3-frequency signal (250 Hz, 1000 Hz, 6300 Hz), first in the left and then in the right channels with a 7 second pause between them and before them for measuring the separation between the adjacent related tracks.

Total signal duration: 6 minutes 37 seconds.

The test_snr.wav test signal (mono, 48 kHz) (it is in the following folder by default: C:\Program Files (x86)\Speech Technology Center\SIS ET\TestSignals) is used to determine Signal-to-Noise Ratio (SNR); it contains:
- 6 seconds of a 1000 Hz harmonic signal with an amplitude of 32 000;
- 14 seconds of silence.

Total signal duration: 20 seconds.

Both signals are recorded at 48 000 Hz sampling frequency, 16 bit/sample, linear law.

These signals are to be copied onto the examined recorder, entered back into the computer, and then they should be analysed with the EdiTracker software module.

Steps to follow:
1. Connect linear inputs and outputs of the examined recorder to linear outputs and inputs of the I/O device (STC-H246).
2. Connect the I/O device (STC-H246) to a PC via USB interface.

To obtain information about structure, technical characteristics and principle of the operation of USB Sound I/O Device STC-H246, as well as about instructions required for device proper performance, intended for measuring characteristics and forming electrical signals in the sound frequency range, refer to the “USB Sound I/O Device STC-H246. User manual”.

3. Run the SIS II software.
4. Open the files with the test1s.wav and test_snr.wav test signals in two different data windows. To perform this, use the File → Open… menu commands and in the dialog window choose necessary files.

For detailed description of operation with the SIS II sound editor see the “SIS II. Sound editor. User’s Guide”.

5. Insert a new clean cassette into the tape recorder and enable the “Record-Pause” mode.

It should be kept in mind that recorders do not usually erase previously recorded information completely. Thus, if the cassette has been used for recording before, its residual noise will affect the measurements (especially those related to noise level).

6. Go to the window with the test1s.wav signal. Play the test1s.wav signal. Using volume and balance controls, ensure that the recording level indicator of the tape recorder shows 0 dB in both channels (maximum recorded signal level without overload).
7. Stop signal playback.
“RECORDER PARAMETERS” TAB

8. Set the recorder to record mode.
9. Start signal playback from the beginning.
10. Disable playback restart.
11. After the test1s.wav signal has been played to the end, do not interrupt the recording process. Go to the window with the test_snr.wav signal.
12. Play the test_snr.wav signal 7-10 times, each time with a slightly raised recording level relative to the position during the recording of the test1s.wav signal, until it reaches the maximum.

⚠️ The position of the recording level control should be changed only after both the harmonic signal and the pause in the test_snr.wav file have been played, i.e. when no signal is being played. This is necessary for correct signal/noise ratio estimation.

After you have changed the slider position of the level control, restart signal playback by pressing the [Go to Start] icon or in the Playback menu click Go to Start (Playback → Go to Start).
13. Stop the tape recorder. The total duration of the recorded signal should be approximately 9-12 minutes.
14. Memorize or take down the recorder’s meter reading.
15. Rewind the tape approximately 20 seconds back. This is the point in the signal with the maximum recording level.
16. In the Service menu click Options (Service → Options).
17. Open the Sound tab. Ensure that the signal will be recorded at 48 000 Hz, stereo, 16 bit. Click OK.
18. The recording has not started yet, but you can monitor the level of the input signal. Enable signal playback on the tape recorder and ensure that the final part of the signal (the loudest) will be recorded without overload (when there is overload the running oscillogram rests on the border of the window).
19. Rewind the tape to the beginning of the recording.
20. Use one of the following methods to start recording:
   - In the File menu click Recording (File → Recording).
   - Click the Recording pictogram on the toolbar.
   - Press Ctrl+R on the keyboard.
21. Start playback on the tape recorder.
22. After the entire signal has been played, stop playback on the tape recorder. Use the meter reading to ensure that the entire signal has been inputted.
23. Use one of the following methods to finish recording:
   1) In the File menu click Recording (File → Recording) again.
   2) Click the Recording pictogram on the toolbar again.
   3) Press Ctrl+R on the keyboard again.
4) In the **Playback** menu click **Stop** *(Playback → Stop)*.

5) Click the **Stop** pictogram on the toolbar.

6) Press **Esc** key on the keyboard.

24. In the **Modules** menu click **EdiTracker** *(Modules → EdiTracker)*.

25. On the **Recorder Parameters** tab click the **Calculate** button.

26. After calculations have been performed, you can operate with the following tabs: **AFR** *(Amplitude-Frequency Response)*, **Noise and interference**, **THD** *(Total Harmonic Distortion)*, **Detonation**, **Histogram of PAM** *(Histogram of Parasitic Amplitude Modulation)*, **Speed**. You will see the respective characteristics displayed.

27. Click the **Analysis result** button, to display all obtained characteristics in the window, shown on Figure 5. To save the data in text file, click **Save**. Click **OK** or **X** to close the window without saving the results.

![Figure 5 – “Recorder Parameters” window](image-url)
3.2 Amplitude-Frequency Response (AFR)

Amplitude-Frequency Response (AFR) is the most important recorder characteristic. It is the primary criterion for determining which of the recorders is better. Other parameters allow us only to say that they are within the admissible limits. If they overstep the limits, it means that the recorder is out of order.

According to the GOST (State Standards of the Russian Federation) relating to measuring recorder characteristics, record-playback channel AFR should be measured while transmitting to the input harmonic signals of varying frequency –20 dB from the nominal level.

To reduce the time needed for recording creation, 11 7-frequency signals have been generated:

1. 3.55; 12.5; 45; 160; 560; 2000; 7100 Hz;
2. 4; 14; 50; 180; 630; 2200; 8000 Hz;
3. 4.5; 16; 56; 200; 710; 2500; 8900 Hz;
4. 5; 18; 63; 220; 800; 2800; 10 000 Hz;
5. 5.6; 20; 71; 250; 890; 3150; 11 000 Hz;
6. 6.3; 22; 80; 280; 1000; 3550; 12 500 Hz;
7. 7.1; 25; 89; 315; 1100; 4500; 14 000 Hz;
8. 8; 28; 100; 355; 1250; 5000; 16 000 Hz;
9. 8.9; 31.5; 110; 400; 1400; 5600; 18 000 Hz;
10. 10; 35.5; 125; 450; 1600; 6300; 20 000 Hz;
11. 11; 40; 140; 500; 1800; 7100; 22 000 Hz.

The frequencies are selected so that each following frequency would be about 1.1 times higher than the preceding frequency. AFR of the signal is shown on Figure 6.

AFR is normalized so that the level would be 0 dB at 1000 Hz.

The more flat is the graph, the better the recorder’s quality. It should also be taken into account that AFR depends not only on the tape recorder, but on the used cassette as well. It should be kept in mind that now cassettes are often manufactured with high frequencies intentionally raised by 1-2 dB, to provide better music quality.

The full frequency range of the recorder is the one within which the signal recorded at –20 dB from the nominal exceeds the noise level by at least 6 dB.
According to GOST, the full effective frequency range is a group of frequencies with AFR ripple not more than 6 dB for tape recorders of the 0 and 1st complexity groups and not more than 7 dB for tape recorders of the 2nd, 3rd and 4th complexity groups. For tape recorders of the 0 and 1st groups AFR ripple in the 250-6300 Hz range must not exceed 3 dB; for tape recorders of the 2nd, 3rd and 4th groups in the 250-4000 Hz range it must not exceed 4 dB.

Table 1. Effective frequency ranges of different complexity groups of tape recorders (according to GOST)

<table>
<thead>
<tr>
<th>Group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3–4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range, HZ</td>
<td>25–20 000</td>
<td>31.5–18 000</td>
<td>40–14 000</td>
<td>63–10 000</td>
</tr>
</tbody>
</table>

Seven frequency ranges are calculated, with AFR ripple changing from 1 to 7 dB.

The obtained characteristic is used to calculate AFR mismatch between channels. According to GOST, the measurements are taken in the 250-6300 Hz frequency range with AFR matched at 1000 Hz. The result of the measurements is the maximum difference between ordinate values in the 250-6300 Hz frequency range. This value should not be more than 2 dB for tape recorders of the 0 and 1st groups, not more than 3 dB for tape recorders of the 2nd group and not more than 4 dB for the 3rd and 4th groups.
3.3 Noise and interference

To estimate noise and interference, calculate average spectrum using 100 seconds of silence recorded on the tape.

The noise spectrum obtained with using EdiTracker is shown below (Figure 7).

It can be seen that noise and interference are mainly concentrated in the low frequency area. This area can be examined in detail (Figure 8).
To calculate the full weighted and unweighted signal/noise ratio, we use a 1000 Hz harmonic signal and the subsequent pause (the 2nd test signal) recorded with varying positions of the recording level control. We search for a harmonic signal with the 3rd harmonic distortion exceeding 3%. Then we calculate weighted (by A-filter) and unweighted ratio between the sum of all spectrum components of the subsequent pause and the sum of all spectral components of the signal.

A-filter reflects auditory signal perception (low and high frequencies are suppressed, while mid frequencies are enhanced).

The pause is taken 6 seconds after the signal end, to prevent the “copying effect” produced by a powerful signal on the adjacent winds of the cassette tape.

The obtained values in dB are presented in Table 2.

Table 2. Weighted signal/noise ratios for various complexity groups of tape recorders (according to GOST)

<table>
<thead>
<tr>
<th>Group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio, dB</td>
<td>60</td>
<td>56</td>
<td>54</td>
<td>48</td>
</tr>
</tbody>
</table>
3.4 Total harmonic distortion (THD)

During sound recording (or playback) on a tape recorder there appear harmonics – multiples of the one recorded. Their value relative to the effective signal is described as total harmonic distortion (THD). To calculate THD, it is sufficient to use the largest harmonics (the 2nd and the 3rd):

\[
\text{THD}_2 = 100\% \times \sqrt{\frac{E_2 + E_3}{E_1 + E_2 + E_3}}
\]

\[
\text{THD}_3 = 100\% \times \sqrt{\frac{E_3}{E_1 + E_3}}
\]

where

- \(E_1\) – energy in the recorded frequency area \(FREQ\),
- \(E_2\) – energy in the \(2 \times FREQ\) area,
- \(E_3\) – energy in the \(3 \times FREQ\) area.

To estimate THD, a 315 Hz harmonic signal is used.

Figure 9 illustrates the spectrum obtained for THD calculation:

![Figure 9 – 315 Hz harmonic signal spectra for the left and right channels](image-url)

Table 3. The 3rd harmonic distortion in various complexity groups of tape recorders (according to GOST)

<table>
<thead>
<tr>
<th>Group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>1.5</td>
<td>2.5</td>
<td>2.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>
3.5 Detonation and Speed

Let us consider in detail a 315 Hz harmonic signal spectrum obtained for harmonic distortion calculation in the 315 Hz main maximum area.

As can be seen in Fig 10, there are other spectral maxima, symmetrical relative to the central one, to the left and to the right of the maximum. This effect is caused by the changes in the tape transport speed and is called Detonation.

According to GOST, for home tape recorders weighted detonation is measured by a special device – detonation meter – at the beginning and at the end of the full cassette during test tape playback. For tape recorders of the 0 and 1st complexity groups it is allowed to take measurements using a 3150 Hz harmonic signal recorded on the examined tape recorder and its subsequent playback on the same tape recorder.

The result is taken to be the arithmetic average of 5 measurements of the same recording. Weighted detonation is the amplitude of the test signal frequency fluctuations with sine frequency modulation equal to 4 Hz, expressed in per cent relative to the test signal average frequency. With other modulation frequencies, this value is weighted by a special filter corresponding to auditory perception.

The waveform of a 3150 Hz harmonic signal is used to calculate the precise frequency value FREQ(I) at each period by interpolated zero-crossing points. It is directly proportional to the speed value. We average it over the whole signal. Let us call the obtained value FREQ(avg.).

\[
(FREQ(\text{av.}) - 3150) / 3150 \times 100\% \text{ is the average (or main) deviation of recorder playback speed from the recording speed.}
\]
Instantaneous deviations from the average frequency are calculated as follows:

\[ \Delta \text{Speed (I)} = \frac{\text{FREQ (I)} - \text{FREQ (av.)}}{\text{FREQ (av.)}}. \]

The diagram (Fig. 11) illustrates the overall speed increase during playback (or decrease during recording). By enlarging the image (Fig. 12), one can see that speed variation is characterized by periodicity.
Now let us calculate the spectrum. According to GOST, the area of interest lies between 0.2 and 200 Hz.

![Spectrum for Detonation estimation (1 kHz)](image1)

Figure 13 – Diagram of a speed deviations spectrum for measuring detonation

One can see characteristic peaks on the spectrum (Fig. 13). The low-frequency area can be examined in detail (Fig. 14).

![Spectrum for Detonation estimation (1 kHz)](image2)

Figure 14 – Diagram of a speed deviation spectrum for measuring detonations; enlarged

Human ear is considered to be the most sensitive to the floating of sound at 4 Hz. We weigh the obtained spectrum by a filter corresponding to auditory perception of detonations (implemented in the detonation meter). This way the weighted detonation value is obtained.
Sometimes detonation is specified in recorder characteristics as ± value. Then, for bringing them into accord, the value of detonations obtained using our software should be divided by 2.

According to GOST, weighted detonation must not be more than 0.16 – 0.8% (or not more than ±0.08 – ±0.4) for tape recorders of various classes.

<table>
<thead>
<tr>
<th>Group</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detonation, %</td>
<td>0.16 (±0.08)</td>
<td>0.24 (±0.12)</td>
<td>0.4 (±0.2)</td>
<td>0.7 (±0.35)</td>
<td>0.8 (±0.4)</td>
</tr>
</tbody>
</table>
3.6 Parasitic amplitude modulation (PAM)

If we examine an enlarged waveform of a harmonic signal inputted from a tape recorder, we can notice changes of signal level in some points of the signal (Figure 15). This phenomenon is called parasitic amplitude modulation (PAM).

![Figure 15 – Diagram of a 3150 Hz harmonic signal waveform](image)

According to GOST, PAM is calculated only for rather prolonged (more than 0.5 seconds long) amplitude changes. Thus average amplitude is taken for each 0.5 seconds of a continuous harmonic signal. Then the minimum and maximum values are taken for 100 seconds.

\[
PAM = \frac{\text{max\_average} - \text{min\_average}}{\text{max\_average} + \text{min\_average}} \times 100\%
\]

The PAM value for various classes of in-car recorders must not exceed 15 – 30% (PAM is not mentioned in GOST for home recorders).

One more issue of interest could be instantaneous changes clearly visible on the waveform. They can be examined in detail using a PAM histogram computed by averaging over short signal fragments (0.01 sec. in duration). The X axis represents amplitude in per cent from the average; the Y axis represents occurrence in the logarithmic scale (Fig 16).
Figure 16 – Diagram of a PAM histogram of the left and right channels
3.6.1 In-phaseness of recorded signals

The figure below demonstrates that the signals recorded via different channels are not quite in phase:

![Diagram of a 1000 Hz harmonic signal waveform (the left and right channels)](image)

Phase displacement between channels results from the incorrectly adjusted tilt of recording and reproducing magnetic heads and is caused by the delay between the channels. This delay is calculated in a 3150 harmonic signal by interpolated zero-crossing points. The obtained value shows the delay in the right channel relative to the left channel and is measured in milliseconds. It must not exceed 0.02 ms in absolute value (modulus), provided adjustment is correct.

3.6.2 Separation between the adjacent related tracks

This parameter is measured at 250, 1000 and 6300 Hz frequencies. A 3-frequency signal recorded first via one channel and then via another one is used. The ratio between spectral energies is calculated at each of these signal frequencies, from clean track to the recorded.
To search for traces of digital processing, it is required that audio recordings is entered into the computer at a sampling frequency (Fs) not below 44 100 Hz.

