BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Orthoepic-prosodic Foundations of the Kazakh Speech Synthesis

Zhumabayeva Zhanar , ***Fazylzhanova Anar, Bazarbayeva Zeinep, Amanbayeva Aisaule*** * , ***Ospangaziyeva Nazgul***

The Institute of Linguistics named after A. Baitursynuly, Almaty 050010, Kazakhstan

ABSTRACT

This article examines the issues related to prosodic and orthoepic norms in Kazakh speech synthesis. To ensure that the synthesizer delivers speech that is both realistic and intelligible, the text must be systematized according to orthoepic standards, with changes in phonemes, vowel reductions, and various sound phenomena described on the basis of linguistic data. The article also outlines the relevance of speech synthesis and the methods employed, while identifying their distinctive features. Furthermore, contemporary speech synthesis programs are discussed, along with their advantages and drawbacks. The article particularly focuses on enhancing speech synthesis by ensuring that narrators read texts in accordance with orthoepic norms and accurately convey prosodic features. The study uses the 11th-grade *Kazakh Literature* textbook, comprising 62 pages, as its source material. The text was internally segmented into syntagms, and the analysis addressed aspects such as vowel harmony, consonant compatibility, changes between roots and affixes, shifts across rhythmic groups, assimilation, dissimilation, reductions, variations and variants, elision, and other relevant features — all presented in accordance with orthoepic norms. The article analyzes articulatory, formant-based, parametric, and neural models used in the implementation of word synthesis. In order to improve the quality of speech synthesis in the Kazakh language, the study highlights the necessity of expanding abbreviations and numerals, adhering to orthoepic norms, and accurately modeling intonation and rhythmic patterrns. The research findings provide a scientific foundation for developing high-quality speech synthesis based on the phonetic and phonological regularities of the Kazakh language.

Keywords: Orthoepy; Prosody; Intonation; Reduction; Variant; Phonetics; Spoken word; Algorithm

*CORRESPONDING AUTHOR:

Amanbayeva Aisaule, The Institute of Linguistics named after A. Baitursynuly, Almaty 050010, Kazakhstan; Email: Aaisaule@mail.ru

ARTICLE INFO

Received: 23 July 2025 | Revised: 28 August 2025 | Accepted: 5 September 2025 | Published Online: 3 November 2025
DOI: https://doi.org/10.30564/fls.v7i12.11233

CITATION

Zhanar, Z., Anar, F., Zeinep, B., et al., 2025. Orthoepic-prosodic Foundations of the Kazakh Speech Synthesis. Forum for Linguistic Studies. 7(12): 36–48. DOI: https://doi.org/10.30564/fls.v7i12.11233

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (https://creativecommons.org/licenses/by-nc/4.0/).

1. Introduction

Today, adapting technology to human needs — specifically, enabling computers to speak written text — has become a rapidly developing trend – in line with contemporary demands. In this context, the need to synthesize speech from Kazakh-language texts has emerged as a pressing requirement. Speech synthesis, after all, refers to the correct articulation of written text in terms of both orthoepy and prosody — that is, its transformation into spoken language.

The history of speech synthesis dates back to the 10th century. In the mid-13th century, the first known example was created by the monk Albertus Magnus (Albert von Bollstädt) and the English philosopher Roger Bacon, who constructed a prototype known as the "talking head." By the end of the 18th century, the German scholar Christian Kratzenstein developed a model for five sustained vowel sounds (a, e, i, o, u) based on real human speech. In 1788, building upon his research, a mechanical-acoustic speaking machine was created using models of the lips and tongue, capable of producing specific sounds. Remarkably, this machine could mimic the voice of a child aged three or four.

Introduction to the Problem of the Study

Subsequently, in 1837, a more advanced version was developed by Charles Wheatstone, capable of producing both vowel and consonant sounds. In 1846, Joseph Faber further extended the possibilities of speech synthesis by demonstrating that machines could be designed not only to produce speech but also to sing. As research expanded, by the end of the 19th century, Alexander Graham Bell and Wheatstone develop a mechanical speaking machine model. The 20th century marked the beginning of the electronic era, in which scientists began using sound wave generators and working toward creating algorithmic models of speech.

In 1920, Homer Dudley, an engineer at a major American company, invented a device called the VOCODER (from *voice* and *coder*), which could be operated using a keyboard. Originally designed for broadcasting purposes at radio stations, the VOCODER was later adapted for use in various musical genres. The earliest phrase-based speech synthesizers produced highly artificial output. However, over time and through ongoing experimentation, these synthesizers were gradually refined to approximate natural speech more closely. Eventually, an optimized version of Dudley's invention, known as the VODER, was presented at the 1939 World's Fair in New York.

By the late 1950s, the first speech synthesis system was developed using computational technology, and in 1968, a "text-to-speech" synthesizer was constructed. At the beginning of the 1960s, speech synthesis began to be studied from an acoustic perspective. In the 1970s, in an effort to improve speech synthesis, scientists from Moscow, Leningrad, Minsk, and Tallinn collaborated to develop a phoneme-formant synthesizer based on the principles of formant synthesis. This resulted in the creation of the first device called *Phonemophone-1*, and later its improved version, *Phonemophone-3*, which eventually entered industrial use.

Researchers examined automated speech synthesis from articulatory and acoustic perspectives. Articulatory synthesis was implemented using mechanical synthesizers that mimicked the way sounds are physically formed in the vocal tract, whereas acoustic synthesis employed various types of electronic synthesizers to generate speech signals. Research in speech synthesis continued to progress and deepen. For instance, from the 1980s onward, advances in modern technology enabled a wide range of experimental investigations [1].

Speech synthesis is the outcome of technological and linguistic research aimed at modeling the natural human voice. Through the artificial construction of speech signals, written texts are transformed into spoken language. To ensure that speech synthesis is both high-quality and realistic, several factors must be taken into account: the texts should represent diverse subject areas, they must be normalized; and the audio recordings of a speaker must be aligned with the corresponding texts and uploaded into the software system. Speech synthesis generally aims to achieve two key objectives: clarity and naturalness.

To achieve this goal, it is essential to consider the following linguistic aspects: the positional patterns of the smallest segments in Kazakh texts, the articulatory, acoustic, and perceptual characteristics of individual sounds; the phonetic changes that occur within words and at word boundaries; the manifestation of intonemes in spontaneous speech; and the prosodic features of continuous speech flow. This is because speech synthesis is based on phonetic laws and aims to imitate human speech patterns with the highest degree of

naturalness and precision. Phonetic laws regulate the functioning and evolution of a language's sound system, ensuring the stability of phonological units while also governing their continuous variation, mutual alternation, and combinatory potential [2].

