International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 37, No. 6, December, 2023, pp. 1527-1535
Journal homepage: http://iieta.org/journals/ria

Creation of An Intelligent System for Uzbek Language Teaching Using Phoneme-Based Speech Recognition

Sayyora Ibragimova*

Digital Technologies and Artificial Intelligence Research Institute of the Ministry of Digital Technologies of the Republic of Uzbekistan, 17A Boz-2, Tashkent 100125, Republic of Uzbekistan

Corresponding Author Email: si6299564@gmail.com

https://doi.org/10.18280/ria.370617

Received: 24 July 2023
Revised: 20 September 2023
Accepted: 26 October 2023
Available online: 27 December 2023

Keywords:
intellectual system, word segmentation, phonetic sequences, sound system, automatic processing

ABSTRACT

The recent surge in interest to learn the Uzbek language among foreigners has underscored the need for innovative teaching tools. Despite the limited studies on intelligent systems for phonemic speech recognition in the Uzbek context, this research aimed to address this gap. The purpose of this study was to create an intelligent system for teaching the Uzbek language as a foreign language based on the technology of phonemic recognition of speech signals. It was developed an intelligent system for Uzbek language instruction using phonemic speech recognition technology. The approach utilized various methods, including pinpointing challenging phonemes, comparative data analyses, and analytical-synthetic breakdowns of linguistic components, all enhanced by the wavelet transform's signal refinement. The system's precision in recognizing speech signals phoneme-by-phoneme, emphasizing difficult sounds for learners, promises broader AI-driven language study applications. Specifically designed for the Uzbek language, the system achieves an accuracy range of 67% to 95%. This breakthrough not only propels AI-driven language processing but offers a robust tool for improving Uzbek language instruction, especially beneficial for the Turkic language group. Future avenues include its use in computer modeling and automatic speech processing for Turkic languages, solidifying its innovative contribution to AI-driven language teaching.

1. INTRODUCTION

Technologies used for automatic processing of the Uzbek language, as a rule, are guided by traditional methods, but today it is relevant to introduce a variety of modern technologies based on the operation of wavelet transforms. The process of speech recognition in Uzbek is limited in terms of an insufficient number of data corpora, since it belongs to low-resource languages. But at the same time, modern methods of automatic processing of speech signals are successfully applied, expanding knowledge about phonetic speech samples and the sound system. The transformation of speech signals into phonetic sequences, the determination of their trajectory and phonetic boundaries, allows taking into account not only the physical, but also the acoustic characteristics of sounds [1]. Therefore, the wider and more voluminous the data corpus, the more effective speech recognition will be, the more speech samples are in the thesaurus, the more accurate the results will be. The relevance of this work is due to the minimum number of publications on the topic of automatic recognition and phonetic segmentation of texts in the Uzbek scientific discourse. Based on the experience of studying high-resource languages (English, Chinese) with large data sets, it is possible to introduce the most effective and efficient technologies when conducting experiments and empirical research on the example of the Uzbek language.

Researchers Mukhamadiyev et al. [2] point to the need for automatic speech recognition in low-resource languages, such as Uzbek, since language model training was mainly used for high-resource languages (English, Chinese, European). The proposed UzLM model based on neural networks, based on 80 million words and 15 million sentences of the Uzbek language, reduced the error rate to 5.26%. In the work of Khuda [3], a software architecture for an intelligent system was developed for the purpose of speech recognition for learning a foreign language, the key advantages and disadvantages of its use were analysed, and technologies for recognizing and converting speech signals were considered. The development of the software architecture was carried out on the basis of the audio data comparison algorithm and its alignment using mathematical methods.

An article by Mamyrbayev et al. [4] shows the use of a multimodal speech recognition system based on coupled Markov models using audiovisual information on the example of the Kazakh language using voice activity detection, linear perception predictions, speech segmentation and lip movements. The study of the Uzbek language in the direction of comparison and analysis of phonological changes between a group of foreign students studying the Uzbek language and a group of Uzbek students studying a foreign language considers Mirzayevna [5], noting the huge role of reading and listening for the development of communication skills through network learning. End-to-end automatic speech recognition systems are characterized by the transformation of certain sequences of acoustic characteristics into text, which is encoded using grapheme subwords. Papadourakis et al. [6] indicate that the developed method of phonetic induction of subwords showed high efficiency, up to 15.21%.

The Uzbek language, categorized as a low-resource language, presents challenges in automatic processing due to the scarcity of extensive data corpora. Most technologies applied to the Uzbek language employ traditional methods, whereas modern techniques, especially those centered around wavelet transforms, have shown promise in phonetic speech samples processing. There's an acute need to transform speech signals into phonetic sequences, considering both their physical and acoustic characteristics, for efficient speech recognition. Despite the success of modern speech signal processing methods in high-resource languages like English and Chinese, their application in the context of the Uzbek language remains underexplored. There is also a discernible void in holistic research targeting the Uzbek language, particularly in terms of phonemic speech recognition, effectiveness of intelligent systems, and their application in language learning.

The purpose of this study was to develop an intelligent system of the Uzbek language for learning, which allows implementing the strategy of phonemic speech recognition, forming an idea of the effectiveness of intelligent systems and various technologies for automatic speech processing, and determining the accuracy of speech recognition. Based on the purpose of this study, the following tasks were set:

- formation of an idea about the operation of intelligent systems;
- study of methods and technologies of phonemic speech recognition and their application in various branches of knowledge;
- creating the intellectual system of the Uzbek language and testing its effectiveness on non-native speakers learning the Uzbek language;
- identification of accuracy in the segmentation of phonetic words and recognition of the sounds of the Uzbek language.