If the signal was digitized at a lower sampling frequency, its average spectrum would as a rule contain a characteristic fall-off in the 0.9…1.0*(Fs/2) area. This fall-off should be sharper than the previous and the next fall-offs. Ideally the picture should look as follows (for Fs = 8000) (Fig. 19):
There is a sharp fall-off in the 3600-4000 Hz area. The fall-off up to 3600 Hz is much smaller and when the fall-off more than 4000 Hz the spectrum is almost flat (the boost in the high frequency area is caused by the recorder's self-noise produced during the copying of a digitised signal onto the tape). In many cases this fall-off cannot be so clear and sometimes it is completely hidden by the characteristics of the used sound recording devices. This depends on the quality of the initial recording and of the recorder used for re-recording (the higher it is, the easier it is to detect digitisation) and on the sampling rate (the smaller it is, the easier it is to reveal digitisation). However, even for most such cases the algorithm maintains efficiency due to the relatively mild threshold values set for the examined parameters.

To start the analysis, click the **Search** button.

While the calculations are being performed, the colour scale in the lower part of the dialog window will be changing.

After the calculations are made, the text of conclusions with commentaries will be displayed in the **Analysis results** window.

The **Periodogram (average spectrum) of the audio recording** diagram in the window will display the averaged signal spectrum. Spectrum areas containing traces of digital editing will be marked in red. Also the position of the frequency corresponding to the assumed half of the sampling frequency will be indicated at which digital processing is performed.

To receive methodical assistance when using this tab, click **Procedure**.
5 “SPECTRUM ESTIMATION” TAB

The Spectrum Estimation analysis is intended for:

a) Detection of quasi-stationary harmonics present in the frequency range specified by the expert. This includes not only detection of a clearly manifested peak in the Instantaneous spectrum window when SNR is in a favorable range for the expert, but also in conditions of strong white noise when in the Instantaneous spectrum window the amplitude of the stationary harmonic is commensurable to the amplitudes of fluctuating frequencies of the white noise.

b) SNR estimation for selecting the mode for subsequent examination of these harmonics’ phase on the Phase Scanning tab.

c) Sufficiently accurate (for subsequent examination of these harmonics’ phase on the Phase Scanning tab) measurement of the middle frequency of these harmonics within a specified time interval (either on the entire signal or, if the file is too long (i.e. longer than 10 minutes), within a certain file fragment).

d) Detection of evident breaks in the frequency of these quasi-stationary harmonics, to be traced in the Indicator of search for a precise frequency value of stationary harmonics in the audio recording window.

e) Detection of the best candidate among stationary harmonics for subsequent examination of its phase on the Phase Scanning tab. The best here means the one having the largest amplitude, as this improves SNR.

To receive methodical assistance when using this tab, click Procedure.

5.1 Analysis objects and their properties

When analysing an audio signal recorded on an analog tape recorder connected to the city power system, the standard objects for examination will be:

a) The 50 Hz power harmonic and its duplicates. The power harmonic frequency is slowly oscillating for one hour with the 2 Hz amplitude. During the analysis this even and slow change of the power harmonic frequency should be distinguished from an abrupt frequency jump observed, for instance, when another fragment of speech is inserted into the recording, with another carrier frequency differing from the original by a considerable value (up to 2 Hz). Duplicates or copies of the power harmonic are new 50 Hz harmonics with another frequency and amplitude, introduced into the audio recording during its re-recording. In this case the analyzed frequency range is from 48 Hz to 52 Hz (always set by default).

b) The second (100 Hz) harmonic from the power harmonic and its duplicates. This harmonic often surpasses the first 50 Hz one in power. This is the only reason why it should be used. Its behavior is wholly identical to the first power harmonic. The analyzed frequency range in this case is from 98 Hz to 102 Hz.

c) Harmonics caused by the mechanical rolls of an analog tape recorder TTM. This is, for instance, the rotation frequency of the reels (1.1 Hz) and its multiple harmonics. The analyzed frequency range in this case is from
0.2 Hz to 6 Hz. For other TTM-related phenomena the range of analyzed frequencies will be from 6 Hz to 26 Hz.

Besides these objects, there can be other harmonics as well, for instance, the monitor frequency (80–112 Hz) of the computer on which the recording was edited, or reversal magnetisation frequencies (19 000–21 000 Hz) etc.

In general, any stationary harmonic within the range from 0.5 Hz to half of the signal sampling frequency is suitable for analysis and an expert can, if desires, specify any frequency range of interest for the detection of stationary harmonics. The only limitation is imposed by cases when the harmonic of interest falls into the frequency range which is perceived by the human ear as parasitic noise (from 120 Hz to 1500 Hz). For this reason it is not recommended to analyse the 3rd and 4th harmonics from the power harmonic.
5.2 Spectrum analysis instruction

Before starting work with the Spectrum Estimation tab, it is useful to obtain a sonogram of the examined file to detect strong non-standard stationary harmonics for their subsequent examination.

The expert's work begins with the choice of an analysis object and the choice of an appropriate frequency range for the analysis (the default one is 48–52 Hz). Then the Search button should be clicked. Once the analysis is complete, the program will select the 4 strongest peaks in the signal in the Instantaneous spectrum window.

When working with this tab, you need not wait until the end of file (especially if the file is very long) – just press the Stop button before the file is over and you will still get the values of the 4 strongest peaks in the Instantaneous spectrum window:

Figure 20 – Spectrum Estimation tab

Figure 20 illustrates the analysis of a 50 Hz harmonic in the presence of strong white noise. The amplitude of 50 Hz here is 0.96 counts, while the speech maximum is approximately 6000 counts, i.e. the noise is over 6000 times stronger than the signal. A more accurate estimate of SNR can be obtained by analysing the data
in the **Instantaneous spectrum** window. By estimating the harmonic’s peak and average level of the ambient noise in this window the expert can see the SNR displayed on a linear scale.

In this case the most effective way of detecting a stationary harmonic will be to use a search indicator window instead of the instantaneous spectrum window, for the search indicator window provides good representation of the 50 Hz harmonic continuity. The break in the yellow line (the one with the largest amplitude) observed in the 10th second in the indicator window is not a break in the 50 Hz harmonic. It can be clearly seen that before the 10th second the 50 Hz harmonic was present as well (the red and orange lines extend the yellow line backwards); they were just weaker than the noise at 49 Hz (represented by the yellow line up to 10 seconds).

In this case only one value of 50 Hz corresponds to the stationary harmonic, while the rest 3 frequency values indicated in the **Analysis results** window are just white noise. Besides, it is always possible to trace the yellow line with the cursor in the search indicator window (provided there are harmonic continuity stretches) along the entire file length; the current values of the harmonic frequency will appear as a result. The current harmonic frequency value should be selected and entered manually into the frequency window on the **Phase Scanning** tab.

With a still worse SNR there can be a situation when neither of the frequency values in the **Analysis results** window is worth entering into the **Phase Scanning** tab, i.e. phase scanning cannot be performed as there is no analysis object.

To handle strong white noise, the expert can select high scanning accuracy by increasing the length of temporal averaging. The file length in this case should not be below 120 seconds.

The precision of frequency resolution in this example is 0.01 Hz (1/400 of the range). The expert can always narrow the range and repeat the analysis to obtain a more precise value for the harmonic frequency.
6 “PHASE SCANNING” TAB

The Phase Scanning tab allows:

a) Detection of presence/absence of phase breaks in a continuous harmonic selected by the expert;

b) Sufficiently accurate estimation of the phase break size, as well as its localization in time;

c) Detection of breaks in the frequency of a continuous harmonic selected by the expert.

To receive methodical assistance when using this tab, click Procedure.

6.1 Physical meaning of the continuity indicator

Before describing the ways of working with this tab, let us clarify the physical meaning of the parameter which is graphically represented in the Audio recording continuity indicator lower window. Suppose we have a continuous harmonic of an unknown frequency $\omega$.

$$S(t) = A \cdot \sin[\omega(t) \cdot t + \theta_0] = A \cdot \sin[\omega_0 \cdot t + \{\Delta \omega(t) \cdot t + \theta_0\}]$$  \hspace{1cm} (1)

Suppose, $\omega_0$ is our estimate of this frequency – presumably absolutely accurate, i.e. $\Delta \omega = 0$

The diagram of the Audio recording continuity indicator window represents a constant (zero-frequency) component of the continuous harmonic phase – i.e. its initial phase $\theta_0$ (in radians). Thus, the graph of such a harmonic without breaks will be a flat horizontal line, provided we know its exact frequency ($\Delta \omega = 0$). If we cut out pieces of this sine wave at two points, the graph will look as shown in Fig. 21.
This clear picture of two breaks in the phase of a continuous harmonic is observed when we know exactly the harmonic frequency.

If our estimate of the harmonic frequency is not accurate, and, for instance, $\Delta \omega > 0$, then, as follows from formula (1), the phase of a continuous harmonic will be determined not only by the constant component $\theta_0$, but also by the $(\Delta \omega \cdot t)$ component linear in time, and the graph will look like a slanting straight line (with $\Delta \omega > 0$ the slant will be directed upwards).

Another and still more interesting example is when we precisely determined the harmonic frequency at the beginning of the file as $F=50$ Hz, while at the end of the file it is not observed – there is another harmonic instead, with a similar frequency $F=51$ Hz. Such a situation often takes place in audio recording editing as a result of insertions and over-recordings.
Then the phase indicator graph (Figure 22) will look not as a stepped line, but as a broken line with a clearly seen jog. Thus, the **Phase Scanning** tab can be used to scan both the phase and frequency of a specified harmonic.

In reality, the frequency of a continuous harmonic is smoothly drifting in time near the constant. This, for instance, happens to the 50 Hz power harmonic.

The picture below (Fig. 23) presents an analysis of a 50 Hz harmonic with a SNR favorable for the harmonic phase analysis. The 50 Hz amplitude here is 37 counts, while the speech maximum approaches 20,000 counts, i.e. the noise is less than 1000 times stronger than the signal. More accurate data about the SNR can be found in the **Instantaneous Spectrum** window on the **Spectrum Estimation** tab. SNR roughly equals 10.
The phase graph shown below (Fig. 24) demonstrates that the harmonic frequency is drifting continuously (without jogs), which means that there are no long insertions. In the middle of the file a typical stepped structure is observed, reflecting three phase breaks of this harmonic. This means that three short cuts have been made in the file.

Figure 23 – Power harmonics assessment
Figure 24 – Detection of phase breaks
6.2 Phase estimation

The expert’s work starts with the choice of an analysis object. Then the Search button should be clicked. The result of work with the tab will be marks designating harmonic phase breaks.

The number of various techniques applied during the analysis is determined mainly by the SNR for the selected harmonic. Provisional assessment of SNR can be obtained by analysing the data in the Instantaneous Spectrum window of the Spectrum Estimation tab.

In this window, the expert, assessing the harmonic peak and average level of ambient noise, can see SNR on a linear scale – as shown in Figure 25. The amplitude of the generated 50 Hz harmonic equals 25 counts. This harmonic was broken and masked with white noise having the amplitude of 10 000 counts.

In Figure 25 SNR, found in the Instantaneous Spectrum window, equals approximately 6. The Phase Scanning tab reveals the following picture in this case (Fig. 26):

![Image of the EdiTracker interface](image-url)
Normal phase accuracy (0.9) is selected in the Frame length for phase estimation window on the tab; although for this case high accuracy (1; the rightmost slider in the Frame length for phase estimation window) is recommended.

Figure 26 – Example of a phase jump against the strong noise background

A phase jump is clearly seen against the strong phase noise background in Figure 26. In this case the phase break is more than phase noise average value and the decision taken by the expert will be unambiguous.

The accuracy of phase break detection depends only on the difference by which the observed phase jump exceeds the average level of phase noise for the specified harmonic.
**“PHASE SCANNING” TAB**

The main instances of using the **Phase Scanning** tab are listed below:

a) In a standard situation (see Fig. 24), when SNR is specified on the previous tab for the selected harmonic: \( 20 \geq \frac{S}{N} \geq 8 \), prior to the analysis, the expert can choose the mode of normal phase smoothing. To do this, move the slider in the range \([0.6; 0.93]\) in the **Frame length for phase estimation** window. Phase jump area is extended from 1 to 4 seconds.

b) In case of weak noise and a very strong harmonic, when SNR is specified on the previous tab for the selected harmonic: \( \frac{S}{N} \geq 20 \), the expert can choose the mode of minimum phase smoothing. To do this, move the slider in the range \([0.01; 0.3]\) in the **Frame length for phase estimation** window. Phase jump area is extended from 0.5 to 2 seconds.

c) In case of strong noise in the harmonic phase, when SNR is specified on the previous tab for the selected harmonic: \( \frac{S}{N} \leq 8 \), prior to the analysis, the expert can choose the mode of maximum phase smoothing. To do this, move the slider in the range \([0.96; 1.0]\) in the **Frame length for phase estimation** window. In this mode both the noise and the phase jump proper will be considerably smoothed, with the phase jump area extended up to 10 seconds. This mode is recommended mainly when analyzing long files (longer than 2 minutes).

It should be taken into account that with any (low or high) SNR, the slider position in the range \([0.96; 1.0]\) yields the most reliable phase estimate. However, two conditions are to be met in this case. Firstly, file length should exceed 30 seconds. Secondly, there must not be more than one break within 20 seconds, otherwise other breaks will probably escape detection.

The slider in the **Frame length for phase estimation** window is controlled by the mouse and by the **Left** and **Right** buttons on the keyboard for more precise slider manipulation (to do this, activate the slider by clicking it with the left mouse button). The slider range is \([0.01; 1]\).
6.3 Additional options

a) Selecting the **Narrow-band filter** check box, you can use the mode of strong frequency filtration that additionally smoothes phase noises and suppresses the influence on the selected harmonic produced by other close harmonics (for instance, its duplicates). This option is recommended only in cases when duplicates or strong phase noises are present (i.e. with a very poor SNR).

b) Additional suppression of phase noises can be enabled by selecting the **Noise cancellation** check box.

c) **Automatic marks’ placement mode.** Selecting the **Auto-Marks** check box, you can use the automatic marks’ placement mode. With the help of the **Thresh** slider, set the minimum threshold for the phase jump according to which a decision will be taken automatically on availability of discontinuous-phase. Otherwise, you should set the marks manually according to the phase curve behaviour. The **Thresh** slider is controlled with the mouse as well as with the **Left** and **Right** buttons on the keyboard for more precise slider control. The slider range is [0; 1], which corresponds to using the threshold marks in the range [0; 3.14].

d) **Manual marks’ placement mode.** It is possible to delete any (or all) automatically set marks, if it is clear that the phase jump value corresponding to this mark is commensurable to the phase noise value. To do this, hover the mouse cursor over the desired mark and, once the cursor sign changes to the “index finger”, press the right mouse button. A menu will appear allowing you to delete this mark or all marks. The expert can set a mark in any place manually if it is clear that at this point a typical phase jump is observed satisfying all listed above criteria, even though this point escaped automatic detection. To do this, hover the mouse cursor over this place in the audio file window and double click it with the left mouse button.

Similar situations can be observed in complex cases of strong phase noise presenting difficulty for automatic analysis.

e) **Phase display mode.** Selecting the **Display phase in the [-3.14; 3.14] range** check box, you can use zoom phase display only on its period (Figure 27).
Figure 27 – “Phase Scanning” tab

Otherwise, zoom phase display will be absolute, i.e., the phase will be displayed on an infinite range $[-\infty, +\infty]$, as shown on Figure 24.

The phase display mode on the period is more noise-stable, and the **Frame length for phase estimation** parameter can be shorter than in the absolute scale. It is recommended in case of large phase noise, when large parasitic phase jumps on the values $(2 \cdot \pi \cdot n)$ can occur in the absolute scale, to verify questionable places of these jumps in the phase display mode on the period.
7 “BACKGROUND SCANNING” TAB

7.1 General information

![Figure 28 – “Background Scanning” tab](image)

The purpose of the analysis procedure on the Background Scanning tab is the detection of unnoticeable on the waveform events (points) of abrupt change of the spectrum background (ambient noise), but related to characteristic events of the recording process or, possibly, to tape editing.

