Analysis and Processing of Audio Signals Using Complex Form Representation

Vladimir Taktakishvili, Alexey Ovchinnikov,Oleg Popov, Valentin Abramov
Moscow Technical University of Communications and Informatics (MTUCI)
Moscow, Russian Federation
vladimir@smartvend.ru, ovchinnikovmsk@gmail.com,
olegp45@yandex.ru, vabramov44@mail.ru

***Abstract*—The subject of study is a complex presentation of audio signals. The aim of the work are proposals for the development of methods for monitoring, processing and compact representation of audio signals based on their complex form presentation. It is shown that almost all channels that transmit a sound signal adaptively change their characteristics in accordance with the properties of the signal, which does not allow preserving of waveform original shape. It is shown that the main difficulties are associated with the need to ensure high accuracy of the formation of an orthogonal signal for wideband audio signals. Attention is drawn to the fact that the accuracy of the synthesis of an orthogonal signal is largely determined by the correct selection of the window function used in the FFT. It is shown that the representation of the audio signal in the form of an analytical envelope and cosine of the instantaneous phase made it possible to develop an original, almost instantaneous method of controlling the levels of the audio signal, making it possible to significantly increase its relative average power without changing the dynamic range of the signal. Based on the use of the averaged instantaneous frequency of the signal, selected as a result of the complex form presentation of this sound signal, a new method of adaptive filtering has been developed.**

I. Introduction

The effectiveness of audio playback, transmission and storage systems, including in audio broadcasting and communication channels, is largely determined by the way the signal is represented. Today, almost all channels that transmit a sound signal in analog or digital form, adaptively change their characteristics in accordance with the properties of the signal, which does not allow to preserve waveform shape. This is manifested both in the preservation of only subjective sound quality and in the deterioration of this quality. Therefore, existing methods of monitoring audio signals, based on the assessment of changes in their waveform shape, do not allow to fully control modern channels and the quality of these signals.

It was found that the algorithms used in the compact representation of the audio signal, its processing and the objective assessment of sound quality, have reached the maximum possibilities achievable with the existing methods of representing the signal in the time and frequency domains.

To date, the algorithms used in the compact representation of the audio signal, its processing and an objective assessment of the sound quality have reached the maximum possibilities achievable with the existing methods of representing the signal in the time and frequency domains. Therefore, development of new ways for representing sound signals, which are based, in particular, on a complex representation of a signal, allows separate description and processing of modulation parameters.

II. Hilbert envelope and instantaneous phase representation of sound signal

To describe sound signal we can use different representations. One of such representations is the analytical (Hilbert) envelope, the instantaneous phase and its first derivative, instantaneous frequency [1]. In the majority of publications on this topic, it is proposed to use an artificially synthesized orthogonal signal for the implementation of a complex presentation, with the main difficulties associated with the need to ensure high accuracy of its formation, which is difficult for broadband broadcast signals [2].

The required accuracy of the implementation of the orthogonal transformation of the original sound signal is determined, ultimately, by the properties of the ear for the noticeable distortion of the signal modulation. This allows us to associate the parameters of permissible distortion of the modulation of sound signals in amplitude and frequency, with a permissible error of the orthogonal transform, accordingly to the perception by human ear [3].

Graphs of a "peripheral auditory analyzer" (which is literal translation of [4] name for human hearing system) threshold sensitivity to amplitude modulation (AM) of a signal with different modulation ratios m are shown in Fig. 1, and those with frequency modulation (FM) with different modulation indices and modulating frequency 4 Hz in Fig. 2. These data allowed us to form estimates of the relative changes in the modulating functions: amplitude envelope (1):

$$\delta A = \frac{\Delta A_{thresh}}{A} \tag{1}$$

and instantaneous frequency (2):

$$\delta \omega = \frac{\Delta \omega_{thresh}}{\omega} \tag{2}$$

from the frequency of the test signal, for the threshold sensitivity of the auditory analyzer at a listening level of 80 dB, Fig. 3. It is noticeable, that the threshold of sensitivity of human ear for frequency modulation is higher than to amplitude modulation. The permissible errors of the formation of an orthogonal signal are refined analytically. White noise was additively mixed in with the test oscillation synthesized

with a given frequency and amplitude, with a certain signal-to-noise ratio (SNR). Then, for each SNR, the relative standard deviations (RSD) of the calculated modulating functions from their specified values in the test signal were calculated. RSD on a sample of N points was determined in accordance with equation (3) for a discrete signal.