2. MATERIALS AND METHODS

The theoretical basis of this work is based on modern research by Uzbek, Kazakh, English, Chinese, Czech, and other authors dealing with computer diagnostics and data modelling, automatic text processing, and the creation of intelligent systems (for example, applications) for learning foreign languages. When developing the intellectual system, the voluminous data corpus of the Uzbek language UZWORDNET, which includes 28140 synsets, 64389 semantic series and 20683 words, represented by samples of native speakers of the Uzbek language with different pronunciation of sounds were used [7]. Using methods for learning with the help of an intelligent system. In this work, several methods of analysis were used, including computer modelling methods (phoneme-based speech recognition, creation of an intelligent system), as well as comparative, graphic, analytical-synthetic, and statistical methods. These methods of analysis were used in a complex way, since only in this way it is possible to systematize and classify the results obtained during the study.

The phonemic speech recognition process starts with the input of a speech signal through a microphone, which undergoes primary processing. Following this, the speech trajectory of the signal is determined. Any pauses in the speech are removed to ensure a smooth analysis. This processed signal is then segmented into phonemic sequences. Subsequent stages involve creating dictionaries and descriptions of speech units based on certain time intervals. Linguistic information derived from this is labeled, and the speech sequence is automatically separated. During the phonemic analysis, the study closely monitored the accuracy of recognizing phonetic features by using various Uzbek words as examples.

The research foundation heavily depended on the extensive UZWORDNET data corpus of the Uzbek language. This dataset is a comprehensive collection of native Uzbek pronunciations. The methodology behind the application aimed to teach Uzbek as a foreign language was grounded in the SPeach phoneme-based speech recognition system. For decomposing signals, the discrete wavelet transform (DWT) was applied over the original signal. Subsequently, the modeling of these results was conducted through the Matlab program, ensuring accurate representation. A significant portion of the system utilized a large database of syntactic patterns, which added depth to the teaching framework.

When considering design elements and challenges, the primary objective was to develop an application that could accurately recognize Uzbek phonemes and subsequently facilitate effective language teaching. Integrating a diverse data corpus like UZWORDNET fortified the system's foundation. Techniques like the application of the DWT over the original signal and modeling through Matlab significantly enhanced the system's precision. Performance metrics were an essential aspect of the research. Tables were crafted to demonstrate the efficiency of the algorithm with the SPeach system. This also showcased the accuracy of phonemic recognition using the wavelet transform system. Another pivotal metric was charting the progress of students over a 6-month period, emphasizing the efficacy of the developed intellectual system.

In conclusion, the study utilized a combination of computational techniques and traditional analytical methods to foster a comprehensive approach to teaching the Uzbek language to non-native speakers.

3. RESULTS AND DISCUSSION

An intelligent system for teaching the Uzbek language allows providing feedback on speech recognition in real-time, and evaluation of the parameters of pronunciation and fluency of speech [8]. The artificial intelligence technology is based on a database of speech samples from different people with different accents, which makes it possible to recognize patterns even of those who are not native speakers. It is also adaptive, as it provides a personalized technology based on the analysis of the performance and behaviour of each student individually, taking into account the daily curriculum.

3.1 Stages of creating an intelligent system

To create an intelligent system for the purpose of teaching the Uzbek language as a foreign language, the creation of computer interfaces based on the system of phonemic speech recognition (SPSR), based on a database of the studied language with a large dictionary, was used. As the basis of the Uzbek language, a "vocabulary network" called UZWORDNET was used, containing 28140 synsets, 64389 semantic series and 20683 words with an estimated accuracy of 75.98% [6]. The study consisted of several stages:

(1) Creating an optimal dictionary and choosing a data storage method.

(2) Segmentation of the speech signal into small units.

(3) Alignment of algorithms for recognition of a speech signal (SS) based on phonemic synthesis.

The schemes of the system of phonemic speech recognition in teaching the Uzbek language as a foreign language can be represented as follows:

(1) Inputting a speech signal from a microphone, recording a speech message using a wav file.

(2) Primary processing of a speech signal through the formation of a trajectory of parameters using the spectral-temporal and spectral-band representation of the speech signal (ST SS and SB SS, respectively).

(3) Removing pauses before and after utterances.

(4) Automatic segmentation of the speech signal with the refinement of phonemic boundaries.

(5) Formation of dictionaries for speech units: dictionaries 1, 2, 3 (D1, D2, D3, respectively).

(6) Description of speech units of the dictionary after smoothing segments of phonemes in temporal areas using bell-shaped functions (Figure 1).

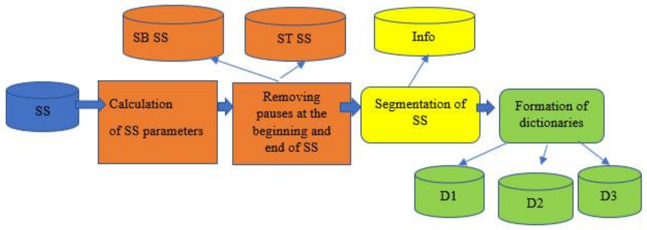
SB SS
ST SS
Info
SS
Calculation of SS parameters
Removing pauses at the beginning and end of SS
Segmentation of SS
Formation of dictionaries
D1
D2
D3

Figure 1. Diagram of the operation of the speech recognition system

Source: Compiled based on Savenkova and Karpov [9]

Table 1. Algorithm for working with the SPeach system

1	SS input from microphone or wav file using WAVReadWrite
2	Spectral analysis of SS and calculation of parameters using SpectrAnalis
3	Filtering and smoothing data with Filters
4	Formation of SU Unit dictionaries using TeachProcFirmirSlovar
5	Building a spline description of the SB using ApproximationProc
6	Construction of the description of ST in the class of bell-shaped functions using Lokony_Anjezi
7	Adding items to a dictionary with Recogn
8	Building a BFS (breadth-first search) algorithm with PoiskVShirinu
9	Building a DFS (depth-first search) algorithm using PoiskVGlubinu
10	Building a heuristic algorithm using Evristics
11	Using graphics with ModDraw3dGraph

Source: compiled based on Savenkova and Karpov [9]

Linguistic information (Info) is represented by main (name+transcription) and auxiliary (number of time points+number and boundaries of segments, tone+noise+pause) information. Improving automatic phonemic speech recognition when teaching Uzbek as a foreign language can combine acoustic and articulatory information, so the study is focused on improving the process of speech stream transformation, taking into account the correlation between articulatory features [8]. The automatic division of the speech sequence into acoustic syllables standards (they can be two-, three-, and four-component ones) occurred with the help of linguistic information marking. Information about a speech unit in dictionaries is presented in the following form:

(1) <SU Number>_<SU Name>_<SU Transcription>_.