Background is a stationary part of the signal with a stable or slowly changing spectrum as compared to the spectrum of non-stationary events (speech, steps, and impulses). Background change corresponds to a transition from a stationary spectrum to a stationary spectrum of a different form. Along with stationary background, an audio recording contains, as a rule, non-stationary events as well.

Unlike the background, non-stationary events are of short duration and are characterised by rapid change of the signal spectrum. The background reveals itself in intervals between non-stationary events. After the end of a non-stationary event the signal spectrum returns to the initial background spectrum.
Background estimation is performed on the averaging interval. The smaller is the interval, the stronger is the influence of non-stationary processes (e.g. speech) on the estimate. Hence, on small averaging intervals the background also becomes unsteady.

It should be kept in mind that the overwhelming majority of detected background jumps correspond to events unrelated to tape editing. E.g., the sound of a window opened in a room or the sound in a moving car can result in a background jump.

Figure 29 provides an example of background scanning for a conversation recorded over a radio channel.

Speech intervals in the channel are filled with powerful interference background. The detector shows background change which, however, is not related to tape editing.
Table 5. Events causing background spectrum changes

<table>
<thead>
<tr>
<th>Background behaviour</th>
<th>Event (process)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrupt level drop</td>
<td>Pause in the recording (absence of a signal)</td>
</tr>
<tr>
<td>Change of the average level</td>
<td>Change of the environment (recording conditions)</td>
</tr>
<tr>
<td>Change of the average level</td>
<td>Radio channel, strong interference in speech intervals</td>
</tr>
<tr>
<td>Impulses</td>
<td>Knocks and overload of the amplifier, speech signal leakage</td>
</tr>
</tbody>
</table>
7.2 Techniques of background spectrum analysis

Various representations (compression degrees) of background spectrum data can be obtained:

– spectrogram (sonogram) of the background;
– current background level (an integral characteristic of the background form and level);
– markers of abrupt change of the integral characteristic of the background spectrum.

**Signal spectrogram analysis.** Signal spectrogram (sonogram) contains the most detailed information about the signal (including background) spectrum and allows you to notice (detect) sufficiently contrastive changes in the recording level, background form change and other events.

**Current background level analysis (the “Background” option).** Spectrogram analysis is a rather laborious process. So it is convenient to use the current background level. The current level and dynamics of the background allows you to detect recording fragments, differing in the form of background spectrum. The presence of speech signals to certain extent affects the current background level.

The current background level allows you to notice (detect) sufficiently contrastive changes in the recording level, background form change and other events.

**Markers of abrupt change of the current background level.** The most convenient for fast analysis are markers of abrupt change in the current background level. Markers are a tool for fast detection of integral characteristic jumps. Markers are displayed as marks set in a waveform. Each marker corresponds to a certain time along the absolute time scale of the signal.

To reduce the number of false detections, adjust the detection settings.

**Control parameters.** Signal (and background) properties (statistics) can be very different. Similarly, background changes (jumps) can also be of a varying character (e.g. duration of recording interruptions). Control parameters are intended for adjusting detection settings to the characteristics of the interference and background changes.

**Control range.** Sets the frequency range for calculating integral (for the frequency) background characteristic. In the extreme case the analysis range covers the whole frequency band from 0 Hz to half the sampling frequency.

It is advisable to specify a frequency area where the influence of speech and other extraneous signals is minimal, while the background is characterised by the highest stability and its change is the most contrastive.

Besides, it is worthwhile to analyse the actual recording band of the signal. In cases when the signal is re-digitised (the upper frequency area contains only uniform quantisation noise), it is effective to set the real upper frequency of signal change or to decrease the sampling frequency.

**Frame Length.** Sets frame size and number of analysis quanta, affects sensitivity to events of respective duration (impulses, speech).

The *minimum frame size* \( N_{\text{min}} \) is determined by the sampling frequency in the following way:

\[
Fs > 11 \text{ kHz}, \quad N_{\text{min}} = 512
\]
“BACKGROUND SCANNING” TAB

Fs > 22 kHz, \( N_{\text{min}} = 1024 \)
Fs > 44 kHz, \( N_{\text{min}} = 2048 \)
Fs > 88 kHz, \( N_{\text{min}} = 4096 \)
Fs > 176 kHz, \( N_{\text{min}} = 8192 \)

*Averaging time.* Sets temporal duration of averaging intervals in which the background level is compared:

\[
\text{AVERAGING INTERVAL} = \text{NUMBER OF FRAMES} \times \text{FRAME SIZE} / \text{SAMPLING FREQUENCY}
\]

The larger is the interval, the higher is the sensitivity of background change detection, and the fewer is the false reactions of the algorithm on short random fluctuations of the background (interference). The same interval is used for setting the minimum duration of a detected event (e.g. a short interruption in the recording).

It is advisable to set \( T_s > 0.5 \) sec. for background change control, and \( -T_s \leq 0.1 \) sec. for controlling signal interruptions.
7.3 Method of signal mean value controlling ("Mean value" method)

The purpose of mean value control is the detection of unnoticeable on the waveform events (points) of abrupt change of the spectrum background (ambient noise), but related to characteristic events of the recording process or, possibly, to tape editing.

Operations with an audio recording having a constant mean value do not result in the change of the mean (fragment deletion, insertion of a fragment taken from a different place of the same audio recording). However, in case of trend or parasitic amplitude modulation of the mean value, operations with an audio recording can be accompanied by abrupt changes of the mean.

Thus, the mean value is an important informative parameter providing information about certain peculiarities of the recording process.

The mean value is its numeric estimate within the user-specified interval (from 1 frame to several seconds). The frame size is user-specified.

The sequence of mean value estimates in the frames is a stochastic process with a sampling rate equal to the inverse of the time step (shift) of analysis frames. Averaging mean estimates over several frames (the averaging interval) is equivalent to low-frequency filtration of the process. The larger is the averaging interval, the narrower is the stochastic process area, the smoother is the function of the mean value estimate. Individual short outbursts of the mean are smoothed and eliminated.

Besides random fluctuation (observed, for instance, in an artificially generated random signal), mean value change is affected by the signal recording process (pause, drift, interference etc.), as well as by operations related to tape editing.

Events causing mean value change are related to the properties of the tape recorder and the audio recording and can be determined by various factors (see Table 6).

<table>
<thead>
<tr>
<th>Mean value behaviour</th>
<th>Event (process)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jumps</td>
<td>Recording start/stop</td>
</tr>
<tr>
<td>Instability (trend)</td>
<td>Instability of the speech signal related to amplifier operation. Some individual portions of the recording have different mean values.</td>
</tr>
<tr>
<td>Periodic vibration</td>
<td>Amplitude modulation at 2.5 Hz frequency</td>
</tr>
<tr>
<td></td>
<td>Amplitude modulation at 25 Hz frequency</td>
</tr>
<tr>
<td>Impulses</td>
<td>Knocks and overload of the amplifier, speech signal leakage</td>
</tr>
</tbody>
</table>
Mean value characteristics. Various representations of the mean value data (with different degree of detail) can be obtained:

– mean value;
– markers of abrupt change of the mean.

Current background level. Allows monitoring the mean value behaviour in time and setting markers by hand.

Markers of abrupt change of the mean. Markers are a tool for a fast automated detection of jumps of the mean. Markers are displayed as blue marks set in a waveform. Each marker corresponds to a certain time along the absolute time scale of the signal.

Since markers characterise jumps of the mean relative to its statistics, they can also reflect random fluctuations of the mean. To reduce the number of false detections, adjust the detection settings.

Control parameters. Control parameters are intended for adjusting the detection settings to the characteristics of the interference and mean value changes.

Minimum frame size \((N_{\text{min}})\) is determined by the sampling frequency in the following way:

\[
\begin{align*}
\text{Fs} & > 11 \text{ kHz}, \ N_{\text{min}} = 512 \\
\text{Fs} & > 22 \text{ kHz}, \ N_{\text{min}} = 1024 \\
\text{Fs} & > 44 \text{ kHz}, \ N_{\text{min}} = 2048 \\
\text{Fs} & > 88 \text{ kHz}, \ N_{\text{min}} = 4096 \\
\text{Fs} & > 176 \text{ kHz}, \ N_{\text{min}} = 8192
\end{align*}
\]

Averaging interval. Sets temporal duration of averaging intervals, in which the background level is compared.

When estimating the mean, the used control parameters are frame size and averaging time (interval).

\[
\text{AVERAGING TIME} = \text{NUMBER OF FRAMES} \times \text{FRAME SIZE} / \text{SAMPLING FREQUENCY}
\]

The larger is the time interval, the higher is the sensitivity of mean value change detections, and the fewer is the false reactions of the algorithm on short random fluctuations of the mean value (interference).

It is advisable to set long averaging times (> 1 sec.) for tracing continuous changes of the mean’s level.

Short events (recorder start/stop) can be best detected with averaging times < 0.2 sec.

The detectability limit of the mean value’s change amplitude (for 16-bit data) is 20. As the noise level increases, the detectability is reduced and equals (on white noise of the maximum amplitude) the amplitude of the mean value (approximately 350).
The method yields many false reactions on clipped (recorded with overload) signals of large amplitude.

The **Background Scanning** tab window provides two graphs:

1. A graph of the audio signal waveform;
2. A graph of the audio recording continuity indicator.

To start audio recording analysis, choose the desired frequency range, frame length and averaging time and click the **Search** button.

The signal appears on the audio recording signal graph. The signal graph is moving while the signal is being processed. The graph of the audio recording continuity indicator is periodically refreshed – simultaneously with the signal graph. If, according to a certain criterion, a discontinuity appears on the continuity indicator, a blue mark will be set in the respective position of the signal waveform.

In the **Points of non-uniformity** list a list of marks will be displayed in the order of their appearance in the signal. Each entry in the list corresponds to one mark and specifies its time position in the signal, as well as the strength of the detected discontinuity.

After clicking the **OK** button blue marks will appear in the waveform, corresponding to those listed in the **Points of non-uniformity** list.
8.1 Guidelines for the detection of audio recording authenticity breaches by means of auditory-linguistic analysis

Detection of audio recording authenticity breaches based on auditory-linguistic analysis requires that all of the sound data be carefully listened to for any signs of recording authenticity breaches.

This type of analysis is based on the assumption that while the original recording is being played, its signal characteristics, communicative circumstances, verbal and non-verbal behaviour of speakers, characteristics of extra sound sources etc. form a holistic structure of mutually related sound and speech events. Auditory analysis of such events, adjusted for the characteristics of the used equipment and recording means known to the expert, allows the detection of integrity breaches, determination (with a varying degree of certainty) of their positions and factors causing such breaches.

The suggested technique is based on a holistic approach to audio contents analysis involving consecutive examination of various signal aspects while constantly resorting to its holistic consideration.

The basic audio recording analysis guidelines include:

a) Detection of manifestations of additional signals (extraneous audio events, artefacts) and sound distortions inconsistent with the known circumstances and conditions of the communicative act.

b) Detection of continuity and integrity breaches in the main sound signal.

c) Detection of continuity and integrity breaches in the background (additional) sound sources.

d) Detection of continuity and integrity breaches in the communicative act environment.

e) Detection of inconsistencies between the established (on the basis of the analysis) information, conditions, circumstances, speaker characteristics and the information about this communicative act known from other sources.

A more detailed description of these guidelines is given below.

In general the technique comprises three basic levels of audio recording examination.

1. Auditory examination of the recording by its careful repeated selective listening and detection of sound objects and their characteristics inconsistent with the communicative circumstances or characterised by changes unjustified from the point of view of the known communicative circumstances.

2. The establishment of linguistic features characterising the speech in the audio recording and their analysis in terms of uniformity, integrity and consistency with the known circumstances.

3. The establishment of a complete set of circumstances accompanying the recorded communicative act (characteristics of the analysed discourse) and analysis of their consistency, semantic and pragmatic content, taking into consideration the information known from other sources.

A stepwise description of the expert’s actions following this technique is given below.
Auditory-linguistic analysis of an audio recording is an autonomous procedure, but due to the requirements for the completeness of expert examination, auditory analysis should be combined with instrumental analysis.

To improve accuracy and objectivity of the analysis findings, it is recommended to use visual and parametric analysis of speech signal waveforms. For the detection of extraneous sound events (artefacts – knocks, crackle, signal dropouts, impulses and signal transitions, jumps of signal level and timbre etc) it is recommended to use the analysis of the signal waveform and dynamic spectrogram. These types of analysis are also extremely useful for determining phonetic and prosodic peculiarities of the speech signal. Pitch properties of the speech signal should be examined by means of dynamic cepstrogram images and voice pitch curves.

When taking an expert decision, each detected breach of audio recording authenticity should be checked using both types of analysis: instrumental and auditory-linguistic.
8.2 Description of the work with the SIS II software when performing auditory linguistic analysis of audio recordings

Basic operations for audio recording processing in the SIS II sound editor are illustrated below.

For detailed description of other features of signal processing in the data windows, its visualization and playback in the SIS II sound editor see “SIS II. Sound editor. User’s Guide”.

8.2.1 Opening sound files

To listen to an audio recording, it is necessary to open the required sound file in the active data window. Please perform the following actions, to open a file:

1) In the File menu click Open… (File → Open…), click the Open pictogram on the toolbar or press Ctrl+O on the keyboard.

2) Choose necessary file in the Open File dialog box and click the Open button (Fig. 30).

Figure 30 – “Open File” dialog window to choose a sound file
8.2.2 Visualization of data to be listened

The area of data visible in the window provides the ample opportunities for browsing both entire data and any part of them.

During signal playback it is often inconvenient to represent all segment data in the data window, particularly if its duration is more than 1 minute. In this case the user does not see the data played once the cursor reaches the boundary of the data window.

It may be changed with:

- the **View** menu;
- the context menu of the data window and the scroll bar, which may be opened by hovering the mouse cursor over them and clicking the right mouse button;
- the mouse wheel when hovering the cursor over the horizontal or vertical scale;
- the horizontal scroll bar;
- the toolbar pictograms.
8.2.3 Audio recordings playback

The audio recordings playback is implemented with the Playback menu, with the proper pictograms on the toolbar and also with the keyboard shortcuts.

To play the whole signal of the active tab, use one of the methods:

1) In the Playback menu click Playback (Playback → Playback).
2) Click the ▶ Playback pictogram on the toolbar.
3) Press the F6 key on the keyboard.

To play the active tab’s signal beginning from the vertical cursor position, in the Playback menu click From Cursor (Playback → From Cursor).

To pause the playback at the position where the cursor is in the certain moment, use one of the methods:

1) In the Playback menu click Pause (Playback → Pause).
2) Click the ⏸ Pause pictogram on the toolbar.
3) Press Ctrl+P on the keyboard.

To resume the playback, repeat the actions to pause playback.

To stop the playback, use one of the methods:

1) In the Playback menu click Stop (Playback → Stop).
2) Click the ■ Stop pictogram on the toolbar.
3) Press the Esc key on the keyboard.

8.2.4 Operations with the marks

To start working with an audio recording, it is necessary to open the required recording in a separate window.

An expert should listen to the audio recording to find the fragments of suspected breaches of authenticity. It is recommended to rely on the list of audio recording authenticity breaches proposed in this user’s guide (see the appendices A and B).

The fragments, which contain features of audio recording authenticity breaches, are marked up with the proper marks during the auditory-linguistic analysis.

To set a mark of the linguistic breach, hover the mouse cursor over a necessary position and press the 9 key (on the main keyboard part, not a numerical one). The window, which contains the tree list of features of authenticity breaches will appear on the screen (see Fig. 32). Find a proper feature in the tree list and choose it by clicking the left mouse button. The mark of this feature will be represented in the data window.
The mark information will appear in the table of the marks list of the data window (Fig. 33). If it is necessary, you can edit the mark comment in the table.
To remove the single mark, use one of the following methods:

1) Move the mouse cursor over the single mark you want to remove and press **Ctrl+Delete**.