$$P = \sqrt{\frac{\sum_{i=0}^{N-1}(x_i - \tilde{x}_i)^2}{\sum_{i=0}^{N-1}(x_i)^2}} \quad (3)$$

where P is relative standard deviation, x_i – signal sample and $\tilde{x}_i$ – sample estimates.

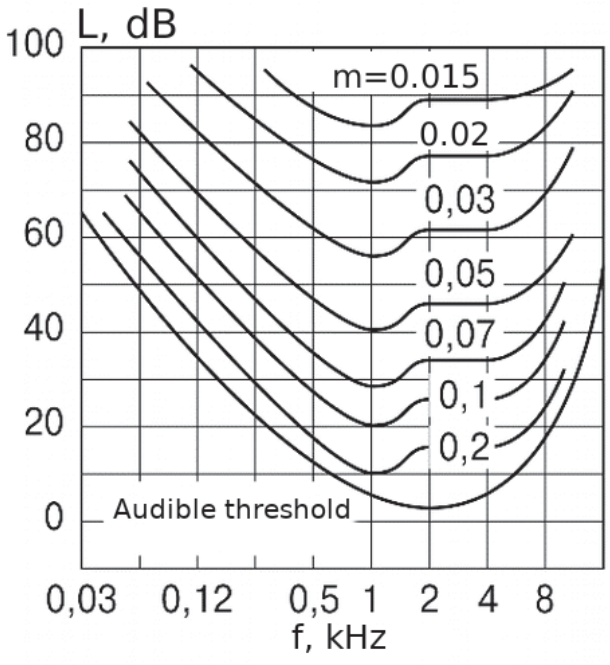

Fig. 1. Curves of thresholds of sensitivity to AM at different modulation ratios m [4]

Using the dependences shown in Fig. 2, the SNRs are selected that correspond to the thresholds of perception by ear for the modulations. The results are shown in Fig. 4. From the graphs it can be seen that the accuracy of the synthesis of an orthogonal signal should be significantly higher if it is necessary to work with an instantaneous frequency, and not with the signal envelope. Allowable error in the synthesis of an orthogonal signal should not exceed 10^{-5}.

III. Hilbert Transform Implementation

To implement the Hilbert transform (HT) of broadband signal, there are two main ways. According to the first, a broadband audio signal is converted into a narrowband by its amplitude modulation on a high-frequency carrier and a further phase rotation of the narrowband sideband [1]. The achievable accuracy is small, the error is 10^{-3}, and in the regions of non-stationarity of the signal it falls down to 10^{-2} error, which is not enough to solve the set tasks.

According to the second method, a broadband audio signal is subjected to HTs directly in the occupied spectrum, without

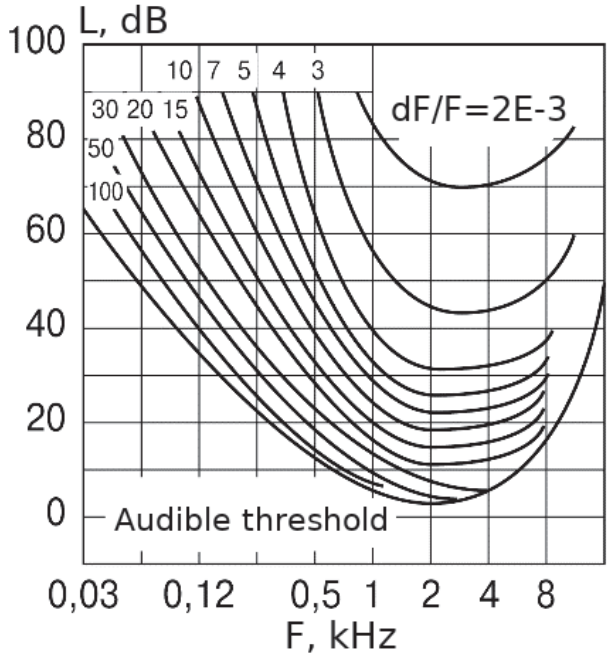

Fig. 2. Curves threshold modulation index values for FM. Modulating frequency is 4 Hz [4]

