Research of Speech Signals Backgrounds of the Ukrainian Language Using the Wavelet Transform

Oleksandr Tymchenko [a,b], Bohdana Havrysh[c], Orest Khamula [b] and Natalya Kustra [c]

[a] University of Warmia and Mazury, Ochapowskiego str,2, Olsztyn, 10-719, Poland
[b] Ukrainian Academy of Printing, Pidholosko st., 19, Lviv, 79020, Ukraine
[c] Lviv Polytechnic National University, Stepana Bandery Street, 12, Lviv, 79000, Ukraine

Abstract

Perception of speech depends on many factors – pace of pronunciation, presence of pauses between words and phrases, surrounding noise. There are three pronunciation styles (normal, fast and slow). Full or normal pronunciation style is characterised by clear articulation of all sounds, syllables, pauses between words. In the full style, the sound of the word features, as well as the characteristic features of each sound in particular, are best revealed. This style is optimal for language perception and recognition. With a fast style of pronunciation, some phonemes are not clearly expressed and do not fully preserve their features. In this case, the recognition of individual words is not enough, both the omission of individual words and the loss of the content of the message are possible. Methods of automatic recognition of individual phonemes of a speech signal use frequency (spectral) representation. However, the Fourier transformation traditionally used for this purpose is not informative enough, does not provide information about the predominant distribution of frequencies in time, and can give incorrect results for signals with areas of sharp change, which is characteristic of phonemes number, in particular fricatives and plosives. The wavelet transformation method, which has significant advantages over the Fourier transformation, is much more promising for the analysis (and synthesis) of phonemes – it localises sharp amplitude-frequency changes in time, which makes it possible to identify the features of the speech signal and carry out its reliable recognition. This article describes the technique of using wavelet transformation to describe a number of phonemes of the Ukrainian language, namely the sounds "a", "b" and "k", which correspond to vocalised, explosive and voiceless sounds. The obtained three-dimensional wavelet diagrams well illustrate which frequency components prevail in the speech signal and at what point in time this occurs. The periodicity of individual components at different levels of wavelet diagrams for the same sounds is clearly visible. It is shown that when the duration of phonemes is changed (the pace of pronunciation is changed), the wavelet diagrams of these sounds retain all their features, which allows us to draw a conclusion about the effectiveness of wavelet analysis even of short recordings of the speech signal.

Keywords 1

Speech signal, phonemes, wavelet transformation, speech signal recognition

1. Introduction

The classic method of signal analysis is the Fourier transform. The result of the Fourier transformation is the amplitude and phase-frequency spectrum, which can be used to determine the presence of a certain frequency in the investigated signal. Methods of determining formant trajectories,

IntelITSIS'2023: 4th International Workshop on Intelligent Information Technologies and Systems of Information Security, March 22–24, 2023, Khmelnytskyi, Ukraine

EMAIL: olexandr.tymchenko@uwm.edu.pl (Oleksandr Tymchenko); dana.havrysh@gmail.com (Bohdana Havrysh); email3@mail.com (Orest Khamula); nataliia.o.kustra@lpnu.ua (Natalya Kustra)
ORCID: 0000-0001-6315-9375 (Oleksandr Tymchenko); 0000-0003-3213-9747 (Bohdana Havrysh); 0000-0003-0926-9156 (Orest Khamula); 0000-0002-3562-2032 (Natalya Kustra)

whose work algorithms are based on Fourier transformation or linear prediction, are often used in the field of speech information recognition.

The apparatus of Fourier transformations provides fairly simple formulas for calculations and a transparent interpretation of the results, but it is not without its shortcomings. So, for example, it does not distinguish a signal that is the sum of two sinusoids from the situation of sinusoids sequential inclusion, does not give information about the predominant distribution of frequencies in time, can give incorrect results for signals with areas of sharp change. Note that the local Fourier transform does not solve these problems, because according to the Heisenber uncertainty principle, it is impossible to simultaneously accurately estimate the spectral components of the signal in the frequency and time domains.

Unlike the traditionally used Fourier transform, wavelet analysis allows simultaneous detection of fluctuations in both the frequency and time domains; to reveal the frequency features of the time series, which precede in time unexpected and single "spikes" in the dynamics, etc. [1, 25]. In other words, when using the traditional spectral analysis of time series based on the Fourier transform, it is only possible to record various cycles (repetitions) that are observed on the entire analysed time series, and make the wrong conclusion that these cycles are present throughout the entire existence of the time series.

In contrast to this, wavelet analysis has its strengths [2-4] - it is a detailed study and classification of observed abrupt changes in the investigated processes – a promising task for further research and phoneme recognition of speech signals. However, each wavelet has characteristic features in physical and spectral space, so sometimes different details of the analysed signal can be detected with the help of different wavelets.

That is why it is advisable to consider the possibility of using the wavelet transform for the recognition of speech information.

The purpose of this work is to investigate the possibility of using wavelet transformation for the analysis of individual phonemes of the Ukrainian language, which is based on wavelet analysis of recorded speech signals of people of different age groups [5, 7].

2. Related works

Natural language recognition is a very complex technical problem, the solution of which lies at the intersection of many branches of science.