2) Move the mouse cursor over the single mark you want to remove, click the right mouse button and select the **Delete mark** item in the context menu.

To remove all the vertical marks, on the **Marks** menu click **Remove Vertical Marks** (**Marks** → **Remove Vertical Marks**).

Set marks appear not only in the data area, visible in the window, but also in the navigation oscillogram, as well as in the list of marks of the group **Mont. traces** → **Linguistic break** (Fig. 34).

![Figure 34 – “Marks” tab of Manager Panel](image)
8.3 Step-by-step user instructions for performing auditory-linguistic analysis of an audio recording for the detection of audio recording authenticity breaches

1. Preparing the data for expert examination:
   a) Open the recording to be analyzed in a “signal window” of the sound editor SIS II.
   b) Create a window with a new table of audio recording authenticity breaches and assign a name to the table.
   c) Open the list of audio recording authenticity breaches (see the appendices A and B).

2. Localization of phenomena suggestive of audio recording tampering and their preliminary analysis.
   a) Listen to the audio recording. It is usually effective to play the entire recording firstly (from the beginning to the end), marking suspicious areas for a more detailed auditory examination.
   b) Determine the text transcript of the recorded conversation, save it as a text file and print if necessary.
   c) If necessary, cut certain fragments from the audio recording for further autonomous analysis and save them as separate files.
   d) It is recommended to listen to the recording several times, each time concentrating on the following aspects:
      – Characteristics of the speaker;
      – Characteristics of the communication environment;
      – Characteristics of the recorded signal.

It is necessary to analyze all basic groups of characteristics.

   e) Start careful selective listening, playing small fragments of the audio recording repeatedly, each time concentrating your attention on one or several similar indicators of recording inauthenticity. For instance, when listening to the recording for the first time, try to detect clicks and energy gaps in the signal. During the next listening session concentrate on background noise jumps in pauses and during word utterance. When listening to the recording for the third time, pay attention to the uniformity and continuity of background music and noise of remote speech and sounds of transport vehicles. The following listening sessions can be directed at the speaker’s state, mood and intentions etc.

Experienced experts can get by with a small number of sessions, however in this case the accuracy of tampering signs detection falls considerably. A tape recording of 10 minutes long should be typically listened to 3-5 times from the beginning to the end, while some of its questionable fragments require up to 20 listenings.

   f) Listen to the recording while conducting auditory analysis. Tick off suspicious points of inauthenticity and enter the respective data into the table.
g) Analyze the text of the audio recording, mark all indicators of audio recording inauthenticity having a linguistic nature and enter them into the table.

After detecting and marking all questionable points, go to the next step of the analysis.

3. Careful examination of detected points in the audio recording containing indicators of its presumptive inauthenticity. Work out in detail the data entered into the table: define or specify the type of each indicator. If necessary, add your own text comments. It is necessary to examine all points marked in the table (Figure 35). If you notice another point containing indicators of audio recording inauthenticity, enter it into the table and analyze it in detail.

To get assistance with the choice and analysis of audio recording authenticity breaches, click the Procedure button. You will get access to these step-by-step instructions for the auditory-linguistic analysis and to the extended list of indicators of recording authenticity breaches arranged as a tree, as well as to a detailed description of each indicator.

Figure 35 – Table of marks of the “Auditory analysis” tab
8.4 Comparison of the results of auditory and instrumental analysis

1. Audio recording fragments with detected inauthenticity indicators should be subjected to the instrumental analysis.

2. Study the results of audio recording instrumental analysis. Analyze thoroughly all areas with detected and instrumentally established indicators of audio recording inauthenticity.

3. Enter the results of your analysis of these fragments into the table. If all experts agree on the presence of an authenticity breach, the respective statement is entered into the Conclusions. In case of contradictions and discrepancy in the findings of audio recording analysis conducted by different experts, each opinion is described in the research part of the Conclusions; the conclusions are formulated on the basis of compromise between the experts.

Verification of audio recording authenticity is a rather complex procedure, and thus the detection of audio recording inauthenticity indicators by means of at least one type of analysis procedure is usually sufficient for drawing a conclusion about audio recording inauthenticity.

4. According to the research findings draw a conclusion about the presence or absence of audio recording authenticity breaches in the audio recording.

5. Drawing up a conclusion. The whole table with the left areas of phonograms with the time marks and the set signs can be placed in the text of an expert conclusion.
On the basis of the analysis findings the **Conclusions** tab formulates conclusions about the presence or absence of inauthenticity indicators in the given audio recording.

You can save obtained data (conclusions) as a text file by clicking the **Save** button.
10 MODULE SHUTDOWN

After the research of the audio recording, study and saving of necessary data, close the EdiTracker module window by clicking **OK**, **Cancel** or **x**.

If the study concerning phase or background scanning was carried out on the relevant tabs of the module (refer to the sections 6 “PHASE SCANNING” TAB, 7 “BACKGROUND SCANNING” TAB), then when closing windows by clicking **Cancel** or **x**, dialog box will appear, notifying you about the closure of the module without saving your work (Figure 37).

To continue working with the module or to save the results of your work, click **Cancel** or **x**.
To exit without saving the results, click **OK**.

![Figure 37 – Dialog box of saving the results](image)

If the study concerning phase or background scanning was carried out, then when clicking **OK**, dialog box will appear, suggesting setting labels (marks) of possible tampering in the signal window (Figure 38).

![Figure 38 – Dialog box of setting the marks (labels) of possible tampering](image)

To continue working with the module, click **Cancel** or **x**.
To exit without setting the marks, click **No**.
To set the labels (marks) of phase or background break in the signal, click **Yes**. They will also appear on the **Marks** tab of the **Manager Panel** in the **SIS II** main window (Figure 39).
Operation with these marks is just the same as with the linguistic break marks (refer to the section 8.2.4 Operations with the marks).
Appendix A: List of audio recording authenticity breaches

1 Change of speaker characteristics
   1.1 Change of the composition of speakers
      1.1.1 Change of the number of speakers
      1.1.2 Change of individual speaker characteristics
         1.1.2.1 Change of appearance-related characteristics of the speaker
            1.1.2.1.1 Change of anatomical/physiological characteristics of the speaker
               1.1.2.1.1.1 Change of auditory assessment of the speaker's bodily constitution
               1.1.2.1.1.2 Change of auditory assessment of the speaker's relatively constant emotional and energy characteristics
               1.1.2.1.1.3 Change of auditory assessment of the speaker's temperament (strength, volatility/stability of the nervous system)
            1.1.2.1.2 Change of auditory assessment of the speaker's emotionality
            1.1.2.1.2.3 Change of auditory assessment of the speaker's energy tone (psychological activity)
            1.1.2.1.3 Change of auditory assessment of the speaker's sensorimotor organization
            1.1.2.1.4 Change of auditory assessment of the speaker's speech production and respiratory system defects
            1.1.2.1.5 Change of auditory assessment of the speaker's intellect
            1.1.2.1.6 Change of auditory assessment of the speaker's character, inclinations and abilities
            1.1.2.1.7 Change of auditory assessment of the speaker's ability for empathy
         1.1.2.2 Change of relatively stable social and behavioral characteristics of the speaker
            1.1.2.2.1 Change of auditory assessment of speaker motivation (organic, material, social, spiritual)
            1.1.2.2.2 Change of auditory assessment of the speaker's social functions/roles
            1.1.2.2.3 Change of auditory assessment of the speaker in terms of extraversion/introversion
            1.1.2.2.4 Change of auditory assessment of the speaker's suggestibility and conformity
            1.1.2.2.5 Change of auditory assessment of the speaker's values/principles/ideals of living
            1.1.2.2.6 Change of auditory assessment of the speaker's status (family, social, work, etc.)
            1.1.2.2.7 Change of auditory assessment of the speaker's knowledge, skills and abilities
            1.1.2.2.8 Change of auditory assessment of the speaker's goals and orientation (interests, volitions, beliefs, world outlook)
            1.1.2.2.9 Change of auditory assessment of the speaker's volitional qualities (goal striving, persistence, self-possession, determination, initiative, orderliness, discipline, courage, dutifulness)
            1.1.2.2.10 Change of auditory assessment of the speaker's curiosity
            1.1.2.2.11 Change of auditory assessment of the speaker's self-appraisal
               1.1.2.2.11.1 Change of auditory assessment of the speaker's self-appraisal quality
APPENDICES

1.1.2.1.2 Change of auditory assessment of manifestation degree of the speaker's self-appraisal

1.1.2.2 Considerable change of the perceived voice quality (timbre) of the speaker

1.1.2.3 Considerable change of the speaker's speech characteristics

1.1.2.3.1 Change of auditory assessment of the speaker's degree of language proficiency, accent/dialect

1.1.2.3.2 Change of auditory assessment of the speaker's speech defects

1.1.2.3.3 Change of individual speech characteristics of the speaker

1.2 Considerable change of the speaker's state

1.2.1 Change of the speaker's physiological state

1.2.1.1 Change of the speaker's disease manifestations

1.2.1.1.1 Change of auditory assessment of manifestations of the speaker's speech and respiratory system disease

1.2.1.1.2 Change of auditory assessment of manifestations of the speaker's mental diseases

1.2.1.1.3 Change of auditory assessment of manifestations of the speaker's general-system diseases

1.2.1.2 Change of the state of the speaker's consciousness

1.2.1.2.1 Change of auditory assessment of the speaker's alcoholic intoxication (presence/absence and degree of manifestation)

1.2.1.2.2 Change of auditory assessment of the speaker's narcotic intoxication (presence/absence and degree of manifestation)

1.2.1.2.3 Change of auditory assessment of the speaker's affect state (presence/absence and degree of manifestation)

1.2.1.2.4 Change of auditory assessment of the speaker's state after taking sedatives (presence/absence and degree of manifestation)

1.2.2 Change of the speaker's psychological state

1.2.2.1 Change of auditory assessment of the speaker's emotional state (change of polarity, degree of manifestation and type of emotion)

1.2.2.2 Change of auditory assessment of expressed feelings (moral, intellectual, aesthetic, activity-oriented)

1.2.2.3 Change of auditory assessment of the speaker's energy tone

1.2.2.4 Change of auditory assessment of the speaker's attention state (activity, direction, scope, switchability, intensity, stability)

1.2.2.5 Change of auditory assessment of the speaker's role policies

1.2.2.5.1 Qualitative change of the speaker's role policy of domination

1.2.2.5.2 Qualitative change of the speaker's role policy of equality

1.2.2.5.3 Qualitative change of the speaker's role policy of subordination

1.2.2.6 Change of the speaker's attitude to the communication act

1.2.2.6.1 Change of auditory assessment of the speaker's interest in the conversation
1.2.2.6.2 Change of auditory assessment of carefulness/carelessness of the speaker's communicative actions
1.2.2.6.3 Change of auditory assessment of unhurriedness/haste in the speaker's communicative actions
1.2.2.6.4 Change of auditory assessment of the speaker's laconism/verbosity
1.2.2.7 Change of attitude towards communication partner
1.2.2.7.1 Change of auditory assessment of the degree of manifestation of the speaker's friendliness towards communication partner
1.2.2.7.2 Change of auditory assessment of the degree of manifestation of the speaker's neutral attitude towards communication partner
1.2.2.7.3 Change of auditory assessment of the degree of manifestation of the speaker's hostility towards communication partner
1.2.2.8 Change of auditory assessment of the interlocutors' familiarity with each other
1.2.2.8.1 Change within the “close people” level of acquaintance
1.2.2.8.2 Change within the “good acquaintances” level of acquaintance
1.2.2.8.3 Change within the “acquaintances” level of acquaintance
1.2.2.8.4 Change within the “superficial acquaintances” level of acquaintance
1.2.2.8.5 Change within the “strangers” level of acquaintance
1.2.2.9 Change of auditory assessment of the direction of communicative actions in the conversation
1.2.2.9.1 Qualitative change of the instructive, persuasive tone of the speaker's communicative actions
1.2.2.9.2 Qualitative change of the narrative, declarative tone of the speaker's communicative actions
1.2.2.9.3 Qualitative change of the disputing, argumentative (everyday/professional conversation) tone of the speaker's communicative actions
1.2.2.9.4 Qualitative change of the expressive tone of the speaker's communicative actions
1.2.3 Change of auditory assessment of the speaker's mental state
1.2.3.1 Change of the speaker's knowledge about (degree of familiarity with) the subjects and topics touched upon in the conversation
1.2.3.2 Change of the goals of the communication act
1.2.3.2.1 Unjustified appearance or disappearance of a communication goal
1.2.3.2.2 Unjustified change of the speaker's attitude to the goal (change of the degree of its relevance to the speaker)
1.2.3.3 Unjustified change of the opinion about the subjects and topics touched upon in the conversation
1.3 Change of peculiarities of discourse behavior
1.3.1 Unjustified change of the speaker's discourse strategy
1.3.2 Unjustified change of the speaker's manner of communication
1.3.2.1 Unjustified change of the speaker's tone of voice (instructive, ingratiating etc.)
1.3.2.2 Unjustified change of distance in communication
1.3.2.3 Unjustified appearance/disappearance of the verbal play, humor
1.3.2.4 Unjustified change of the used idiolect/argot
1.3.2.5 Unjustified change of the degree of selectiveness in information perception

1.3.2.6 Other features of speaker communication

1.3.3 Change of the person (grammatical category) used by the speaker (a. Speaks in the first person singular (“for himself”); b. Speaks in the first person plural (“for a group”); c. Speaks for another person or persons; d. Uses impersonal utterances)

1.3.4 Change of preparedness degree for communication (a. The communicating parties are not prepared; b. Prepared by force of custom, habit; c. Specifically prepared for this conversation)

1.3.5 Change of the psychological style of communication (individually typical peculiarities of interaction)

1.3.5.1 Change within the psychological style “Efforts towards getting information”

1.3.5.2 Change within the psychological style “Efforts towards giving information”

1.3.5.3 Change within the psychological style “Efforts towards expressing personal attitude”

1.3.5.4 Change within the psychological style “Efforts towards getting along with the interlocutor”

1.3.5.5 Change within the psychological style “Efforts towards conflicting with the interlocutor”

1.3.5.6 Change within the psychological style “Efforts towards getting appreciation from the interlocutor, producing an impression”

1.3.6 Breach of integrity and continuity of the extra-linguistic discourse component

1.3.6.1 Change of the speaker’s manner of gesticulation

1.3.6.2 Change of the speaker’s mimicry

1.3.6.3 Change of proxemics (personal space in communication)

1.3.7 Breach of integrity and continuity of the paralinguistic discourse component

1.3.7.1 Breach of integrity and continuity of acoustic phenomena related to nonverbal activities (laughter, cry, cough etc.)

1.3.7.2 Breach of integrity and continuity of the emotional voice change

1.3.7.3 Breach of integrity and continuity of breathing

1.3.7.3.1 Unjustified change of the overall character of breathing

1.3.7.3.2 Unjustified change of speech breathing pattern

2 Considerable change of communication environment

2.1 Discontinuity of place of communication

2.1.1 Breach of integrity and continuity of the process of speaker localization change

2.1.2 Unjustified change of speaker localization

2.1.2.1 Unjustified change of auditory characteristics of an open space where the recording takes place

2.1.2.2 Unjustified change of auditory characteristics of an enclosed space where the recording takes place

2.1.2.3 Unjustified change of auditory characteristics of a large-sized room where the recording takes place

2.1.2.4 Unjustified change of auditory characteristics of a medium-sized room where the recording takes place

2.1.2.5 Unjustified change of auditory characteristics of a small-sized room where the recording takes place

2.1.2.6 Unjustified change of auditory characteristics of a vehicle where the recording takes place

2.1.2.7 Unjustified change of auditory characteristics of a specialized room where the recording takes place
2.2 Breach of integrity and continuity of the acoustic environment
   2.2.1 Unjustified change of background noise environment
      2.2.1.1 Unjustified change of noises in the acoustic environment
      2.2.1.2 Unjustified change of the microphone and signal transmission channel noise
      2.2.1.3 Unjustified change of tape and recorder noise
   2.2.2 Unjustified change of the acoustic background
      2.2.2.1 Breach of integrity and continuity of acoustic events (music, speech, etc.)
      2.2.2.2 Instability of periodic sound sources (clock ticking, etc.)