At present, there are several efficient speech synthesis systems in operation. One notable example is the demo program titled *Kazakh Text-to-Speech 2 (Kazakh TTS2)* developed by the Institute of Smart Systems and Artificial Intelligence (ISSAI) at Nazarbayev University. This program features three female and two male voices, with each speaker contributing over 25 hours of recorded speech data. The demo version is available on the Institute's official website. However, it currently lacks the capability to process numbers and abbreviations.

In parallel, the Institute of Informatics and Mathematics has developed *Kazakh ASR*, a speech recognition program functioning within a Telegram bot, which performs the opposite task by converting spoken language into written text. Another speech synthesis system developed by the Til-Qazyna National Scientific and Practical Center also operates within a Telegram bot and is characterized by its reliance on orthographic pronunciation, meaning it vocalizes words exactly as they are written.

Considering these available systems, the issue of ensuring phonetic accuracy — especially in accordance with normative Kazakh orthoepy — remains a pressing priority in the development of Kazakh-language speech synthesis. To this end, the phonetic and prosodic properties of the Kazakh language play a vital role.

The function of speech synthesis extends to a variety of fields, including audio advertising and marketing (where TTS can generate voice messages tailored for media channels); education and learning (TTS assists in the creation of audio materials and educational resources to facilitate effective learning and information acquisition); interactive voice systems (TTS simplifies client interaction through voice menus and automated voice responses); and multimedia platforms (TTS enhances user engagement by creating realistic voice characters for video games, animations, and audio tours) [3].

In order for synthesized speech to sound natural — with coherent prosody and authentic sound combinations and modifications — it is essential to first account for the phonetic characteristics of the language, the positional behavior of phonemes, and the orthoepic norms of words. This approach improves the overall quality of speech synthesis, ensures efficient performance, preserves the phonological harmony and melodic flow of spoken language, and results in a more authentic sound that reflects the unique features of the Kazakh language.

In the current era of rapid advancement in information technologies and artificial intelligence, the development and enhancement of speech technologies across various languages have become one of the key priorities. Among them, the development of Kazakh speech synthesis plays a crucial role in promoting and expanding the use of the state language in the digital space.

However, the creation of natural, authentic, and orthoepically accurate Kazakh speech synthesis still requires comprehensive research. Although existing synthesizers are capable of producing basic vocal output, they often fail to fully reflect prosodic features such as rhythm and intonation, as well as phonetic-phonological phenomena like vowel and consonant harmony, assimilation, and reduction [4].

In this regard, the issue addressed in this article is highly relevant to the current technological demands. High-quality speech synthesis is not merely the conversion of text into voice, but a complex linguistic-acoustic process aimed at generating speech that is both intelligible and natural for listeners. Furthermore, the development of accurate Kazakh speech synthesizers opens up broad opportunities for application in fields such as education, assistive technologies, mobile applications, and voice-based user interfaces.

The goal is to develop high-quality and efficient synthesis technologies for Kazakh speech synthesis that produce natural, authentic, and orthoepically accurate speech by taking into account phonetic and phonological rules. This aim is to expand the use of the state language in the digital space and enhance the capabilities of Kazakh speech synthesis in fields such as education, assistive technologies, mobile applications, and voice-based interfaces [5].

The theoretical significance of this article lies in its phonetic investigation of modern Kazakh speech, both at the segmental and suprasegmental levels, as well as in its comprehensive study of Kazakh word orthoepy and intonation. The harmony of vowels and the compatibility of consonants in Kazakh, along with phonetic phenomena such as reduction, assimilation, and dissimilation, contribute to

systematizing the orthoepic norms of Kazakh speech. In order for Kazakh speech synthesis to preserve its natural form and produce authentic vocalizations — i.e., to maintain its orthoepic norms — it is necessary to formulate clear rules.

The practical significance of the article lies in its potential to support the work of researchers and software developers engaged in speech synthesis. Studies in this field, aligned with current advances in technology, contribute to the enhancement of artificial intelligence performance.

2. Materials and Methods

Materials. The demo program "Kazakh Text-to-Speech Conversion — 2 (Kazakh TTS2)," developed by Nazarbayev University's Institute of Smart Systems and Artificial Intelligence (ISSAI), and the Kazakh ASR program created by specialists from the Institute of Informatics and Mathematics in the field of speech recognition, were utilized alongside theoretical studies and textbook materials to substantiate the phonetic characteristics of speech synthesis. The research material involved the creation of an algorithm for modeling the changes occurring within words, at word boundaries, and between rhythmic groups, as well as the annotation of prosodic features, to ensure authentic synthesis of speech in accordance with orthoepic and prosodic norms. Specifically, the text of the 10th-grade *Kazakh Literature* textbook was analyzed from the perspectives of orthoepy and prosody to provide a reliable phonetic foundation for speech synthesis. The aim was to address the prosodic labeling of the text and capture the nuanced sound variations that emerge naturally in spoken language. In constructing speech synthesis, two different approaches are employed.

Linguistic sources — such as textbooks, language programs, and linguistic studies in Kazakh, Russian, and other languages — serve as essential foundations for understanding key linguistic concepts like sound, phoneme, variant, variation, assimilation rules, and orthoepy. These materials contribute to the linguistic substantiation of speech synthesis and provide the necessary data for modeling and implementing it accurately in the Kazakh language.

The concatenative synthesis method is based on recording and storing individual sound samples, voiced and read by a speaker (a voice talent or announcer). These samples are adapted according to the variants and variations of sounds and their phonetic changes in actual speech. This method pays particular attention to the dynamic behavior of vowels and consonants, making it suitable for modeling natural speech variability.

The formant-based synthesis method operates primarily through an intonational model. In this approach, speech synthesis emphasizes the prosodic annotation of the text, which includes melody (pitch contour), tempo (speech rate), and pause marking. It also involves segmenting the text into meaningful prosodic units, or syntagmas, based on semantic and syntactic structure.

Parametric synthesis is another method characterized by high-quality speech output in contexts with limited and repetitive message sets. However, its applicability is restricted to predefined messages and cannot be flexibly adapted to novel or spontaneous text inputs.

Rule-based full speech synthesis allows the generation of speech from previously unknown texts by controlling all parameters of the speech signal. This method includes two distinct subtypes:

– *Articulatory synthesis*, which models the physical movements of human speech organs (tongue, lips, glottis), simulating how speech is produced anatomically.