A Wavelets study published in 1992, Ingrid Daubechies suggests that the human ear is designed to process a sound signal and transmit a wavelet image of the signal to the brain. According to the proposed presentation, sound vibrations spread along the entire length of the convolution of the inner ear. Since the curl has the form of a spiral when it is straightened, it can be seen that the resulting sound transformation will coincide with the wavelet transformation with accuracy up to a constant.

Modern models of the sound vibrations perception process by the human ear are more often based on cepstral transformations of fine-scale coefficients. However, it is obvious that wavelet analysis of speech signals is much more appropriate for recognizing continuous speech, as it allows better separation of the speech signal individual segments both in terms of frequency and time.

This study is aimed at solving the problems of automatic segmentation of the speech signal into phonemes and removing noise and artifacts from it.

Speech segmentation algorithms are conventionally divided into three classes. Algorithms that segment the speech signal under the condition of the phonemes sequence a priori knowledge uttered in the given speech signal, for example, based on Markov models, belong to the first class. The second class includes algorithms that determine the boundaries of segments based on the degree of change in the acoustic characteristics of the signal and that do not use a priori information about the sequence of phonemes. As a rule, these are methods implemented on the basis of the Fourier transform and the short-time Fourier transform. The third class consists of combined algorithms that use the methods of the first and second classes.

However, for automatic speech recognition systems, it is important to segment the speech signal according to the language phonetic transcription. Therefore, before the stage of recognition, it is necessary to perform primary processing of the speech signal: remove areas with noise and divide the merged speech into separate words, and only then separate individual phonemes.

For segmentation into phonemes, the places of interphoneme transitions should be identified. Using Fourier transform methods, it is difficult to do this, because Fourier transform reflects the signal in the frequency domain and completely loses information about time, that is, taking into account the non-stationarity of the speech signal, it leads to significant information loss. The wavelet transform displays the signal in the time-frequency domain, which greatly facilitates the task of finding interphoneme transitions. The speech signal at the junction of phonemes undergoes significant amplitude-frequency changes, which is easily manifested by a change in the value of the wavelet coefficients. At stationary sites, this change is insignificant.

3. Ukrainian language phonemes

A major condition for language perception is the distinction between the length and intensity of individual sounds or their combinations. Perception and understanding of language messages significantly depend on the pace of their transmission, the presence of pauses between words and phrases, and other factors. It is considered optimal that the signal exceeds the noise level by more than 6.5 dB at a speech rate of 120 words/minute.

An important characteristic of speech, closely related to tempo, is pronunciation style [6, 8-10]. Depending only on sound features, three styles of pronunciation are distinguished (normal, fast and slow). Full or normal pronunciation style is characterized by clear articulation of all sounds, syllables, pauses between words. In the full style, the sound features of the word, as well as the characteristic features of each sound in particular, are best revealed. With a fast style of pronunciation, some phonemes are not clearly expressed and do not fully preserve their features. In some cases, not only the reduction of vowels is possible, but also the loss of whole syllables, strong assimilation, etc.

According to phonetic studies [11, 13, 16], the average duration of one sound ranges from 65 to 112 ms. At the maximum rate to preserve all the necessary sound units in the language, it is 50-65 ms [12, 28].

Depending on the vocal cords tension degree, either short pulsed portions of air are formed at their outlet (the air flow is modulated by the vibrating vocal cords) or a turbulent air flow (the vocal cords are not tense and the air passes freely through the constriction in the vocal tract), or there is a short explosive process caused by the creation of increased air pressure at the site of the vocal tract closure. These three types of air flow are necessary for forming, respectively:

- Vocalized sounds (these include vowels and sonorous consonants {*а, о, у, е, и, и, в, й, л, р, м, н*}), which are created as a result of periodic air vibrations passing through the vocal tract, in as a result, the resonant properties of the vocal tract are excited by impulses - portions of air modulated by tense vocal cords.
- Non-vocalized or fricative sounds (deaf consonants {*ф, с, ш, х, з, ж, г, ц, ч*}), which are formed when the vocal tract narrows in some place (usually at the end of the oral cavity) and pushing air through a narrowed place at a speed sufficient to form a turbulent flow.
- Plosive sounds {*б, п, д, т, к*}, which are formed when the vocal tract is completely closed (as a rule, at the beginning of the tract). Under the pressure of the air flow, the closure is forcibly broken, and at the same time, an explosion specific to this articulation is created.

The given classification of sounds is connected with the corresponding classification of phonemes, which are defined as "the smallest sound unit that can be a meaningful carrier in a language and serves in it to recognize words and their forms" [14, 15-17]. From the point of view of the phonetic division of the language stream, a phoneme is a minimal indivisible unit, usually there are about 40 of them in the languages of different nations. Generalized classifications of phonemes of the Ukrainian and English languages are given in Table 1.

When, in the process of speech generation, the vocal cords are stretched, the frequency of the main component of the speech signal spectrum increases, which is expressed in an increase in the pitch of the main tonality of the voice [18].

