

AN APPROACH TO ISOLATED WORD AND PHONEME-LEVEL SPEECH RECOGNITION IN THE AZERBAIJANI LANGUAGE

Elchin R. Ismayilov*

*Institute of Control Systems of the Ministry of Science and Education of the Republic of Azerbaijan
Received 17 July 2025; accepted 22 August 2025*

<https://doi.org/10.30546/209501.102.2025.2.3.524>

Abstract

This article presents an empirical approach for isolated word and phoneme-level speech recognition in the Azerbaijani language. The method integrates three complementary techniques: Dynamic Time Warping (DTW) and spectral analysis based on Fourier (FT) and wavelet (WT) transforms applied to continuous speech signals. Each technique captures distinct temporal and frequency-domain features, enabling a comprehensive representation of phonetic and lexical information. Recognition is performed using an algorithm that evaluates informative features and selects an appropriate similarity measure for one-dimensional signals, coordinating time-domain alignment with multiresolution spectral features. The framework is particularly suited for languages with limited annotated speech resources, supporting effective recognition at both phoneme and word levels. Experimental results demonstrate that the combined use of DTW with Fourier- and wavelet-based features improves recognition accuracy over individual methods. These findings confirm the feasibility of the proposed approach for Azerbaijani speech recognition and highlight its potential applicability to other low-resource languages and speech processing tasks.

Keywords: speech phoneme, speech recognition, digital signal, signal proximity measure, recognition method

Mathematics Subject Classification (2020): 68T10, 68T50, 94A12, 65T60

* Corresponding author.

E-mail address: elchinismayilov@gmail.com (Fidan Isgandarli)

1. Introduction

Owing to extensive empirical experience, traditional speech recognition methods, including those based on spectral analysis, have proven to be effective tools for recognizing acoustic signals, particularly isolated words and phonemes pronounced by native speakers with individual pronunciation characteristics and varying tonalities. Classical approaches to speech analysis rely on time–frequency representations of speech signals, such as the Fourier transform, which enables the extraction of global spectral features widely used in early and contemporary recognition systems [1, 2].

A phoneme (ancient Greek φώνημα “sound”) is the smallest distinctive unit of a language. Although a phoneme does not possess independent lexical or grammatical meaning, it serves to distinguish and identify meaningful linguistic units such as morphemes and words. Accurate phoneme recognition remains a fundamental task in automatic speech recognition, as phoneme-level modeling directly influences word-level recognition performance [3].

Of particular interest is the study of the factors and characteristics that influence recorded speech signals. The direct processing of speech recordings in the form of one-dimensional representations of sound wave oscillations with a given sampling frequency enables the recognition of speech signals while accounting for pronunciation variability and tonal features using classical recognition techniques. In this context, Dynamic Time Warping (DTW) has been extensively applied to align speech signals with different temporal structures and speaking rates, demonstrating robust performance in isolated word and phoneme recognition tasks [4, 5].

In addition to time-domain alignment methods, multiresolution spectral analysis techniques, such as the wavelet transform, have gained increasing attention in speech processing. Unlike the Fourier transform, wavelet-based approaches provide simultaneous localization in both time and frequency domains, allowing the analysis of nonstationary speech signals and transient phonetic events [6, 7]. These properties make wavelet-based features particularly suitable for phoneme recognition and for capturing fine-grained speech dynamics.

Advances in modern linguistics, computer science, and mathematical statistics have significantly contributed to the development and improvement of speech

recognition algorithms aimed at solving applied problems. Today, there is a growing demand for speech recognition systems that ensure a high degree of correspondence between recognition results and the words and phonemes actually pronounced by native speakers. Such systems must be capable of identifying and compensating for pronunciation variability, including phonetic deviations and letter-by-letter articulation effects, across different tonalities and speaking styles.

The objective of this study is to develop an algorithmic approach for the automatic recognition of words and phonemes in the Azerbaijani language. The proposed approach integrates classical spectral analysis methods with DTW and wavelet-based feature extraction, forming a unified framework for phoneme- and word-level speech recognition that accounts for the specific phonetic and prosodic characteristics of Azerbaijani speech.

2. Problem definition

It is assumed that a time sequence of speech signal samples SP_k ($k=1\div n$), taken with a certain sampling interval Δt , is fed to the input of a hypothetical recognition system. As an example, two words “book, notebook” are chosen, sequentially pronounced by a speaker in the Azerbaijani language. Through a sound device, these spoken words are converted into analogy signals. Then, by quantization, these signals are transformed into the corresponding digital signals, which are shown in Fig. 1.