– *Formant synthesis*, which is based on computing speech formants — acoustic resonances of the vocal tract — modeling sounds using acoustic parameters such as formant frequencies, taking into account the phonetic specifics of the language.

Neural network-based synthesis, exemplified by Tacotron and VITS, incorporates stress patterns, intonation, and grammatical features of the language. These approaches are also used by developers to encode linguistic material into algorithms and computational models for text-to-speech (TTS) applications.

The article mainly draws upon *articulatory and formant-based approaches* for adapting linguistic material into speech synthesis. In addition, a *phoneme-based synthesis method* is employed, in which words are segmented into phonemes, and their possible transformations in speech are identified. The input text for speech synthesis is divided into multiple hierarchical units, such as syntagmas, rhythmic groups, prosodic patterns, and intonational contours.

The *intonational structure of the text* is particularly important in enhancing speech naturalness. In this regard,

the input text is segmented into *rhythmic groups* and *syntagmas*. A rhythmic group refers to the smallest semantic unit of speech, unified by a primary stress falling on its final syllable. A syntagma is the smallest spoken unit of speech that demonstrates the interrelation between syntactic structure and intonation.

Previous research on speech synthesis has primarily focused on technical and algorithmic aspects. Specifically, these studies emphasize engineering solutions such as speech signal processing, voice data collection, and model training. In contrast, the present study approaches the issue from a linguistic perspective, aiming to lay the theoretical foundation for speech synthesis based on the phonetic and phonological features of the Kazakh language as well as its orthoepic norms. In this regard, the study highlights the interconnection between earlier technical research and linguistic approaches, paving the way for the development of a high-quality and natural-sounding Kazakh-language speech synthesizer.

Considering all the above methodologies, the article examines the material used in speech synthesis through the lenses of *prosody*, *orthoepic norms*, and *phonetic characteristics*, with the aim of developing a linguistically grounded and acoustically natural Kazakh TTS system.

3. Results and Discussion

Research on speech synthesis has been widely explored across the globe, particularly in Western countries such as the United States, Japan, and France. In Russian linguistics, investigations into speech synthesis began as early as 1971, with foundational works by M.F. Derkach[6], V.N. Sorokin[7], and G. Fant[8]. Subsequent studies conducted in laboratories at the Royal Institute of Technology in Stockholm further advanced the field through the contributions of L.V. Bondarko[9], S.V. Golubtsova[10], L.V. Zlataustov[11], J.L. Flanagan[12], and S. Rybin[13], laying the groundwork for resolving various aspects of speech synthesis. Research in experimental phonetics related to speech synthesis has also utilized the work of scholars such as B.M. Lobanov[14], L.I. Tsirulnik, and E.Yu. Kyunnap[15].

In Kazakh linguistics, research on speech synthesis originates from the works of Ä. Jünisbek, Z.M. Bazarbayeva, and A. Fazylzhan. The results of these scholars' studies form the theoretical basis for the advancement of contemporary speech synthesis in the Kazakh language. Accordingly, the current article relies on several key works, including Ä. Zhunisbek's[16] *Issues in Kazakh Linguistics*, Z.M. Bazarbayeva's *Foundations of Kazakh Phonology*[17], and the research of A. Fazylzhan. In particular, Ä. Jünisbek's research was used to describe the acoustic and articulatory properties of sounds, while Bazarbayeva's conceptualizations were instrumental in outlining the prosodic features of speech synthesis. Discreteness also plays a vital role in the development of Kazakh speech synthesis, and in this respect, A. Fazylzhan's research on the discrete nature of speech synthesis provided a crucial foundation.

When considering speech synthesis from a phonetic-phonological perspective, several theoretical sources were consulted, including N. Wali's *Orthography, Orthoepy, and Script*, Ä. Jünisbek's *Kazakh Phonetics*, Z.M. Bazarbayeva's[18], *Intonology* and *Kazakh Phonology*, Q. Kúderinova's[19] monograph *History of Kazakh Script*; and A. Fazylzhan's[20] *Melody of Speech and Intonation*. These works collectively informed the theoretical grounding of the article.

For the implementation of speech synthesis, the article also draws upon material from the educational platform designed for public schools, specifically the 11th-grade textbook *Kazakh Literature* for the socio-humanitarian track[21].

The study of speech synthesis at the orthoepic and phonetic-phonological levels is closely tied to the oral speech culture of the Kazakh language and to its authentic sound patterns. The preservation of orthoepic norms provides an opportunity to understand national identity, cognition, and uniqueness. Therefore, in Kazakh speech synthesis, it is essential to maintain the principle of sound harmony, which serves as the core of oral speech. The implementation of speech synthesis is based on both the orthoepic norm and speech prosody. In preserving the orthoepic standard, the method of concatenative synthesis is employed.

To achieve a natural-sounding and high-quality speech synthesis, authentic spoken samples from a human speaker are required as foundational data. Naturally, it is impossible to synthesize speech by modeling and inputting each phoneme individually into the synthesizer. This is because when isolated sounds are combined, the resulting synthesized output lacks naturalness and sounds robotic and unintelli-

gible. This occurs due to the dynamic changes that sounds undergo in natural speech. For this reason, it is crucial to first identify and analyze the phonetic changes that occur at the boundaries between syllables and between roots and affixes. Recognizing these contextual variations is essential to developing a synthesis model that accurately reflects the natural flow and prosodic structure of spoken Kazakh.

The speaker reads the written text not according to orthography, but in accordance with orthoepic norms. To achieve this, the speaker must be instructed on the phonetic changes that occur during speech – namely, vowel harmony and consonant assimilation, morphophonemic alternations between root and suffix, assimilation, dissimilation, reduction, variants and variations, elision, and other relevant phenomena[21]. For example, when the verb *жан* (‘to burn’) is used in *жанды* (‘burned’), its original form is preserved. However, in combinations such as *жанбады* (‘did not burn’) and *жанған жоқ* (‘has not burned’), it undergoes changes and becomes *жамбады*, *жаңған жоқ*, respectively. Similarly, the rules of variant and variation in Kazakh must be incorporated into the speech synthesis program. For example: *айтшы–айтшы* (‘say!’), *айтса–айтса* (‘if (someone) says’), *өтсе–уөтсө* (‘if (someone) passes’), *кетсе–кетсе* (‘if (someone) leaves’), *басшы–башшы* (‘leader’), *қобызшы–қобышшы* (‘kobyz player’), *қабақ–қаб̂ақ* (‘eyebrow’), *обал–уоб̂ал* (‘sin’), *қыпшақ–қып̂шақ* (‘Kipchak’) and others[22].