(2) <Number of time samples>_<Number of segments>_.

(3) <Addresses of segment boundaries>_.

(4) <Group membership of segments>.

The creation of an intelligent system for teaching the Uzbek language was carried out using the SPeach phoneme-based speech recognition system (Table 1).

Combining acoustic and articulatory information, as indicated in the study, points towards a holistic approach in language learning systems. Future technologies might merge visual feedback (like articulatory animations) with audio feedback to provide learners with a comprehensive understanding of speech production. The specific algorithms described, including BFS, DFS, and heuristic algorithms, demonstrate the potential for optimized, individualized learning pathways. As AI technology advances, these systems can adapt in real-time to a learner's progress, ensuring the most efficient trajectory for language acquisition.

3.2 Using wavelet transforms

At the present stage of computational linguistics, various types of wavelet analysis are widely used: wavelet transforms, wavelet frames, wavelet series, discrete, stationary and analytical wavelet transforms, and wavelet packets. The wavelet transform is considered to be an effective tool used in time-frequency analysis to extract phonetic features in the process of phonemic speech recognition, allows converting a speech signal into a form that makes the original signal values more amenable to study, and also helps to compress the original data set [10].

Segmentation of a speech signal is characterized by the selection of certain sections of the signal that correspond to its specific substructures. Per-phoneme speech recognition is represented by segmentation with the definition and tracking of inter-phoneme transitions. Wavelet transforms help to ensure the process of phoneme recognition in extended quasi-stationary sections of a speech signal. A characteristic feature of the signal at interphonemic transitions: active changes in significant areas of the study and an increase in the level of detail with an increase in wavelet coefficients should be noted [11]. At the same time, the stationary parts of phonemes demonstrate the grouping of wavelet coefficients directly within small areas. The search for interphonemic boundaries takes into account the increase in wavelet coefficients when tracking different zoom levels.

When applying the wavelet transform, it is possible to notice a high level of data redundancy, while it is necessary to determine the most informative level at which the input speech signal is decomposed. Wavelet coefficients are used for both low and high frequencies, so it is possible to evaluate the behaviour of a speech signal in different frequency ranges. The nature of human speech covers the range from 150 to 4000 Hz, so 6 levels of decomposition are sufficient. There are many directions in the theory of wavelet analysis. For example, using multiscale wavelet analysis, a signal can be represented as a sequence of images with different levels of detail, which helps to identify signal characteristics in certain areas and classify signals according to their degree of intensity. The analysis is based on the decomposition of the signal into functions that form an orthonormal basis [12]. Each function can be decomposed at a certain given resolution level (scale) j_n:

$$f(x) = \sum_{k=0}^{2M-1} s_{j_n,k}\varphi_{j_n,k} + \sum_{j\geq j_n}^{j_{max}} \sum_{k=0}^{2M-1} d_{j_n,k}\psi_{j_n,k} \qquad (1)$$

where, $\varphi_{j_n,k}$ and $\psi_{j_n,k}$ -scaled and shifted versions φ (scale function) and ψ ("mother wavelet"); $s_{j_n,k}$ -approximation coefficients; $d_{j_n,k}$-coefficients of detail.

Below are analytical graphs of functions φ and ψ Daubechies wavelet (Figure 2).

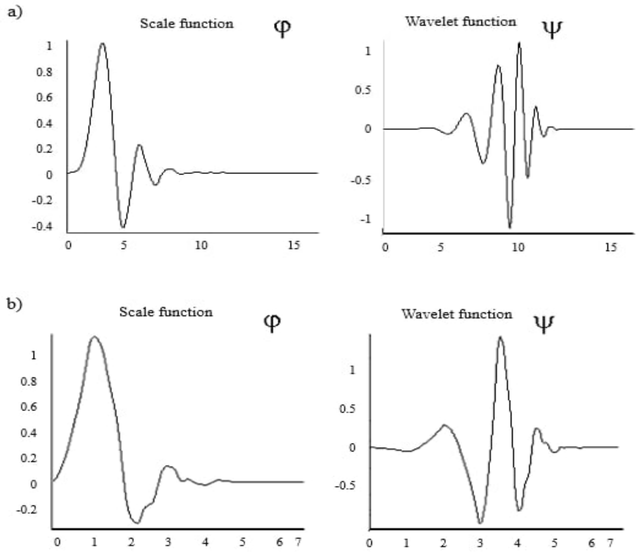

Figure 2. Analytic graphs of functions φ and ψ: a) Daubechies 8 wavelet; b) Daubechies wavelet 4 (screenshots)
Source: compiled based on Novikov [13]

Scaling and offset functions φ and ψ happen according to the law:

$$\varphi_{j,k} = 2^{\frac{j}{2}}\varphi(2^j x - k);\ \psi_{j,k} = 2^{\frac{j}{2}}\psi(2^j x - k), \qquad (2)$$

where, $\varphi_{j_n,k}$ and $\psi_{j_n,k}$ -scaled and shifted versions φ (scale function) and ψ ("mother wavelet").

In turn, the functions themselves φ and ψ defined as:

$$\varphi(x) = \sqrt{2}\sum_{k=0}^{2M-1} h_k\varphi(2x-k)\,;\ \psi(x) = \sqrt{2}\sum_{k=0}^{2M-1} g_k\psi(2x-k), \qquad (3)$$

where, $g_k = (-1)^k h_{2M-k-1}$.

Using the orthogonality properties of scaling functions and setting the scale of values, M it is possible to calculate specific values of the coefficients, $h_k$defining orthogonal wavelets. For example, when $M = 2$ a series of coefficients h_k will be obtained that determines the Daubechies wavelet 4. Thus, orthogonal wavelet analysis is reduced to finding the approximation coefficients $s_{j,k}$ and detail factors $d_{j,k}$ when decomposing the signal *f(x)*. Below is an example location of points of local maxima in the wavelet transform for the phoneme "a" (Figure 3).