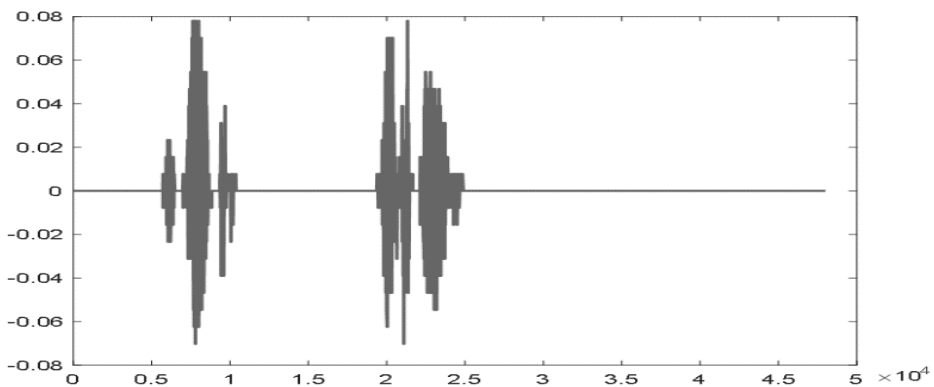


Fig. 1. A digital signal representing the phrase “book, notebook”

To create a recognition algorithm, a priori information about the input words in the form of an audio database is required. As a rule, the standards of the words as structural speech units are specified in the form of classified training samples. In this regard, the audio database of three words is considered [8]: “book”, “notebook” and “pencil”, reproduced by a speaker in the Azerbaijani language as [k'it'a: b], [dæf'tər] and [gɟæ'læm]. The corresponding digital signals are shown in Fig. 2.

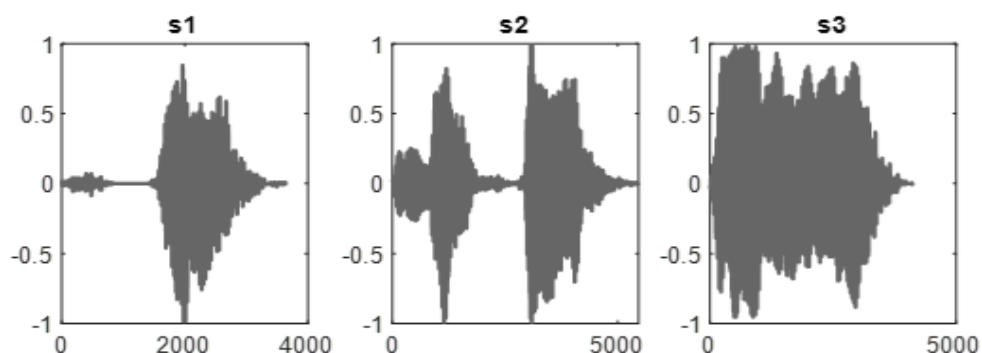


Fig. 2. Audio database consisting of digital signals s_1 , s_2 and s_3

An approach is required to identify a mechanism that processes voiced Azerbaijani speech and determines the assignment of words, viewed as indivisible structural units, to given reference standards with an acceptable recognition accuracy.

3. An iterative approach to speech phoneme recognition

To compare structural speech units with standards, word segmentation is performed, which at the recognition stage allows eliminating redundant decision-making procedures for signals that do not carry speech information. Signal segmentation ensures the division of speech into indivisible structural units and the assessment of their time boundaries. In particular, segmentation of the digital signal from Fig. 1 divides it into two separate words, “book” and “notebook”, the digital signals of which are presented in Fig. 3 as s_4 and s_5 .

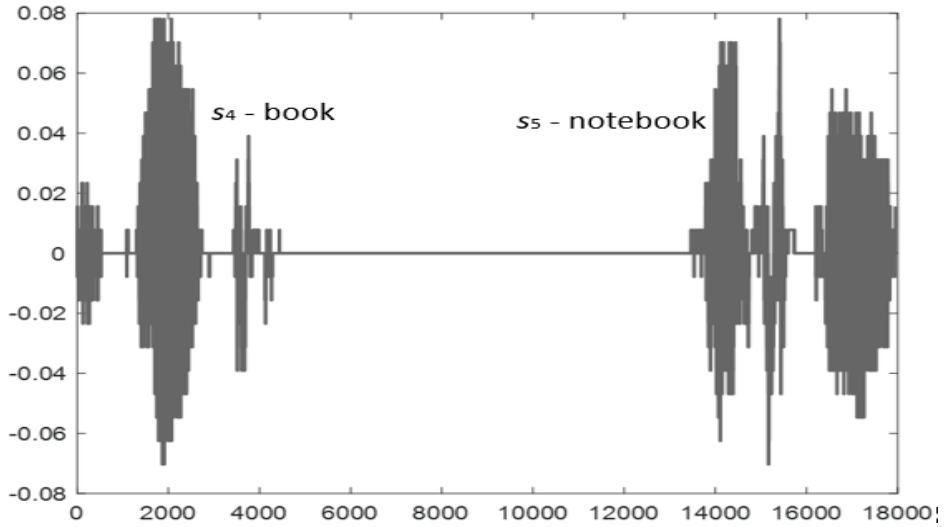


Fig. 3. Segmentation of a digital signal reflecting the phrase “book, notebook”

To perform the pairwise comparison procedure using the selected recognition methods and a chosen similarity measure for one-dimensional signals, a set of discrete signals $S=\{s_1, s_2, s_3, s_4, s_5\}$ is considered, as illustrated in Fig. 4.