In addition, in Kazakh, sounds such as *қ–ғ*, *п–б*, *б–п* undergo changes at the junctions of words under the influence of neighboring sounds during actual pronunciation. For instance, while in writing we have *ақ ешкі* (‘white goat’), *ақ жүн* (‘white wool’), *қара қой* (‘black sheep’), *жас бала* (‘young child’), in speech they are pronounced as *ағешкі*, *ағжүн*, *қарағой*, *жаспала*. Thus, when submitting written text to a synthesis program, it is necessary to account for these specific phonological patterns of Kazakh and incorporate orthoepic rules to ensure natural-sounding synthesized speech.

In other words, in order for the synthesizer to distinguish between written text and the phonetic transformations occurring in actual speech, it is necessary to adapt the orthoepic norms and morphophonemic changes (especially between root and suffix) into computer-readable form. Furthermore, in Kazakh, close vowels are often weakened and become obscure in pronunciation, undergoing reduction. For example, the words *адыраңдау* (‘to fidget’), *аңызақ* (‘dry wind’), *бадырақ* (‘bulging’), *дәрігер* (‘doctor’), *еділеңдеу* (‘to swagger’), *жадырап* (‘cheerfully’), *жұмылу* (‘to unite’), *жұдырық* (‘fist’), *иірім* (‘eddy’), *көбірек* (‘more’), *көбінесе* (‘mostly’), *көкірек* (‘chest’), *қатынас* (‘relation’), *қасірет* (‘grief’), *құдірет* (‘power’), *тәжірибе* (‘experience’) are pronounced in speech as *ад'раңдау*, *аң'зақ*, *бад'рақ*, *дәр'гер*, *ед'реңдеу*, *жад'рап*, *жұм'лу*, *жұд'рық*, *и'рім*, *көб'рек*, *көк'рек*, *көб'несе*, *қат'нас*, *қас'рет*, *құд'рет*, *тәж'рибе*, etc.

Thus, in order to create effective speech synthesis, it is necessary to systematically codify the rules governing reduction in Kazakh and integrate them into the synthesizer’s programming.

In addition, when the final sound of a word ends in a vowel and the initial sound of the following word also begins with a vowel, the phenomenon of elision — where one of the adjacent vowels is dropped — must also be accounted for by the speech synthesizer as a regular phonetic change in spoken language. For example: *қара ала* (‘black piebald’) becomes *қарала*; *ала алмады* (‘couldn’t take’) becomes *алалмады*; *айта алмайды* (‘cannot say’) becomes *айталмайды*; *жоғары іл* (‘hang high’) becomes *жоғаріл*; *жақсы өнер* (‘good art’) becomes *жақсөнер*; *жақсы өлең* (‘good poem’) becomes *жақсөлең*; *жиырма алты* (‘twenty-six’) becomes *жиырмалты*; *мына адам* (‘this person’) becomes *мынадам*; *ала ешкі* (‘spotted goat’) becomes *алешкі*, and so forth. In other words, although these word pairs are written separately in orthography, in actual pronunciation they are articulated in a single breath, and one of the consecutive vowels is elided.

Moreover, in contrast to elision, there exists an opposite phonological phenomenon known as apheresis, where both adjacent vowels are preserved and pronounced together in one breath to retain the integrity of the word stem. This too holds a significant place in speech synthesis. In Kazakh linguistics, such examples are rare and usually concern functional words such as *де*, *же*, *бұ(л)* (‘this’), *о(л)* (‘that’), *не* (‘what’), and the particles *да*, *де*, *та*, *те*. For example, we write and pronounce *не алдың?* (‘what did you take?’), *не істейін?* (‘what should I do?’) with both vowels intact. However, in the case of particles, the pronunciation differs. For instance, *жазып та алды* (‘wrote and took’) is pro-

nounced as *жазыпт'алды*, *тұрып та ішті* ('stood and drank') becomes *тұрыпт'ішті*, and *мыс па екен* ('was it copper?') becomes *мысп'екен*. Thus, while the meaning remains unchanged, one of the two consecutive vowels is dropped in speech[23].

Furthermore, when the final sound of a word is *с* or *з* and it is followed by a suffix beginning with *с* or *ш*, the sound *з* changes into *с*, and *с* changes into *ш* in spoken Kazakh. For example: *ауызша* ('oral') becomes *ауышша*; *жұмысшы* ('worker') becomes *жұмышшы*; *басшы* ('leader') becomes *башшы*; *сөзсіз* ('undoubtedly') becomes *сөссіз*; *жазсын* ('let him/her write') becomes *жассын*; *көзсіз* ('blind') becomes *көссіз*, and so on. Similarly, when the final sound of a word is *н* and it is followed by a suffix beginning with *г*, *г*, or *б*, the initial consonant of the suffix changes due to phonetic assimilation. For example: *күнге* ('to the sun') becomes *күңге*, *түнге* ('to the night') becomes *түңге*; *жанбайды* ('doesn't burn') becomes *жамбайды*; *сөнбейді* ('doesn't extinguish') becomes *сөмбейді*, *жанға* ('to the soul/person') becomes *жаңға*, and so forth[24].

Thus, in constructing a speech synthesis model, such phonological transformations governed by Kazakh's orthoepic norms must be encoded and simulated accurately in the system to produce natural-sounding synthesized speech.

In formant-based — or acoustic model-based — speech synthesis, neural networks are employed to model the signal. Since this method is directly tied to voice quality, it relies on a set of key prosodic parameters, such as pitch contour (i.e. the fundamental frequency), the number of formants, and the frequency of each formant. Due to the continuous change of sounds in natural speech, these parameters also change dynamically during articulation, with the exception of the formant count, which remains stable. Consequently, in constructing a high-quality synthesis, intoneme models are also integrated, based on how a human speaker would naturally read the text. For synthesized speech to achieve prosodic (intonational) quality, it must account for typical melodic contours (such as rising-falling, falling-rising, or level pitch patterns) as well as tempo variation (fast or slow pacing).

First, the text intended for the speaker is divided into syntagmatic units according to meaning, and then its corresponding intonational structures are identified and modeled. These include final, non-final, general question, specific question, exclamatory sentence, strict command, polite command, and parenthetical segment intonemes[24]. In Kazakh, intonemes indicate whether a sentence is complete or incomplete, what emotional stance it conveys, and whether the utterance represents a general or specific question. Therefore, the speech synthesizer must adapt to certain prosodic norms — for instance, a sentence-final drop in pitch or a mid-sentence rise in pitch to signal incomplete thought[25].

