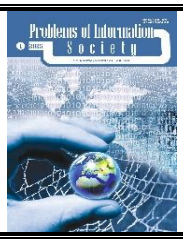

www.jpis.az

An empirical analysis of traditional recognition methods using examples of identifying words spoken by native speakers

Elchin Ismayilov

Institute of Control Systems, Baku, Azerbaijan

elchin.ismayilov1@mail.ru

orcid.org/0009-0000-4554-0084

ARTICLE INFO

http://doi.org/10.25045/jpis.v16.i1.08

Article history:
Received 18 September 2024
Received in revised form
25 November 2024
Accepted 31 January 2025

Keywords:

Signal recognition
Recognition method
Audio database
Sound recording
Adequacy criteria
Distance metric
Pairwise comparison of signals

ABSTRACT

Many users now interact with a form of artificial intelligence on a daily basis through search engines, social media, and voice recognition software. As the field matures, it is likely to permeate our lives in ever more surprising ways, so it will be important to create new governance structures to ensure its fair and transparent use. Along with machine vision algorithms for processing photo and video information, as well as natural language techniques for semantic analysis of texts, working with audio information is also the most demanded procedure for conducting business analytics. The article considers the problem of speech signal recognition using the example of an audio database formed on the basis of words reproduced by a native speaker in different tonalities with his characteristic pronunciation. In the proposed approach, the sound signal is considered as a one-dimensional representation of sound wave oscillations with a certain sampling frequency. To implement the task, classical DTW and DDTW methods, as well as methods based on the Fourier transform, discrete and continuous wavelet transforms are used. A computational experiment with the recognition of speech signals reproduced in the Azerbaijani language revealed the advantages of the continuous wavelet transform as the most accurate recognition method in the context of solving the problem.

1. Introduction

Today, thanks to accumulated empirical knowledge, traditional recognition methods, including those based on spectral analysis, have become quite effective tools for recognizing sound signals, in particular, words pronounced by a native speaker with his or her own pronunciation in various tonalities (Geler et al.; Itakura, 1975; Linh et al., 2014; Zhi-Qiang et al., 2024; Jiang and Chen; Novozhilov, 2016; Rajeev and Abhishek, 2019; Afouras et al., 2018; Haridas et al., 2018; Elmir and Abdeslam, 2019). In this regard, the study of various factors and characteristics of a speech sound recording is of particular interest. At the same time, the possibility of directly processing a sound recording in the form of a one-dimensional representation of sound wave oscillations with a certain sampling frequency makes it possible to recognize sound signals taking into account the features of pronunciation and tonality using traditional recognition methods. The following methods are used as tools for voice signal recognition (Dynamic Time Warping, DTW) and (Derivative Dynamic Time Warping, DDTW) which have proven themselves in the contextual field, as well as recognition methods based on the Fourier transform and wavelet transform, which is used in continuous and discrete cases. Obviously, the most important condition for the efficiency of a computational experiment related to voice signal identification is the accuracy of the recognition method used.

2. Related works

Speech recognition has been a widely studied field with numerous methodologies proposed for improving accuracy and efficiency. Several studies have explored the use of spectral analysis techniques such as Fourier Transform, Wavelet Transform, and Dynamic Time Warping (DTW) for speech processing. This section highlights some of the key works referenced in this paper that have significantly contributed to the development of speech recognition technologies. Today, there are two main areas of development in speech recognition: Automatic Speech Recognition (ASR) and Natural Language Processing (NLP). It should be noted that each of these methods has unique features, advantages, and disadvantages.

The ASR is a combination of computer hardware and software technologies that directly identify and process the human voice (Yu and Li, 2016). The operating principle of this technology can be defined as automatic transcription of spoken language into readable text. Voice recognition occurs in real time based on pre-set sound patterns. Thus, the ASR method provides the computer with the ability to identify words from human speech and translate them into electronic text.