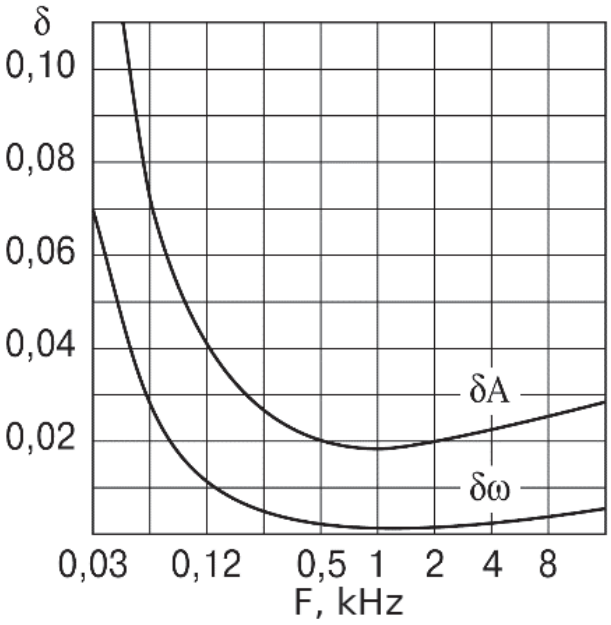

Fig. 3. Relative changes in the modulating functions corresponding to the threshold sensitivity of the auditory analyzer at the listening level of 80 dB

transfer to the high-frequency spectrum. In the considered variant, it is proposed to form an orthogonal signal in the process of direct and inverse fast Fourier transform (FFT and IFFT), with the phase rotated by 90° before the IFFT.

In this case, the accuracy of the synthesis of an orthogonal signal (OS) is largely determined by the correct selection of the window function used in the FFT. Obviously, the lower the

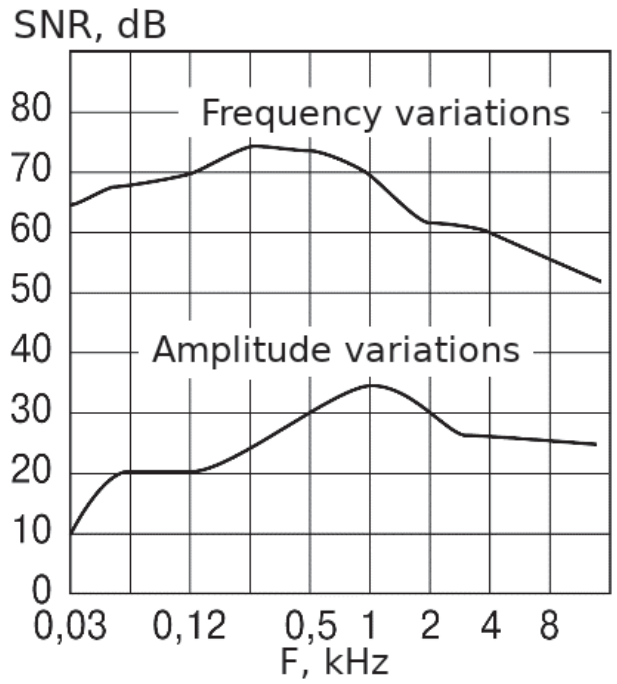

Fig. 4. Thresholds of conspicuity for an orthogonal signal synthesis error in frequency and amplitude