Table 1
Classification of the Ukrainian language phonemes

Vowels	Consonants					
	Vocalized		Non-vocalized		Plosive	
	Sonorous	Nasal	Fricative	Affricate	Voiced	Unvoiced
а	в	м	ф	ц	б	т
о	й	н	с	ч	д	п
у	л	н’	ш	ц’	б’	т’
е	р		с’	дз	г	к
і	р’		х	дж		
и	л’		з	дз’		
			ж			
			з’			
			г			

With a confidence level of 0.95, the frequency of the fundamental tone of the male voice lies in the range of 97-195 Hz and has an average value of 139 Hz. The frequency of the main tone of the female voice is in the range of 195-320 Hz with an average value of 249 Hz. In special cases, with strong accents, the frequency of the main tone can reach 480 Hz [19, 20].

3.1. Research methodology

We will describe the mathematical apparatus that was used in the analysis of phonemes. The calculation of the wavelet coefficients $f(t)$ of the signal was carried out according to the formula of the continuous wavelet transformation (1):

$$C(a,b) = \sqrt{a}\int_{-\infty}^{+\infty}\psi\left(\frac{t-b}{a}\right)f(t)dt \qquad (1)$$

However, it is necessary to carry out this transformation in a discrete form. Therefore, let's consider the method of constructing a wavelet diagram in more detail [21, 22].

A Gaussian wavelet function was used to obtain this diagram. 1024 discrete values of the Gaussian wavelet Ψ were taken (Fig. 1). According to them, compression or stretching of the Gaussian function took place depending on the value of the $"a"$ coefficient, the essence of which is to change the number of discrete readings representing this wavelet [23-25].

The number of counts at different levels of transformation (2):

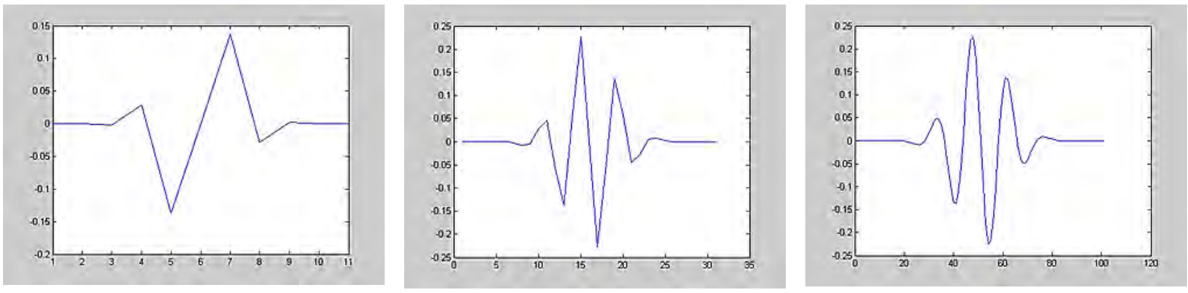
$$T = (a \cdot 10) + 1 \qquad (2)$$

and the reading values themselves were selected evenly over the interval (1...1024) according to the formula (3):

$$\psi_k(a) = \Psi(1 + floor((k-1) * 10/(a * 10/1024))) \qquad (3)$$

where $k = 1..(a \cdot 10) + 1$.

Several Gaussian wavelets for different values of the coefficient 'a', i.e. the level of the wavelet transform are shown below [26, 28-30].



$a = 1$ $a = 3$ $a = 10$

Figure 1: The shape of the Gaussian wavelet for different *a* coefficients

Next, the wavelet coefficients were determined by the discrete convolution method, which corresponds to the shift of the wavelet in the time domain, that is, the change in the coefficient b:

$$C(a,b) = -sqrt(a) * (dif\,f\,(conv(f,\psi(a))) \tag{4}$$

where conv(x,y) – discrete convolution function,

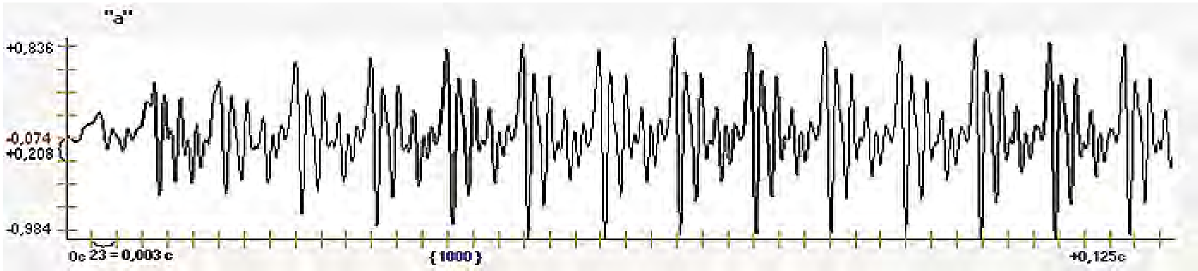
diff(x) – corresponds to the difference formula $y_i = x_{i+1} - x_i, i = 1..n-1$

Based on the values of the C(a,b) coefficients, we obtain a three-dimensional wavelet diagram.