2.3 Change of auditory assessment of the speaker's closeness to the microphone

2.4 Change of auditory assessment of the peculiarities of the recording environment (directed at the speaker, microphone placement, sound transmission channel)
   2.4.1 Change of auditory assessment of the type of sound transmission channel (transceiver, telephone, microphone, etc.)
   2.4.2 Considerable change of auditory assessment of the recording channel properties

2.5 Breach of integrity and continuity of the background chain of events

3 Inconsistency between the communicative act and circumstances known to the speakers or experts
   3.1 Inconsistency between the contents of the utterances and circumstances known to the speakers or experts
   3.2 Inconsistency between the speakers' discourse behaviour and circumstances known to the speakers or experts
   3.3 Inconsistency between the speakers' behaviour and circumstances known to the speakers or experts
   3.4 Inconsistency between the established circumstances or communication act environment and the circumstances known to the speakers or experts

4 Considerable change of the recorded signal properties
   4.1 Considerable change of speech signal integrity and continuity pattern
      4.1.1 Breach of integrity and continuity of oral discourse contents
         4.1.1.1 Breach of integrity and continuity of the rhematic and thematic structure
         4.1.1.2 Inadequacy of speech act structural components (locution, illocution, perlocution)
            4.1.1.2.1 Inconsistency between locution and illocution
            4.1.1.2.2 Illocutionary inadequacy
               4.1.1.2.2.1 Breach of intention adequacy and integrity within the discourse
               4.1.1.2.2.2 Breach of intention adequacy and integrity within an utterance
            4.1.1.2.3 Perlocutionary inadequacy
               4.1.1.2.3.1 Inadequacy of reaction to the interlocutor's illocution
               4.1.1.2.3.2 Inadequacy of reaction to the interlocutor's locution
4.1.1.3 Breach of topic continuity of oral discourse (presence or absence of situationally unconditioned changes of the conversation topic)
  4.1.1.3.1 Breach of macro-topic integrity and continuity
  4.1.1.3.2 Breach of micro-topic integrity and continuity

4.1.1.4 Unjustified change of the conversation topic fixity degree (a. Free topic; b. Fixed topic area; c. Highly fixed specific topic)

4.1.1.5 Breach of logical and semantic integrity and continuity of utterances and oral discourse in general
  4.1.1.5.1 Breach of logical and semantic integrity and continuity of conversation in general
  4.1.1.5.2 Breach of logical and semantic integrity and continuity of dialogic entity
  4.1.1.5.3 Breach of logical and semantic integrity and continuity of individual utterances

4.1.2 Non-standard way of oral discourse realisation
  4.1.2.1 Non-standard grammar of utterances
    4.1.2.1.1 Unjustified non-standardness of word order in the utterance
    4.1.2.1.2 Unjustified non-standardness of grammatical agreement
    4.1.2.1.3 Unjustified non-standardness of grammatical government
    4.1.2.1.4 Unjustified non-standardness of negation
    4.1.2.1.5 Unjustified non-standardness of conjugation
    4.1.2.1.6 Unjustified non-standardness of declension
    4.1.2.1.7 Unjustified non-standardness of the plural formation
  4.1.2.2 Breach of lexical integrity and continuity of oral discourse
  4.1.2.3 Breach of stylistic integrity and continuity
    4.1.2.3.1 Inconsistency between the used stylistically marked structures, idioms, expressions and the functional speech style in the given communicative situation
    4.1.2.3.2 Unjustified change of speech style
    4.1.2.3.3 Unjustified change of language code
  4.1.2.4 Breach of phonetic integrity and continuity (violation of phonetic laws)
    4.1.2.4.1 Incorrectly realised combinatorial changes of consonants
    4.1.2.4.2 Incorrectly realised positional changes of consonants
    4.1.2.4.3 Incorrectly realised combinatorial changes of vowels
    4.1.2.4.4 Incorrectly realised positional changes of vowels
  4.1.2.5 Breach of prosodic integrity and continuity of oral discourse
    4.1.2.5.1 Unjustified change of speech loudness
    4.1.2.5.2 Unjustified change of speech tempo and rhythm
    4.1.2.5.3 Unjustified change of speech timbre
    4.1.2.5.4 Unjustified change of discourse pitch patterns
      4.1.2.5.4.1 Unjustified change of word pitch pattern
      4.1.2.5.4.2 Unjustified change of syntagm pitch pattern
      4.1.2.5.4.3 Unjustified change of utterance pitch pattern
APPENDICES

4.1.2.5.4.4 Unjustified change of phonetic paragraph pitch pattern
4.1.2.5.5 Incorrect pausing in oral discourse
  4.1.2.5.5.1 Non-standard breathing pauses
  4.1.2.5.5.2 Non-standard pausing between utterances
  4.1.2.5.5.3 Non-standard pauses between syntagms
  4.1.2.5.5.4 Non-standard pauses between phonetic words (rhythmic groups)
  4.1.2.5.5.5 Non-standard character of situational conditioned pauses
  4.1.2.5.5.6 Non-standard hesitation pauses
4.1.2.5.6 Incorrect stress/accent placement in oral discourse
  4.1.2.5.6.1 Non-standard placement of logical accent
  4.1.2.5.6.2 Non-standard placement of phrasal accent
  4.1.2.5.6.3 Non-standard placement of syntagmatic accent
  4.1.2.5.6.4 Non-standard placement of word stress
  4.1.2.5.6.5 Non-standard placement of emphatic accent
4.1.2.5.7 Unjustified change of the used ways of prosodic patterning, phrasing, highlighting of oral discourse elements

4.2 Presence of extraneous sound signals and distortions related to the sound recording and transmission process
  4.2.1 Presence of “clicks”, “crackle”, “knocks”
  4.2.2 Change of acoustic depth of the recording
  4.2.3 Change of sound level
    4.2.3.1 Sound signal dropouts
    4.2.3.2 Jumps of loudness
    4.2.3.3 Signs of application of AGC systems
    4.2.3.4 Signs of application of voice-activated recording means
  4.2.4 Recording speed fluctuations
  4.2.5 Change of the recording spectral range and Amplitude-Frequency Response
  4.2.6 Change of the compression way – conversion of the recorded signal
  4.2.7 Change of the distortions’ nature in the recorded signal
  4.2.8 Change of the background interference nature in the sound transmission channel
Appendix B: Description of indicators used in auditory-linguistic analysis of audio recordings

1 Change of speaker characteristics

Change of perceptually relevant (to the expert’s ear) substantial characteristics of the speakers participating in the conversation, which can indicate unjustified (by the circumstances known to the expert) change in the composition of the speakers or state(s) of the speaker(s), which in turn can indicate a breach of audio recording authenticity.

1.1 Change of the composition of speakers

Changes in the composition of the speakers taking part in the recorded conversation can be considered an indicator of audio recording inauthenticity if it is situational unjustified, i.e.:

– If the composition of speakers is not consistent with the known case data, circumstances known to the expert and the logical structure of the recorded conversation;
– If the change of the communicative act contents is not consistent with the change in the composition of speakers (e.g. the interlocutors show no reaction to the appearance of a new speaker);
– If there are unjustified changes in the individual characteristics of the speaker, i.e. the speaker personality changes (e.g. when some of the utterances are pronounced by an imitating speaker (mimic, impersonator) instead of the original speaker).

1.1.1 Change of the number of speakers

Disappearance of one of the participating speakers from the conversation or the appearance of a new speaker. If the change in the number of speakers is situational unjustified, it may be considered an indicator of audio recording authenticity breach. However, it should be taken into account that a new speaker could have been there all along from the beginning of the conversation, but just kept silence, and that the reaction to the appearance of a new speaker could also be non-verbal (e.g. a handshake, a nod etc.).

1.1.2 Change of speaker individual (stable) characteristics

Stable characteristics are those which do not change for a specific speaker or change only over a considerable period of time. An abrupt and unjustified change of such characteristics can indicate either that the speech belongs to different speakers or that the speech belongs to one speaker but was recorded at different periods of time. An example could be a different emotional or physical state of the speaker, the appearance of signs of fatigue, disease, effect of alcohol or narcotics, etc. in his speech.

1.1.2.1 Change of appearance-related characteristics of the speaker

Change of voice- and speech-derived individual characteristics of the speaker related to his anatomical-physiological, mental, psychological, socio-cultural peculiarities.

1.1.2.1.1 Anatomical-physiological characteristics of the speaker

Speaker characteristics reflected in the speech signal and related to the geometrical dimensions, acoustic properties of tissues of the speech and respiratory organs, as well as to his breathing system activity.
First of all, it is the size (length) of the speaker’s vocal tract (distance from the larynx to the lips), equivalent acoustic dimensions of the thoracic cavity, the degree of acoustic pliancy of muscular walls of the oral cavity and the larynx, the degree of sound attenuation in the bronchial tree of the lungs (related to the characteristic tone of the smooth muscles of the bronchi), the degree of acoustic connection of the oral and nasal cavity with the lungs through the vocal folds, the degree of acoustic ties between the oral and nasal cavity and skull sinuses, the kind and degree of acoustic relations between the nasal and oral cavities, the presence of defects in the structure of the soft and hard palate, teeth, lips, the size (diameter) of the larynx, the size, form and typical tone of the vocal folds etc. These parameters are inherently manifested in the speaker-specific speech timbre (both voice timbre and phonetic timbre), in the mean and minimum voice pitch, in the mean distance between the formants, in the mean position of the first formant etc.

Speaker-specific breathing (in particular, speech breathing) pattern is a separate individual characteristic. Speech data allows the assessment of frequency, regularity and depth of the breathing, degree of its standardness, typical breathing efforts of the speaker, sufficiency of breathing for speech and other activities. Other individual characteristics of the speaker are related to the specific character of his nervous system manifesting itself in the specific temporary and dynamic organisation of utterances.

Relying on the expert’s experience, it is considered possible to determine, with a high degree of probability, the speaker’s age, basic parameters and type of bodily constitution, neuro-dynamic and sensory-motor organisation (nervous system typology), manifested in speech specific structure of his speech and breathing system.

1.1.2.1.1.1 Bodily constitution

Constitution is the average basic dimensions of the human body (assessment by speech of height and weight – clothes size of the speaker). In general, these characteristics are assessed by his voice pitch properties.

1.1.2.1.1.2 Neuro-dynamic characteristics

Properties of the speaker’s nervous system reflecting its strength, volatility, steadiness (balance), speed of response. In speech these are manifested in the speed of speech stimuli change, their stability, accuracy of realisation, dynamic span, stability in maintaining once chosen tempo and rhythm, breathing intensity, articulation accuracy, diversity of the used rhythmic patterns etc.

1.1.2.1.1.2.1 Temperament

Temperament is a generalized property reflecting the strength, volatility, stability and speed of the nervous system response. A relatively stable individual feature having as a physiological foundation a certain type of the speaker’s nervous system. Usually 4 basic temperament types are distinguished: melancholic, sanguine, phlegmatic and choleric.

1.1.2.1.1.2.2 Emotionality

Emotionality is emotional responsiveness (ability to express feelings, attitudes) of a person depending on the type of his nervous system, upbringing, manners and some other factors.
Individuals are distinguished by their emotionality both from the point of view of ability to feel certain emotions of certain strength, and from the point of view of emotional lability, i.e. tendency of different emotions to alternate at a certain rate.

Emotions are means of conveying information about the state of the individual and have specific manifestations: motor and sound responses, mimic, changes in breathing, the activity of ductless glands, heart, lumen of the blood vessels etc. Emotions are manifested in speech both at the global level of discourse (change of speech tempo and rhythm, timbre, loudness etc.) and in the choice of discourse means, ways of discourse organisation and textual contents of utterances.

1.1.2.1.1.2.3 Energy tone

Energy tone is the level of psychological and functional (activity-related) speaker activity which is often connected with the degree of speaker domination in the conversation. It is usually assessed along the low – high scale and has both constant (status (energy)) components and those related to the speaker current state (degree of fatigue, vivacity).

1.1.2.1.1.3 Sensory-motor coordination

Sensory-motor coordination is the ability to assess the position and state of one's body, muscles and the ability to fully control the muscles for performing the desired movement.

1.1.2.1.1.4 Defects of the speech and respiratory system

They are determined by the state of the articulatory and breathing organs of the speaker. Among the defects of the articulatory system the first to be mentioned are occlusion (bite) peculiarities, lack of teeth, presence of extraneous objects in the mouth, oedemata (swellings), areas of insensibility after anaesthesia, paresis, defects of the lips, hard and soft palate, inadequacy in the activity of the velum when opening/closing the passage into the nasal cavity, huskiness, hoarseness, unmusicalness of the voice, failures in vocal folds control (shrillness, change to whisper etc.). Among the defects of the respiratory system the first to be mentioned are insufficient depth and too small volume of in-breathing, noisiness, hoarseness, irregularity of breathing, inadequacy in breathing pauses placement, hollowness of the speech due to high tonus of the bronchi.

1.1.2.1.1.5 Intellect

Intellect is a complex of cognitive abilities of a person determining the level (of development) of his thinking and ability to solve complex problems. The development of the intellect is assessed by the profoundness of knowledge and the ability of the person not only to store it in his memory, but also to make effective use of it. The intelligence quotient is a relatively constant characteristic of a personality which manifests itself in his speech – usually in the ability to adequately select discourse means for the realization of a set goal and in logical, consistent, uncontroversial, and correlating with circumstances contents of speech.

1.1.2.1.1.6 Character, Inclinations, Abilities

Character is an integrated and stable individual make-up of a person's mental life, its type, the sum of a person's basic mental traits, reflected in his behavior, manners, habits, way of thinking and emotional circle.
Character is the result of a fusion of a person's temperament and his education, breeding, direction of his will and mind.

Inclinations are selective direction of a person at a certain kind of activity in a certain field, which finds reflection in his speech.

Abilities are person's individual mental traits, necessary for mastering a certain kind of activity and its successful realization.

All these peculiarities of a person are relatively fixed and find their reflection in his speech.

1.1.2.1.7 Ability for empathy

Ability for empathy is the ability to respond emotionally to the feelings of other people and to understand other people's motives and feelings as they are expressed, and the ability to compassion. It is manifested in speech in the adequacy of reactions to the emotions and state of the partner.

1.1.2.1.2 Relatively fixed socio-behavioral characteristics of the speaker

Characteristics of a person that are in the course of communication and include: motivation, social functions-roles, extraversion/introversion, suggestibility, conformity, values, living ideals, status, knowledge, skills, goals and direction, speaker's volitional qualities, curiosity and self-esteem.

1.1.2.1.2.1 Motivation

Motivation is psychological stimuli attaching goals and directions to the people's actions. Motivation can be organic, material, social, spiritual.

Motivation can be constant with regard to a specific communicative situation or with regard to a certain stage in person's life.

1.1.2.1.2.2 Social functions-roles

Social roles are a set of demands imposed by the society on an individual and acknowledged by this individual as obligatory for his behavior in the society. This is a complex of actions to be performed by a person occupying a certain position in a social system. A person can have many roles depending on the communicative situation type. There are institutionalizing (institute of marriage, family, social roles of mother, daughter, wife, etc.) and conventional (accepted by convention) demands, although a person can refuse to accept them.

A social role consists of a role expectation and a role implementation. Parsons proposes 5 characteristics of any role:

1. Emotional: changes along the restraint-relaxedness scale.
2. Way of acquisition: prescribed and gained.
4. Formalization: activity within strictly fixed limits or arbitrary.
5. Motivation: personal profit, social benefit, etc.

