Scientific Visualization, 2025, volume 17, number 5, pages 1 - 9, DOI: 10.26583/sv.17.5.01

Means of Analysis and Visualization of Speech Signals in the Preparation of Engineering Students

Yu.G. Gorshkov[1]

Bauman Moscow State Technical University, Russian Federation

[1] ORCID: 0000-0003-0483-4603, y.gorshkov@npo-echelon.ru

Abstract

The paper considers the foreign experience of teaching engineering students to analyze speech signals using instrumental methods. Examples of obtaining spectrograms are given, as well as the capabilities of speech analysis software used in undergraduate and graduate engineering courses at the University of California (Los Angeles, USA). The disadvantages of speech analysis and visualization based on Fourier transform are shown. New solutions for processing and visualizing speech signals based on multilevel wavelet analysis are proposed. The main characteristics of the developed WaveView and WaveView-MWA programs that provide increased time-frequency resolution of vowel sounds are considered. For the first time, the results of high-precision analysis and visualization of consonant sounds - non-stationary signals inaccessible to spectral analysis using the Fourier transform are presented. A comparative analysis of the time-frequency resolution of spectrograms and wavelet sonograms in the visualization of English speech is performed. The developed technology of high-precision analysis and visualization of speech signals is used in the training of specialists of the Department of «Information Security» of the Faculty of «Informatics and Management» of the Bauman Moscow State Technical University during laboratory work on the course «Forensic study of phonograms».

Keywords: visualization of speech signals, spectral analysis, multilevel wavelet analysis, sonogram.

1. Introduction

Leading domestic and foreign universities attach great importance to the analysis of speech signals in the training of engineering students. For example, the work [1] presents the results of implementing methods for teaching «reading» spectrograms in undergraduate and graduate engineering courses at the University of California, Los Angeles (ranked by Forbes magazine as one of the best colleges in the world in 2019). Spectrogram analysis is an essential skill for studying speech acoustics and is necessary for visualizing the causal relationships between the movements of the speech articulator and the resulting sound. Reading spectrograms is often a challenging task for students who do not have prior training in acoustic phonetics.

The mathematical apparatus for constructing spectrograms was developed in the 1960s. The definition of the short-term Fourier transform (whose logarithmic value is displayed on the spectrogram) was formulated in 1967, and its calculation using a computer became possible in 1970 [2]. Some of the original spectrographic parameters and terms have survived to the present day, including the use of «broadband» and «narrowband» to describe spectrograms calculated using «short» and «long» analysis windows, respectively [3]. To obtain spectrograms of a speech signal or «visible speech» images, programs such as Praat [4] and Audacity Team [5] are used in the University of California and several other universities.

The purpose of this study is to evaluate the frequency-time resolution of speech signal spectrograms obtained using Fourier transformation, as well as to explore the possibilities of a new solution for processing and visualizing vowel and consonant sounds in speech using multilevel wavelet analysis.

2. Software for obtaining speech signal spectrograms

2.1. The Praat Program

Praat computer program (doing phonetics by computer), developed by the Institute of Phonetic Research at the University of Amsterdam, Netherlands, 2009. Its purpose is to perform Fourier analysis and visualize the formant characteristics of a speech signal [6, 7]. A formant is a band of the speech tract's transfer function characterized by its frequency, Fi, amplitude, Ai, and bandwidth, Bi [8, 9]. In the amplitude-frequency spectrum, formants appear as prominent peaks on vowel sounds (Figure 1).

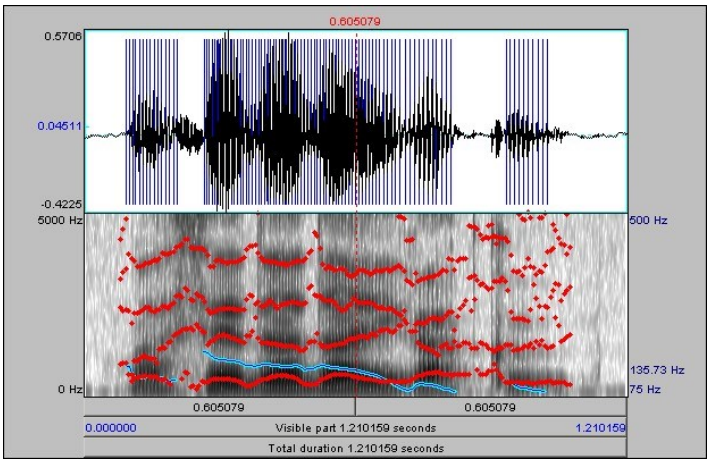

Fig. 1. Praat. The «Sound» window. The formant trajectories of vowel sounds are marked in red