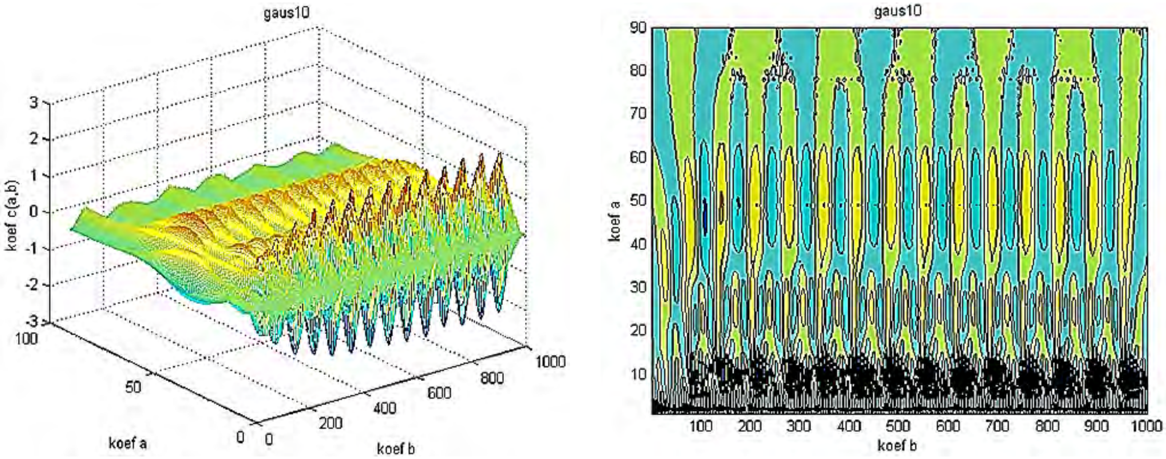
3.2. Results of the study of phonemes of the Ukrainian language

As an example of the research carried out, here are several wavelet diagrams for speech signals corresponding to three sounds of the same duration (125ms), namely the sounds "a", "б" and "к", which correspond to vocalized, plosive and voiceless sounds. This speech signal was recorded with a sampling frequency of 8 kHz and a word size of 16 bits. A male voice under the age of 30 was analysed.

The given wavelet diagrams illustrate well which frequency components prevail in the speech signal and at what point in time this happens. The periodicity of individual components at different levels of wavelet diagrams for the same sounds is clearly visible.

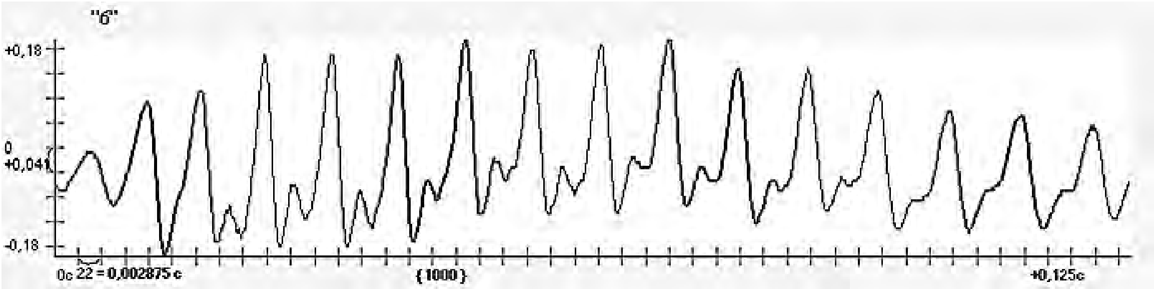


a) “a” phoneme form

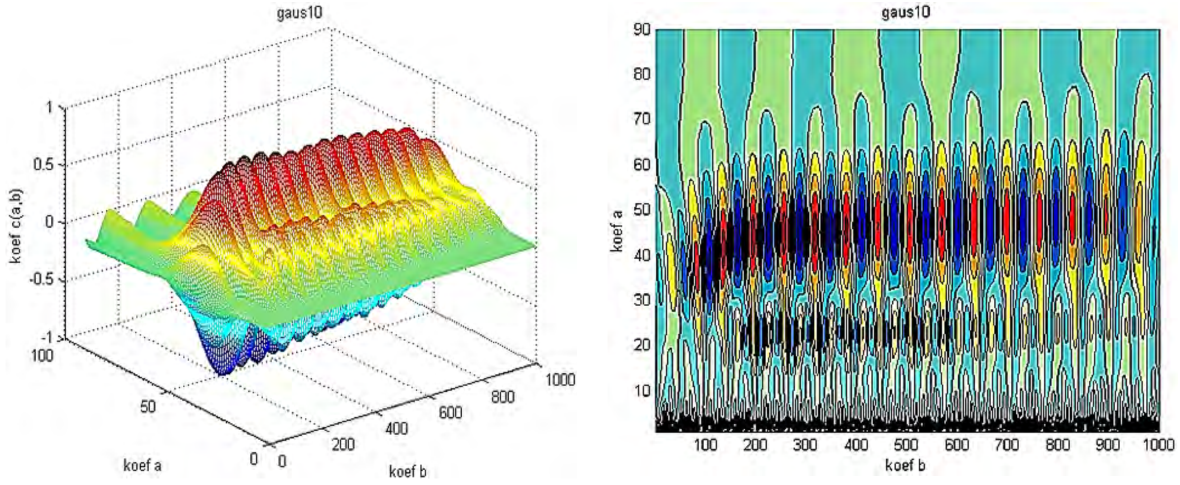


b) c)

Figure 2: b) Three-dimensional wavelet diagram of the phoneme "a", c) contour representation of the wavelet diagram of the phoneme "a"