NLP is an innovative method of voice recognition. It is a current direction in the field of machine learning, in which voice processing is performed based on intelligent algorithms (Khurana et al., 2023). In this case, one of the options for the voice recognition process in a voice assistant based on NLP may look like this: recording a person's speech; machine conversion of words from audio into electronic text; parsing the text into its main components to understand the context of the conversation and the person's goals; based on the results of the work, the system determines the command to execute.

3. Material and Methods

Speech recognition technology has seen remarkable advancements in recent years, yet significant challenges remain in the accurate recognition of underrepresented languages such as Azerbaijani. The complexity of speech signals, variations in pronunciation, and tonal differences necessitate robust methodologies for effective recognition. Developing accurate recognition methods for Azerbaijani is crucial for integrating AI-powered solutions into various applications, including education, customer service, and assistive technologies. Furthermore, speech-based systems that cater to diverse linguistic structures enhance inclusivity and accessibility for users worldwide. The present study addresses these challenges by evaluating and comparing multiple recognition methods to determine the most effective approach for processing Azerbaijani speech signals.

3.1. Problem Definition

Several words, such as “book”, “notebook” and “pencil” are reproduced by a person in Azerbaijani as [k'ıt'a: b], [dæf'tər] and [gjæ'læm], and are converted into analog signals through an appropriate audio device. Further, by quantization, these signals are transformed into corresponding digital signals s_1, s_2 and s_3, which form the current audio database shown in Fig. 1.

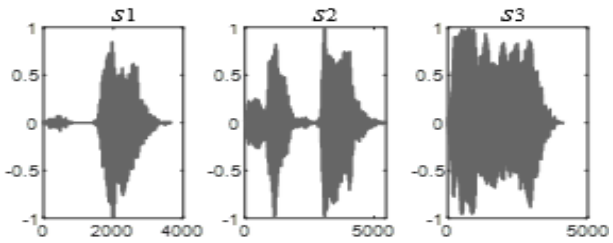

Fig. 1. Audio database including digital signals s_1, s_2, s_3

Assume that the word “pencil” is reproduced by a native speaker with his characteristic pronunciation three times in succession in different tonalities. After the appropriate transformations, this sound recording is represented as three corresponding digital signals (see fig. 2), which are reflected in expanded form in fig. 3.

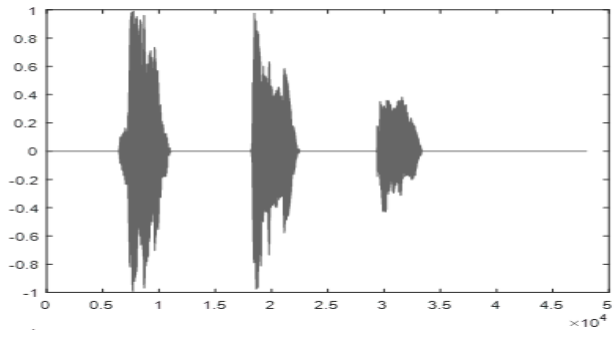

Fig. 2. Sound recording of the Azerbaijani word “pencil” pronounced three times in succession

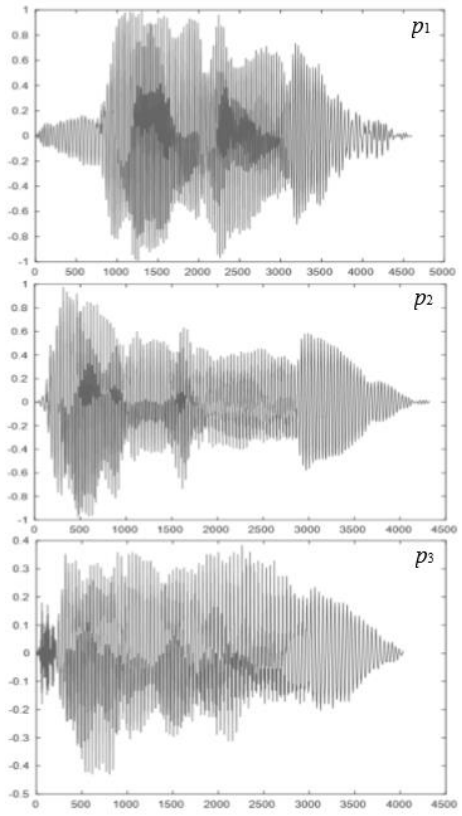

Fig. 3. The Azerbaijani word "pencil", pronounced in three different tonalities: p_1, p_2 and p_3