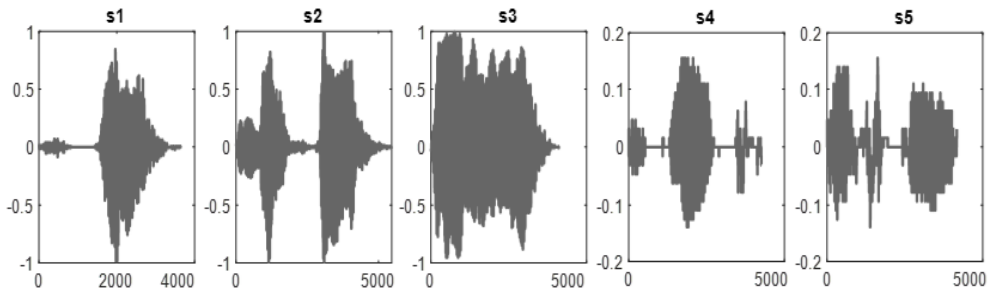


Fig. 4. A set of discrete signals S

For the pairwise comparison of one-dimensional signals from the set S , the DTW method is employed, along with recognition methods based on the Fourier transform and the wavelet transform.

The DTW method is based on an algorithm that is used to compare two numerical sequences (or time series) that are not necessarily synchronized in time in order to find the optimal match between them [9]. For example, for numerical

sequences $\{f_1, f_2, \dots, f_n\}$ and $\{g_1, g_2, \dots, g_m\}$ with lengths n and m , respectively, local deviations between the components of these sequences are first calculated in absolute value using, for example, the Euclidean metric. Further, a matrix of size $n \times m$ is formed, consisting of squared deviations of the form $d_{ij} = (f_i - g_j)^2$, $i=1 \div n$, $j=1 \div m$, and the minimum distance $DTW(f_i, g_j)$ is calculated using the following equalities:

$$\begin{cases} DTW(f_i, g_j)^2 = d_{ij} + \min\{DTW(f_i, g_{(j-1)})^2, DTW(f_{(i-1)}, g_j)^2, \\ DTW(f_{(i-1)}, g_{(j-1)})^2\}; \\ DTW(f_1, g_1)^2 = d_{11}. \end{cases}$$

By iteratively calculating the $DTW(f_i, g_j)$ indicator, the final distance norm is established as $D_2 = \sqrt{DTW(f_n, g_m)}$, which determines the correspondence between sequences.

The application of the DTW method implies the fulfillment of the following conditions:

- monotonicity – both indices i and j increase sequentially.
- continuity – in one step the indices i and j increase by no more than unit.
- the sequential construction of “paths” starts in the lower left and ends in the upper right corner.

It should be noted that the DTW algorithm is applied with “limitation” and “without limitation” on the size of the so-called “window” w , which determines the number of allowed samples for comparing the components of the sequences both on the right and on the left. In this case, the total number of samples is $2w+1$, and the procedure for comparing the sequences f_i and g_j by the i -th sample point at the first sequence and the j -th sample point at the second sequence must be performed taking into account the inequality $|i - j| \leq w$.

The Fourier transform (FT) is the mathematical basis of spectral analysis as the main method of signal processing. FT connects a spatial or temporal signal (or some model of this signal) with its representation in the frequency domain [10]. In other words, the Fourier transform of a real-valued function defined on the time axis of the variable t , as an integral representation provides information only about the frequency that is present in the signal and does not provide any

information about the time interval in which this frequency is present in the signal. The usual Fourier transform, by its nature, cannot distinguish a stationary signal from a non-stationary one, which is a big problem for its applicability.

Therefore, further the following windowed Fourier transform is used

$$F(t, w) = \int_{-\infty}^{+\infty} f(t)W(\tau - t)e^{-i\tau w} d\tau,$$

where $W(\tau - t)$ is the so-called window function, which can be a Gaussian, Hamming, Hann or Kaiser window. Unlike the usual Fourier transform, the windowed Fourier transform is already a function of time, frequency and amplitude. That is, it allows one to obtain the characteristic of the signal frequency distribution (with amplitude) in time. Thus, it is considered that the usual Fourier transform is a windowed Fourier transform with a window of infinity. As the window width increases (its resolution decreases), the accuracy relative to frequency increases, but the accuracy relative to time decreases. The question arises: how to select the window width value to achieve the optimal ratio of accuracies? The wavelet transform answers this question.