2.2. Audacity Program

Audacity is an open-source audio editor and recording application software available for Windows, macOS, Linux, and other Unix-like operating systems; developed by Carnegie Mellon University, USA, 2000. Purpose - instrumental Fourier analysis and visualization of audio signals. Fig. 2 shows the appearance of the Audacity program interface, the Multi-view mode.

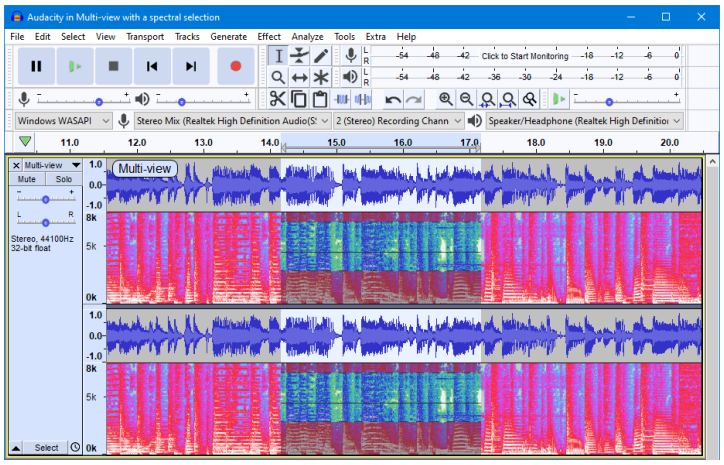

Fig. 2. The appearance of the Audacity program interface, Multi-view mode

2.3. Examples of speech signal spectrograms

Let's consider the spectrograms of two phrases: «We saw the goal to win boons» and «We sue the bowl to bin beans», obtained using the Praat program and reviewed by Bruce Hayes, UCLA [6]. (Bruce Hayes is a Distinguished Professor of Linguistics at the University of California, Los Angeles).

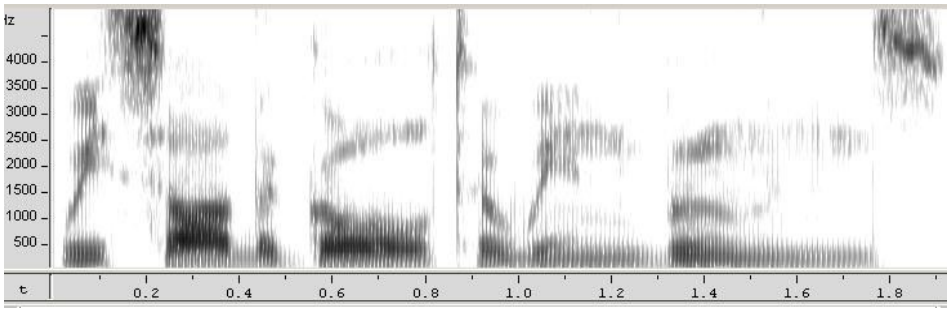

Fig. 3. Spectrogram of a phrase «We saw the goal to win boons»