Based on this trivial example, it is necessary to formulate an approach to recognizing sound signals voiced by a native speaker. Based on the selected recognition mechanism, it is necessary to formulate and experimentally test an approach to forming an audio database that adequately reflects words from the Azerbaijani lexicon.

3.2. Problem Solution

To address the challenges associated with recognizing Azerbaijani speech signals, a combination of traditional and modern recognition techniques is applied. This study utilizes Dynamic Time Warping (DTW) and Derivative Dynamic Time Warping (DDTW) to account for variations in pronunciation. Additionally, spectral analysis methods such as Fourier Transform (FT), Discrete Wavelet Transform (DWT), and Continuous Wavelet Transform (CWT) are employed to enhance accuracy. By comparing the effectiveness of these methods, the study aims to identify the optimal approach for speech signal recognition. The integration of these techniques allows for improved handling of tonal shifts and pronunciation inconsistencies, ultimately leading to a more robust speech recognition system for Azerbaijani language applications.

3.3. Methods of Recognition and Criteria of Their Adequacy

Within the framework of the proposed approach, the mechanism for recognizing voice signals involves the use of DTW and DDTW methods, as well as recognition methods based on spectral analysis of signals, in particular, on the Fourier transform and wavelet transform, which is used for continuous and discrete cases.

DTW Method. This method has been studied quite well (Sakoe, 1978) and has been used in signal processing practice for a long time. Using a trivial example, the essence of the DTW method is as follows. Let $\{f_1, f_2, \ldots, f_n\}$ and $\{g_1, g_2, \ldots, g_m\}$ with lengths n and m, respectively, be numerical sequences. At the initial stage, local deviations between the components of these sequences are calculated in absolute value using, for example, the Euclidean metric. As a result, a matrix of size $n \times m$ is formed, consisting of square 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:

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

As a result of the iterative calculation of the DTW(f_i, g_j) indicator, the norm of the distance between signals is set in the following form

$$D_1 = \sqrt{\mathrm{DTW}(f_n, g_m)}\,. \qquad (1)$$

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

- Monotonicity – both indices i and j increase consistently.
- Continuity – in one step the indices i and j increase by no more than one.
- The sequential construction of the "paths" starts in the lower left and ends in the upper right corner.

The DTW algorithm is applied with "limitation" and "without limitation" on the size of the so-called "window", the size of which w determines the number of samples, allowing comparison of signal components both on the right and on the left. In this case, the total number of samples is $2w+1$, where the procedure for comparing f_i and g_j by the i-th sample point of the 1st and the j-th sample point of the 2nd must correspond to the inequality $|i-j| \le w$.

DDTW Method (Keogh and Pazzani, 2024). Unlike DTW, which uses amplitude values at the sample points, the DDTW method uses their first derivatives. As is known, for the discrete case, the 1st-order derivative at the sample point is determined as: $\dot{a}(i) = [a(i) - a(i-1)]/T$, where $a(i) = a(iT)$, $i = 0, 1, \ldots, N$; T is the sampling period of the analog signal a. Therefore, replacing f_i and g_j in the representation of the minimum distance in the DTW method with the corresponding derivatives p_i and q_j (i=1÷n, j=1÷m), for the DDTW method we have, respectively:

$$\begin{cases} \mathrm{DDTW}(p_i, q_j)^2 = d_{ij} + \min\{\mathrm{DDTW}(p_i, q_{j-1})^2, \\ \qquad\qquad \mathrm{DDTW}(p_{i-1}, q_j)^2, \mathrm{DDTW}(p_{i-1}, q_{j-1})^2\} \\ \mathrm{DDTW}(p_1, q_1)^2 = d_{11}. \end{cases}$$