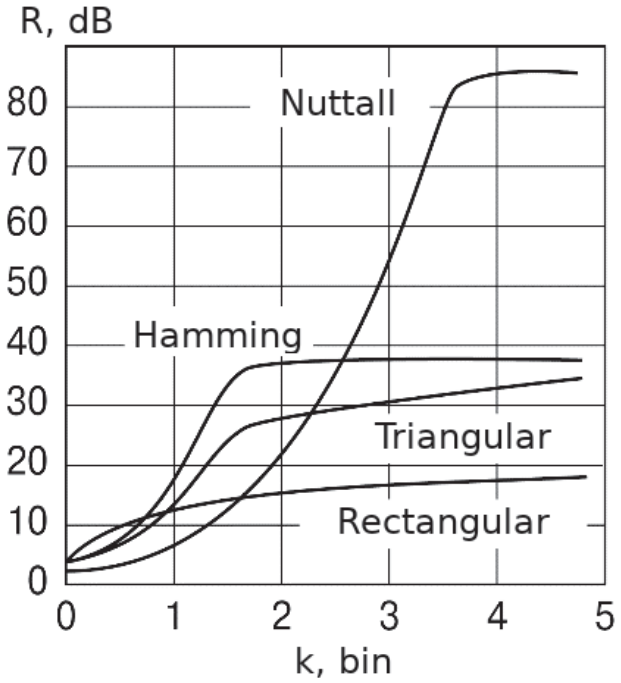

Fig. 5. R_k for four window functions [5]

sidelobe level of the window function in the frequency domain, the higher the conversion accuracy. However, a decrease in side lobes is accompanied by an increase in the width of the main lobe.

Since the coefficients of the real and conjugate parts of the spectrum vary in phase in opposite directions, the signals displayed by the first and last FFT coefficients will be subjected to the greatest distortions. Of the known functions, the Nuttall window (or the minimum 4-term Blackman–Harris) has a minimal level of side lobes. However, since this window does not provide a single transfer coefficient for any overlap factor, it is necessary to introduce an additional compensating function after IFFT [3]. The results of measurements of the ratio of the energy of a given coefficient to the energy of the remaining coefficients located higher in frequency R_k are shown in Fig. 5, which shows the results of the analysis of distortions for the first five coefficients for the four most widely used window functions. R_k can be calculated using equation (4).

$$R_k = \frac{\sum_{i=1}^{N/2-1} E_i}{\sum_{i=k+1}^{N/2-1} E_i} \tag{4}$$

where E_i is the energy of the i-th coefficient and k is the coefficient number, to which the energy of all coefficients referred to, in means of the spectral energy.

R_k determines the maximum theoretical error of OS synthesis, for a sinusoidal signal, corresponding to the k-th coefficient, and determined by the properties of the window. In the graphs in Fig. 5, R_k for four common windows is shown: Nuttall, Hamming, triangular and rectangular. The intermediate points (corresponding to parts of bin) are obtained by interpolation. For the last coefficients of the real part of the spectrum, the graphs will be similar. For the first two coefficients, the Nuttall window is inferior to other windows in the signal-to-noise protection ratio, but significantly exceeds them in the coefficients of large numbers [3]. At the 4'th coefficient, the value of R_k reaches approximately 85.7 dB and then almost does not change.

Fig. 6 presents the illustrations of using the window function for overlapping sequences, smoothed by the window, used in algorithm 1. This algorithm is used for synthesizing an orthogonal signal. The errors in the formation of an orthogonal signal, presented in Fig. 7, were calculated for a sinusoidal signal with a frequency equal to the frequency of a given FFT component.

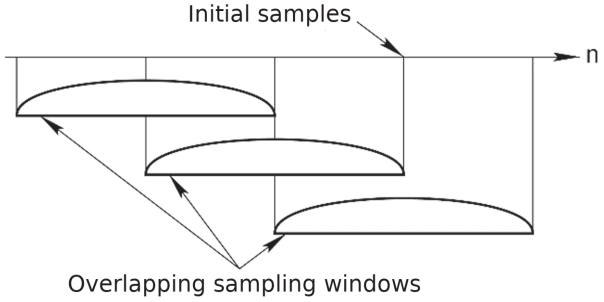

Fig. 6. Sampling Scheme for the Orthogonal Signal Synthesis Algorithm

It can be seen from the figure that the SINR within the limits of the bin changes, reaching a minimum between the bins. The minimum SINR is implemented at extreme coefficients, which limits the possibility of using this method