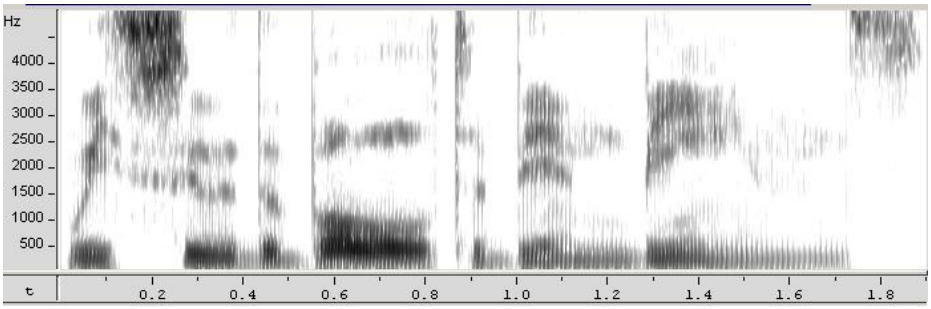

Fig. 4. Spectrogram of a phrase «We sue the bowl to bin beans»

The paper [6] also provides comments on the analysis of the spectrograms in Figures 3 and 4:

saw/**sue**, vowels at about 800 msec. in both spectrograms: note high F1 for the low vowel (IPA open o) in «saw» (about 600 Hz.) vs. low F1 for [u] in «sue» (about 300 Hz). High vowels have low F1, low vowels have high F2.

goal/bowl: initial stop at about 1100 msec. in both spectrograms. First spectrogram: [g] has a velar pinch in F2/F3. Second spectrogram: formants rise out of the stop closure for [b].

win/**bin**: First spectrogram: [w] at about 1500 msec. has a gentle decrease/increase in amplitude; it is a sonorant consonant. Second spectrogram: [b], a stop, has a sudden increase in amplitude, with a small burst, at the moment of release.

boons/**beans**: First spectrogram: [u] at about 1800 msec. has F2 around 1300. Second spectrogram: [i] at around 1800 msec. has F2 around 2200 Hz.; back vowels, and round vowels, have lower F2 than front/unrounded vowels.

From the data obtained, it follows that the considered technique of formant estimation by spectrograms obtained based on the Fourier transform allows to perform a primary analysis of the speech signal, and only vowel sounds (in English, 6 vowels, 21 consonants). At the same time, for solving such tasks as automatic speech recognition, speaker authentication, determination of the dialect of the language and the accent of a foreign language speaker, new high-precision solutions for signal analysis are required [10, 11].

3. Analysis and visualization of a speech signal based on wavelet transformation

3.1. The wavelet analysis program WaveView

Speech signals are complex non-stationary signals, so using wavelet transforms for speech analysis will provide a more accurate representation of speech in the frequency-time domain [12]. The first version of WaveView was developed in 2003 [13]. The WaveView wavelet analysis program provides the following features [14]: display of a signal oscillogram; analysis of a signal section with the ability to select a frequency band and time-frequency resolution; visualization - display of analysis results in the form of a wavelet sonogram (Morlet wavelet is used); obtaining a frequency section at a specified time; support for a large number of audio file formats. PC requirements: Windows operating system.

Wavelet sonograms can also be constructed remotely using the Acustocardiograph portal (http://acustocard.ru) [15]. Figure 5 shows an example of the visualization of the vowel sounds А, Э, И, О, У, Ы, as well as the points of the phonogram's editing, obtained by 6th-year students of the Department of Information Security at the Bauman Moscow State Technical University during the laboratory work «Determination of the Authenticity of Phonograms Using FFT and Wavelet Analysis Technologies» in the course «Forensic Analysis of Phonograms» [16].