1.1.2.1.2.3 Extraversion/introversion
Extraversion denotes person’s individual psychological peculiarities whereby the person’s interests are directed outwards at the expense of his own interests.

Introversion denotes person’s individual psychological peculiarities whereby the person’s attention is fixed on his own interests, on his inner world. The personality of an introvert is directed to the events in his own subjective world.

This property is usually very stable.

1.1.2.1.2.4 Suggestibility and conformity

The degree of personality stability against the impact/influence upon his will and conscience. Suggestibility is usually understood as the ability to change one’s opinion under directional influence of other people – persuasion, argumentation, compulsion. Conformity is the ability of a person to adopt (by himself, without any direct influence from the outside) norms, habits and values and to change his original judgments under the influence of other people’s opinion (neighbors, colleagues, bosses, his nation, crowd etc.). Conformity can be both external and internal. Internal conformity is an actual transformation of individual policies of a person as a result of his acceptance of the environment’s position. External conformity is demonstrative submission of a person to a group in order to avoid the group’s pressure, whereby internally the person continues to resist the common point of view.

1.1.2.1.2.5 Values/principles/living ideals

Values are reality phenomena socially and culturally significant to a person.

Principles are internal beliefs of a person determining his attitude to the reality, norms of behavior and activity, reflected in his speech.

Ideals are conceptions of superior perfection in any respect; the object of aspiration regardless of whether the opportunity of its implementation is close or distant.

1.1.2.1.2.6 Status

Correlative position of an individual or a group determined by social, natural characteristics, as well as by the prestige and place in the control structure of the given social sphere. They are family, social, work and other statuses.

1.1.2.1.2.7 Knowledge, skills, attainments

Knowledge is the result of reality cognition reflected in the consciousness of a person in the form of ideas, conceptions, judgments and theories.

Skills are abilities to do something, acquired through knowledge, experience.

Attainments are the automatism ability to solve certain kinds of tasks.

1.1.2.1.2.8 Goals and orientation. Interests, volitions, beliefs, world outlook.

Goals are an ideal, mental anticipation of the activity result and ways of achieving it using certain means.

John Holland distinguishes 6 types of personality’s social orientation:

1. Realistic
A person is unsocial, present-day-reality-oriented, and emotionally stable, has good skills in dealing with material objects (things, tools, machines), and prefers occupations requiring motor and practical skills. He/she has good mathematical and non-verbal abilities.

2. Investigative (intellectual)
A person is unsocial, rational, independent, and non-typical, has analytical mind. Values are predominantly theoretical. A person is good at solving problems requiring abstract thinking. An intellectual with equally developed verbal and non-verbal abilities.

3. Social
A person has social attainments and seeks contact with people. His/her personality traits include an inclination to instruct and teach; direction at another person, humanity. A representative of this personality type tries to avoid intellectual problems; he is active, but often dependent on the opinion of a group of people. He solves problems; he is guided by emotions, feelings, social skills, has good verbal skills.

4. Conventional
A person prefers rigorously structured activities. He/she is stereotypical, straightforward, practical, he does not show criticism, originality; he is conservative, dependent, does not like change of activity. A person has weak organisational skills; predominantly math skills.

5. Enterprising
A person selects goals that require energy, enthusiasm, impulsivity, likes adventures. A representative of this personality type is dominant, tries to gain public recognition, and likes to lead people. He dislikes practical activities, as well as occupations requiring intellectual effort and diligence. He is good at solving problems connected with leadership, social status and power. He is aggressive and enterprising, has good verbal abilities (sociable).

6. Artistic
A person relies on emotions, imagination, and intuition; has complex life views. He/she is independent in decision making, original, creative, has highly developed motor and verbal abilities. Representatives of this personality type typically have a high life ideal with his “ego” assertion. He is unsocial in the sense that he does not adhere to the conventionalities accepted in the society.

1.1.2.1.2.9 Volitional qualities of the speaker
Speaker qualities related to his will, i.e. ability to achieve set goals and to realize aspirations. Volitional qualities include goal striving, persistence, self-possession, determination, initiative, independence, orderliness, discipline, courage, dutifulness.

1.1.2.1.2.10 Curiosity
Efforts towards gaining new knowledge, towards investigating the environment; inquisitiveness, expression of interest extending beyond the limits of everyday problems.

1.1.2.1.2.11 Speaker’s self-esteem (self-appraisal)
Person’s assessment of himself, his abilities, qualities and place among other people, which manifests itself in the speaker’s speech, in particular, in words describing himself, in speech (grammatical person used, degree of assertiveness of his utterances etc.). It can be differentiated both by quality and by expression degree.

1.1.2.1.2.11.1 Quality of self-esteem: high, low.

1.1.2.1.2.11.2 Degree of self-esteem expression (manifestation): manifested strongly, clearly; avoidance of self-esteem expression.

1.1.2.2 Change of auditory assessment of the speaker’s voice quality (timbre)
Subjective impression of speaker change (for two compared fragments of the audio recording) without clear understanding (by the expert) which of the perceptually assessed individual characteristics of the speaker or of his voice form the basis of this impression.

1.1.2.3 Change of the basic speech properties
The change of such identification characteristics as: level of language command, expression of accent/dialect, individual peculiarities of speech and breathing.

1.1.2.3.1 Level of language command, markedness of accent/dialect
Level of language command is a relatively fixed speaker characteristic, although in different situations language command can manifest itself to a greater or lesser degree.

Accent is peculiar pronunciation characteristic of a speaker knowing to a smaller or greater extent pronunciation standards of different languages and consisting in involuntary use of speech stereotypes of one language when speaking a different language.

Dialect is a variety of a specific language used as means of communication among people belonging to the same geographical, social or professional group. There are social and territorial (regional) dialects.

1.1.2.3.2 Change of auditory assessment of the speaker’s speech defects
Assessment of constant peculiarities of pronunciation anomalies (lisping, stutter, dyslalia, alalia, etc.).

1.1.2.3.3 Change of individual speech characteristics of the speaker
Change of the general typical articulatory pattern of utterances, general style of oral speech pronunciation, basic set of varying speech elements, strategy and tactics in the use of speech means.

1.2 Change of the speaker’s state
Speaker state is understood as a clearly expressed set of his individual qualities and properties changeable at the expense of external and internal circumstances. States can be fast changing and relatively stable.

1.2.1 Change of the speaker’s physiological state
Physiological state of a speaker is a set of temporary, less fixed than stable, characteristics related to his physiology. Physiological state of a speaker includes presence or absence of diseases and state of vivacity and lucidity, degree of fatigue, energy tone.

1.2.1.1 Change of the speaker’s disease manifestations
The state of disease or health is relatively stable. Thus, an abrupt change of disease manifestations is unnatural and can be indicative of audio recording inauthenticity. Diseases are subdivided into general-
system, mental and diseases of the speech and respiratory system. A disease is an anomaly in the organism's normal living activity, conditioned by functional and/or morphological changes. Disease appearance is connected with the influence of harmful environmental factors upon the organism.

1.2.1.1 Speech and respiratory system diseases
Manifestations of a cold, cough, nasopharynx edema, loss of voice, breathing deficiency, etc.

1.2.1.2 Mental diseases
A mental disease is a disease or a disease state related to mental anomaly or disorder. They manifest themselves in the inadequate use of speech means for utterance formation or in the inadequate (for the given situation) choice of communicative types of utterances.

1.2.1.3 General-system diseases
Diseases of the organism in general. Diseases with the general lowering of the speaker tonus related to breathing impediments, muscle control disorders, severe pains, paralysis, edema of organs, impact upon the speaker’s hearing and mind can be relevant for manifestation in speech.

1.2.1.2 Change of the state of the speaker's consciousness
Consciousness is the topmost form of reality reflection. It is a set of mental processes allowing a person to find his way in the surrounding world, time, and his personality. In certain kinds of disease clarity and adequacy of consciousness can change due to alcoholic, psychotropic, drug-induced intoxication (poisoning), due to certain mental and somatic diseases, fatigue, change of awake degree, strong emotions.

1.2.1.2.1 Alcoholic intoxication
Intoxication is an impairment of physical functions of the organism, first of all of his central nervous system, caused by taking alcohol or other similar narcotic drugs. It manifests itself in the groundless joyful, elevated mood (euphoria), arousal (in grave cases turning into depression), inability to realistically assess the situation and the role of his own personality, disorders of memory, attention, speech and coordination of movements.

Alcohol contained in alcoholic drinks – the most frequent reason for intoxication (drunkenness) – is rapidly absorbed by the blood (20% from the stomach and 80% from the bowels) and selectively accumulates in the brain where its concentration is by 75% higher than in blood.

There are distinguished 3 degrees of alcoholic intoxication:

– Weak intoxication can be observed after taking approximately 30-40 g of pure alcohol (pure spirits); it is characterized by moderate arousal, warm sensation in the body and worsening of work quality (both physical and mental).

– Average intoxication is observed after taking 50-100 g of pure alcohol and is characterized by clearly expressed and prolonged arousal state, during which the intoxicated person is persistently anxious to communicate with others, demands attention etc.; law offence, traumas, accidents are frequent. Arousal gives place to depression (dullness, sleepiness).

– Strong intoxication is observed after taking 100-300 g of pure alcohol and is characterized by a short (30 minutes – 1 hour) but violent arousal passing into deep sleep.
The influence of alcohol on speech is manifested first of all in the decrease or increase of speech tempo, in voice pitch change, in the growing number of misarticulations, decreasing speech accuracy, simplification of the used speech structures, in the superficial expression of the speech contents, in the appearance of hallucinations and in the situational inadequacy of the speaker.

1.2.1.2.2 Narcotic intoxication

*Intoxication* is an impairment of physical functions of the organism, first of all of his central nervous system, caused by taking narcotic drugs. It manifests itself in the groundless joyful, elevated mood (euphoria), arousal (in grave cases turning into depression), and inability to realistically assess the situation and the role of his own personality, disorders of memory, attention, speech and coordination of movements.

1.2.1.2.3 Affect state

Affect is a special kind of emotions characterized by great strength, rapid development, explicit vegetative symptoms (e.g. fear, rage). When in the state of affect, a person acts impulsively; he is not guided by logic and existing limitations. Affect is usually expressed as very expressive active emotion in speech.

1.2.1.2.4 State after taking sedatives

*Sedatives* are tranquillizing medicines. Their effect in speech is manifested in the lowering of tempo, speech dynamics, in passive behavior of the speaker.

1.2.2 Change of the speaker's psychological state

Psychological state is energy and emotional state of a speaker, expressed feelings, state of attention, role policy, attitude towards the communication act and communication partner, direction of communicative actions.

Under the influence of situational factors the psychological state can change during the communicative act.

1.2.2.1 Emotional state of the speaker. Sign (polarity), manifestation degree and type of emotion.

*Emotions* are subjective evaluative responses of a person to the influence of internal and external stimuli (irritants) manifested as content or discontent, increase/decrease of anxiety degree (positive/negative prognosis for situation development), subjective attitude to what is happening (joy, fear, surprise, anger, apathy, interest, etc.).

Each emotion’s manifestation in speech depends on its energy and evaluation factors, as well as on the speaker’s behavioral stereotypes and upbringing.

Emotions are determined by physiological processes taking place in the brain and in the body as a whole, as a result of integrated cortical-subcortical arousal centres formed on the basis of inherited and acquired experience. Brain structures of various levels, including neocortex, limbic system (hippocampus, tonsil, hypothalamus) take place in the realization of emotions. The subjective sensation appears as a result of this system arousal. Any brain injuries changing the arousal integration lead to the anomalies in the emotional response, first of all in the ability to adequately evaluate the results of performed actions. Emotions, as one of the means of conveying information about the state of the individual, have typical outward expression: motor and sound reactions, mimicry, change of breathing, activity of the stomach, bowels, urinary bladder,
ductless glands, heart, lumen of blood vessels, etc.). Some reactions (movements, breathing, etc.) can be voluntarily controlled by the person, while others (blood pressure, heart beat) normally cannot be voluntarily controlled. In situations of conflict emotions can turn into emotional stresses, with emotional arousals spreading continuously from the brain centres via the vegetative nervous system and ductless glands to the peripheral processes (especially involuntary). This can lead to the change of functions of various internal organs and to the development of neuroses, ischemic heart disease, hypertensive disease, stomach ulcer, eczema etc. Although the emotional activity of a person is to a certain extent dependent on his individual (typological) properties, the critical importance belongs to the right upbringing, especially in the early childhood.

1.2.2.2 Expressed feelings

*Feelings* are a special kind of emotional sensations characterized by clearly expressed object orientation and relative stability.

There are moral, intellectual, aesthetic, praxical (activity-oriented) feelings.

1.2.2.3 Change of auditory assessment of the speaker's energy tone

Assessment of the level of behavioral and discourse activity of the speaker, degree of domination and level of general response to external circumstances.

1.2.2.4 State of attention

Activity, direction, scope, switchability, intensity, stability

1.2.2.5 Speaker's role policies

Role policy is the speaker's role with respect to his communication partner in the given situation. Role policy can be towards domination, equality or subordination.

1.2.2.5.1 Domination

Role policy towards subordination, suppression of a communication partner.

1.2.2.5.2 Equality

Role policy towards communication with an equal (by status).

1.2.2.5.3 Subordination

Role policy towards subordination to the communication partner.

1.2.2.6 Attitude to the communication act

Speaker's attitude towards the communication act is expressed in the degree of interest in the conversation, carefulness/carelessness of the speaker's communicative actions, in haste and laconism or, conversely, slowness and verbosity.

1.2.2.6.1 Degree of interest in the conversation

Degree of interest, involvement in the conversation can be changed along with the change of conversation topic.

1.2.2.6.2 Carefulness/carelessness of the speaker's communicative actions
Communicative action is a symbolically mediate action controlled by the norms and directed at achieving interaction. Carefulness of communicative actions can be manifested in increased reductions and substitutions of articulatory close sounds, inaccuracy in word choice, unfinished utterances, inconsistency of utterance contents, etc.

1.2.2.6.3 Haste-laconism/unhurriedness-verbosity

Haste and slowness are determined by the communicative situation and the attitude of communication participants towards the communication act.

The speaker can be in a hurry along the entire dialogue, or in some part of it – when the discussion concerns something irrelevant for the interlocutors or when the speaker is in a hurry to give some information to his interlocutor, as well as when there are some external limitations on the conversation duration. Haste often goes with laconism, unhurriedness – with verbosity, although other combinations are also quite possible.

1.2.2.7 Attitude towards a communication partner

The attitude towards a communication partner can change in the course of the conversation, but it never happens without reason. However, the reason for attitude change can be, for example, words having for a conversation participant a different meaning than for the expert.

1.2.2.7.1 Friendliness

1.2.2.7.2 Neutrality

1.2.2.7.3 Hostility

1.2.2.8 Extent of the speaker's familiarity with his interlocutor

In the course of the conversation it is often possible to figure out the degree of familiarity of speakers with each other, the peculiarities of their relations, the degree of confidentiality, common knowledge background and attitude to some events and phenomena in life or in the environment. Such peculiarities should usually retain uniformity in the course of the entire conversation.

1.2.2.8.1 Close people

1.2.2.8.2 Good acquaintances

1.2.2.8.3 Acquaintances

1.2.2.8.4 Superficial acquaintances

1.2.2.8.5 Strangers

1.2.2.9 Direction of communicative actions in the conversation

4 basic directions in communicative actions can be distinguished: instructive, narrative, disputing and expressive.

The direction of communicative actions is related to the communication goal. If the goal is to impel someone to certain actions, the direction is instructive, persuasive. If the goal is to convey certain information, to tell the interlocutor about something, the direction is narrative, declarative. If the goal consists in a discussion, in
reaching an agreement etc., the direction is disputing, argumentative. If the goal is the expression of one’s emotions, feelings, attitude, the direction is expressive.

The dialogue is often comprised of communicative actions of various directions, but each direction change is situational determined. Situational undetermined change in the direction of communicative actions in the conversation can be an indicator of audio recording authenticity breach.