In this case, the norm of the distance between signals is formed similarly in the form

$$D_2 = \sqrt{\mathrm{DDTW}(p_n, q_m)}\,. \qquad (2)$$

Fourier transform (FT) (Hindarto). The Fourier transform is the main 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. That is, 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. By its nature, the conventional Fourier transform cannot distinguish a stationary signal from a non-stationary one, which is a major problem for its applicability. Therefore, further in the article, it is applied the windowed Fourier transform of the form

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

where $W(\tau\text{-}t)$ is the so-called window function, which can be a Gaussian, Hamming window, Hann window 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 to obtain the characteristic of the distribution of the signal frequency (with amplitude) over time.

Thus, the usual Fourier transform is considered to be a windowed Fourier transform with a window of infinite width. 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.

Wavelet transform (WT) (Zhao et al., 2009; Saraswat et al., 2024). The method based on the wavelet transform was created as a tool that solves the Heisenberg uncertainty problem for constructing the time-frequency characteristics of a signal. Unlike the windowed Fourier transform, which has a constant scale at any time for all frequencies, the wavelet transform 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 wavelet transform 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 shift, and s represents the scale.

Further, in the process of processing sound signals voiced by a native speaker, discrete (DWT) and continuous (CWT) wavelet transforms are used. It should be noted that FT, DWT and CWT, being methods of spectral analysis, allow determining the powers of the detected spectra.

In (Kerimov, 2022; Rzayev and Kerimov, 2023), 4 criteria for assessing the adequacy of one-dimensional signal recognition methods for their accuracy were proposed and experimentally substantiated. In this case, S artificial families of signals were used as test examples, formed by uniformly shifting to the right or diagonally relative to the selected base signal (Rzayev and Kerimov, 2024).

For the sake of integrity of further discussions, the first three of these criteria are given below in a somewhat abbreviated form, taking into account the following concepts and notations: *MV* is the mean value; *SD* is the standard deviation; $D_k^h(s_i, s_j)$ is the distance between two signals s_i and s_j (i, j = 1÷N_s), where k = 1, 2, … is the serial number of the recognition method; N_s is the number of signals in S; h is signal shift step when forming S.

Criterion C_1 (*uniformity*): as the recogniseble signals become more distant, their distances from the base signal (etalon) should increase uniformly, rather than varying abruptly. Formally, this criterion is described as follows:

$$MCV1_k^h = \max_j \{CV1_k^h(j)\}, \quad (3)$$

where

$$CV1_k^h(j) = \frac{D_k^h(s_0, s_{j+1}) - D_k^h(s_0, s_j)}{\Delta t} \ (j = 1 \div N_s). \quad (4)$$

$MCV1_k^h$, as the maximum deviation of the derivatives of the distances, quantitatively reflects the adequacy of the k-th method according to the criterion C_1.

Criterion C_2 (*symmetry*): for a specific recognizable signal, the Euclidean distances from the left-hand and right-hand signals must be approximately equal, i.e. their ratio must be approximately equal to one. If the left-hand and right-hand signals are symmetrical relative to the given signal, then, obviously, these distances will be absolutely equal, i.e. their pairwise ratios will be identical to one. Formally, this criterion is described by the following ratio:

$$CV2_k^h(i) = \frac{D_k^h(s_i, s_{i+1})}{D_k^h(s_{i-1}, s_i)} \ (i = 1 \div N_s), \quad (5)$$

which determines the proportion between adjacent distances, in particular, between the distances from the right-hand located (i+1)-th and the left-hand located (i-1)-th signals to the i-th signal. In this case, for $i = 1 \div N_s$ we similarly have:

$$MCV2_k^h = \max_i \{CV2_k^h(i)\}. \quad (6)$$

Criterion C_3 (*operating speed*): as the recognizable signals "approach" the etalon, the speed of convergence of distance values increases. Here, the speed of convergence of distance values is understood as the ratio of the next to the current distance, when considering the sequence of distances in reverse order. Formally, this criterion is described by the following relationship:

$$CV3_k^h(i) = \frac{D_k^h(s_1, s_{i+1})}{D_k^h(s_1, s_i)} \ (i = N_s \div 1). \quad (7)$$

In this case, the value

$$MCV3_k^h = CV3_k^h(N_s) \quad (8)$$

determines the speed of convergence of distances $D_k^h(s_i, s_j)$.