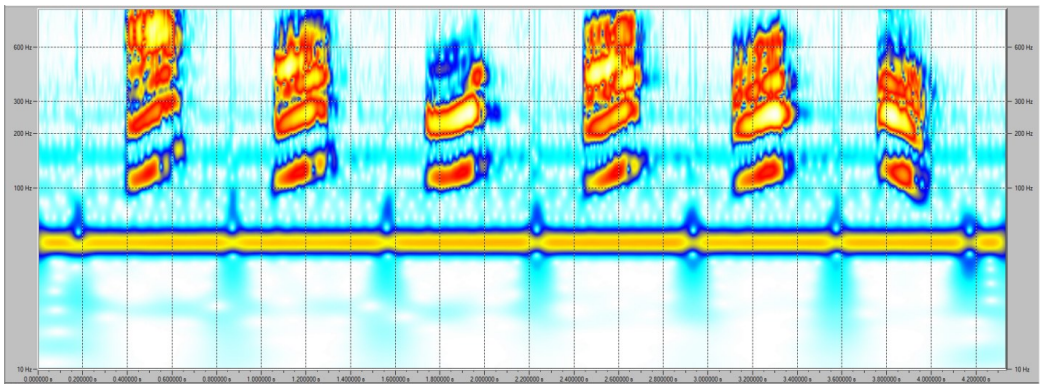

Fig. 5. Wavelet sonogram of vowel sounds А, Э, И, О, У, Ы. In the lower part of the image, you can see the mounting points – the phase breaks of the 50 Hz power supply background signal

3.2. Software WaveView-MWA

The WaveView-MWA software [17] implements several algorithms for constructing wavelet sonograms, performing wavelet filtering, and calculating the phase of a speech signal. A wavelet sonogram is a diagram on which time is plotted along the abscissa axis, frequency is plotted along the ordinate axis, and the amplitude of the corresponding frequency component is marked by the color intensity at a given point on the graph. When constructing a sonogram, the values of the signal spectrum are calculated for each time point according to the specified parameters of the wavelet transform. The obtained amplitude data is the value of one column of the graph. A wavelet sonogram provides high frequency-time resolution of the signal under study.

3.2.1. Description of the main menu

Figure 6 shows the main menu of the WaveView-MWA software settings for building a wavelet sonogram.

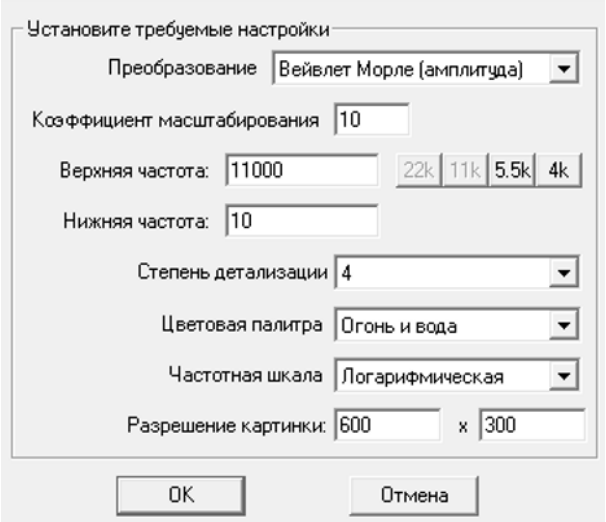

Fig. 6. Main menu view of the WaveView-MWA software

The following processing modes and parameter settings are available:

1.«Transform» - selection of the mother wavelet and the wavelet transform component displayed on the screen (amplitude, phase).

2. «Scaling factor» - a parameter that specifies the degree of localization of the wavelet sonogram in frequency and time. Small values result in high localization in time but low localization in frequency. Conversely, large values result in the opposite.

3. «Upper frequency»/«lower frequency» - the frequency range in which the wavelet sonogram will be constructed. To set the value of the upper frequency, the following designations are used: «22k», «11k», «5.5k», and «4k», which correspond to 22050 Hz, 11025 Hz, 5500 Hz, and 4000 Hz, respectively.

4. «Detail level» - a measure of the frequency-time resolution of the wavelet sonogram.

5. «Color Palette» - select the color representation of the wavelet sonogram.

6. «Frequency scale» can be presented in logarithmic or linear scale. «Logarithmic» provides a detailed representation of the low-frequency region of the signal, while «Linear» provides a detailed representation of the high-frequency region.

7. «Image resolution» sets the size of the sonogram image in pixels.

Figure 7 shows the structure of the wavelet sonogram image of the speech signal, with the word «hook».