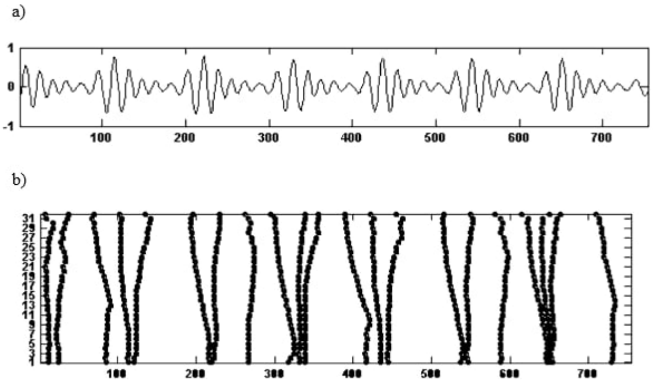

Figure 3. Points of local maxima of the phoneme "a": a) the line of coefficients of the wavelet transform of the phoneme "a"; b) points of local maxima of the phoneme "a" (screenshots)
Source: compiled by the author

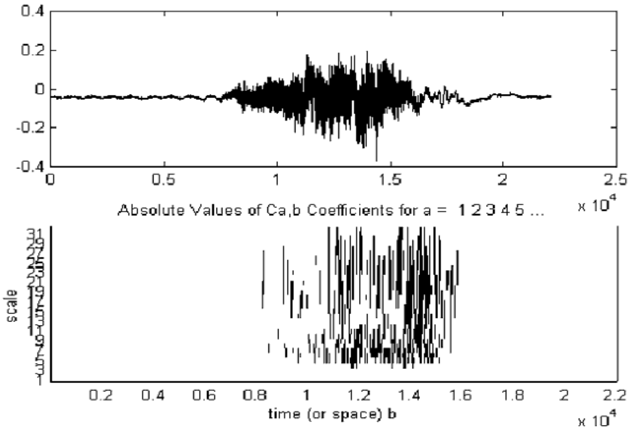

Figure 4. Signal and its wavelet coefficients (screenshot)
Source: compiled by the author

A set of points on the plane (x, a), where the local maxima of the wavelet transform are located, form a skeleton of maxima. At these points, that is, in the values of the local maxima of the wavelet transform, there are informative signs of the signal (Figure 4). The use of this mechanism makes it possible to isolate phonemes from the speech stream more efficiently.

Modelling of the results obtained during the work of the intellectual system for teaching the Uzbek language as a foreign language was carried out using the Matlab program. The output of the percentage of similarity of the incoming signal occurred through a microphone with a signal on the database, which stores the phoneme parameters (phonemic alphabet) of the Uzbek language (Table 2).

During phonemic recognition of Uzbek speech, errors were observed due to the similarity in the pronunciation of some similar-sounding sounds: "g"-[ge], "g"- the average between the sounds [ge] and [he], [h]- soft [he] without effort, "x"- hard [xe], "ng"- a sound consisting of two letters; "e"- [e], "y"- [ye]; "o'"- the average between the sounds [o] and [u], "u"- [u]; "j" - [dje] or sometimes [je], "sh"- [she], "ch"- [che]. Pairs of sounds were also not clearly recognized: “z” and “s”, “t” and “d”, “b” and “p”, “v” and “f”.

Table 2. Accuracy of phoneme recognition of a speech signal using wavelet transforms to segment into phonemes

Lexeme	Accuracy Percentage, %							
“O ‘ttiz”	“o”	“t”	“t”	“i”	“z”			
	75	87	67	97	91			
“Erkaklar”	“e”	“r”	“k”	“a”	“l”	“a”	“r”	
	93	99	89	100	98	99	97	
“Ma’lumot”	“m”	“a”	“ ’ ”	“l”	“u”	“m”	“o”	“t”
	95	99	89	99	83	97	95	98
“Go ‘zallik”	“g”	“o”	“z”	“a”	“l”	“l”	“i”	“k”
	78	79	91	97	95	96	92	83

Source: compiled by the author

The process of isolating and analyzing phonemes using wavelet transforms can be broken down into a series of steps. A speech signal is first captured, typically using a microphone. This raw audio data serves as the primary input for further analysis.

The captured speech signal is subjected to wavelet decomposition. Wavelets can decompose a signal into its time-frequency components, which is particularly useful for analyzing non-stationary signals like speech. The decomposed signal can be segmented into specific substructures or phonemic units. This is particularly important for languages like Uzbek where phonemic transitions play a significant role in meaning. By examining wavelet coefficients, it was identified transitions between phonemes. These interphonemic transitions are characterized by active changes in certain parts of the signal and an increase in wavelet coefficient values. Stationary parts of phonemes, on the other hand, show wavelet coefficients grouped within smaller areas. The wavelet transform allows decomposition of the speech signal across multiple frequency ranges, which is crucial for speech as the frequency range of human speech varies from around 150 to 4000 Hz.

Six levels of decomposition are usually sufficient for capturing the required information for phonemic recognition. The decomposed signal is used to extract specific features that are relevant for phonemic recognition. This includes both low and high-frequency features. Local maxima in the wavelet transform for specific phonemes can be pinpointed, providing a detailed representation of the phoneme's characteristics. The points where local maxima occur form the "skeleton of maxima," which contain informative traits of the signal. This feature extraction step is crucial for differentiating phonemes in the speech stream. Once the features have been extracted, they are compared against a pre-existing database containing parameters of phonemes for the language in question (e.g., the phonemic alphabet of the Uzbek language). By comparing the extracted features with the database, phonemes can be recognized with varying degrees of accuracy.