1.2.2.9.1 Instructive, persuasive
The direction of communicative actions is instructive, or persuasive, in cases when the speaker’s goal is to impel (persuade) his interlocutor to certain actions.

1.2.2.9.2 Narrative, declarative
The direction of communicative actions is narrative, declarative, in cases when the speaker’s goal is to convey certain information, to tell his interlocutor about something.

1.2.2.9.3 Disputing, argumentative (everyday/professional conversation)
The direction of communicative actions is disputing, argumentative in cases when the speaker’s goal is to discuss a certain disputed question, to investigate a problem with each party opposing the interlocutor’s opinion, provides arguments in favor of its point and aspires to reaching its goal. Argument is a discussion phase characterized by the irreconcilability of the parties’ opinions, predominance of the desire to win at any cost, including manipulations with words and concepts, confusing the interlocutor. An argument is often characterized by a transition to the emotional level of discussion at the expense of logic.

1.2.2.9.4 Expressive
The direction of communicative actions is expressive in cases when the speaker's goal is to express his emotions, feelings, attitudes.

1.2.3 Change of the speaker’s mental state
It is manifested in the speaker’s knowledge and opinion about the subject of conversation and its goals.

1.2.3.1 Familiarity with the discussed subject
Familiarity with (knowledge about) the discussed subject is a certain amount of information the speaker possesses at the moment of communicative act and which should remain unchanged, except for the changes and additions to the background introduced in the course of the communicative act.

There are three levels of knowledge of the speaker: expert, advanced in the subject study, beginner.

1.2.3.2 Change of the goals of communication act
The interlocutors start the dialogue with a certain goal (e.g., to reach an agreement about something, to discuss a certain question, to get certain information etc.). Usually in the course of the communicative act (before the set goal is reached) both the goal and the attitude to it are retained the same. Unjustified change in the goal of the communication act or in the attitude to it can indicate an audio recording authenticity breach.

1.2.3.2.1 Presence of a goal
The interlocutors start a dialogue with a certain goal (e.g., to reach an agreement about something, to discuss a certain question, to get certain information etc.) which is retained without unjustified change until full completion of the set task or until the rejection of the set goal upon realizing the impossibility of task fulfillment in the given communicative situation.

1.2.3.2.2 Attitude to the goal
Usually in the course of the communicative act both the goal and the attitude to it (i.e. its relevance to the speaker) remain unchanged. Abrupt unjustified change of speaker attitude to the communication goal can indicate an audio recording authenticity breach.

1.2.3.3 Change of the speaker’s opinion about the subject
The speaker’s opinion about the subject can change in the course of the conversation, but this change should be situational justified. Thus, if the conversation is about the weather and the speaker expresses his negative attitude to it, his opinion can change only if either the weather changes or if his interlocutor gives his arguments to the contrary, etc. Situational unjustified opinion change can be indicative of audio recording inauthenticity.

1.3 Change of peculiarities of discourse behavior
Discourse behavior of a speaker includes his discourse strategy, manner of communication, psychological style of communication, the person in which the speaker speaks (i.e. who he speaks for), extra- and paralinguistic discourse components.

1.3.1 Change of the speaker’s discourse strategy and tactics
Speaker’s possession of latent manipulation methods and techniques, i.e. leading the communication partner to the desired goal, topic, reactions. The person can either possess or not possess such methods and techniques, or possess to a certain extent – which should remain unchanged within a single act of communication.

1.3.2 Change of the speaker’s manner of communication
*Manner of communication* is the usual (for this person) habitual way of actions during a communicative act under the given circumstances. Tone, communication distance, verbal play, humour, used idiolect, argot, selectiveness in information perception, etc belong to communication manner.

1.3.2.1 Tone (instructive, ingratiating, etc.)
Tone belongs to the elements of the speaker’s manner of communication. The tone can be instructive, ingratiating, etc. It is usually retained in the entire course of the communication act, although the tone itself, as well as its manifestation degree can change.

1.3.2.2 Distance in communication
A striking feature of communicative behavior manifesting itself in such peculiarities of interaction as the way of handshake and touching, distance between the interlocutors, their allocation.

1.3.2.3 Verbal play, humour
Playful attitude expressed by the speaker to the discourse subject, use of words and expressions in their figurative meanings.

1.3.2.4 The used idiolect/argot

*Idiolect* is an individual language, language skills of a person (speaker) at a certain period of time. Idiolects are associated with various speech communities: professional, social and territorial (geographical) dialects and argots. Idiolect is a conventional term, as the same person, as a rule, uses different language means in different communicative situations.

*Argot* is the language of socially or professionally isolated groups and communities, consisting of arbitrarily selected modified elements of one or several natural languages. Argot does not form its own language system and is reduced to specific word usage within a common language. Argot interacts with jargon and vernacular, forming a specific lexical stratum called slang. Argot is used, as a rule, with the purpose of concealing the subject of communication, as well as means of conscious or unconscious isolation of a group from the rest of the society.

1.3.2.5 Selectiveness in information perception

Information can be perceived by the speaker with a varying degree of selectiveness which usually remains unchanged for the entire communication. Situational unjustified change of selectiveness degree in information perception can be indicative of audio recording inauthenticity.

Change can be considered situational unjustified if it is associated with a change of the conversation topic, attention attraction, emphasizing the interlocutor’s attention, etc.

1.3.2.6 Other

Other peculiarities of manner of communication are also possible.

1.3.3 Change of the (grammatical) person used by the speaker

The speaker can speak for himself (“I want…”, “I need…” etc.), for a group of people (“We want…”, “We are interested in…” etc.), for another person or persons (“He wants…”, “He needs…” etc.), impersonally, i.e. not using any particular grammatical person.

This characteristic of the speaker’s speech is relatively stable and its change can be indicative of an audio recording authenticity breach.

There are the following types of the (grammatical) person used by the speaker:

- Speaks in the first person singular (“for himself”).
  For instance, “I want…”, “I need…” etc.
- Speaks in the first person plural (“for a group”)
  For instance, “We want…”, “We are interested in…” etc.
- Speaks for another person or persons
  For instance, “He wants…”, “He needs…” etc.
- Uses impersonal utterances
  The speaker does not use any particular grammatical person.
1.3.4 Degree of preparedness for communication

The degree of preparedness for communication is retained within one topic. Unjustified change of preparedness degree can indicate an audio recording authenticity breach.

The following degrees of preparedness for communication are distinguished:

– The communicating parties are unprepared
– Prepared by force of custom, habit
– Specifically prepared for this conversation

1.3.5 Change of the psychological style of communication

Individually typical peculiarities of interaction, first of all those characterized by the speaker’s efforts in the communication act. The following psychological styles of communication are distinguished:

– Efforts towards getting information
– Efforts towards giving information
– Efforts towards expressing personal attitude
– Efforts towards getting along with the interlocutor
– Efforts towards conflicting with the interlocutor
– Efforts towards getting appreciation from the interlocutor, producing an impression

1.3.6 Breach of integrity and continuity of the extra-linguistic discourse component

The extra-linguistic discourse component includes gestures, mimicry, proxemics. It is necessary to remember that the extra-linguistic component can change the meaning or even contradict verbally expressed information.

1.3.6.1 Gestures

Physical movements, pose or facial expression carrying, according to the cultural traditions existing in the society, a certain meaning.

1.3.6.2 Mimicry

Expressive movements of face muscles representing one of the forms of manifestation of certain feelings of a person. It can be natural, involuntary and artificial, consciously simulated.

1.3.6.3 Proxemics

Personal space requirements.

According to E.T. Hall, there exist four clearly distinguished proxemic zones within which a person acts:

– Intimate zone (15-45 cm);
– Personal zone (50-130 cm);
– Social zone (130-350 cm);
– Public zone (350 cm and more).

1.3.7 Breach of integrity and continuity of the paralinguistic discourse component
The *paralinguistic discourse component* includes means accompanying speech, but not related to language: laughter, cry, cough, emotional voice change, breathing, etc.

1.3.7.1 Laughter, cry, cough, etc.

Laughter, cry, cough are processes having the beginning, development and completion stages. Breach of integrity and continuity of these processes can indicate an audio recording authenticity breach.

1.3.7.2 Emotional voice change

Emotional voice change must be situational justified. Situational unjustified emotional change can indicate an audio recording authenticity breach.

1.3.7.3 Breathing

Two types of breathing are distinguished: speech and non-speech. Speech breathing coincides with speech and is related to the process of speech generation. Non-speech breathing is not related to speech, i.e. it is observed when the person is not speaking. Anomalies in the general breathing pattern or in the speech breathing structure can be indicative of an audio recording authenticity breach.

1.3.7.3.1 The overall character of breathing

The overall character of breathing must be situationally determined. Thus, a change in the person’s breathing is natural if it is caused by physical activity or an emotional state. Situational unjustified change of breathing or inconsistency between the overall character of breathing and the current situation can indicate an audio recording authenticity breach.

1.3.7.3.2 Speech breathing pattern

Unjustified breach of the speech breathing pattern (i.e. placement and duration of breathing pauses in speech) can be indicative of an audio recording authenticity breach.

2 Change of the communication environment

The communication environment includes everything related to the discourse apart from the speech proper, i.e. circumstances, conditions under which the communicative act is taking place. The environment can remain practically unchanged, but more often they do change to a certain extent. However, these changes should be characterized by integrity and continuity, as well as they should be situational justified.

2.1 Discontinuity of the place of communication

The place of communication is where the speakers are located during the communication act. Usually the *place of communication* can be determined by the sounds accompanying the speech and by some speech properties (reverberation etc.).

The speakers’ locations can change or remain the same in the course of the entire communication act. The speakers’ locations determined on the basis of an authentic recording must correspond to the case data and either remain unchanged or change without integrity and continuity breaches. Integrity and continuity of the process of speaker location change consists in its consistency and explicit reflection in the audio recording.
2.1.1 Integrity and continuity of the process of speaker localization change

Integrity and continuity of the process of speaker localization change is the consistency and explicit reflection in the audio recording of any changes of the speakers' locations.

2.1.2 Constancy of speaker localization

If, according to the case data, speaker locations remain unchanged in the course of the entire communication act, any changes of speaker locations reflected in the audio recording can be considered as indicators of an audio recording inauthenticity. By the sounds accompanying the speech and by the speech character itself it is possible to establish where the speaker was during the communication act:

– In an open space;
– In an enclosed space;
– In a large-sized room;
– In a medium-sized room;
– In a small-sized room;
– In a vehicle;
– In a specialized room.

Inconsistencies in space characteristics can also be indicative of tampering (e.g. if during the entire communication act the speaker is in a vehicle, but the type of vehicle changes in the course of the conversation and this change is not situational justified).

2.1.2.1 An open space

Situational unjustified change of open space characteristics.

2.1.2.2 An enclosed space

Situational unjustified change of enclosed space characteristics.

2.1.2.3 A large-sized room

Situational unjustified change of large-sized room characteristics.

2.1.2.4 A medium-sized room

Situational unjustified change of medium-sized room characteristics.

2.1.2.5 A small-sized room

Situational unjustified change of small-sized room characteristics.

2.1.2.6 A vehicle

Situational unjustified change of vehicle characteristics.

2.1.2.7 A specialized room

Situational unjustified change of specialized room characteristics.

2.2 Breach of integrity and continuity of the acoustic environment

*Acoustic environment* includes sounds accompanying the speech signal: background noise and acoustic background sounds.
Background noise is the noise of the acoustic environment, the noise of the microphone and the signal transmission channel, as well as tape and recorder noise.

Acoustic background includes acoustic events and periodic sound sources recorded on the carrier along with the speech signal proper.

2.2.1 Background noise

*Background noise* is the noise of the acoustic environment, the noise of the microphone and the signal transmission channel, as well as tape and recorder noise.

There are the following types of noise:
- Noises in the acoustic environment
- Microphone and signal transmission channel noise
- Tape and recorder noise

2.2.2 Acoustic background

*Acoustic background* includes acoustic events and periodic sound sources recorded on the carrier along with the speech signal proper.

2.2.2.1 Integrity and continuity of acoustic events

An *acoustic event* is the sound representation of events taking place simultaneously with the speech signal and recorded along with the speech signal onto a carrier (e.g. music, noise of a moving car, working TV etc.). Possible audio recording tampering can be signaled either by a situational unjustified acoustic event detected in the recorded data, or by breaches in the integrity and consistency of the recorded events.

2.2.2.2 Stability of periodic sound sources

Sounds accompanying recorded speech and produced with a certain periodicity as, for instance, clock ticking.

Possible audio recording tampering can be signaled either by a situational unjustified periodic sound signal detected in the recorded data, by rhythm anomalies or by the unjustified appearance or disappearance of sounds.

2.3 Change of the speaker's closeness to the microphone

Situational unjustified change of the speaker's closeness to the microphone can indicate an audio recording authenticity breach.

2.4 Change of the peculiarities of the recording environment

Peculiarities of the recording environment include direction at the speaker, microphone placement, and sound transmission channel. Situational unjustified change of these characteristics can indicate an audio recording authenticity breach.

2.5 Breach of integrity and continuity of the background chain of events

Events coinciding with the recorded speech can have their sound representations recorded along with the speech signal. The background chain of events can include, e.g., the sound of a moving car. Continuity in this
case means adequacy of this sound to the current situation, as well as the consistency and completeness of event development (e.g. the car sound is first gaining strength as it is approaching and then starts to lose its intensity as it is moving away).

3 Inconsistency between the communicative act and circumstances known to the speakers or experts
3.1 Inadequacy of the utterance contents
Inconsistency between the contents of the utterances and the circumstances known to the speakers or experts can indicate an audio recording authenticity breach.

3.2 Inadequacy of the speakers’ discourse behavior
Inconsistency between the speakers’ discourse behavior and the circumstances known to the speakers or experts can indicate an audio recording authenticity breach.

3.3 Inadequacy of the speakers’ behavior
Inconsistency between the speakers’ actions or behavior and the circumstances known to the speakers or experts can indicate an audio recording authenticity breach.

3.4 Inconsistency between the established circumstances or communication act environment and the circumstances known to the speakers or experts.

Pragmatic inadequacy. In this case detected inconsistencies between the utterance contents, discourse behavior or speakers’ actions on the one hand and the circumstances known to the speakers or experts on the other can indicate an audio recording authenticity breach.

4 Change of the recorded signal properties
A recorded signal includes the effective speech signal proper (analyzed by auditory and linguistic means), as well as the sounds related to the processes of sound transmission and recording and analyzed using auditory and instrumental techniques.

4.1 Change of the speech signal integrity and continuity pattern
The effective speech signal is the speech of the speakers, oral discourse, characterized by text-dependent and text-independent properties.

4.1.1 Text-dependent analysis
Analysis of the contents of oral discourse recorded on the tape.

4.1.1.1 Adequacy and continuity of the thematic and rhematic structure
Inadequacy of utterances with respect to information structure can indicate an audio recording authenticity breach.

4.1.1.2 Adequacy of speech act structural components
Structural components of a speech act are locution, illocution and perlocution.
The locution is the utterance proper characterized by its articulation, speech rate, accuracy, etc. without taking into account the speaker's intentions and the achieved effect. It is one of the components of the speech act along with illocution and perlocution.
The *illocution* is the speaker’s goal. It is one of the components of the speech act along with locution and perlocution. It is an action performed by means of speech – imperative (request, command), question, hesitation, declaration, promise. The illocutionary aspect includes the speaker’s intentions and various circumstances of the speech act.

The *perlocution* is the result of verbal influence on the hearer; the effect achieved as a result of the illocution. It is one of the components of the speech act along with illocution and locution.

*Intention* is the speaker’s intent to say something, to convey a certain subjective meaning in the utterance. Intention is included into the illocutionary aspect. If the intention as a directional act is not meant (by the speaker) for verbal expression, it is not a communicative intention and, consequently, is not an object of linguistic analysis. The expressed communicative intention should not necessarily coincide with the actual intention of the speaker, and the speaker does not always try to let the hearer understand his real intention. In cases of communicative failures or deliberate attempts to confuse the hearer, the real intention of the speaker is often different from his communicative intention expressed in the utterance and accepted by the hearer.