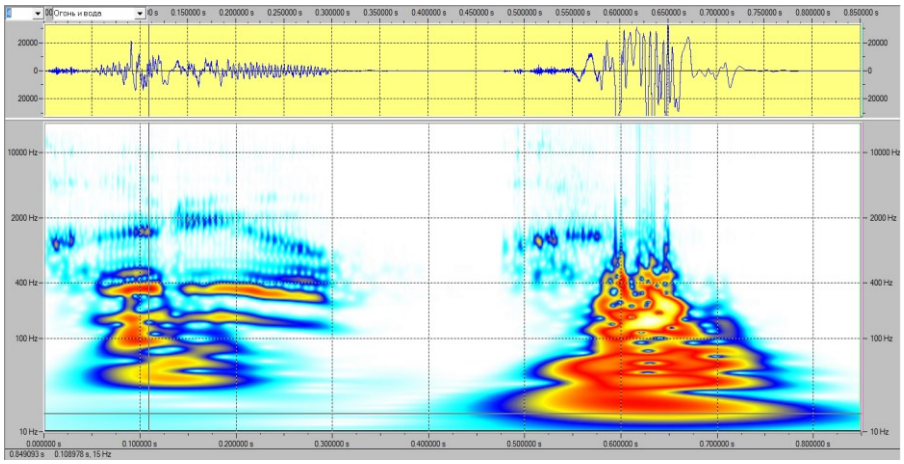
Fig. 7. Structure of a wavelet sonogram image of a speech signal, the word «hook»

The conducted testing of the WaveView-MWA software showed the possibility of extracting and visualizing non-stationary signals of low level (up to -60 dB). The use of the «Sound Microscope» mode allows to obtain the characteristics of vowel speech sounds with increased

frequency-time resolution. Obtained the parameters of consonant sounds, not available to software analysis tools using Fourier transform.

3.3. Examples of Wavelet Sonograms of a Speech Signal

Fig. 8 shows the wavelet sonogram of the phrase: «We saw the goal to win boons», obtained using the WaveView-MWA software. (The spectrogram for comparison is shown in Fig. 3).

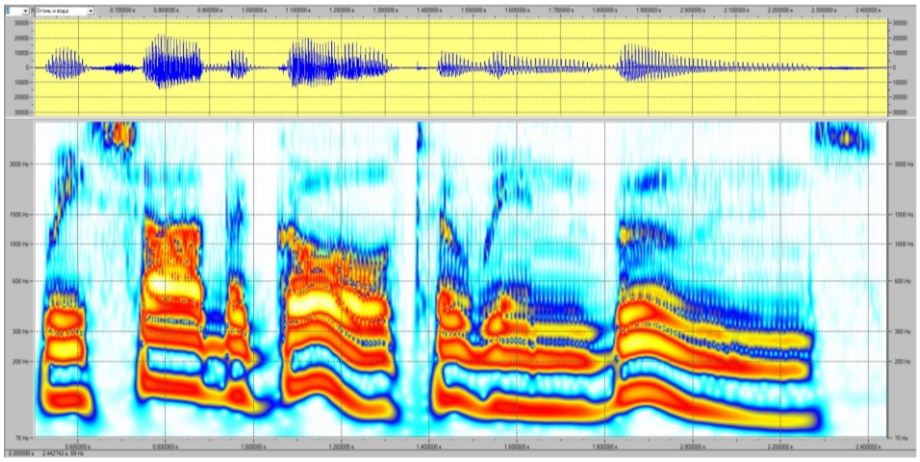

Fig. 8. Wavelet sonogram of a phrase «We saw the goal to win boons»

In contrast to the spectrogram of Fig. 3, we see an increased frequency-time resolution of the tonal sections of the speech signal - vowel sounds. The main characteristics - the digital values of formants, the period of the fundamental tone, and its dynamics - are displayed using a real-time cursor.

Let's consider the possibility of extracting and visualizing consonant sounds in speech. In Fig. 9, the wavelet sonogram shows the combination of consonant-vowel sounds **«go»** (**goal**).