The non-final intoneme, of course, also depends on sentence type: in complex sentences, the first clause is usually pronounced with a non-final intonation, while the second clause concludes with a final intoneme. When reading compound sentences, the speaker starts with a non-final intoneme, transitions into a level intonation mid-sentence, and ends with a final intoneme.

For example:

↑алғашқы білімін / ауыл молдасынан алған →/ сәкен сейфуллийнді / әкесі ↑/ бір мың тоғыжжүз бес / бір мың тоғыжжүс сегізінші жылдары →// нілді зауытындағы / уорыс-қазағ мектебінде уоқытады↓//

('↑He received his first education / from a village mullah →/ Saken Seifullin / was sent by his father ↑/ in 1905 / and again in 1908 →// to study at the Russian-Kazakh school / in the town of Nildy↓//')

In complex sentences, the pitch rises and falls at the beginning, moves through a non-final intonation in the middle, and decreases toward the end. For example:

↑айша дастанының / бас кейіпкері →// батыр ғызымыз / мәншүғ мәметова болса↑// сұлтан дастаны да / қазақтың батыр тұлғалы →// намыст'ұлұ тұуралы↓//

('↑If the main character of the Aysha epic is →// our heroic daughter / Manshuk Mametova↑// then the Sultan epic is also →// about a brave Kazakh warrior↓//')

A general question intoneme is characterized by a rising pitch at the end of the syntagm. For example:

↑намысқа шабар алысымыз / болып тұр ма↑// біз ғазір / апан-апаңға геліп →// тығылған / бөрү тәріздіміз↓//

('↑Are we heading into a fight for our honour↑// Now we're like / wolves hiding →// in our dens↓//')

A specific question intoneme also exhibits a rise in pitch. For example:

↑ал генерал-ғүубөрнаторлар / алғашқы қойған тілектерін / қабыл йетпесе ↑// не істейді↑// уонда /

шешіңген сұудан тайымбайды →// соғұс ашады↓//

(‘↑And if the governors / do not accept / the initial demands↑// what will they do↑// then / they won’t hesitate to get wet once undressed →// they will declare war↓//’)

An exclamatory intoneme, reflecting heightened emotion, is marked by a noticeably raised pitch. For example:

↑сонсоң / көзүнө құйұлған / қара терді →// сұқ̊ саусағын / йекі б̄үктөп // сыпырып тастады д̄а↑// ішінен тағы / ах↑// деді↓//

(‘↑Then / with his index finger / he wiped away the dark sweat →// that had poured into his eyes / folded his finger in two↑// and again, from within / said *Ah*↑//↓’)

The strict command intoneme implies that the order must be fulfilled and is pronounced with a low tone. For example:

мені / ханымыз деб̄ұқ̊сұн↓/// ↑айтқаныма гөнсүн↑// айдауұма жүрсүн↓//

(‘Let them / acknowledge me as their khan↓/// ↑Let them obey↑// and go where I order↓//’)

The polite command intoneme is not as obligatory in tone and implies a request, wish, or suggestion, rather than a strict demand. For example:

гаврыйловтұң уөзүнүң / басын алұу герек↑// деді //↓ кенет қатұуланып кетіп↑// және уөзгөлөрге сабақ полсұн↑// бұл үкүмдү уосұндағы / қашқ̊ын солдаттардың г̄өзүншө // уорұндаған дұрұс↓//

(‘He said↑// Gavrilyov must be beheaded↓// then suddenly hardened↑// Let it be a lesson to others↑// it is better to carry out this sentence / in front of the fugitive soldiers in the camp↓//’)

Finally, the parenthetical segment intoneme adds explanatory or supplementary information and is usually uttered quickly with a raised pitch, conveying that the material is secondary in importance. For example:

↑шынында да↑// ресей патшйалығына б̄ағындым деп // бітім істей тұрұп / кенесарының уөзүн / хаң гөтөртүүү↑// йел ғамын йемес↑// уөз ғамын уойлауұ йеді↓//

(‘↑Indeed↑// although he pretended to have submitted to the Russian Empire // and made peace / Kenesary sought to ↑be enthroned as khan↑// not for the good of the people↑// but to secure his own power↓//’)

Here is the full translation with all examples left in Cyrillic and their translations provided in brackets nearby, as requested:

Texts provided with such orthoepic norms and intonation are read aloud by a narrator, and the audio recordings are input into the speech synthesis program, upon which the speech synthesis is performed.

Additionally, the text to be synthesized undergoes several stages. First, the material to be included in the speech synthesis is selected. The synthesizer cannot “read” every word in the orthographic text in its initial form.

There are the following types of non-standard writing commonly found in spoken language: numbers; special characters that are neither letters nor numbers; abbreviations and acronyms, etc. These symbols must be converted into “normal” standard words or similar forms (for example, transliteration from another alphabet) in order to be synthesized — that is, the prepared text must be adapted for reading by expanding abbreviated words and numbers. This process occurs in the following order:

Expanding Abbreviations

Any abbreviated word in the texts for speech synthesis must be fully expanded. For example: *б.з. VI ғасырынан Түркі кезеңі басталады* (‘from the 6th century AD the Turkic period begins’). The abbreviated phrase here would be expanded as: *біздің заманымыздың алтыншы ғасырынан Түркі кезеңі басталады* (‘from the sixth century of our era the Turkic period begins’). To ensure that the synthesizer works correctly — that is, to vocalize any text fully and naturally — abbreviations must be converted into full words.

When expanding abbreviated words, the following problems should also be considered. That is, abbreviated words (acronyms) are used in various fields. In this regard, abbreviations in texts from different fields appear in various forms and meanings, so they must be correctly expanded in accordance with their context.

For example, the following are abbreviations related to education and science:

ҚР БҒМ – Қазақстан Республикасының Білім және ғылым министрлігі (‘Ministry of Education and Science of the Republic of Kazakhstan’);

ЖОО – Жоғары оқу орны (‘institution of higher education’);

ҰБТ – Ұлттық бірыңғай тестілеу (‘Unified National Testing’);

ҒЗИ – Ғылыми-зерттеу институты (‘scientific re-

search institute');

МЖМБС – Мемлекеттік жалпыға міндетті білім стандарты ('State Compulsory Education Standard').

Similarly, the following are abbreviations related to medicine:

ҚР ДСМ – Қазақстан Республикасының Денсаулық сақтау министрлігі ('Ministry of Health of the Republic of Kazakhstan');

ЖРВИ – Жедел респираторлы вирустық инфекция ('Acute Respiratory Viral Infection');

ҚҚС – Қан қысымы (blood pressure);

КТ – Компьютерлік томография ('CT – computed tomography');

МРТ – Магниттік-резонанстық томография ('MRI – magnetic resonance imaging').