of conversion in the region of low and high frequencies, where the dependencies are similar. In practice, when working with a real audio signal, the sample length are increased and the coefficients located outside the AS spectrum are not used, to ensure minimal signal distortion at the edges of the frequency range. The data in Fig. 7 were obtained using the Nuttall window for a duration of 8192 points. Note that since this window does not provide a single transfer coefficient for any overlap factor, it is necessary to introduce a compensating function (see algorithm 1). The extension of the main lobe, typical fot Nuttall window, can be compensated by an increase in the sample duration.

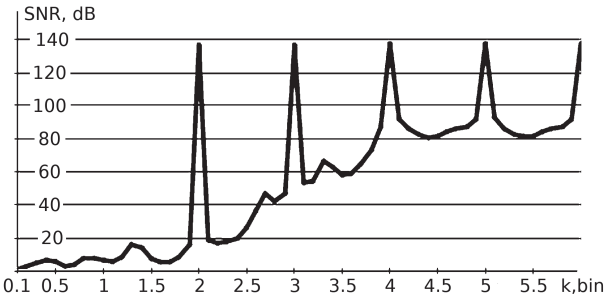

Fig. 7. OS Formation Error for Sinusoidal Signal

Algorithm 1 Synthesis of orthogonal signal [3]

0: Get input signal s
0: Split s into window-sized overlapping samples s_i
0: Apply windowing function w to s_i and get p_i
0: Apply FFT to p_i and get $\dot{S}_i$
0: Multiply first half of values of $\dot{S}_i$ to j
0: Multiply second half of values of $\dot{S}_i$ to $-j$
0: Apply IFFT to $\dot{S}_i$ and get $\hat{S}_i$
0: Join overlapping $\hat{S}_i$ to get $\hat{S}$
0: Multiply $\hat{S}$ by some constant factor to compensate window irregularity
0: Use resulting samples as OS

The error in the synthesis of an orthogonal sound signal using an FFT and the modified Nuttall window can be reduced to 10^{-5}, with a sample duration of at least 4000 points. Such an error is sufficient for the implementation of efficient algorithms for analyzing, processing, and compact representation of a AS based on the modulation parameters of wideband audio signals such as amplitude envelope and instantaneous frequency.

The representation of the audio signal in the form of an analytical envelope and cosine of the instantaneous phase made it possible to develop an original, almost instantaneous method of controlling sound signal levels, which makes it possible to significantly increase its relative average power (RAP) without changing the dynamic range of the signal, which is unattainable for existing analogues [6]. Such regulation allows to compensate some of the undesirable changes made to the signal by existing methods of compact representation and processing in the transmission channel.

Reference [7] provides a detailed analysis of the distortions inherent in dynamic range converters (DRC's), implemented on the basis of processing the modulating functions of a signal — the analytical envelope and the instantaneous frequency.

It is shown that while processing a test single-tone signal, excellent results were achieved, but at the same time checking the performance of the inertialess dynamic range converter (DRC) on a real audio signal gives unexpected results, namely, often observed very poor sound quality of the non-linear processed sound. The result is quite expected and repeatedly met when reviewing the projects of similar automatic gain control (AGC) systems at the Department of Television and Sound Broadcasting (TV&SB) of MTUCI. In fact, it is difficult to imagine how a broadband analytic envelope that has passed nonlinear processing when a signal is restored, by multiplying the instantaneous phase by the cosine, will not introduce distortions.

IV. Hardware Usage of Method

In [7], a detailed analysis of the arising distortions is carried out using a two-tone signal as an example involving the tabular values of the gamma functions of the cosine Fourier transform and the conclusion is drawn that it is impossible to implement inertial-free processing of the real sound broadcasting signal. However, at the TV&SB MTUCI department, an ARGO audio processor (from russian abbreviature for "automatic Hilbert envelope regulator") was developed, which safely and efficiently realizes inertialess regulation of the real broadcast signal. The high quality of the signal at the output of the device is confirmed by independent subjectively statistical examinations and prolonged (since the end of the 90's years of the last century) trial operation in the paths of the primary and secondary distribution of the sound broadcasting signal. In total, about 60 units were used in Moscow and Novosibirsk, several units still used in Ufa. Of course, pilot production is difficult for the department. It can't compete with firms like Orban, especially in terms of operational maintenance of devices, which in fact limited the volume of production. The appearance of the device shown in Fig. 8 and the waveforms of the signal before and after adjustment is shown in Fig. 9.