3.3 Language learning technology using an intelligent system

The following systems are used to teach Uzbek as a foreign language: MemoWord, Learn Uzbek by voice, HelloTalk, Lingualeo, as well as phrasebooks. MemoWord offers a large vocabulary base, contains transcription, translation, and voicing of words, there is a function for entering entire sentences. Learn Uzbek by voice tools allow entering phrases by voice and convert them to text with translation, the application contains information about the pronunciation of phrases. HelloTalk allows entering voice messages and listening to audio files, translating text and voice messages into another language, and correcting errors. The intellectual system Lingualeo offers exercises for pronunciation, reading speed and understanding of textual information [14].

The created intellectual system for teaching Uzbek as a foreign language offers many tasks to improve phonetic speech, depending on the level of knowledge of the Uzbek language:

- choosing the correct pronunciation of the word;
- listening and repeating words and phrases and feedback on pronunciation;
- using a chatbot to communicate on various topics with the practice of reading, rhythm, lexical stress and training phonetic skills.

Lessons include a variety of thematic blocks: online dating, travel, famous people, recreation, and entertainment. In addition to pronunciation practice, the application effectively affects grammar, lexical, syntactic, and stylistic skills, and improves reading and listening fluency. First, it is possible to study individual words and phrases, practice spelling and pronunciation, and then move on to reading long sentences with intonation [15]. The conversations used are presented in the form of speech samples in the mode of real communication, which allows improving communication skills. Effective learning of Uzbek as a foreign language combines blended learning, in which traditional methods are used together with digital ones, including the use of various applications. Students’ communication skills are divided into listening, speaking, reading, and writing. Productive competence refers to the speaking and writing skills needed to express thoughts and work effectively in formal and informal settings. Receptive competence includes listening and reading skills, emphasizing the importance of phonological, syntactic and semantic interpretation in parallel with cognitive processing [16].

Students communicate in both source and target languages using the program, developing correct pronunciation skills, which is a prerequisite for interpreting any utterance. Speech skills in teaching Uzbek as a foreign language using an intellectual system are developed in parallel with mastering the vocabulary, including the practice of memorizing new lexemes is combined with the assimilation of grammatical patterns through personalized learning. The results of the effectiveness of the created intellectual system for teaching the Uzbek language are presented in the table (Table 3).

Table 3. Percentage improvement in skills after using the training program for 6 months

Phonetic speech	32
Receptive Reading Competence	25
Productive Speaking Competence	23
Digital literacy	25
Receptive ability of listening	17
Vocabulary usage	15
Productive Writing Competence	9
Grammar Skills	7

Source: compiled by the author

The integrity of a phonetic utterance is one of the key parameters for communicating with native speakers, so the created intelligent system is aimed at improving phonetic pronunciation. After using the app, students were able to improve phonetic pronunciation by 32%. Phonetic success in teaching Uzbek as a foreign language can be influenced by such factors: increasing the level of awareness in terms of the impact of literate pronunciation on the communication process, the practice of using the International Phonetic Alphabet; understanding of stress at the level of words and sentences, the formation of ideas about the rhythm of the Uzbek language and its intonation patterns [17]. Thus, the created intellectual system allows combining active learning without long-term memorization of phonetic rules and theoretical material, using practical methods and feedback. At the same time, students practice not only phonetic, but also communication skills, learn intonation and put the correct logical stress in sentences. After completing the training, the results are visible in terms of clarity of pronunciation, in terms of lexical and grammatical parameters.

The results provided show the potential advancements in language learning technology, especially with the integration of intelligent systems that focus on phonetic training. The significant 32% improvement in phonetic speech indicates that students highly benefit from targeted phonetic training. It suggests that future language learning platforms should emphasize phonetic training, not only for Uzbek but for any language. The combination of traditional methods and digital applications (blended learning) has proven effective, as seen from the varied improvements in different skills. Future platforms should adopt a similar blended approach to optimize learning outcomes.

The fact that the system addresses various skills, including reading, writing, listening, and speaking, shows the importance of holistic language learning. This integrated approach can be a blueprint for other languages, ensuring that learners become proficient in all aspects of the language.

3.4 Using different technologies for phonemic speech recognition: Efficiency and results

The phoneme segmentation technique, based on the analysis of the spectra of the discrete wavelet transform, determines the localization of phoneme boundaries in the process of speech recognition, information about which is based on the definition of subranges. For analysis, Ziółko et al. [18] used Polish words (16,425 utterances) that were manually segmented. Speech segmentation results showed an F-score of 72.49%. With phoneme-based speech recognition of the Uzbek language, the accuracy of sound parameters was close to the maximum indicators: from 67% to 95%, the average values were 83-90%. The problem of developing a technique for identifying voices with a slight deviation of the voice using wavelet packet transformation, a multilayer perceptron and an artificial neural network is relevant in speech signal recognition. When analysing Morikawa et al. [19] 74 audio files, the following results were obtained: the accuracy of 99.75% and 99.56% for the Symlet 2 family, 91.17% and 70.01% for the Daubechies 2 family. improve the quality of the signal, which can significantly increase the accuracy in recognizing low-resource languages.

In the article by Miao et al. [20], it is developing an algorithm for recognizing digitized English speech and building a model for recognizing its features based on an analysis of the advantages and disadvantages of traditional methods, for example, time-frequency analysis of chaotic and speech signals is used. In this study, when processing speech signals, spectral-temporal and spectral-band analysis were used in parallel, which made it possible to form a trajectory of phonetic features. A speech recognition model with a recurrent neural network using a connectionist temporal classification algorithm is used to ensure the alignment of incoming speech signals [21]. The results of the study Wang [22] confirmed that the accuracy of speech recognition depends on the number of training samples used. Thus, the more samples tested, the better the learning rates. Teaching the Uzbek language with the help of intelligent systems is undergoing significant changes, since today not only new technologies are being actively developed, but databases are also expanding, including terminological ones, so the efficiency of applications is becoming higher.