The following are abbreviations related to economics and finance:

ҚР ҰБ – Қазақстан Республикасының Ұлттық Банкі ('National Bank of the Republic of Kazakhstan');

ҚҚС – Қосылған құн салығы ('value-added tax');

ІЖӨ – Ішкі жалпы өнім ('gross domestic product');

КТС – Корпоративтік табыс салығы ('corporate income tax');

АҚ – Акционерлік қоғам ('joint-stock company').

The following are abbreviations related to transportation and logistics:

ҚТЖ – Қазақстан темір жолы ('Kazakhstan Railways');

ЖЖЕ – Жол жүру ережелері ('traffic rules');

АҚШ – Автомобиль көлік шаруашылығы ('automobile transport enterprise');

ЖЖМ – Жанар-жағармай материалдары ('fuel and lubricants');

ҚЖО – Құрғақ жүк кемесі операторы ('dry cargo ship operator').

When expanding abbreviations, it is necessary to consider the suffixes attached to them and to write the correct form. For example, the abbreviation *км.* ('km.') can be expanded differently depending on the number preceding it: *1 км = 1 километр* ('1 kilometer'), *2 км = 2 километрді* ('2 kilometers' [accusative case]), *5 км = 5 километрдің* ('of 5 kilometers' [genitive case]), *5 км-ге = 5 километрге* ('to 5 kilometers' [dative case]).

Spelling out Numbers

Not only must abbreviated words but also numbers that appear in the text must be rendered in words for speech synthesis. This is because speech synthesis must read numerical data such as dates and years correctly. It is particularly important to render numbers accurately in texts related to mathematics. Additionally, some words in a text may be written in Latin script. In such cases, they need to be correctly transliterated into Cyrillic. To do this, one uses either a database of frequently encountered foreign words or substitution rules that match each Latin letter with its Cyrillic equivalent.

To achieve high-quality speech synthesis, the most important requirement is real human speech. When a speaker reads a written text strictly according to spelling, the result lacks naturalness and sounds artificial. Therefore, in order for speech to sound pleasant and melodious to the ear, Kazakh orthoepy – that is, the norms of spoken language – must be followed. When the law of vowel harmony is violated in spoken language, it usually stems from an inability to distinguish between spoken and written forms. This is because writing cannot fully represent the way words are pronounced. From this point of view, the spoken texts used in synthesis must observe orthoepic norms. Therefore, texts used for synthesis are edited according to orthoepic rules.

For example: *йертістің үлкөн сұуұна ͡ғарай үңүлгән йеңіс // құлабе͡рісте пар ат ͡шеккен жеңіл трашпеңке зырлап ͡келеді /// ұзақ күңгу жүрүстөн талмаған йекі жарау ͡баран ат ͡шйоқұта шаб͡ады /// бұл жүргүншүлер уосұ ͡бетімен ағындап ͡парып // йертіс сұуұна түсүп кетердей /// бірағ ͡жотадан сондай жақың ͡ғөрүңгөн йертіс уонша тақау йемес йекен /// жаңағы ͡ғезеңнен соң ͡да сәл жәйылатын алаң бар ͡боп ͡шықты /// сол алаңд'аттардың желігі ͡басылып, жәй бүлкөкк'ауұстұ ///*

('Down the sloping descent leaning toward the great waters of the Irtysh, a light carriage drawn by a pair of horses is speeding along. After a long day's travel, the two strong bay horses keep up a steady trot. At that pace, the travelers seem as though they might rush straight ahead and plunge into the Irtysh itself. Yet the river, which appeared so close from the ridge, was not actually that near. Beyond the rise there turned out to be a small open plain. On that stretch, the horses' excitement subsided, and they shifted into an easy canter.')

In order to maintain orthoepic norms, the synthesis process is guided by established orthoepic rules. According to N. Uali, the codification of orthoepic norms should rely not only on practical experience but also on theoretical foundations. The scholar emphasizes that as the social importance of oral communication grows, it becomes even more essential to define orthoepic norms theoretically. He notes that orthoepy theories are closely connected with the theories of phoneme and phonology[26]. In this context, texts used in synthesis were analyzed from the perspective of phoneme theory.

Phoneme <a>

It occurs in all positions within a word and combines with all consonants. When following [ш] and [й], [ж] and [й], or when placed after a palatalized syllable and before another palatalized syllable that begins with [й], the <a> phoneme has the acoustic realization [ə], that is, it becomes palatalized. Below are examples of words in which the <a> phoneme appears in its [ə] variant:

Ж<a>й – ж[ə]й, ш<a>й – ш[ə]й, м<a>йсөк – м[ə]йсөк, бид<a>й – бид[ə]й, ауж<a>й – [ə]уж[ə]й, д<a>укес – д[ə]укес, д<a>укестік – д[ə]укестік, ж<a>йбарақат – ж[ə]йбарақат, ж<a>йбарақаттану – ж[ə]йбарақаттану, ж<a>йбасар – ж[ə]йбасар, ж<a>йғастыру – ж[ə]йғастырұу, ж<a>йғасу – ж[ə]йғасұу, ж<a>йғату – ж[ə]йғатұу, ж<a>йғызу – ж[ə]йғызұу, ж<a>йдақ – ж[ə]йдақ, ж<a>йдарлы – ж[ə]йдарлы, ж<a>йдарман – ж[ə]йдарман, ж<a>йдары – ж[ə]йдары, ж<a>й-күй – ж[ə]й-гүй, ж<a>йқалу – ж[ə]йқалұу, ж<a>йлы – ж[ə]йлы, ж<a>йма – ж[ə]йма, ж<a>йма-шуақ – ж[ə]йма-шұуақ, ж<a>йсаң – ж[ə]йсаң, ж<a>йша – ж[ə]йша, ж<a>йшылық – ж[ə]йшылық, ж<a>йсыз – ж[ə]йсыз, ж<a>йнау – ж[ə]йнау, ж<a>йт – ж[ə]йт.