The method based on the wavelet transform (WT) was created as a tool that solves the Heisenberg uncertainty problem for constructing the time-frequency characteristics of a signal [11, 12]. Unlike the windowed Fourier transform, which has a constant scale at any time for all frequencies, WT has a better time representation and a worse frequency representation at low signal frequencies, and a better frequency representation with a worse time representation at high signal frequencies.

In the discrete case, wavelets are represented by samples. The continuous WT (CWT) maps a real-valued function defined on the time axis of the variable t into the following function of two variables

$$\gamma(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t) \gamma^* \left(\frac{t - \tau}{s} \right) dt,$$

where τ denotes parallel translation and s represents scale.

Further, in the process of voice signal processing, which involves pairwise comparison of digital signals from the set S at sample points, the DWT method and recognition methods based on the FT and CWT are used. Tables 1, 2 and 3

show the results of pairwise comparison of discrete signals from the set S .

Table 1. Pairwise comparison of signals from S using the DWT

	s_1	s_2	s_3	s_4	s_5
s_1	0	0.000199	0.002048	0.000546	0.000543
s_2	0.000199	0	0.001861	0.000721	0.000718
s_3	0.002048	0.001861	0	0.002581	0.002579
s_4	0.000546	0.000721	0.002581	0	0.000003
s_5	0.000543	0.000718	0.002579	0.000003	0

Table 2. Pairwise comparison of signals from S using the FT

	s_1	s_2	s_3	s_4	s_5
s_1	0	0.01571	0.09910	0.0330	0.0329
s_2	0.01571	0	0.08338	0.0488	0.0487
s_3	0.09910	0.08338	0	0.1321	0.1320
s_4	0.03305	0.04876	0.13215	0	0.0001
s_5	0.03294	0.04865	0.1320	0.0001	0

Table 3. Pairwise comparison of signals from S using the CWT

	s_1	s_2	s_3	s_4	s_5
s_1	0	1.35579	7.00774	1.029	1.0193
s_2	1.35579	0	5.70845	2.3294	2.3188
s_3	7.00774	5.70845	0	7.9978	7.9875
s_4	1.02879	2.32939	7.99781	0	0.0110
s_5	1.01932	2.31882	7.9875	0.0110	0

It can be observed that the results of the pairwise comparison of signals, presented in Tables 1, 2, and 3 (highlighted in grayscale), are insufficient for accurately identifying the structural units of the phoneme corresponding to “book” and “notebook.” Therefore, in the next iteration, the digital signals representing these structural units are processed with amplitudes scaled by a factor of 2 (2A), as illustrated in Fig. 5.

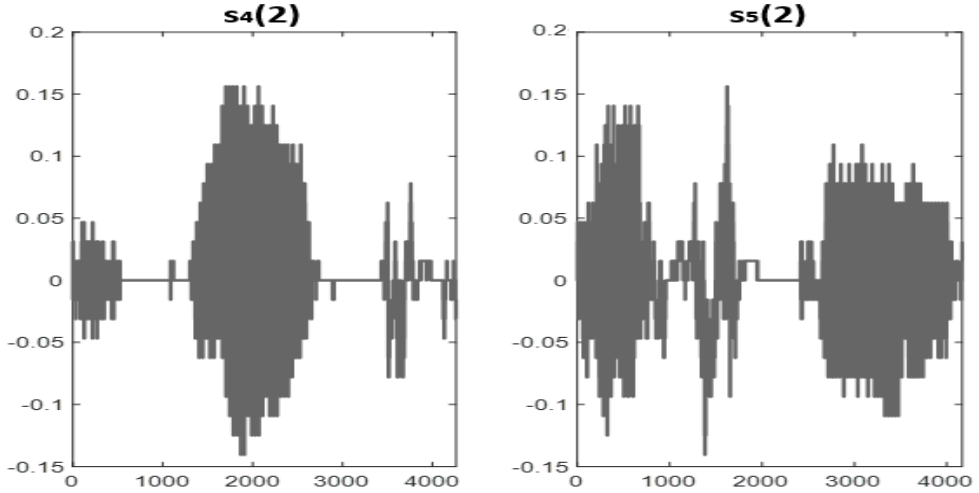


Fig. 5. Digital signals $s_4(2)$ and $s_5(2)$ with doubled 2A amplitudes

Tables 4, 5 and 6 show the results of pairwise comparison of discrete signals from the new set $S(2) = \{s_1, s_2, s_3, s_4(2), s_5(2)\}$.