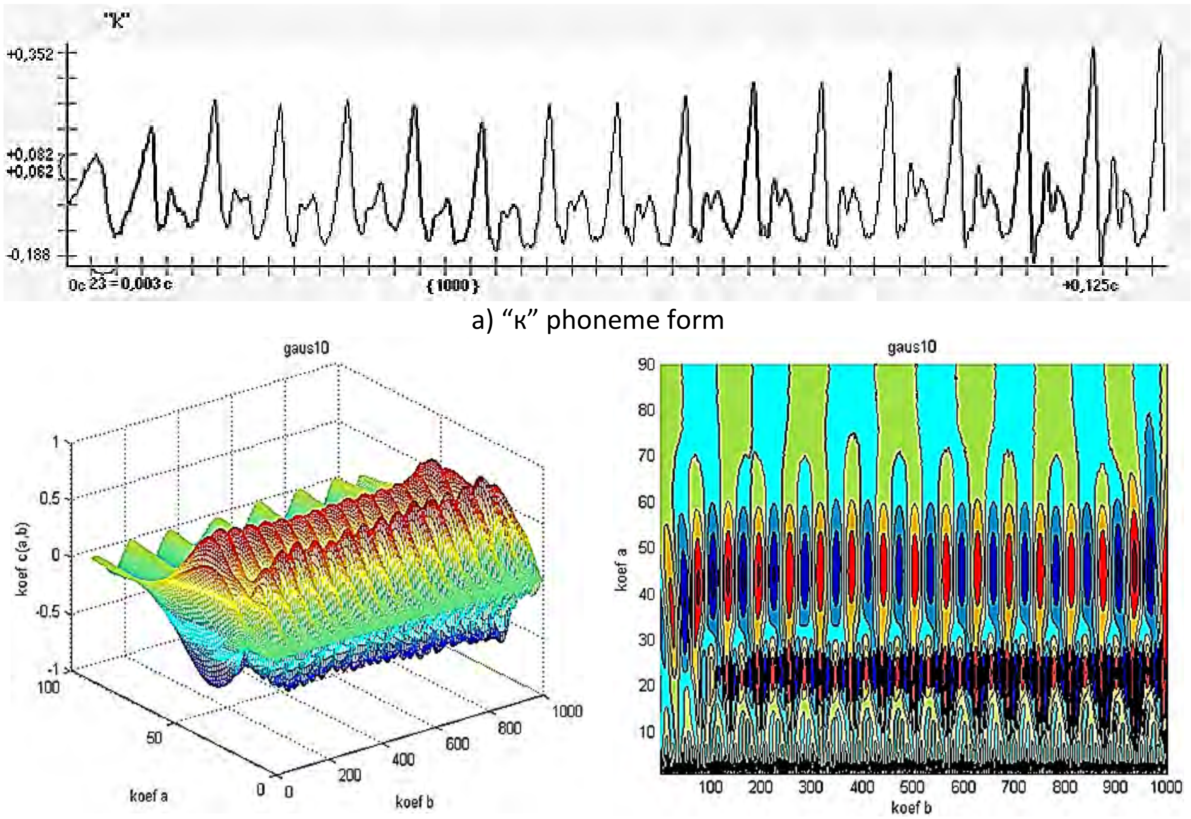


a) “б” phoneme form



b) c)

Figure 3: b) Three-dimensional wavelet diagram of the phoneme "б", c) Contour representation of the wavelet diagram of the phoneme "б"



a) “к” phoneme form

b) c)

Figure 4: b) Three-dimensional wavelet diagram of the phoneme "к", c) Contour representation of the wavelet diagram of the phoneme "к"

The wavelet diagram illustrates the localization of the sound "a" - it has a large number of high-frequency components, which, however, are masked by the significant amplitude of the main tone signal [27].

Comparison of Fig. 3 and 4 allows you to compare and distinguish the sounds "б" and "к". Although at first glance the contour diagrams of the sounds "б" and "к" seem identical, it can be seen that in the sound "к" there are two dominant components of the same magnitude, and in the sound "б" there is only one.

For better recognition of short sounds corresponding to a high pace of speech and greater visibility, wavelet diagrams of phonemes with a duration of 25 milliseconds are presented below (Figs. 5-7). The wavelet diagrams of these sounds have preserved all the features of the previous diagrams in Fig. 3 – 5, which allows us to conclude about the effectiveness of the wavelet analysis even of short phoneme recordings.