4.1.1.2.1 Inconsistency between the locution and the illocution

Inconsistency between the speaker’s goal and the spoken utterance can indicate an audio recording authenticity breach.

4.1.1.2.2 Illocutionary adequacy

Inconsistency between the speaker’s goal and the situation can indicate an audio recording authenticity breach. For instance, the speaker’s goal is to make the interlocutor switch on the light, whereas in reality (situation) the light is already on.

4.1.1.2.2.1 Intention adequacy and integrity within the discourse

Inadequacy of the speaker’s intentions and their unjustified change within the entire discourse can indicate an audio recording authenticity breach.

4.1.1.2.2.2 Intention adequacy and integrity within the utterance

Inadequacy of the speaker’s intentions and their unjustified change within an individual utterance can indicate an audio recording authenticity breach.

4.1.1.2.3 Perlocutionary adequacy

Inadequacy of the speaker’s reaction to the words of his interlocutor can indicate an audio recording authenticity breach.

4.1.1.2.3.1 Adequacy of reaction to the interlocutor’s illocution

Inadequacy of reaction to the interlocutor’s illocution can indicate an audio recording authenticity breach.

4.1.1.2.3.2 Adequacy of reaction to the interlocutor’s locution

Inadequacy of reaction to the interlocutor’s locution can indicate an audio recording authenticity breach.

4.1.1.3 Topic continuity
A situational unjustified change of the conversation topic can indicate an audio recording authenticity breach.

4.1.1.3.1 Macro-topic integrity and continuity

Macro-topic integrity and continuity is the preservation of the main topic/topics of the conversation, its/their natural development and change.

4.1.1.3.2 Micro-topic integrity and continuity

Micro-topic integrity and continuity is natural appearance, development and change of the conversation sub-topics.

4.1.1.4 Conversation topic fixity

The degree of conversation topic fixity is its relatively stable characteristic. A situational unjustified change of this characteristic can indicate an audio recording authenticity breach.

The conversation can be:
– with a free topic;
– with a fixed topic area;
– with a highly fixed specific topic.

4.1.1.4.1 Free topic

A situational unjustified appearance of a strictly fixed topic within a free-topic conversation can indicate an audio recording authenticity breach.

4.1.1.4.2 Fixed topic area

A situational unjustified digression from a fixed topic area can indicate an audio recording authenticity breach.

4.1.1.4.3 Highly fixed specific topic

A situational unjustified digression from a highly fixed specific topic can indicate an audio recording authenticity breach.

4.1.1.5 Logical and semantic integrity and continuity

Logical and semantic integrity and continuity is the naturalness and adequacy of the conversation in terms of its meaning and consistency. Any breaches of speech integrity and continuity in terms of its meaning and consistency can indicate an audio recording inauthenticity.

4.1.1.5.1 Logical and semantic integrity and continuity of the conversation in general

4.1.1.5.2 Logical and semantic integrity and continuity of the dialogical entity

Breaches of logical and semantic integrity and continuity of the dialogical entity can indicate an audio recording inauthenticity.
A *dialogical entity* is comprised of closely related utterances: stimulus and subsequent response forming one relatively complete unit.

A dialogical entity consists of two, less likely three or four utterances closely related both in structure and meaning. The form and content of the first utterance determines the form and content of the following utterance etc., so that only in combination the utterances exhibit the completeness of dialog necessary for its adequate understanding.

4.1.1.5.3 Logical and semantic integrity and continuity of certain utterances

Breaches of logical and semantic integrity and continuity of an utterance can indicate an audio recording inauthenticity.

An *utterance* is a unit of communication characterized by the meaning integrity. It is minimum output of discourse activity including mental, physiological, intellectual and linguistic ability of the speaker. An utterance can coincide with a sentence, word, remark in a dialogue etc.

4.1.2 Text-independent analysis

*Text-independent analysis* is the analysis of oral discourse implementation means.

4.1.2.1 Non-standardness of grammar

Non-conformity of oral discourse grammar to the accepted oral speech standards can be indicative of an audio recording authenticity breach. The standards in this case are understood as accepted requirements for oral discourse realization in the given communicative situation.

*Grammar* is part of linguistics studying regularities in the formation and usage of word forms.

4.1.2.1.1 Word order

Anomalies in the standard word order for conversational speech can indicate an audio recording authenticity breach.

*Word order* is a certain pattern of word ordering in a sentence or a syntactic group.

4.1.2.1.2 Agreement

Anomalous (against the conversational speech standard) agreement can indicate an audio recording authenticity breach.

Agreement is a subordinating relation between clause components making the dependent word repeat all or part of the principal word grammemes. If the principal word changes its from, so does the dependent word.

4.1.2.1.3 Government

Anomalous (against the conversational speech standard) government can indicate an audio recording authenticity breach.

*Government* is a subordinating relation whereby the principal component of a clause requires that the dependent component take a certain grammatical form. The principal word modification does not entail the modification of the dependent word.

4.1.2.1.4 Negation
Anomalous (against the conversational speech standard) negation can indicate an audio recording authenticity breach.

*Negation* is an element of sentence meaning showing that the relation established between the sentence components, in the speaker’s opinion, does not really exist or that the respective statement (affirmative sentence) is rejected (negated) by the speaker as false. Most often a negative utterance is made in a situation when the respective affirmative sentence either was previously uttered or is part of the common presumption (background knowledge) of the speakers.

4.1.2.1.5 Conjugation

Anomalous (against the conversational speech standard) conjugation can indicate an audio recording authenticity breach.

Conjugation is the creation of derived forms of a verb from one basic form. It covers the entire verb paradigm and expresses respective grammatical categories.

4.1.2.1.6 Declension

Anomalous (against the conversational speech standard) declension can indicate an audio recording authenticity breach.

*Declension* is the inflection of nouns, pronouns and adjectives in categories such as case, number, and gender.

4.1.2.1.7 Plural formation

Anomalies (against the conversational speech standard) in plural formation can indicate an audio recording authenticity breach.

4.1.2.2 Lexical integrity and continuity

Consistency of the used vocabulary in different parts of the audio recording.

4.1.2.3 Stylistic integrity and continuity

*Stylistics* is a part of linguistics studying the system of styles within a language and describing language norms and ways of using the literary language under various conditions of language communication and in various spheres of social life.

The subject of stylistics is language in the broad sense of the word (including also speech as a form of language existence). Stylistics deals with ways of expressing additional (stylistic) information which accompanies the content of speech. Stylistics is characterized by its own principle of grouping language material which can be different from the division into vocabulary, phraseology, morphology, syntax.

Stylistics can deal with the stylistic properties of separate language levels, but the researcher always considers the stylistic system of a language as a whole.

4.1.2.3.1 The consistency between the used stylistically marked structures, idioms, expressions and the functional speech style in the given communicative situation

Stylistically marked structures are non-neutral conversational structures, marked stylistically and distinguished by their evaluative, emotional, aesthetic and other speech characteristics.
4.1.2.3.2 Change of speech style

4.1.2.3.3 Change of language code

*Language code* is a means of communication – language, dialect, jargon, a stylistic language variety, etc. Code-switching is the change from one code (language, dialect, style) to another code during speech communication due to the change of communication circumstances: change of the addressed party, change of the speaker’s role, change of the conversation topic, etc.

4.1.2.4 Phonetic integrity and continuity (observance of phonetic laws)

Phonetics studies the sound system of language, i.e. ways of formation and acoustic properties of sounds, as well as their modification in the speech flow. Phonetics distinguished several types of units. Segmental units include phrase, syntagm, phonetic word, syllable, sound. Supra-segmental (super-segmental) phonetic units exist in parallel with the segmental ones. These are stress and intonation.

Phonetic law is phonetic modification in the area of the consonants and vowels sounds: positional and combinatorial changes.

4.1.2.4.1 Combinatorial changes of consonants

These are changes under the influence of adjacent sounds.

*Assimilation* is the process by which a sound is modified so that it becomes phonetically similar to an adjacent or nearby sound in the flow of speech. Assimilation is full (complete) if the assimilating sounds become absolutely identical and often merge into one sound. In the case of partial assimilation only some of the sound properties undergo assimilation, while others remain different so that the sounds do not become identical.

*Accommodation* is the adaptation of some sounds to others. Usually either consonants accommodate to vowels or vice versa.

*Ellipsis* is partial or complete deletion of sounds.

*Epenthesis* is the insertion of a non-etymological sound (vowel or consonant) into a word (most often as a result of dissimilation, to make the pronunciation of a word easier).

*Haplology* is the loss of one of two identical or similar adjacent syllables in a word.

*Metathesis* is the transposition of sounds or syllables within a word.

*Dissimilation* is the process by which one of two similar or identical sounds (adjacent or nearby) in a word becomes different from or less like the other.

4.1.2.4.2 Positional change of consonants

It is a change of a consonant under the influence of its position, e.g. devoice in the word final position.

4.1.2.4.3 Combinatorial changes of vowels

These are changes in vowels under the influence of adjacent consonants.

4.1.2.4.4 Positional change of vowels

This is a change of a vowel under the influence of its position.
Reduction is the change of a sound's articulatory and perceptual properties caused by the shortening of its duration, decreased articulation accuracy (leveling, stronger co-articulation) or weakening of its articulatory tenseness (refers primarily to vowels in unstressed positions).

4.1.2.5 Prosodic integrity and continuity

Prosody is a system of suprasegmental (supersegmental) articulatory and perceptual categories of oral discourse relating to its pitch, intensity and temporal structure realized in speech and performing structuring, meaning formation and expressive-emotional functions (phrasing, patterning and highlighting of utterance elements). The following prosodic elements are distinguished: at the level of physical features – pitch, temporal properties (rhythm, tempo), pauses, loudness, prosodic timbre; at the functional level – accentedness/unaccentedness, intonation, pauses.

4.1.2.5.1 Loudness

Loudness is the perceptual impression of difference in the physical force of uttered sounds registered both subjectively and instrumentally.

– Normal loudness is accepted to be 50-80 dB (with constant background noise up to 10 dB).
– Moderate raising of the voice: 80-90 dB;
– Considerable raising of the voice: 90-110 dB;
– Shouting: over 110 dB;
– Moderate lowering of the voice: 80-90 dB;
– Considerable lowering of the voice: 90-110 dB;
– Whisper: below 20 dB.

Loudness can change in the course of a dialog either gradually or abruptly, but it is always determined by the situation. Thus, increase in loudness can be related to the increase of distance between the interlocutors or to the appearance of interference in the communication channel.

4.1.2.5.2 Tempo (speech rate)

Tempo is the speed of uttering speech elements: sounds, syllables, words. It is estimated either by the number of sounds, syllables etc. pronounces per time unit, or by their average duration.

4.1.2.5.3 Timbre

Timbre is subjective quality or coloring of a speech sound auditorily perceived as the result of and interrelations between various levels of spectral components. Supra-laryngeal timbre (sometimes referred to as sound timbre) is a sound quality depending on the position of various non-phonation articulatory organs and acoustic processes in them. Laryngeal timbre is sound quality determined by the activity of phonation organs.

4.1.2.5.4 Pitch

Pitch is an intonation component realized as raising and lowering the voice in the utterance.

4.1.2.5.4.1 Word pitch
A phonetic word is part of a syntagm united by one stress. In the speech flow relational (auxiliary) words, as a rule, merge with notional words forming a single phonetic word.

4.1.2.5.4.2 Syntagm pitch

A syntagm is a uniform (by its speech properties) meaning entity, usually forming a part of some larger speech unit and conveying one meaning in the given context and in the given situation. A syntagm can be comprised by one word, by a group of words or by the entire sentence. Division of text into syntagms is based on meaning, and the syntagmatic pattern of the text serves to convey meaning relations. A syntagm is marked by a syntagmatic accent which is intensified if it falls on the last word of the syntagm.

4.1.2.5.4.3 Utterances

An utterance is a unit of communication characterized by the meaning integrity. It is the minimum output of discourse activity including mental, physiological, intellectual and linguistic ability of the speaker.

4.1.2.5.4.4 Phonetic paragraph

A phonetic paragraph is a group of utterances united by meaning.

4.1.2.5.5 Pauses

4.1.2.5.5.1 Breathing pauses

4.1.2.5.5.2 Pauses between utterances

4.1.2.5.5.3 Pauses between syntagms

4.1.2.5.5.4 Pauses between phonetic words (rhythmic groups)

4.1.2.5.5.5 Situational conditioned pauses

4.1.2.5.5.6 Hesitation pauses

Hesitation is indecision, uncertainty, doubts reflected in speech by means of pauses (either filled or not).

4.1.2.5.6 Stress/accent

Highlighting of certain discourse units by the following means: intensified articulatory efforts, in particular, increased loudness, duration and accuracy of articulation, change of voice pitch and some other means. There are the following types of accent: logical, phrasal, syntagmatic and emphatic; the notion “stress” usually refers to the word stress.
Active tab
Tab of active (selected) data window, used as a data source. The tab is usually displayed over other tabs.

Amplitude (magnitude) (lat. *amplitudo* – size)
The maximum deviation value (from the equilibrium position) of an oscillating quantity, for example, the deviation from zero of an in-circuit electric current voltage, sound pressure intensity, etc. It represents the size of vibration (deviation value). In strictly periodic vibrations, the amplitude is a constant.

In the research of harmonic sound vibrations, the amplitude means sound pressure in a signal expressed by the amplitude of a current, voltage or other electrical quantity on the output of sound converting equipment (microphone). In the signal waveform figure, the amplitude represents the deviation size of an image up or down from zero position.

Audio/sound record (phonogram)
Speech signal, pre-recorded in the file.

Data
A graphical image in the data window, gathered while recording audio, reading files, operating with the program. A representation of oscillograms (waveforms), spectrograms, histograms and other graphical images.

(Data) box
Data box in the program is the limited by frame independent rectangular area within the central area of the main program window, which displays certain data (waveform, spectrogram and formants, histogram, etc.) as a graphic representation.

Data tab
Independent data that together with other data is stored in one data window while operating with program

Fragment
In SIS, the part of data which is singled out in some way from the segment, but has not lost its connection with the remaining data. It can be, for example, part of a segment limited by temporary marks or part of a segment included in the highlighted interval between permanent marks or part of a segment visible in the box.

Mark/Marker
A tool to highlight specific data areas in the data window.
Noise (interference)
Unwanted sound (different kinds of noise, background and other signals without non-useful information) that complicates the useful signal determination and use.

Pause (lat. *pausa*, gr. *pausis* – stop, termination)
A break in speech, which acoustically corresponds to the absence of sound, and physiologically – to the stop in the activity of speech organs.
When playing back recordings – a break in playback.

Sound
A mechanical oscillation travelling through elastic mediums or bodies (solids, liquids and gases), composed of frequencies within the limits of human hearing (between about 17-20 Hz and 20 000 Hz). The heightened sensibility of human ear is detected in the frequency range from 1 kHz to 5 kHz. Mechanical oscillation which is lower in frequency than 17 Hz is called *infrasound*, while *ultrasound* is an oscillation with a frequency greater than the upper limit of human hearing (20 000 Hz).

Sound spectrum
An acoustic representation of complex sound providing information about the frequency of sound source, pitch harmonics and relative intensity of all its frequency components.

Waveform (oscillogram)
Waveform of the speech signal is a graphic representation of the signal vibration amplitude as a function of time. Waveforms can be obtained using signal processing equipment: loop waveform viewers, signal level recorders and electronic waveform viewers. Waveforms can be used to extract fragments of data for further research.
<table>
<thead>
<tr>
<th>Software version</th>
<th>Document modification date</th>
<th>Document version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.15.25</td>
<td>12072012</td>
<td>1.01</td>
</tr>
<tr>
<td>1.15.25</td>
<td>12.05.2012</td>
<td>1.01</td>
</tr>
</tbody>
</table>