The following aspects are considered in the work of Kaliev [23]: modern approaches to speech synthesis, the development of a prosodic method for low-resource languages, an acoustic processing method to improve the accuracy of acoustics and smoothness of speech, and the creation of software tools for deep machine learning in the Kazakh language. It should be noted that the rhythmic features of the language play an important role in determining sounds, for example, the fusion of pronunciation of words or fluent speech requires the use of additional resources to identify phonetic and other features [24]. TriNNOnto hybrid approach to automatic speech recognition combines various other methods such as language, acoustic, and feature modelling based on the use of the Deep Neural Network. The accuracy of the data acquisition strategy Deepak et al. [25] was estimated at 98.15% and 95.18% for two datasets: CMUKids and TIMIT respectively, word errors were low. The percentage of recognition accuracy of phonetic features in the lexemes of the Uzbek language is in the range of 75-100%.

Aldarmaki et al. [26] use an automatic speech recognition system for training, which can provide high performance when applied to numerous input data with manual transcription of speech. This paper discusses the limitations and requirements for speech recognition that could be used to optimize automatic speech recognition resources. In order to improve performance in recognizing the Uzbek language, it is necessary to expand the base of speech samples, including non-native speakers, as well as pay attention to different accents, intonations, and features of lexical stress [27]. Documenting languages helps prevent and halt the process of extinction of endangered languages. The use of uncontrolled segmentation of words is associated with cutting fragments of the utterance into smaller sound segments. In the article by Boito et al. [28] the results on the application of Bayesian models for Finnish, Romanian, Hungarian, and Russian are provided. The Uzbek language is not at risk of extinction, but is of little resource, since the creation of Uzbek language corpora takes time and resources. That is why text recognition for the purpose of phonetic analysis is one of the key tasks of automatic speech processing.

The work of Naik [29] uses a linear predictive coding method, which is being developed to provide a speech recognition system using a hidden Markov model for robotics. After testing the machine, problems related to the accuracy and estimation of recognition parameters were fixed. After training with the help of the intelligent system of the Uzbek language, it is necessary to analyse data on the effectiveness in terms of the formation of phonetic skills [30]. The speech recognition process is based on converting the input sound into a phonemic sequence, and then outputting the data into text format using language models. In the article by Oh et al. [31], a hierarchical method for clustering phonemes and identifying generated phoneme groups is proposed. The performance of models using these methods improved by 3% for fricatives, 2.1% for affricates, 6.0% for stops, and 2.2% for nasals. The basis of competent recognition is the correct distinction between phonemes that have similar characteristics. Even the most efficient phoneme classification methods cannot provide complete accuracy in identifying sounds [32].

Gros et al. [33] use a phonetically balanced subset of sentences method, since speech cues play an important role in the design of a speech corpus in both recognition and synthesis of a speech stream. Teaching Uzbek as a foreign language showed that the semantic mark-up of the speech corpus is important for identifying lexemes and phrases, which has a positive effect on the process of automatic processing of incoming speech signals. Today, many opportunities have appeared related to the processing and recognition of high-resource languages (English, Chinese), but at the same time, for languages with a low resource, the possibilities remain limited due to the lack of databases, spelling, and pronunciation [34]. The work of Du et al. [35] investigates the Uighur, Kazakh, and Kyrgyz languages belonging to the Altaic language family, and identifies methods for speech recognition that are effective for each of them individually and for all together, since they have their phonological and acoustic properties. The creation of an intelligent system for teaching the Uzbek language is aimed at the automatic processing of the student's speech signals, taking into account phonological and acoustic parameters, as well as feedback through a chatbot [36].

Kipyatkova [37] implements in her research the method of statistical and syntactic analysis of texts, which makes it possible to take into account the grammatical relationships that arise between words during recognition, expanding the language model. In addition, the achievement of the researcher was the recognition of continuous speech phrases in the form of an audio signal coming from a microphone or from a database. The results of the research showed high efficiency in recognizing similar sounds and productivity in segmenting the speech stream into phonemes and converting them into phonemic sequences. The method of non-parametric Bayesian models, according to Ondel et al. [38], in order to automatically detect acoustic data in unallocated corpora, offers alternatives for outputting sounds, which has a positive effect on the processing of large databases. The Uzbek language is limited in terms of data set, since it could not develop naturally for a long time, but the use of Bayesian models can be useful in terms of building rhythmic and intonational parameters. Polák et al. [39] develop a speech recognition pipeline consisting of an acoustic model and a phoneme-grapheme model. Such a system is superior to automatic speech recognition. The development of an intelligent system for teaching the Uzbek language was implemented on the basis of speech signal recognition technology using wavelet technologies, which allow eliminating noise by shifting sound waves [40, 41].

Thus, in the course of comparing the results of the research with the results of other modern researchers, it was concluded that the technologies and methods for converting speech into text are limitless, but the indicator of high or low resource of the language plays the most important role in the creation of intelligent systems. The more phonetic features are known, the clearer the classification parameters, the easier and faster it is possible to obtain information about the speech signal in its decomposition into syntaxemes, lexemes and phonemes.

4. CONCLUSIONS

The study's results, emphasizing the intelligent system's accuracy in phoneme recognition of Uzbek speech, carry implications not only for Uzbek language learning but also for the broader realm of automated speech recognition (ASR), particularly for low-resource languages. Achieving an accuracy range of 74-100% (with average values between 85-90%) for a low-resource language like Uzbek sets a promising precedent for other under-represented languages. This implies that with the right techniques, ASR systems can achieve high performance even when they don't have access to extensive linguistic data.

The fact that a high level of accuracy was achieved despite the scarcity of resources for the Uzbek language signifies the potential of using innovative techniques and technologies in ASR. This might inspire developers to look beyond data quantity and focus on optimizing methodologies, potentially leveraging synthetic data generation or data augmentation techniques. The research indicates that technologies and methodologies from high-resource languages can be successfully integrated and adapted to low-resource contexts. This suggests that advancements in ASR for languages like English and Chinese can pave the way for improvements in less commonly studied languages.

The study highlights the potential for advanced phonetic segmentation and noise reduction in ASR for low-resource languages. Effective phonetic segmentation can lead to more accurate transcription, while noise reduction can make the system more robust in varied real-world scenarios.