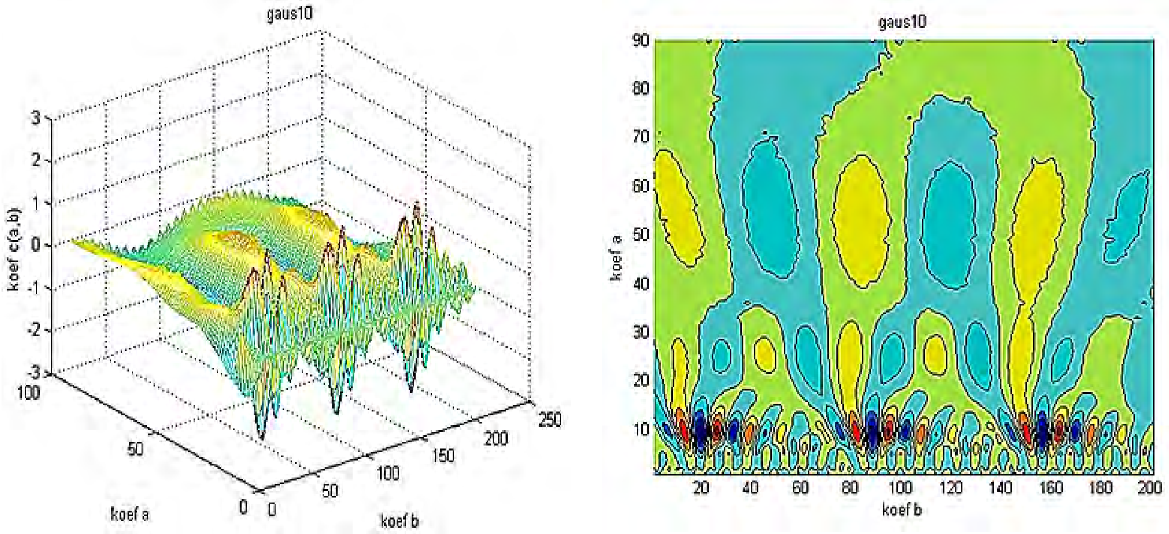


Figure 5: Wavelet diagrams for the 25 ms duration "a" phoneme

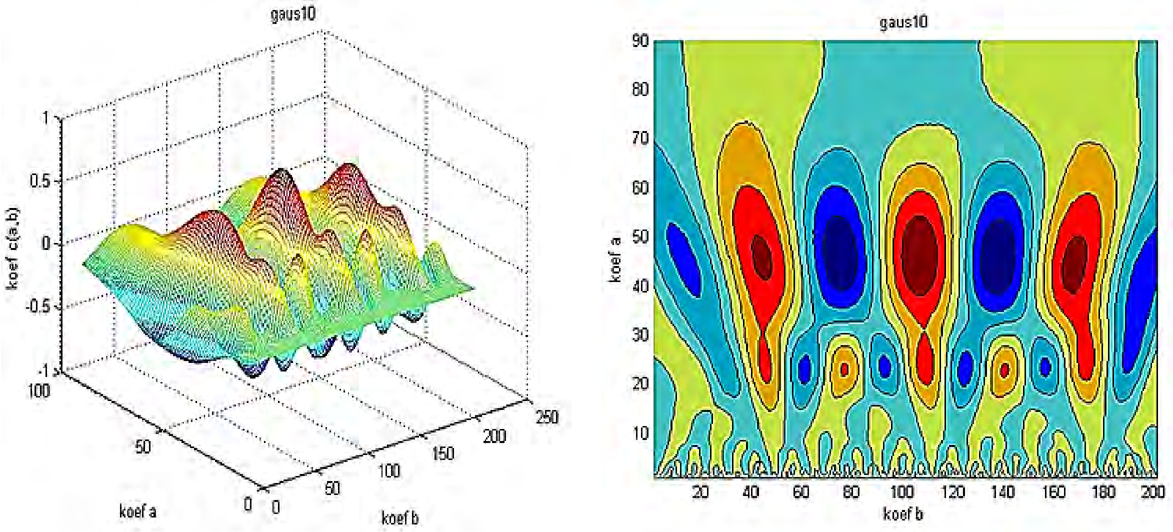


Figure 6: Wavelet diagrams for the 25 ms duration "6" phoneme

4. Conclusions and discussion

The article describes the technique of using wavelet transformation to describe the phonemes of the Ukrainian language corresponding to vocalized, plosive and voiceless sounds. The obtained three-dimensional wavelet diagrams well illustrate which frequency components prevail in the speech signal and at what point in time this occurs. The periodicity of individual components at different levels of wavelet diagrams for the same sounds is clearly visible. It is shown that when the duration of phonemes is changed (the pace of pronunciation is changed), the wavelet diagrams of these sounds retain all their features, which allows us to draw a conclusion about the effectiveness of wavelet analysis even of short recordings of the speech signal.

The considered methods allow conducting research to determine the minimum duration of phonemes for the purpose of their reliable recognition, which, with a given probability of correct recognition, will allow building a sufficiently fast speech signal analysis-synthesis algorithm. Of course, in order to make a final conclusion about the quality of using the wavelet transformation to recognize the phonemes of the Ukrainian language, it is necessary to use more statistics for male and female voices. Note the speed of this method and the high probability of correct recognition of individual sounds. It can be said that this method has a number of prospects, among which one should highlight such as a huge number of types of wavelets that can complement each other, for a more accurate determination of the patterns and features of the speech signal.