3.4. Signal Recognition

The classical procedure of signal recognition is implemented by comparing the recognizable signals with the etalon by calculating the pairwise distances between them based on the selected metric. Comparisons of the recognizable signals p_1, p_2 and p_3 with the signals from the formed audio database are carried out for different window sizes w. Thus, for the case w = 5, the results of the pairwise comparison of signals from the family S = {s_1, s_2, s_3, p_1, p_2, p_3} using the DTW, DDTW, FT, DWT and CWT methods are summarized in tables 1-5.

Table 1. Results of pairwise comparison of signals from the S family using DTW

Signal	s_1	s_2	s_3	p_1	p_2	p_3
s_1	0	7.5780	15.0797	11.4542	12.4801	7.9502
s_2	7.5780	0	11.3362	10.2564	10.1183	9.4793
s_3	15.0797	11.3362	0	9.5579	7.8630	14.5444
p_1	11.4542	10.2564	9.5579	0	11.8500	13.0588
p_2	12.4801	10.1183	7.8630	11.8500	0	9.5309
p_3	7.9502	9.4793	14.5444	13.0588	9.5309	0

Table 2. Results of pairwise comparison of signals from the S family using DDTW

Signal	s_1	s_2	s_3	p_1	p_2	p_3
s_1	0	3.6679	6.6530	4.2606	5.3159	4.0468
s_2	3.6679	0	5.4151	4.2208	4.7423	4.3953
s_3	6.6530	5.4151	0	5.0817	4.3954	5.6842
p_1	4.2606	4.2208	5.0817	0	5.7146	5.1279
p_2	5.3159	4.7423	4.3954	5.7146	0	3.7159
p_3	4.0468	4.3953	5.6842	5.1279	3.7159	0

Table 3. Results of pairwise comparison of signals from the S family using FT

Signal	s_1	s_2	s_3	p_1	p_2	p_3
s_1	0	0.0157	0.0991	0.0687	0.0408	0.0151
s_2	0.0157	0	0.0834	0.0530	0.0251	0.0308
s_3	0.0991	0.0834	0	0.0304	0.0583	0.1142
p_1	0.0687	0.0530	0.0304	0	0.0279	0.0838
p_2	0.0408	0.0251	0.0583	0.0279	0	0.0559
p_3	0.0151	0.0308	0.1142	0.0838	0.0559	0

Table 4. Results of pairwise comparison of signals from the S family using DWT ($\times 10^{-5}$)

Signal	s_1	s_2	s_3	p_1	p_2	p_3
s_1	0	0.084	8898	5134	3184	1678
s_2	0.084	0	8888	5122	3172	1684
s_3	8898	8888	0	3766	5717	10572.0
p_1	5134	5122	3766	0	1951	6806
p_2	3184	3172	5717	1951	0	4855
p_3	1678	1684	105720	6806	4855	0

Table 5. Results of pairwise comparison of signals from the S family using CWT ($\times 10^{-5}$)

Signal	s_1	s_2	s_3	p_1	p_2	p_3
s_1	0	0	0.393	0.077	0.009	0.127
s_2	0	0	0.540	0.172	0.073	0.080
s_3	0.393	0.540	0	0.096	0.210	1.044
p_1	0.077	0.172	0.096	0	0.021	0.500
p_2	0.009	0.073	0.210	0.021	0	0.316
p_3	0.127	0.080	1044	0.500	0.316	0

The following tables 6, 7 and 8 present numerical scores reflecting the satisfaction of the recognition methods in terms of their compliance with criteria C_1, C_2 and C_3 in the context of studying signals from the *S* family.