Table 4. Pairwise comparison of signals from $S(2)$ using the DWT

	s_1	s_2	s_3	$s_4(2)$	$s_5(2)$
s_1	0	0.000199	0.002048	0.000540	0.000531
s_2	0.000199	0	0.001861	0.000715	0.000705
s_3	0.002048	0.001861	0	0.002576	0.002566
$s_4(2)$	0.000540	0.000715	0.002576	0	0.000010
$s_5(2)$	0.000531	0.000705	0.002566	0.000010	0

Table 5. Pairwise comparison of signals from $S(2)$ using the FT

	s_1	s_2	s_3	$s_4(2)$	$s_5(2)$
s_1	0	0.015715	0.099100	0.032373	0.031927
s_2	0.015715	0	0.083385	0.048088	0.047642
s_3	0.099100	0.083385	0	0.131472	0.131027
$s_4(2)$	0.032373	0.048088	0.131472	0	0.000445
$s_5(2)$	0.031927	0.047642	0.131027	0.000445	0

Table 6. Pairwise comparison of signals from $S(2)$ using the CWT

	s_1	s_2	s_3	$s_4(2)$	$s_5(2)$
s_1	0	1.355790	7.007736	1.015443	0.977949
s_2	1.355790	0	5.708453	2.314678	2.272457
s_3	7.007736	5.708453	0	7.984484	7.943390
$s_4(2)$	1.015443	2.314678	7.984484	0	0.044116
$s_5(2)$	0.977949	2.272457	7.943390	0.044116	0

The results of the pairwise comparison of signals remain unsatisfactory for identifying the structural units of the phonemes corresponding to “book” and “notebook” (see cells highlighted in grayscale). Therefore, the procedure proceeds to the next iteration, in which the amplitudes of digital signals s_4 and s_5 are increased by a factor of 4. The resulting signals, with amplitudes scaled to $4A$, are shown in Fig. 6.

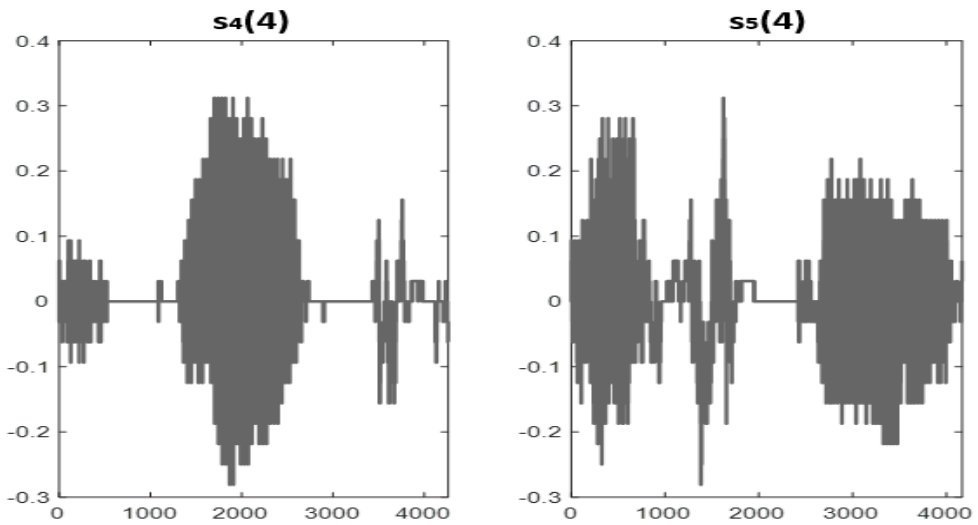


Fig. 6. Digital signals $s_4(4)$ and $s_5(4)$ with $4A$ amplitudes

Tables 7 and 8 show the results of pairwise comparison of discrete signals from the new set $S(4) = \{s_1, s_2, s_3, s_4(4), s_5(4)\}$ using the DWT- and FT-based recognition methods.

Table 7. Pairwise comparison of signals from $S(4)$ using the DWT

	s_1	s_2	s_3	$s_4(4)$	$s_5(4)$
s_1	0	0.000199	0.002048	0.000518	0.000480
s_2	0.000199	0	0.001861	0.000695	0.000654
s_3	0.002048	0.001861	0	0.002555	0.002514
$s_4(4)$	0.000518	0.000695	0.002555	0	0.000042
$s_5(4)$	0.000480	0.000654	0.002514	0.000042	0

Table 8. Pairwise comparison of signals from $S(4)$ using the FT

	s_1	s_2	s_3	$s_4(4)$	$s_5(4)$
s_1	0	0.015715	0.099100	0.029665	0.027884
s_2	0.015715	0	0.083385	0.045380	0.043599
s_3	0.099100	0.083385	0	0.128765	0.126984
$s_4(4)$	0.029665	0.045380	0.128765	0	0.001781
$s_5(4)$	0.027884	0.043599	0.126984	0.001781	0

Observing that these results are still unsatisfactory (see cells highlighted in grayscale), we proceed to the next iteration, in which the amplitudes of digital signals s_4 and s_5 are increased by a factor of 10. The corresponding signals, with amplitudes scaled to 10A, are presented in Fig. 7.