(‘Жай – [jæj], шай – [ʃæj], майсөк – [mæjsøk], бидай – [bidæj], аужай – [æuʒæj], даукес – [dæukes], даукестік – [dæukestik], жайбарақат – [jæjbɑrɑqɑt], жайбарақаттану – [jæjbɑrɑqɑttɑnu], жайбасар – [jæjbɑsɑr], жайғастыру – [jæjɣɑstɯru], жайғасу – [jæjɣɑsuu], жайғату – [jæjɣɑtu], жайғызу – [jæjɣɯzuu], жайдақ – [jæjdɑq], жайдарлы – [jæjdɑrly], жайдарман – [jæjdɑrmɑn], жайдары – [jæjdɑry], жай-күй – [jæj-güj], жайқалу – [jæjqɑluu], жайлы – [jæjly], жайма – [jæjmɑ], жайма-шуақ – [jæjmɑ-ʃuuɑq], жайсаң – [jæjsɑŋ], жайша – [jæjʃɑ], жайшылық – [jæjʃylyq], жайсыз – [jæjsɯz], жайнау – [jæjnɑu], жайт – [jæjt]’)

Additionally, after the consonant [к], the acoustic realization of the <a> phoneme is [ə], meaning it is also palatalized:

қ<a>дір – қ[ə]дір, қ<a>дірлеп – қ[ə]дірлеп, қ<a>дірлі – қ[ə]дірлі, қ<a>стерлеу – қ[ə]стерлеу, қ<a>стерлі – қ[ə]стерлі, қ<a>жет – қ[ə]жет, қ<a>жетсіз – қ[ə]жетсіз.

(‘qadir – [qædir], qadirlep – [qædirlep], qadirli – [qædirli], qasterleu – [qæsterleu], qasterli – [qæsterli], qazhet – [qæʒet], qazhetsiz – [qæʒetsiz]’)

Besides the typical velar timbre of [к], a palatalized variant with a soft timbre appears, which, at the level of variation, may be regarded as assimilation. In addition, in the word *қауесет*, the [ə] is pronounced with a palatalized color under the influence of the final syllable [е]. However, the acoustic impression of this vowel lies between a regular [a] and a typical [ə][27].

Between the consonants [н] and [с], the acoustic realization of the <a> phoneme is also [ə]: *н<a>сихат – н[ə]сійхат, н<a>сихатшы – н[ə]сійхатшы, н<a>сихаттау – н[ə]сійхаттау.* (‘νασιηατ – [næsijxat], νασιηατσηψ – [næsijxatʃy], nasihattau – [næsijxattaw]’)

In borrowed words with a palatalized second syllable, the <a> phoneme also appears as [ə]: *ділд<a> – ділд[ə], тілм<a>ш – тілм[ə]ш, тілм<a>штық – тілм[ə]штық, дінд<a>р – дінд[ə]р.* (‘dilda – [dildæ], tilmash – [tilmæʃ], tilmashtyq – [tilmæʃtyq], dindar – [dindær]’)

In addition, in the Kazakh language, high vowels are often weakened, pronounced indistinctly, and undergo reduction. For example, the words *адыраңдау* (‘to fidget, to move restlessly’), *аңызақ* (‘dry hot wind, scorching wind’), *бадырақ* (‘bulging, protruding’), *дәрігер* (‘doctor, physician’), *еділеңдеу* (‘to swagger, to strut’), *жадырап* (‘cheerful, radiant, delighted’), *жұмылыңқы* (‘closed, shut’), *жұдырық* (‘fist’), *көбірек* (‘more, rather more’), *көбінесе* (‘mostly, mainly’), *көкірек* (‘chest, breast, also figurative: pride’), *қатынас* (‘relation, communication’), *қасірет* (‘grief, sorrow’), *құдірет* (‘might, divine power’), and *тәжірибе* (‘experience, practice, experiment’) are realized in pronunciation as *адыраңдау, аңызақ, бадырақ, дәрігер, йедіреңдеу, жадырап, жұмұлұңқы, жұдұрұқ, көбүрөк, көкүрөк, көбүнесе, қатынас, қәсірет, құдүрет,* and *тәжірійбе* (respectively ‘ to behave arrogantly, dry hot wind, bulging, doctor, to flounce, to smile with delight, slightly closed, fist, more,

mostly, chest, communication, grief, divine power, and experience'). In order to generate accurate word synthesis, it is necessary to systematize such words that undergo reduction in Kazakh and to formulate specific rules to be integrated into the speech synthesizer software [28].

It is not feasible to model and input each individual sound into the synthesizer separately when performing word synthesis. This is because when isolated sounds are simply combined, the output from the synthesizer does not sound natural — it produces speech that is robotic and incomprehensible. This occurs because, during actual speech, sounds continuously undergo various phonetic changes. Therefore, it is imperative to first identify and distinguish the sound changes that occur between syllables and at the juncture of root and suffix. In other words, it becomes necessary to specify and distinguish those changes that arise during speech production — namely, elisions [29].

For instance, ала алмады — ал'алмады ('could not take'), ала ешкі — ал'ешкі ('spotted goat'), айта алмайды — айт'алмайды ('cannot say'), жоғары іл — жоғар'іл ('lift it up'), жақсы өнер — жақс'өнөр ('good art'). Words that are written separately in orthography are often pronounced in a single breath in spoken language, and when two vowel sounds come into contact between words, one of them tends to be elided.

4. Conclusions

Contemporary TTS (Text-to-Speech) technologies are continuously evolving and improving, offering users new possibilities for interaction with digital devices. This ongoing progress opens the way to developing increasingly sophisticated and natural-sounding voice systems that can be applied across various aspects of daily life. The implementation of speech synthesis in the Kazakh language is not merely a technical process, but also a comprehensive scientific inquiry grounded in the phonetic, phonological, prosodic, and orthoepic principles of the language.

In speech synthesis, the orthoepic norms of the Kazakh language play a decisive role, as the differences between written text and spoken language directly influence the quality of synthesis. Only when changes in phonemes, vowel reduction, elision, and assimilation rules are accounted for does the synthesized speech sound natural. Therefore, the models incorporated into the synthesizer must adhere to the phonetic and phonological laws of the Kazakh language. The outcomes of this research determine the scientific foundations necessary for producing high-quality Kazakh speech synthesis. The phonetic and phonological examination of speech synthesis is crucial for the construction of effective linguistic algorithms and models.

Given the importance of representing orthoepic norms and prosodic features in the construction of Kazakh speech synthesis, texts were segmented into syntagmas. In addressing the issue of orthoepic norms in Kazakh, attention was given to vowel harmony and consonant compatibility, sound changes between root and suffix, as well as within rhythmic groups; processes such as assimilation, dissimilation, reduction, variants and variations, elision, etc., were studied in order to adapt them for speech synthesis and to develop an appropriate model. For the speech synthesis component within an educational platform, the textbook *Kazakh Literature* for the 11th grade was selected, and each sentence was analyzed for its orthoepic characteristics and prosodic patterns (based on eight intonemes). During the research process, both the concatenative synthesis method and the formant-acoustic method were employed, and corresponding models were provided for dictor's reading.