Fig. 8. ARGO audio processor front panel

The solution that made it possible to remove the problem of the distortions in the course of regulation consists in the separation of the analytical envelope into low-frequency (LF) and high-frequency (HF) components. A cut-off frequency of about 10 Hz is used, which corresponds to the maximum repetition rate of sound objects in the sound signal, while the LF envelope is transformed using the modified μ-characteristic, and the HF envelope is adjusted proportionally to LF [3]. Such adjustment allows us to rearrange the volume of sound objects without changing the overall dynamic range of the signal. The relative average power increases by $40..100\%$ when transmitting audio broadcast signals and can be increased up to 350% in mass notification systems.

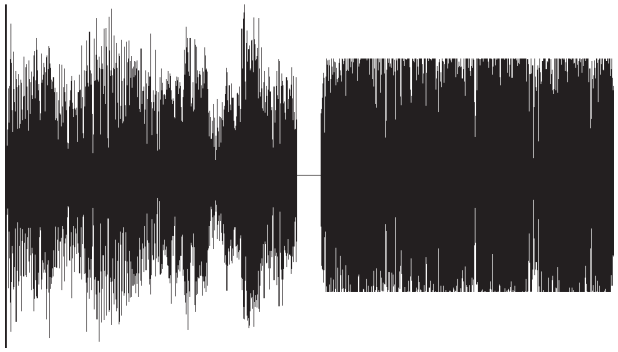

Fig. 9. Waveforms of signal before and after processing with ARGO

Naturally, this regulation does not take into account the systematic delay of the signal as a whole, necessary for accumulating the data array for the FFT and evaluating the OTM signal that determines the change in the shape of the amplitude characteristic of the control. In this task, the waveform is deliberately changed, therefore, the quality of regulation is assessed by the results of subjectively statistical tests and signal spectrum enrichment using a specially developed spectral analysis algorithm with an accuracy close to the accuracy of a human hearing analyzer [3].

The same approach was used in the development of an algorithm for the non-distorting companding of an audio broadcasting signal (ABS) in the transmission channels with an insufficient dynamic range. The low-frequency envelope (which always corresponds to the most powerful components of the signal) is compressed before transmission, while the high-frequency component (corresponding to the weak ones), is expanded. At the reception the reverse processing takes place.

Since the system in question retains the waveform, the noise component at its output can be determined by directly subtracting the original signal from the transmitted program model of the transmission path, in which white noise with a given level was mixed into the compressed signal. This allows you to use the relative root-mean-square value of noise (RMS) to assess the performance of the algorithm. Namely, the processing efficiency can be assessed by the difference SINR of the signal that has undergone compander processing and the signal that has not been processed. An important indicator of the effectiveness of compression of the ABS is also the degree of increase in RAP.

With the above optimized processing parameters, as a result of modeling, the relative gain in SINR was obtained $2.3..2.5$ times as compared to the unprocessed signal at a noise level of -46 and -40 dB. The results of the change in RAP after compression of audio braodcasting signals are given in table IV [3].

TABLE I. RESULTS OF COMPRESSION OF AUDIO BRAODCASTING SIGNALS

Signal type	increase in RAP, times
speech	2.5
symphonic music	2.2
pop music	1.8

Thus, the algorithm of non-distorting companding allows to reduce the level of the transmitted signal, at least twice, with a margin for distortion of the real channel, i.e. reduce its power by four times. At the same time, the RAP of the compressed signal will increase by 2.5 times, therefore, the resulting change in signal power will be 1.5 times.