Table 6. Assessments reflecting the satisfaction of methods relative to criterion C_1

Methods	Intermediate values obtained using (4)					Assessments (3)
DTW	7.578	7.502	3.626	1.026	4.5299	7.578
DDTW	3.6679	2.9851	2.3923	1.0553	1.26913	3.6679
FT	0.0157	0.0834	0.0304	0.0279	0.0256	0.0834
DWT	$8 \cdot 10^{-7}$	$8.81 \cdot 10^{-5}$	$3.76 \cdot 10^{-5}$	$1.95 \cdot 10^{-5}$	$1.51 \cdot 10^{-5}$	$8.81 \cdot 10^{-5}$
CWT	0	$3.9 \cdot 10^{-6}$	$3.2 \cdot 10^{-6}$	$0.7 \cdot 10^{-6}$	$1.2 \cdot 10^{-6}$	$3.9 \cdot 10^{-6}$

Table 7. Assessments reflecting the satisfaction of methods relative to criterion C_2

Методы	Intermediate values obtained using (5)				Assessments (6)
DTW	3.7582	1.7783	2.2921	2.3191	3.7582
DDTW	1.7472	0.3333	0.6329	1.9987	1.9987
FT	0.0677	0.05296	0.0025	0.0280	0.0677
DWT	$8.8 \cdot 10^{-5}$	$5.12 \cdot 10^{-5}$	$1.82 \cdot 10^{-5}$	$2.9 \cdot 10^{-5}$	$8.8 \cdot 10^{-5}$
CWT	$0.54 \cdot 10^{-5}$	$0.44 \cdot 10^{-5}$	$0.08 \cdot 10^{-5}$	$0.3 \cdot 10^{-5}$	$0.54 \cdot 10^{-5}$

Table 8. Assessments reflecting the satisfaction of methods relative to criterion C_3

Методы	Intermediate values obtained using (7)				Assessments (8)
DTW	1.5698	0.9178	1.3165	0.5025	0.5025
DDTW	1.3136	0.8015	1.5615	0.5513	0.5513
FT	2.6947	1.6839	1.4431	0.1586	0.1586
DWT	1.8974	1.6125	1.7332	0.0095	0.0095
CWT	0.0748	8.1310	5.09096	0.00076	0.00076

As can be seen from the obtained results, the recognition methods based on spectral analysis of signals are the best in terms of compliance with criteria C_1, C_2 and C_3, which is confirmed by the corresponding significantly smaller numerical estimates in the last columns of tables 6, 7 and 8. Moreover, in the selected example of recognizing a voice signal voiced by a native speaker, the continuous wavelet transform demonstrates the best result in terms of recognition accuracy.

4. Discussion

Future advancements in speech recognition for the Azerbaijani language can be achieved through several approaches. One potential solution is the application of deep learning models such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), which have demonstrated high accuracy in other languages. Expanding the dataset with diverse speakers and tonal variations can also enhance model robustness. Additionally, the integration of self-supervised learning techniques can reduce the dependency on large labeled datasets. Another promising avenue is the use of hybrid models that combine traditional methods like DTW with machine learning-based approaches to improve recognition accuracy. Finally, developing context-aware recognition systems that incorporate natural language processing (NLP) could significantly refine recognition results by understanding speech in a broader linguistic context.

5. Conclusion

Using a trivial example of reproducing one word in the Azerbaijani language in three tones, the article formulates an approach to recognizing voice signals based on the combined use of five recognition methods. A promising system that combines different recognition methods will be able to respond to voice commands and perform tasks based on user actions. This is just one example of how one of the AI tools can be integrated into ordinary devices to make them more intuitive and able to interact with Azerbaijani citizens in natural language.

The calculations and results presented in the article were obtained using the author's software of Associate Professor A.B. Kerimov, which was used in (Kerimov, 2022; Rzayev and Kerimov, 2023; Rzayev and Kerimov, 2024) and in writing a number of other peer-reviewed articles under the supervision of Professor R.R. Rzaev.