One of the implications of the study is the emphasis on creating a more substantial corpus for the Uzbek language and building lexical and grammatical dictionaries. Such resources can significantly enhance ASR systems, making them more comprehensive and accurate. It also underscores the importance of resource development for any language seeking advancements in ASR. The study points towards the development of all-encompassing ASR systems that not only recognize speech but also offer reading, learning, and evaluative features for the Uzbek language. This holistic approach could be a template for ASR systems for other low-resource languages.

The comparative analysis with works of computational linguists from English, Chinese, Uzbek, and Czech backgrounds underlines the universal applicability of ASR research. This could encourage more cross-linguistic and interdisciplinary collaborations, leading to global advancements in the field. With the demonstrated effectiveness of the intelligent system, there's potential for its commercial application. As more people seek to learn languages, especially less commonly taught ones, such systems can find a market, especially in ed-tech and communication sectors.

In conclusion, the findings of this research significantly contribute to the domain of ASR for low-resource languages. They underscore the potential for high performance in speech recognition, even with limited linguistic resources. Moreover, the study lays a foundation for future research and development in the area, emphasizing the creation of linguistic resources, advanced segmentation techniques, and comprehensive ASR solutions.

REFERENCES

[1] Lemyk, I. (2022). Innovative methods of teaching classical languages (using the example of Latin). Scientific Bulletin of Mukachevo State University. Series “Pedagogy and Psychology”, 8(2): 18-24. https://doi.org/10.52534/msu-pp.8(2).2022.18-24

[2] Mukhamadiyev, A., Mukhiddinov, M., Khujayarov, I., Ochilov, M., Cho, J. (2023). Development of language models for continuous uzbek speech recognition system. Sensors, 23(3): 1145. https://doi.org/10.3390/s23031145

[3] Khuda, A.O. (2020). An intelligent speech recognition system for learning foreign words based on machine learning. Tesis. National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

[4] Mamyrbayev, O.Z., Alimhan, K., Amirgaliyev, B., Zhumazhanov, B., Mussayeva, D., Gusmanova, F. (2020). Multimodal systems for speech recognition. International Journal of Mobile Communications, 18(3): 314-326. https://doi.org/10.1504/IJMC.2020.107097

[5] Mirzayevna, B.S. (2023). Phonological features of the uzbek language. O'zbekistonda Fanlararo Innovatsiyalar Va Ilmiy Tadqiqotlar Jurnali, 2(17): 233-237.

[6] Papadourakis, V., Müller, M., Liu, J., Mouchtaris, A., Omologo, M. (2021). Phonetically induced subwords for end-to-end speech recognition. Interspeech, 1992-1996. https://www.isca-speech.org/archive/pdfs/interspeech_2021/papadourakis21_interspeech.pdf.

[7] Abudaqa, A., Hilmi, M.F., AlMujaini, H., Alzahmi, R.A., Ahmed, G. (2021). Students’ perception of e-Learning during the Covid Pandemic: A fresh evidence from United arab emirates (UAE). Journal of E-learning and Knowledge Society, 17(3): 110-118. https://doi.org/10.20368/1971-8829/1135556

[8] Pushkar, O., Hrabovskyi, Y. (2019). Methodology for developing an intelligent user interface for educational publications in the e-learning system. Development Management, 17(3): 23-34. http://doi.org/10.21511/dm.17(3).2019.03

[9] Savenkova, O.A., Karpov, O.N. (2008). Technology for building an intelligent speech recognition system. Piece Intelligence, 4: 785-795. http://dspace.nbuv.gov.ua/handle/123456789/7672.

[10] Sakhipov, A., Yermaganbetova, M., Latypov, R., Ualiyev, N. (2022). Application of blockchain technology in higher education institutions. Journal of Theoretical and Applied Information Technology, 100(4): 1138-1147.

[11] Nurdaulet, I., Talgat, M., Orken, M., Gulzat, Z. (2018). Application of fuzzy and interval analysis to the study of the prediction and control model of the epidemiologic situation. Journal of Theoretical and Applied Information Technology, 96(14): 4358-4368.

[12] Karhina, N. (2023). Using interactive teaching methods in English lessons. Scientific Bulletin of Mukachevo State University. Series “Pedagogy and Psychology”, 1(9): 9-15. https://doi.org/10.52534/msu-pp1.2023.09

[13] Novikov, L.V. (1999). Fundamentals of wavelet analysis of signals. Publishing House OOO “MODUS+”, Saint Petersburg.

[14] Tserklevych, V., Prokopenko, O., Goncharova, O., Horbenko, I., Fedorenko, O., Romanyuk, Y. (2021). Virtual museum space as the innovative tool for the student research practice. International Journal of Emerging Technologies in Learning (iJET), 16(14): 213-231.

[15] Balykbayev, T.A.K.I.R., Bidaibekov, E.S.E.N., Grinshkun, V.A.D.I.M., Kurmangaliyeva, N.U.R.G.U.L. (2022). The influence of interdisciplinary integration of information technologies on the effectiveness of it training of future teachers. Journal of Theoretical and Applied Information Technology, 100(5): 1265-1274.

[16] Odeh, A.H., Odeh, M., Odeh, H., Odeh, N. (2023). Using natural language processing for programming language code classification with Multinomial Naive Bayes. Revue d'Intelligence Artificielle, 37(5): 1229-1236. https://doi.org/10.18280/ria.370515

[17] Kerimkhulle, S., Koishybayeva, M., Slanbekova, A., Alimova, Z., Baizakov, N., Azieva, G. (2023). Created and realization of a demographic population model for a small city. Proceedings on Engineering Sciences, 5(3). https://doi.org/10.24874/PES05.03.003

[18] Ziółko, B., Manandhar, S., Wilson, R.C., ZIÓŁKO, M. (2011). Phoneme segmentation based on wavelet spectra analysis. Archives of Acoustics, 36(1): 29-47. http://doi.org/10.2478/v10168-011-0003-2