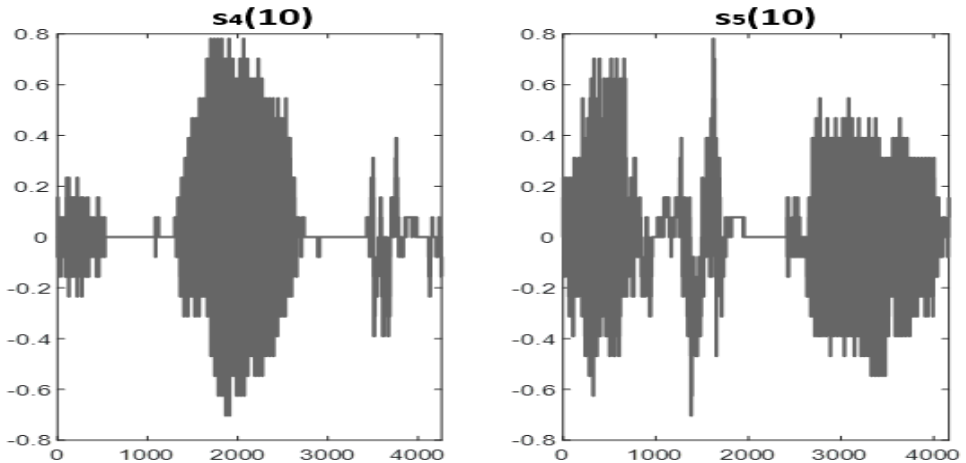


Fig. 7. Digital signals $s_4(10)$ and $s_5(10)$ with 10A amplitudes

Tables 9, 10 and 11 show the results of pairwise comparison of discrete signals from the new set $S(10) = \{s_1, s_2, s_3, s_4(10), s_5(10)\}$.

Table 9. Pairwise comparison of signals from $S(10)$ using the DWT

	s_1	s_2	s_3	$s_4(10)$	$s_5(10)$
s_1	0	0.000199	0.002048	0.000367	0.000138
s_2	0.000199	0	0.001861	0.000552	0.000294
s_3	0.002048	0.001861	0	0.002412	0.002155
$s_4(10)$	0.000367	0.000552	0.002412	0	0.000261
$s_5(10)$	0.000138	0.000294	0.002155	0.000261	0

Table 10. Pairwise comparison of signals from $S(10)$ using the FT

	s_1	s_2	s_3	$s_4(10)$	$s_5(10)$
s_1	0	0.015715	0.099100	0.010712	0.000417
s_2	0.015715	0	0.083385	0.026427	0.015298
s_3	0.099100	0.083385	0	0.109812	0.098682
$s_4(10)$	0.010712	0.026427	0.109812	0	0.011130
$s_5(10)$	0.000417	0.015298	0.098682	0.011130	0

Table 11. Pairwise comparison of signals from $S(10)$ using the CWT

	s_1	s_2	s_3	$s_4(10)$	$s_5(10)$
s_1	0	1.355790	7.007736	0.688563	0.923746
s_2	1.355790	0	5.708453	1.867485	0.945183
s_3	7.007736	5.708453	0	7.566563	6.567884
$s_4(10)$	0.688563	1.867485	7.566563	0	1.102899
$s_5(10)$	0.923746	0.945183	6.567884	1.102899	0

These results remain unsatisfactory. Therefore, the procedure advances to the next iteration, in which the amplitudes of digital signals s_4 and s_5 are increased by a factor of 15. The corresponding signals, with amplitudes scaled to 15A, are presented in Fig. 8.