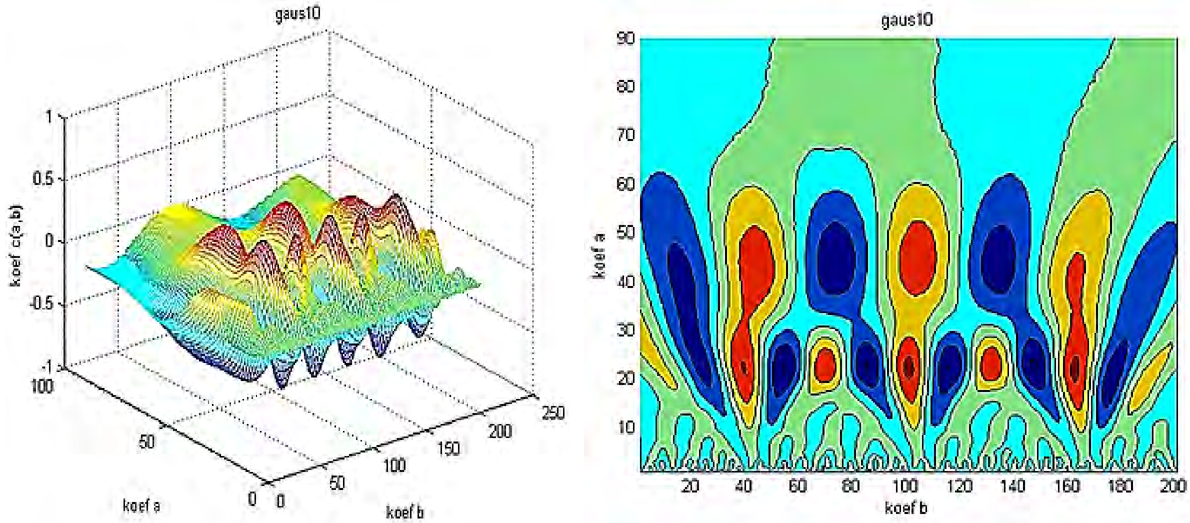


Figure 7: Wavelet diagrams for the 25 ms duration "к" phoneme

5. Acknowledgements

The authors are appreciative of colleagues for their support and appropriate suggestions, which allowed to improve the materials of the article.

6. References

[1] L. Sun, "Using End-to-end Multitask Model for Simultaneous Language Identification and Phoneme Recognition," 2022 16th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2022, pp. 46-50, doi: 10.1109/ICSP56322.2022.9965220.

[2] L. Sun, "Spoken Language Identification with Deep Temporal Neural Network and Multi-levels Discriminative Cues," 2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP), Shanghai, China, 2020, pp. 153-157, doi: 10.1109/ICICSP50920.2020.9232093.

[3] B. Durnyak, B. Havrysh, O. Tymchenko, M. Zelyanovsky, O. O. Tymchenko and O. Khamula, "Intelligent System for Sensor Wireless Network Access: Modeling Methods of Network Construction," 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Lviv, Ukraine, 2018, pp. 93-97, doi: 10.1109/IDAACS-SWS.2018.8525792.

[4] S. Hara and H. Nishizaki, "Acoustic modeling with a shared phoneme set for multilingual speech recognition without code-switching," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, 2017, pp. 1617-1620, doi: 10.1109/APSIPA.2017.8282284.

[5] Z. Wang et al., "Building Robust Spoken Language Understanding by Cross Attention Between Phoneme Sequence and ASR Hypothesis," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 7147-7151, doi: 10.1109/ICASSP43922.2022.9747198.

[6] S. Zhang, J. Yi, Z. Tian, Y. Bai, J. Tao and Z. wen, "Decoupling Pronunciation and Language for End-to-End Code-Switching Automatic Speech Recognition," ICASSP 2021 - 2021 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 6249-6253, doi: 10.1109/ICASSP39728.2021.9414428.

[7] J. Zhao and W. -Q. Zhang, "Improving Automatic Speech Recognition Performance for Low-Resource Languages With Self-Supervised Models," in IEEE Journal of Selected Topics in Signal Processing, vol. 16, no. 6, pp. 1227-1241, Oct. 2022, doi: 10.1109/JSTSP.2022.3184480.

[8] Tymchenko, O.O., Havrysh, B., Khamula, O., Lysenko, S., & Havrysh, K. (2020). Risks of Loss of Personal Data in the Process of Sending and Printing Documents. CITRisk, CEUR Workshop ProceedingsVolume 2805, Pages 373 - 3842020 1st International Workshop on Computational and Information Technologies for Risk-Informed Systems, CITRisk 2020Virtual, Kherson15 October 2020 through 16 October 2020Code 166724

[9] M. Yu et al., "Multilingual Grapheme-To-Phoneme Conversion with Byte Representation," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 8234-8238, doi: 10.1109/ICASSP40776.2020.9054696.

[10] X. Liu et al., "Leveraging Pre-trained BERT for Audio Captioning," 2022 30th European Signal Processing Conference (EUSIPCO), Belgrade, Serbia, 2022, pp. 1145-1149, doi: 10.23919/EUSIPCO55093.2022.9909761.

[11] Ivan Izonin, Roman Tkachenko, Michal Gregus, Khrystyna Zub, Nataliia Lotoshynska "Input Doubling Method based on SVR with RBF kernel in Clinical Practice: Focus on Small Data", Procedia Computer Science, vol. 184, 2021,pp. 606-613. https://doi.org/10.1016/j.procs.2021.03.075.

[12] B. Zhang, S. Khorram and E. M. Provost, "Exploiting Acoustic and Lexical Properties of Phonemes to Recognize Valence from Speech," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019, pp. 5871-5875, doi: 10.1109/ICASSP.2019.8683190.