[19] Morikawa, M., Hernane Spatti, D., Dajer, M.E. (2022). Wavelet packet transform and multilayer perceptron to identify voices with a mild degree of vocal deviation. Revista de Investigación e Innovación en Ciencias de la Salud, 4(1): 16-25. https://doi.org/10.46634/riics.126

[20] Miao, Y., Liu, H., Gu, S. (2022). English speech feature recognition-based fuzzy algorithm and artificial intelligent. Wireless Communications and Mobile Computing, 2022: 1-10. https://doi.org/10.1155/2022/4421520

[21] Opitasari, O., Yaddarabullah, Y., Sensuse, D.I. (2023). Employee welfare financing system with support vector machine and Naïve Bayes to Syariah banking. In AIP Conference Proceedings. AIP Publishing, 2482(1). https://doi.org/10.1063/5.0111450

[22] Wang, S. (2023). Recognition of English speech-using a deep learning algorithm. Journal of Intelligent Systems, 32(1): 20220236. https://doi.org/10.1515/jisys-2022-0236

[23] Kaliev, A. (2019). Speech synthesis based on deep machine learning. ITMO University, Saint Petersburg.

[24] Sakhipov, A., Yermaganbetova, M. (2022). An educational portal with elements of blockchain technology in higher education institutions of Kazakhstan: Opportunities and benefits. Global Journal of Engineering Education, 24(2): 149-154.

[25] Deepak, G., Surya, D., Trivedi, I., Kumar, A., Lingampalli, A. (2022). An artificially intelligent approach for automatic speech processing based on triune ontology and adaptive tribonacci deep neural networks. Computers & Electrical Engineering, 98: 107736. https://doi.org/10.1016/j.compeleceng.2022.107736

[26] Aldarmaki, H., Ullah, A., Ram, S., Zaki, N. (2022). Unsupervised automatic speech recognition: A review.

Speech Communication, 139: 76-91. https://doi.org/10.1016/j.specom.2022.02.005

[27] Sandra, L., Heryadi, Y., Suparta, W., Wibowo, A. (2021). Deep learning based facial emotion recognition using multiple layers model. In 2021 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA), Surabaya, Indonesia, pp. 137-142. https://doi.org/10.1109/ICAMIMIA54022.2021.9809908

[28] Boito, M.Z., Yusuf, B., Ondel, L., Villavicencio, A., Besacier, L. (2021). Unsupervised word segmentation from discrete speech units in low-resource settings. arXiv Preprint arXiv: 2106.04298. https://doi.org/10.48550/arXiv.2106.04298

[29] Naik, A. (2021). HMM-based phoneme speech recognition system for the control and command of industrial robots. Technical Transactions, 118(1). https://doi.org/10.37705/TechTrans/e2021002

[30] Abudaqa, A., Al Nuaimi, S., Buhazzaa, H., Al Hosani, S. (2021). Examining the significance of internal mobility hiring in determining the individual vs organizational outcomes: An empirical investigation from ADNOC FURSA platform during recent pandemic of COVID-19. In Abu Dhabi International Petroleum Exhibition and Conference, SPE, pp. D031S080R001. https://doi.org/10.2118/207535-MS

[31] Oh, D., Park, J.S., Kim, J.H., Jang, G.J. (2021). Hierarchical phoneme classification for improved speech recognition. Applied Sciences, 11(1): 428. https://doi.org/10.3390/app11010428

[32] Humeniuk, O.I. (2018). Digital communication space research in the education reform context. Ов Суший Соціальна Психологія Націєтворення: Концептуальні Засади І Методологічні Принципи Дослідження, 41(44): 202. https://lib.iitta.gov.ua/id/eprint/715297.

[33] Gros, J.Z., Vesnicer, B., Dobrisek, S. (2022). A method for selection of phonetically balanced sentences in read speech corpus design. In Proceedings of the 30th European Signal Processing Conference, EUSIPCO. 1136-1139.

[34] Balykbayev, T., Issabayeva, D., Rakhimzhanova, L., Zhanysbekova, S. (2021). Distance learning at KazNPU named after Abai: Models and technologies. In 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, pp. 1-6. https://doi.org/10.1109/SIST50301.2021.9465980

[35] Du, W., Maimaitiyiming, Y., Nijat, M., Li, L., Hamdulla, A., Wang, D. (2022). Automatic speech recognition for uyghur, kazakh, and kyrgyz: An overview. Applied Sciences, 13(1): 326. https://doi.org/10.3390/app13010326

[36] Aliaskar, M., Mazakov, T., Mazakova, A., Jomartova, S., Shormanov, T. (2022). Human voice identification based on the detection of fundamental harmonics. In 2022 IEEE 7th International Energy Conference (ENERGYCON), Riga, Latvia, pp. 1-4. https://doi.org/10.1109/ENERGYCON53164.2022.9830471

[37] Kipyatkova, I.S. (2011). Methods and software for phonetic-linguistic modelling in automatic Russian speech recognition systems (Unpublished doctoral dissertation). ITMO University, Saint Petersburg.

[38] Ondel, L., Burget, L., Černocký, J. (2016). Variational inference for acoustic unit discovery. Procedia Computer Science, 81: 80-86. https://doi.org/10.1016/j.procs.2016.04.033

[39] Polák, P., Sagar, S., Macháček, D., Bojar, O. (2020). CUNI neural ASR with phoneme-level intermediate step for~non-native~SLT at IWSLT 2020. In Proceedings of the 17th International Conference on Spoken Language Translation, pp. 191-199. http://doi.org/10.18653/v1/2020.iwslt-1.24

[40] Lend'el-Syarkevych, A.A. (2016). Modernization of higher pedagogical education in conditions of educational space globalization. Scientific Bulletin of Mukachevo State University. Series "Pedagogy and Psychology", 2(1): 46-51.

[41] Dilekh, T., Boulahia, M.A., Benharzallah, S. (2023). Assessing semantic similarity measures and proposing a WuP-Resnik hybrid metric for enhanced Arabic language processing. Revue d'Intelligence Artificielle, 37(5): 1311-1322. https://doi.org/10.18280/ria.370524