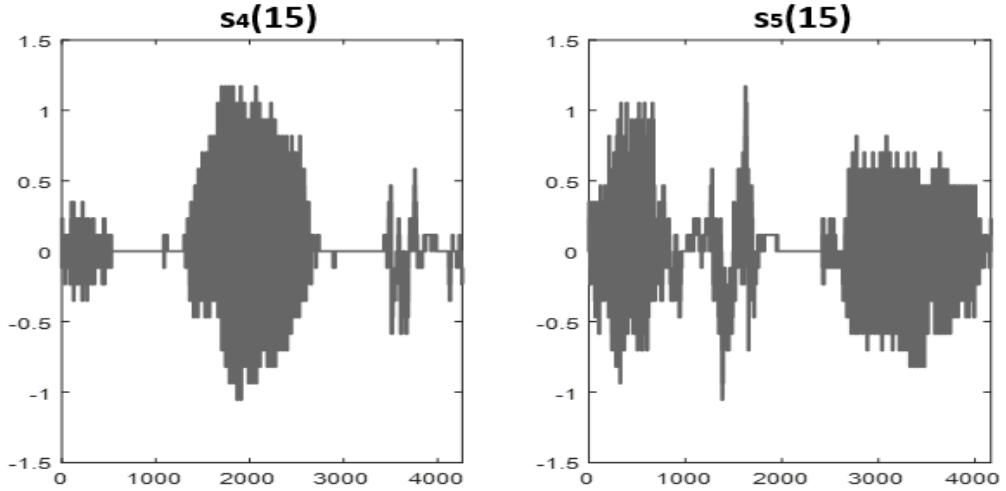


Fig. 8. Digital signals $s_4(15)$ and $s_5(15)$ with 15A amplitudes

Tables 12, 13 and 14 show the corresponding results of pairwise comparison of discrete signals from the new set $S(15) = \{s_1, s_2, s_3, s_4(15), s_5(15)\}$.

Table 12. Pairwise comparison of signals from $S(15)$ using the DWT

	s_1	s_2	s_3	$s_4(15)$	$s_5(15)$
s_1	0	0.000199	0.002048	0.000161	0.000429
s_2	0.000199	0	0.001861	0.000360	0.000242
s_3	0.002048	0.001861	0	0.002202	0.001621
$s_4(15)$	0.000161	0.000360	0.002202	0	0.000586
$s_5(15)$	0.000429	0.000242	0.001621	0.000586	0

Table 13. Pairwise comparison of signals from $S(15)$ using the FT

	s_1	s_2	s_3	$s_4(15)$	$s_5(15)$
s_1	0	0.015715	0.099100	0.017491	0.042533
s_2	0.015715	0	0.083385	0.001776	0.026818
s_3	0.099100	0.083385	0	0.081608	0.056567
$s_4(15)$	0.017491	0.001776	0.081608	0	0.025042
$s_5(15)$	0.042533	0.026818	0.056567	0.025042	0

Table 14. Pairwise comparison of signals from $S(15)$ using the CWT

	s_1	s_2	s_3	$s_4(15)$	$s_5(15)$
s_1	0	1.355790	7.007736	0.816888	2.847010
s_2	1.355790	0	5.708453	1.403835	1.648926
s_3	7.007736	5.708453	0	7.050458	4.953125
$s_4(15)$	0.816888	1.403835	7.050458	0	2.481524
$s_5(15)$	2.847010	1.648926	4.953125	2.481524	0

The results of the pairwise comparison at this iteration are sufficiently satisfactory, allowing for the correct identification of the structural units of the phonemes corresponding to “book” and “notebook” (see cells highlighted in grayscale). The calculation results presented in Tables 15, 16, and 17 indicate that a further increase in amplitude, for example by a factor of 20 (see Fig. 9), leads to a deterioration in recognition accuracy (see cells highlighted in grayscale). These findings suggest that, for the phonemes “book” and “notebook,” amplitude values of 15A for the digital representations of their structural units are optimal for accurately identifying the corresponding signals.

Table 15. Pairwise comparison of signals from $S(20)$ using the DWT

	s_1	s_2	s_3	$s_4(20)$	$s_5(20)$
s_1	0	0.000199	0.002048	0.000235	0.001174
s_2	0.000199	0	0.001861	0.000258	0.000991
s_3	0.002048	0.001861	0	0.001917	0.000877
$s_4(20)$	0.000235	0.000258	0.001917	0	0.001043
$s_5(20)$	0.001174	0.000991	0.000877	0.001043	0

Table 16. Pairwise comparison of signals from $S(20)$ using the FT

	s_1	s_2	s_3	$s_4(20)$	$s_5(20)$
s_1	0	0.015715	0.099100	0.056976	0.101495
s_2	0.015715	0	0.083385	0.041261	0.085780
s_3	0.099100	0.083385	0	0.042124	0.002395
$s_4(20)$	0.056976	0.041261	0.042124	0	0.044518
$s_5(20)$	0.101495	0.085780	0.002395	0.044518	0

Table 17. Pairwise comparison of signals from $S(20)$ using the CWT

	s_1	s_2	s_3	$s_4(20)$	$s_5(20)$
s_1	0	1.355790	7.007736	1.638061	5.682953
s_2	1.355790	0	5.708453	1.262056	4.394570
s_3	7.007736	5.708453	0	6.394759	3.539179
$s_4(20)$	1.638061	1.262056	6.394759	0	4.411597
$s_5(20)$	5.682953	4.394570	3.539179	4.411597	0