[13] I. -T. Hsieh, C. -H. Wu and Z. -H. Zhao, "Selection of Supplementary Acoustic Data for Meta-Learning in Under-Resourced Speech Recognition," 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 2022, pp. 409-414, doi: 10.23919/APSIPAASC55919.2022.9979997.

[14] Kovtun V., Kovtun O., Semenov A. "Entropy-Argumentative Concept of Computational Phonetic Analysis of Speech Taking into Account Dialect and Individuality of Phonation," Entropy, vol. 24, no. 7, 2022; 1006. https://doi.org/10.3390/e24071006.

[15] B. Durnyak, O. Tymchenko, O. Tymchenko and B. Havrysh, "Applying the Neuronetchic Methodology to Text Images for Their Recognition," 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 2018, pp. 584-589, doi: 10.1109/DSMP.2018.8478482.

[16] S. Khanal, M. T. Johnson and N. Bozorg, "Articulatory Comparison of L1 and L2 Speech for Mispronunciation Diagnosis," 2021 IEEE Spoken Language Technology Workshop (SLT), Shenzhen, China, 2021, pp. 693-697, doi: 10.1109/SLT48900.2021.9383574.

[17] P. Gupta and H. Patil, "Effect of Speaker-Microphone Proximity on Pop Noise: Continuous Wavelet Transform-Based Approach," 2022 13th International Symposium on Chinese Spoken Language Processing (ISCSLP), Singapore, Singapore, 2022, pp. 110-114, doi: 10.1109/ISCSLP57327.2022.10038157.

[18] S. Ding, G. Zhao, C. Liberatore and R. Gutierrez-Osuna, "Learning Structured Sparse Representations for Voice Conversion," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 343-354, 2020, doi: 10.1109/TASLP.2019.2955289.

[19] T. Fujimoto, S. Takaki, K. Hashimoto, K. Oura, Y. Nankaku and K. Tokuda, "Semi-Supervised Learning Based on Hierarchical Generative Models for End-to-End Speech Synthesis," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 2020, pp. 7644-7648, doi: 10.1109/ICASSP40776.2020.9054466.

[20] R. Ranjan, A. Thakur and Y. Narayan, "Acoustic Feature Extraction and Isolated Word Recognition of Speech Signal Using HMM for Different Dialects," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9848374.

[21] F. Zhang, M. Tu, S. Liu and J. Yan, "ASR Error Correction with Dual-Channel Self-Supervised Learning," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal

Processing (ICASSP), Singapore, Singapore, 2022, pp. 7282-7286, doi: 10.1109/ICASSP43922.2022.9746763.

[22] D. Li, T. I, N. Arivazhagan, C. Cherry and D. Padfield, "Sentence Boundary Augmentation for Neural Machine Translation Robustness," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 7553-7557, doi: 10.1109/ICASSP39728.2021.9413492.

[23] D. Le, X. Zhang, W. Zheng, C. Fügen, G. Zweig and M. L. Seltzer, "From Senones to Chenones: Tied Context-Dependent Graphemes for Hybrid Speech Recognition," 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Singapore, 2019, pp. 457-464, doi: 10.1109/ASRU46091.2019.9003972.

[24] D. D. Tannus, D. D. G. B. Cruz and O. A. Z. Sotomayor, "Output-only Based Identification of Modal Parameters of Linear and Nonlinear Structures by Wavelet Transform," in IEEE Latin America Transactions, vol. 19, no. 01, pp. 124-131, January 2021, doi: 10.1109/TLA.2021.9423855.

[25] M. Sakthi, M. Desai, L. Hamilton and A. Tewfik, Keyword-spotting and speech onset detection in EEG-based Brain Computer Interfaces," 2021 10th International IEEE/EMBS Conference on Neural Engineering (NER), Italy, 2021, pp. 519-522, doi: 10.1109/NER49283.2021.9441118.

[26] J. Yu, K. Markov and T. Matsui, "Articulatory and Spectrum Information Fusion Based on Deep Recurrent Neural Networks," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 4, pp. 742-752, April 2019, doi: 10.1109/TASLP.2019.2894554.

[27] S. Sajikumar and S. Vinod, "Ramanujan Sums-Wavelet Transform: A New Approach to Signal Processing," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, pp. 179-183, doi: 10.1109/ICICICT54557.2022.9917837.

[28] A. Kobayashi, "Neural Networks Using Multiplicative Features Based on Second-Order Statistics for Acoustic and Speech Applications," 2022 Asia Conference on Advanced Robotics, Automation, and Control Engineering (ARACE), Qingdao, China, 2022, pp. 121-126, doi: 10.1109/ARACE56528.2022.00029.

[29] B. Durnyak, B. Havrysh, O. Tymchenko, O. Tymchenko and D. Anastasiya, "Research of image processing methods in publishing output systems," 2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH), Lviv, Ukraine, 2018, pp. 178-181, doi: 10.1109/MEMSTECH.2018.8365728.

[30] S. Sankar, D. Beautemps and T. Hueber, "Multistream Neural Architectures for Cued Speech Recognition Using a Pre-Trained Visual Feature Extractor and Constrained CTC Decoding," ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022, pp. 8477-8481, doi: 10.1109/ICASSP43922.2022.9746976.