The basic principles underlying the considered control algorithm make it possible to preserve an objective signal quality. In this case, the depth of regulation is limited to tolerances regulated by GOST R 52742-2007 [8]. In particular, the value of permissible signal distortion, which should not exceed $1.4..2.8\%$ when transmitting a signal with a frequency band up to 15 kHz. It should be noted that the method of estimating the permissible distortion as an error between the input and output real broadcast signals leads to an overestimation of the requirements for the device. This is due to the fact that permissible, according to GOST [8], distortions are determined on the basis of the perceptibility of distortions of a single-component test signal with minimal masking effect, while the permissible distortions of a real broadcast signal, by the criterion of perceptibility of its changes for a listener, are several times higher. Note that the Dolbi noise suppressors provide a gain of about 20 dB (according to an advertisement) in terms of subjective perception without saving the waveform. With non-distorting companding, a gain of about 10 dB is ensured, but according to objective measurements of the waveform, Fig. 9 and Fig. 10.

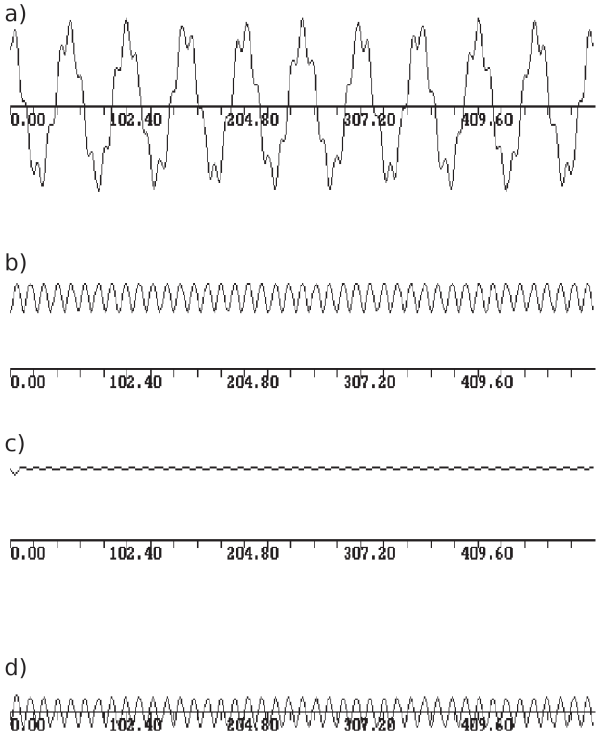

Fig. 10. Formation of the OS and its on the LF and HF components

Based on the use of the averaged instantaneous frequency of the signal, selected as a result of the complex form presentation of this sound signal, a new method of adaptive filtering has been developed. It allows you to tune the narrow-band

filters in frequency, tracking the local maxima of the envelope of the amplitude spectrum of the signal, consistent with the properties of the peripheral auditory analyzer.

V. Conclusion

The method provides a high concentration of energy in a small set of subband signals and allows you to save the waveform. This, in turn, allows its use in assessing the quality of transmission based on existing techniques.

Using the complex form signal representation, a method of compact representation of the audio signal was developed based on the generation of quasi-permanent and variable signals on the first, second and third stages of modulation decomposition from the signal associated with the instantaneous frequency parameters and the Hilbert amplitude envelope of the signal. The parameters, selected at three stages of modulation decomposition, after digitization are transferred to the receiving side, where the sound signal is restored from them [9]. The method provides a fairly high compression ratio (roughly 4 times), while maintaining the waveform, and hence maintaining the quality of the audio signal.

The use of modulation parameters, with a complex representation of the signal, improves the efficiency of algorithms for an objective assessment of the quality of transmission the sound broadcasting signal in the paths without saving the waveform, which are all modern analog and digital paths [3].

It was found that the permissible error in the synthesis of an orthogonal signal should not exceed 10^{-5}.