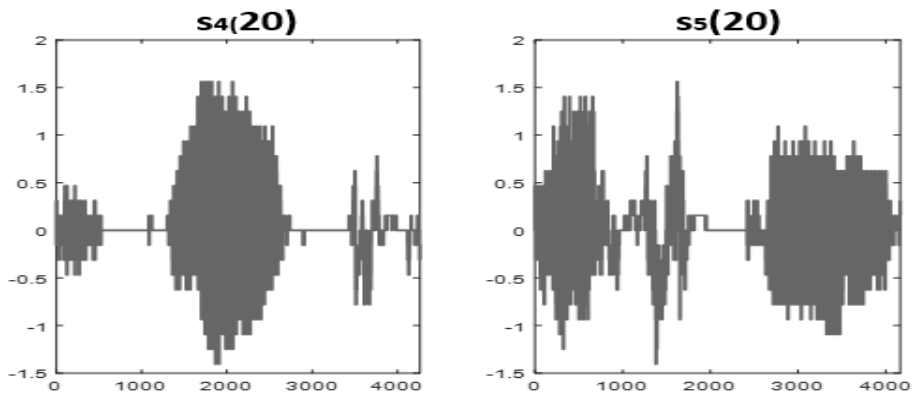


Fig.9. Digital signals $s_4(20)$ and $s_5(20)$ with 20A amplitudes

4. Conclusion

Each word consists of one or more syllables, which in turn are composed of one or more phonemes. A phoneme is the minimal linguistic unit that distinguishes meaning; it does not carry independent lexical or grammatical significance but enables native speakers to perceive and understand the elementary units of language – words. Modern automatic speech recognition (ASR) systems primarily rely on methods that segment a spoken signal into phonemes and analyze their amplitude–frequency characteristics to identify individual phonemes based on predefined sets of spectral features. In such approaches, each phoneme is treated as a single, indivisible unit of the speech signal with quasi-stationary frequency characteristics. However, these methods do not fully account for the dynamic temporal variations of phonemes that naturally occur in fluent speech.

This article proposes an iterative approach for recognizing spoken word sequences in the Azerbaijani language, based on systematically searching for amplitude values that are optimal for identifying the corresponding digital signals. The proposed method enables more precise recognition of phonemes by taking into account their dynamic characteristics and by empirically adjusting the amplitude of digital representations of speech segments.

Beyond speech recognition, this approach has broader applications. It can facilitate the analytical study of phonemes, support the construction of mathematical models of speech, and contribute to the synthesis of sound that closely replicates the original human voice. By incorporating empirical amplitude adjustment and iterative refinement, the method also provides a foundation for overcoming challenges posed by variations in voice quality, including emotional coloring and prosodic modulation. These extensions, however, constitute the subject of our future research.

Overall, the proposed approach demonstrates that iterative amplitude-based analysis of digital speech signals offers a robust framework for phoneme- and word-level recognition, bridging the gap between traditional quasi-stationary models and the inherently dynamic nature of natural speech.

The authors express their sincere gratitude to Dr. Ramin Rzayev, Doctor of Technical Sciences and Professor, for his invaluable scientific guidance in the preparation of this article. In particular, aspects of system analysis discussed by R.R. Rzayev in [13] and [14] provided essential support and insight that greatly contributed to the development of this work.

References

- [1] Proakis J.G., Manolakis D.G. *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed., Pearson Prentice Hall, 2007.
- [2] Rabiner L.R., Schafer R.W. *Theory and Applications of Digital Speech Processing*, Prentice Hall, 2011.
- [3] Rabiner L.R. "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 1989, v. 77 (2), p. 257–286.

- [4] Sakoe H., Chiba S. "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1978, v. 26 (1), p. 43–49.
- [5] Rabiner L., Juang B.-H. *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [6] Mallat S. *A Wavelet Tour of Signal Processing*, 3rd ed., Academic Press, 2009.
- [7] Quatieri T.F. *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall, 2002.
- [8] Ismayilov E.R. "An empirical analysis of traditional recognition methods using examples of identifying words spoken by native speakers," *Problems of Information Society*, 2025, v. 16 (1), p. 68–74.
- [9] Sakoe H., Chiba S. "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1978, v. ASSP-26 (1), p. 43–49.
- [10] Hindarto H., Anshory I., Efiyanti A. "Feature extraction of heart signals using fast Fourier transform," <https://jurnal.unej.ac.id/index.php/prosiding/article/view/4187>, last accessed 13/08/2025.
- [11] Zhao M., Chai Q., Zhang Sh. "A method of image feature extraction using wavelet," in *Proceedings, International Conference on Intelligent Computing (ICIC), Emerging Intelligent Computing Technology and Applications*, 2009, p. 187–192.
- [12] Saraswat S., Srivastava G., Sachchidanand N. "Wavelet transform based feature extraction and classification of atrial fibrillation arrhythmia," <http://biomedpharmajournal.org/?p=17470>, last accessed 13/08/2025.
- [13] Rzayev R.R. *Analytical Decision Support in Organizational Systems*, Palmerium Academic Publishing, Saarbruchen, 2016 (in Russian).
- [14] Rzayev R.R. *Neuro-Fuzzy Modeling of Economic Behavior*, Lambert Academic Publishing, Saarbrücken, 2012 (in Russian).