References

Afouras, T., Chung, J.S., Senior, A., Vinyals, O., Zisserman, A. (2018). Deep audio-visual speech recognition. IEEE Trans. Pattern Anal. Mach. Intell., 44(12): 8717–8727.

Elmir, B.S., Abdeslam, Y.D. (2019). A study on automatic speech recognition. Journal of Information Technology Review, 10, 77–85.

Geler, Z., Kurbalija, V., Ivanović, M., Radovanović, M., Dai, W. (2024). Dynamic time warping: Itakura vs Sakoe-Chiba, In Proceedings of International Symposium on Innovations in Intelligent Systems and Applications, https://ieeexplore.ieee.org/document/8778300.

Haridas, A.V., Marimuthu, R., Sivakumar, V.G. (2018). A critical review and analysis on techniques of speech recognition: the road ahead. Int. J. Knowl. Base. Intell. Eng. Syst, 22, 39–57.

Hindarto, H., Anshory, I.; Efiyanti, A. (2024). Feature extraction of heart signals using fast Fourier transform. https://jurnal.unej.ac.id/index.php/prosiding/article/view/4187

Itakura, F. (1978) Minimum prediction residual principle applied to speech recognition. Transactions on Acoustics, Speech and Signal Processing, 23(1), 67–72.

Jiang, Sh., Chen, Z. (2024). Application of dynamic time warping optimization algorithm in speech recognition of machine translation. Research Article, 9(11), https://doi.org/10.1016/j.heliyon.2023.e21625.

Keogh, E.J., Pazzani M.J. Derivative Dynamic Time Warping. (2001). In PROCEEDINGS of the 2001 SIAM International Conference on Data Mining, https://doi.org/10.1137/1.9781611972719.1

Kerimov, A.B. (2022). Accuracy comparison of signal recognition methods on the example of a family of successively horizontally displaced curves. Informatics and Control Problems, 42(2), 80–91.

Kerimov, A.B. (2022). Accuracy comparison of signal recognition methods on the example of a family of successively horizontally displaced curves. Informatics and Control Problems, 42(2), 80–91.

Linh, L.H., Hai, N.T.; Thuyen, N.V., Mai, T.T., Toi, V.V. (2014). MFCC-DTW algorithm for speech recognition in an intelligent wheelchair. In PROCEEDINGS of 5th International Conference on Biomedical Engineering, pp. 417–421.

Novozhilov, B.M. (2016). Calculation of the derivative of an analog signal in a programmable logic controller. Aerospace Scientific Journal of Moscow State Technical University, 4, 1–12 (In Russian).

Rajeev, R., Abhishek, T. (2019). Analysis of feature extraction techniques for speech recognition system. Int. J. Innovative Technol. Explor. Eng, 8, 197–200.

Rzayev, R.R., Kerimov, A.B., Garibli, U.G., Salmanov, F.M. (2024). Criteria for assessing the adequacy of image recognition methods and their verification using examples of artificial series of signals. Problems of Information Society, 15(1), 10–17.

Sakoe, H., Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(1), 43–49.

Saraswat, S., Srivastava, G., Sachchidanand, N. (2024). Wavelet transform based feature extraction and classification of atrial fibrillation arrhythmia. http://biomedpharmajournal.org/?p=17470

Zhao, M., Chai, Q., Zhang, Sh. (2009). A method of image feature extraction using wavelet. In PROCEEDINGS, International Conference on Intelligent Computing, ICIC, Emerging Intelligent Computing Technology and Applications, pp. 187–192.

Zhi-Qiang, U., Jia-Qi, Z., Xin, W., Zi-Wei, L., Yong, L. (2024). Improved algorithm of DTW in speech recognition. IOP Conference Series: Materials Science and Engineering, 563(5), 24–36.

Yu, Dong, and Li, Deng. (2016). Automatic Speech Recognition. Springer London limited.

Khurana, D., Koli, A., Khatter, K. et al. (2023). Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl 82, 3713–3744.