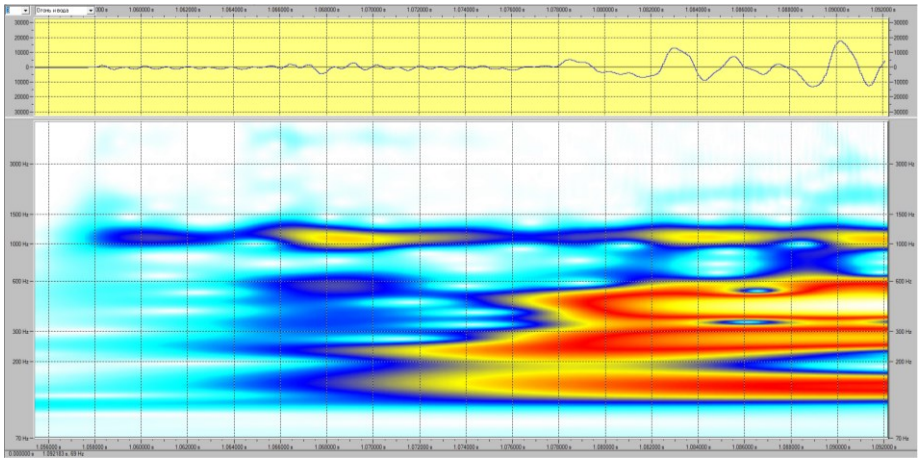

Fig. 9. Wavelet sonogram «go» (goal), 1.058-1.092 sec. Frequency band «g»: 1000-1300 Hz; «o»: 100-600 Hz

Figure 10 shows a wavelet sonogram of a consonant-vowel combination **«to»** (**to win**).

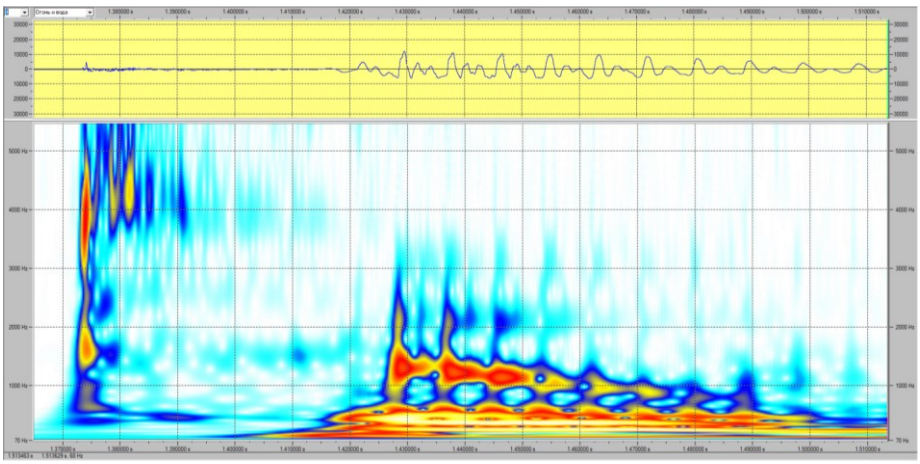

Fig. 10. Wavelet sonogram **«to»** (**to win**), 1.37-1.51 sec. Frequency band **«t»**: 295-5405 Hz; **«o»**: 88-3070 Hz. Duration **«t»**: 0.025 sec; **«o»**: 0.115 sec

Figure 11 shows the wavelet sonogram of the second phrase analyzed: «We sue the bowl to bin beans», also obtained using the WaveView-MWA software. (The spectrogram for comparison is shown in Figure 4).

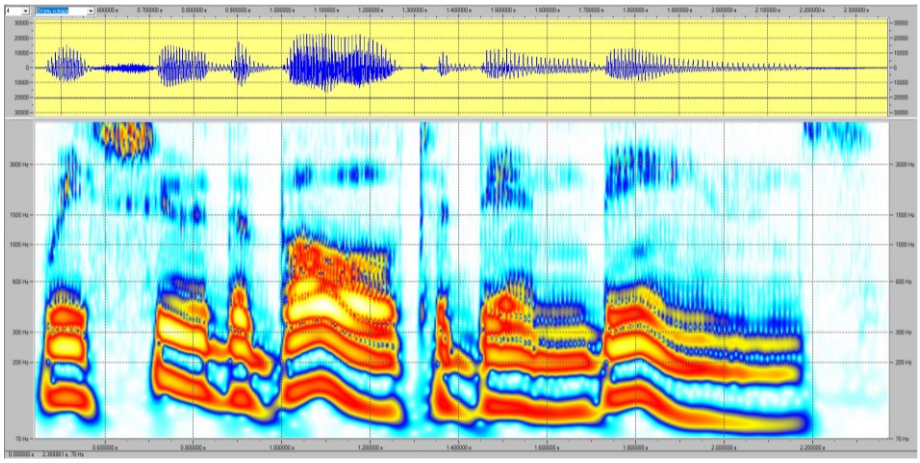

Fig. 11. Wavelet sonogram of the phrase «We sue the bowl to bin beans»

Compared to the spectrogram in Figure 4, the increased frequency-time resolution of vowel sounds is also noticeable.