To ensure that the synthesizer produces speech that is both natural and intelligible, all texts were pre-processed: abbreviated words and numbers were expanded into their full forms, special symbols were clarified, and words written in Latin script were transliterated into Cyrillic. The synthesis texts underwent these procedures and were subjected to linguistic analysis.

In conclusion, the development of Kazakh speech synthesis requires a careful distinction between orthographic and orthoepic conventions, with all relevant phonological and phonetic changes encoded as formal rules within the memory of the computer program.

Author Contributions

All authors have equal contributions to this article. All authors have read the published version of the manuscript and agreed with it.

Funding

The article was prepared as part of the program-specific financing of the [IRN BR24993111], *Developing the Speech Synthesis System Based on the Orthoepic Norm of the Kazakh Language*.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Acknowledgments

We would like to express our gratitude to the authors for their contributions to each section of the article. By conducting research related to the topic and making significant efforts in writing the article, the authors have made a substantial contribution.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Mussakhojayeva, S., Khassanov, Y., Varol, H.A., 2022. KazakhTTS2: Extending the Open-Source Kazakh TTS Corpus With More Data, Speakers, and Topics. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, 20–25 June 2022; pp. 5404–5411.

[2] Kaliyev, G., 2005. Explanatory Dictionary of Linguistic Terms. Dictionary: Almaty, Kazakhstan. p. 440. (in Kazakh)

[3] Abilbekov, A., Mussakhojayeva, S., Yeshpanov, R., et al., 2024. KazEmoTTS: A Dataset for Kazakh Emotional Text-to-Speech Synthesis. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italia, 20–25 May 2024; pp. 9626–9632.

[4] Bazarbayeva, Z., Ospangaziyeva, N., Karshigayeva, A., 2024. Syllable Theory and Diachronic Phonology: Vocalism and Consonantism in Turkic Languages. Eurasian Journal of Applied Linguistics. 10(1), 50–59. Available from: https://ejal.info/menuscript/index.php/ejal/article/view/685/224 (cited 11 May 2025)

[5] Zhumabayeva, Z.T., Ospangaziyeva, N., Bazarbayeva, Z.M., et al., 2024. The Historical Change of the Vowels a/ə/e in Turkic Languages. Theory and Practice in Language Studies. 14(7). DOI: https://doi.org/10.17507/tpls.1407.10

[6] Derkach, M.F., Gumetsky, R.Y., Gura, B.M., et al., 1983. Dynamic Spectra of Speech Signals. The Higher School, Publishing House at Lviv University: Lviv, Ukraine. p. 168. (in Russian)

[7] Sorokin, V.N., 1992. Speech Synthesis. Nauka: Moscow, Russia. p. 392. (in Russian)

[8] Fant, G., 1964. The Acoustic Theory of Speech Production. Mouton Publishers: The Hague, Netherlands. p. 284.

[9] Bondarko, L.V., 1967. The Structure of the Syllable and the Characteristics of Phonemes. Issues of Linguistics. 1, 34–46. (in Russian)

[10] Golubtsov, S.V., 1969. Speech Synthesis. In: Proceedings of the All-Union School-Seminar ARSO-4. Tallinn: Kyiv, Ukraine. pp. 107–130. (in Russian)

[11] Zlatoustova, L.V., 1997. Acoustic and Perceptual Characteristics of Spontaneous Speech. Govor. XIV(1–2), 77–87.

[12] Flanagan, J.L., 1972. Speech Analysis, Synthesis, and Perception, 2nd ed. Springer-Verlag: Berlin, Germany. p. 394.

[13] Rybin, S., 2014. Speech Synthesis. ITMO University: St. Petersburg, Russia. p. 92. (in Russian)

[14] Lobanov, B.M., Tsirulnik, L.I., 2008. Computer Speech Synthesis and Cloning. Belarusian Science: Minsk, Belarus. p. 342. (in Russian)

[15] Kjunnap, E.Y., 1975. Speech Signal Synthesizers. Valgus: Tallinn, Estonia. p. 240. (in Russian)

[16] Zhunisbek, A., 2018. Issues of Kazakh Linguistics: Kazakh Phonetics. Abzal-ai: Almaty, Kazakhstan. p. 368. (in Kazakh)

[17] Bazarbayeva, Z.M., 2022. Intonology. Everest: Almaty, Kazakhstan. p. 440. (in Kazakh)

[18] Bazarbayeva, Z., 2022. Fundamentals of Kazakh Phonology. Everest: Almaty, Kazakhstan. p. 460. (in Kazakh)

[19] Kuderinova, K., 2013. History and Theory of Kazakh Writing. Eltanym: Almaty, Kazakhstan. (in Kazakh)

[20] Fazylzhanova, A.M., 2022. Melody of Speech and Intonation: An Experimental Phonetic Study. Bolashak: Almaty, Kazakhstan. p. 208. (in Kazakh)

[21] Aktanova, A.S., et al., 2020. Kazakh Literature: Part 1. Textbook for the Social-Humanitarian Track of Grade

11 in General Education Schools. Atamura: Almaty, Kazakhstan. p. 144. (in Kazakh)
[22] Badanbekqyzy, Z., 2001. Phoneme Sound Inventories in the Kazakh Language. Gylym: Almaty, Kazakhstan. p. 134. (in Kazakh)
[23] Uali, N.M., 2018. Graphics. Orthography. Orthoepy. Evero: Almaty, Kazakhstan. p. 250. (in Kazakh)
[24] Kazakh Grammar: Phonetics, Word Formation, Morphology, Syntax, 2002. Gylym: Astana, Kazakhstan. p. 784. (in Kazakh)
[25] Zhumabayeva, Z., 2016. Speech Synthesis in Kazakh Linguistics. Tiltanym. 2, 91–94. (in Kazakh)
[26] Bazarbayeva, Z., 2008. Kazakh Language: Intonology, Phonology. Zhibek Zholy: Almaty, Kazakhstan. p. 324. (in Kazakh)
[27] Amanbayeva, A., 2016. Speech Synthesis: Formant and Concatenative Methods. Til-tanym. 2, 88–90. (in Kazakh)
[28] Orthoepic Dictionary, 2007. Arys Publishing House: Almaty, Kazakhstan. p. 800. (in Kazakh)
[29] Bazarbayeva, Z.M., Sadyk, D., Amanbayeva, A., et al., 2025. Segmental-Prosodic Foundations of Kazakh Speech Synthesis. Eurasian Journal of Applied Linguistics. 11(2), 69–80.