A complex presentation of the audio signal allows you to expand the capabilities of the algorithms of compact representation, processing and evaluation of the quality of the transmission of an audio signal via a communication channel. The main obstacle for using such a representation is the low accuracy of the synthesis of an orthogonal signal, which is overcome during the synthesis of an orthogonal signal using an FFT and the Nuttol window function, supplemented with a compensating window at the output of the converter. The error in the synthesis of an orthogonal signal does not exceed 10^{-5} at a sample duration as small as 4000 points. Such accuracy is sufficient to create algorithms and devices for processing, analyzing and compact representation of the audio signal.

The exact complex representation of the audio signal allows you to realize its compact representation as a sum of narrowband signals generated by frequency-tuned filters that track the local maxima of the envelope of the amplitude spectrum of the signal.

Analysis of the properties of the analytical envelope and the instantaneous frequency allows us to develop new ways to objectively evaluate sound quality in transmission channels without maintaining the waveform, providing good estimation of both subjectively high enough and degraded sound quality [3].

The developed algorithm of non-distorting companding allows to reduce the level of the transmitted signal at least twice – with a margin for distortion of the real channel, i.e. reduce its power by four times, at the same time the compressed signal relative average power will increase by 2.5 times.

Inertialess regulation of the signal allows you to maintain the signal level almost at the speed of following sound objects, which is unattainable for existing analogs [6] and makes it possible to reduce signal degradation determined by the used methods of reducing the transmission speed. The preservation of the waveform, when it is represented by localized modulating functions, ensures high transmission quality, while reducing the signal volume.

In the process of research and development of practical applications of a complex presentation of the audio signal, an original method was proposed for measuring the instantaneous and average values of the absolute and relative power of sampled audio signals over short time intervals [10]. A method was also proposed for forming a spectrum estimate with an accuracy close to the capabilities of a peripheral auditory analyzer [11]. In addition, a method has been developed for transmitting and receiving signals represented by the parameters of stepwise modulation decomposition [9]. The original software was developed to evaluate the parameters of the signal, presented in a complex form and allowing to predict the evaluation of the transmission quality by the Estim listener [8].

References

[1] Yu.M. Ishutkin and V.K. Uvarov, *Basics of modulation transformations of audio signals*. Saint-Petersburg: SPbGUKiT, 2004, 102 pages.

[2] V.K. Uvarov and A.Yu. Redko, “Modulation analysis – synthesis of sound signals and the prospects for its use in noise reduction”, Moscow: *Fundamental research*, 2015, #6-3, pp. 518-522.

[3] O.B. Popov and S.G. Richter, *Digital processing and measurement of signals in the channels of audio broadcasting*. Moscow: Insvyazyzdat, 2010, 292 pages.

[4] E. Zwicker and R. Feldtkeller, *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel, 1967, 232 pages.

[5] S. L. Marple. *Digital Spectral Analysis: With Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1986, 492 pages.

[6] O.B. Popov and S.G. Richter, “A method of automatic control of peak values of electric broadcasting signals to a predetermined level while stabilizing the relative average power and a device for its implementation”. RF patent No. 2408976 BI n1. 10.01.2011.

[7] Yu.K. Zirova. “Research and development of methods for improving the inertive converters of the dynamic range of sound signals”, *Instruments* #3, 2009, Moscow: Union of Public Associations “International Scientific and Technical Society of Instrument-makers and Metrologists”, ISSN: 2071-7865, pp. 50-57.

[8] V.A. Abramov, G.M. Ozhdikhin, O.B. Popov, K.V. Chernikov and A.V. Malov, “Parameters analysis of sound signals ESTIM”. Certificate of registration of software No. 2013616645, 15.07.2013.

[9] V.A. Abramov and O.B. Popov, “A method of transmitting and receiving signals represented by the parameters of a stepped modulation decomposition and a device for its implementation”. RF patent No. 2584462 BI n14. 20.06.2016.

[10] V.A. Abramov, O.B. Popov and S.G. Richter, “A method for measuring the instantaneous and average values of the absolute and relative power of acoustic signals and a device for its implementation”. RF patent No. 2458340 BI n10. 10.04.2012.

[11] V.A. Abramov and O.B. Popov, “The method of measuring the spectrum of information acoustic signals of broadcasting and device for its implementation”. RF patent No. RU2573248 C2, BI n2. 20.01.2016.