Figure 12 shows a wavelet sonogram of a consonant-vowel combination **«to»** (**to bin**).

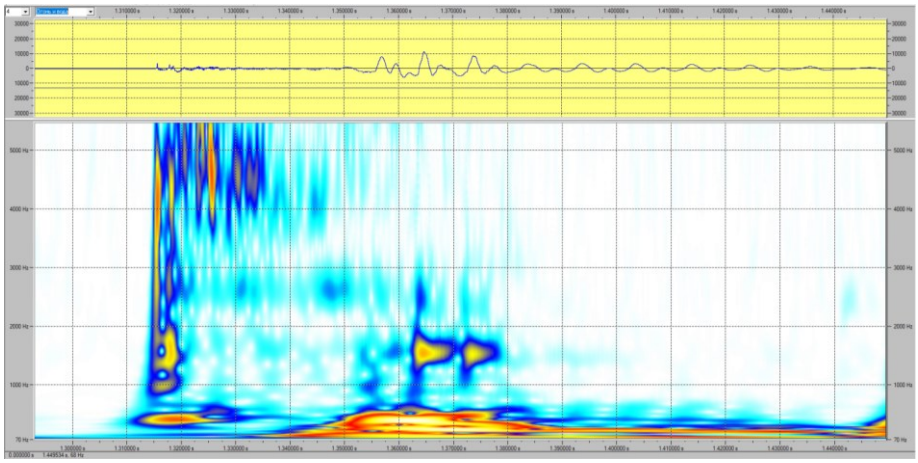

Fig. 12. Wavelet sonogram **«to»** (**to bin**), 1.31-1.335 sec. Frequency band **«t»**: 327-5613 Hz; **«o»**: 88-2885 Hz. Duration **«t»**: 0.025 sec; **«o»**: 0.06 sec

From the analysis of spectrograms (Fig. 3, 4) and sonograms (Fig. 8-12), it follows that wavelet sonograms have an increased frequency-time resolution compared to Fourier spectrograms. Visualization of data from multi-level wavelet analysis of non-stationary sections of a low-level speech signal allows for obtaining frequency-time characteristics of consonant sounds.

4. Conclusion

The developed software tools for multi-level wavelet analysis and visualization, WaveView and WaveView-MWA, allow for a frequency-time description of the speech signal with a resolution that exceeds the capabilities of Fourier analysis programs such as Praat and Audacity Team. Wavelet sonograms provide a visual, objective, and comprehensive representation of the parameters of vowel sounds. For the first time, the possibility of obtaining characteristics of consonant sounds in speech has been demonstrated.

The proposed speech signal visualization technology has confirmed its high efficiency in the performance of laboratory works by 6th-year students of the Department of Information Security of the Faculty of Computer Science and Management of the Bauman Moscow State Technical University in the course «Forensic Analysis of Phonograms» [18].

The WaveView-MWA software tools have also been used in the creation of new generation telemedicine systems [19, 20], high-precision visualization of heart sounds [21], pulmonary sounds [22], and human emotional tension based on speech signals [23].

References

1. Alexander Johnson. An integrated approach for teaching speech spectrogram analysis to engineering students. The Journal of the Acoustical Society of America. 152. 2022. pp. 1962-1969. (https://doi.org/10.1121/10.0014172)
2. Flanagan, J.L. Speech Analysis, Synthesis and Perception. Springer, New York, 1972. 446 p.
3. Sean A. Fulop. The beginning of time-frequency analysis. The Journal of the Acoustical Society of America. 152. 2022. R9-R10. (https://doi.org/10.1121/10.0014987)
4. P. Boersma and D. Weenink. Praat: Doing phonetics by computer (version 6.2.14) [computer program], 2022. (http://www.praat.org/)
5. Audacity Team Audacity(r): Free audio editor and recorder (version 3.0.0) [computer application], 2021. (https://audacityteam.org/)
6. B. Hayes. Spectrogram reading practice, 2021. (https://linguistics.ucla.edu/people/hayes/103/SpectrogramReading/index.htm)
7. Stolbov M.B. Fundamentals of speech signal analysis and processing. St. Petersburg: ITMO Research Institute. 2021. 101 p.
8. Lobanov B.M. Speech interface of intelligent systems: a textbook / B.M. Lobanov, O.E. Eliseeva; under the scientific editorship of prof. V.V. Golenkov. Minsk: BGUIR, 2006. 152 p.
9. Eliseeva O.E. Speech interface. Laboratory workshop: an educational and methodological guide for students of institutions providing higher education in the specialty «Artificial Intelligence»: in 2 hours, Part 1 / edited by prof. V.V. Golenkov. Minsk: BGUIR, 2008. 44 p.
10. Tampel I.B., Khitrov M.V. Automatic speech recognition. Textbook on the discipline «Speech recognition». St. Petersburg: SPbNIU ITMO. 2014. 119 p.
11. Leonov A.S., Sorokin V.N. Formant analysis of a speech signal in the phase domain. Information processes. 2021. Vol. 21. No. 2. pp. 125-134. (http://www.jip.ru)
12. Gorshkov Yu.G. Processing speech and acoustic biomedical signals based on wavelets. Scientific edition. Moscow: Radio Engineering, 2017. 240 p.
13. Gorshkov Yu.G., Kuzin A.Y. Application of the Wavelet transform in solving speech signal analysis problems. Problems of information security in the higher school system. X All-Russian Scientific Conference. Moscow, 2003. p. 51.

14. Gorshkov Yu.G., Kaindin A.M., Markov A.S., Cirlov V.L. WaveView. Wavelet analysis of speech and acoustic biomedical signals. Certificate of registration of the computer program RU 2017662425, 07.11.2017. Application No. 2017619325 dated 09/14/2017.
15. The portal «Acoustocardiograph» (http://acustocard.ru)
16. Gorshkov Yu.G. Forensic examination of phonograms: Guidelines for laboratory work. Educational and methodical manual. Moscow: Publishing House of Bauman Moscow State Technical University, 2017. 32 p.
17. Gorshkov Yu.G., Kaindin A.M., Markov A.S., Cirlov V.L. WaveView-MWA. Multilevel wavelet analysis of speech and acoustic biomedical signals. Certificate of registration of the computer program RU 2017662094, 27.10.2017. Application No. 2017619124 dated 08.09.2017.
18. Gorshkov Yu.G. Visualization of multilevel wavelet analysis of phonograms. Scientific visualization. 2015. Vol. 7. No. 2. pp. 96-111.
19. Gorshkov Yu.G. New visualization solutions for biomedical signals in telemedicine systems. Scientific visualization. 2019. Vol. 11. No. 2. pp. 56-72. DOI: 10.26583/sv.11.2.05.
20. Gorshkov Yu.G. Visualization of power supply network interference in telemedicine mobile electrocardiography systems. Scientific visualization. 2021. Vol. 13. No. 1. pp. 44-53. DOI: 10.26583/sv.13.1.04.
21. Gorshkov Yu.G. Visualization of heart sounds. Scientific visualization. 2017. Vol. 9. No. 1. pp. 97-111.
22. Gorshkov Yu.G. Visualization of Lung Sounds Based on Multilevel Wavelet Analysis. Scientific Visualization. 2022. Vol. 14. No. 2. pp. 18-26. DOI: 10.26583/sv.14.2.02.
23. Gorshkov Yu.G. Visualization of human emotional tension by speech signal. Scientific visualization. 2023. Vol. 15. No. 2. pp. 102-112. DOI: 10.26583/sv.15.2.09.