EURASIAN NATIONAL UNIVERSITY

Turkic Studies Journal 3 (2026) 212-235

Journal homepage: https:tsj.enu.kz

Article

The implementation of orthoepic norms in the Kazakh speech synthesis system[1]

Zh.T. Zhumabayeva[a], N. Ospangazieva[b], *A.M. Zhalalova[c]

[a]*Akhmet Baitursynuly Institute of Linguistics, Almaty, Kazakhstan*
ORCID: 0000-0002-0025-5201
(E-mail: zhanar.zhumabayeva.80@bk.ru
[b]*Akhmet Baitursynuly Institute of Linguistics, Almaty, Kazakhstan*
ORCID: 0000-0001-8875-5747
(E-mail: nb_ospangazievan@inbox.ru)
[c]*Temirbek Zurgenov Kazakh National Academy of Arts, Almaty, Kazakhstan*
ORCID: 0000-0003-1216-3889
*(E-mail: akshaim01@mail.ru). *Corresponding author: akshaim01@mail.ru*

ARTICLE INFO

Keywords:
speech synthesis, orthoepy, prosody, model, orthoepic norm, phonetic regularities, borrowed word, rhythmic group, synharmonism, reduction, phoneme.

IRSTI 16.21.3

DOI: https://doi.org/10.32523/2664-5157-2026-3-212-235

ABSTRACT

The article discusses the issues of modeling the speech synthesis system in the Kazakh language based on orthoepic norms. As a result of the rapid development of digital technologies, the importance of speech synthesis systems aimed at the natural-sounding of text in linguistics is increasing. In this context, the article analyzes the scientific foundations of the use of phonetic, orthographic, and prosodic regularities of the Kazakh language in the speech synthesis system. In particular, the importance of maintaining orthographic norms when converting written text into spoken speech is highlighted. The quality of a speech synthesis system is determined not only by the grammatically correct processing of the text, but also by its natural and literary sounding. Accordingly, sound interaction and the accurate representation of intonational characteristics are essential for producing natural-sounding synthesized speech. Ignoring orthoepic norms can reduce the naturalness of speech and negatively affect listeners' perception and comprehension. In this regard, the introduction of the sound patterns of the Kazakh language into the speech synthesis system is one of the current scientific issues. The paper highlights the history of studying orthoepy in Kazakh linguistics and current research trends, as well as reviews both domestic and international studies. Based on the analysis, the necessity of systematically modeling the phonetic-phonological and prosodic features of text to improve the quality of synthesized speech is emphasized. The research draws upon the works

[1] This article is implemented within the framework of the targeted funded project BR24993111 "Development of a speech synthesis system based on the orthoepic norms of the Kazakh language"

© Zh.T. Zhumabayeva, N. Ospangazieva, A.M. Zhalalova 2026

of scholars such as I. Kenesbayev, M. Balakayev, S. Myrzabekov, R. Syzdyk, A. Zhunisbek, and Z. Bazarbayeva, while theoretical issues of speech synthesis are considered in comparison with international studies. In addition, the characteristics of modern Kazakh speech synthesis systems (Kazakh TTS2, Kazakh ASR, and other projects) are analyzed, and the level of orthoepic norm implementation in these systems is described. The study considers major speech synthesis approaches, including compiler-based, parametric, formant-based, and neural network methods, and examines their applicability to the modeling of linguistic material. The necessity of considering phonemic changes, positional and combinatorial features of sounds, intonational structure, and rhythmic groups in speech synthesis systems is emphasized. Based on the conducted research, the systematic use of an orthoepic algorithm and prosodic annotation is proposed for improving the Kazakh speech synthesis system. The article also analyzes the adaptation of loanwords from an orthoepic perspective and outlines ways of their phonetic integration in accordance with the phonological laws of the Kazakh language. The results of the study demonstrate that systematic modeling of orthoepic norms and prosodic features enhances the naturalness of synthesized speech and facilitates listener comprehension. Furthermore, the development of the Kazakh speech synthesis system contributes to the broader use of the national language in the digital environment.

Ж.Т. Жұмабаева[a]
[a]*Ахмет Байтұрсынұлы атындағы Тіл білімі институты*
Алматы, Қазақстан Республикасы
ORCID: 0000-0002-0025-5201
(E-mail: zhanar.zhumabayeva.80@bk.ru)

Н. Оспанғазиева[b]
[b]*Ахмет Байтұрсынұлы атындағы Тіл білімі институты*
Алматы, Қазақстан Республикасы
ORCID: 0000-0001-8875-5747
(E-mail: nb_ospangazievan@inbox.ru)

***А.М. Жалалова**[c]
[c]*Т. Жүргенов атындағы Қазақ ұлттық өнер академиясы*
Алматы, Қазақстан Республикасы
ORCID: 0000-0003-1216-3889
(E-mail: akshaim01@mail.ru)
*Байланыс үшін автор: akshaim01@mail.ru

Received 10 June 2026. Revised 13 July 2026. Accepted 16 August 2026. Available online 30 September 2026.

For citation: Zh.T. Zhumabayeva, N. Ospangazieva, A.M. Zhalalova The implementation of orthoepic norms in the Kazakh speech synthesis system // Turkic Studies Journal. 2026. V. 8. № 3. P. 212-235. DOI: http://doi.org/10.32523/2664-5157-2026-3-212-235

Қазақ сөз синтезі жүйесінде орфоэпиялық норманың берілуі

Аннотация. Аталмыш мақалада қазақ тіліндегі сөз синтезі жүйесін орфоэпиялық норма негізінде модельдеу мәселелері қарастырылады. Цифрлық технологиялардың қарқынды дамуы нәтижесінде тіл білімінде мәтінді табиғи дыбыстауға бағытталған сөз синтезі жүйесінің маңызы артып отыр. Осыған байланысты мақалада қазақ тілінің фонетикалық, орфоэпиялық және просодикалық заңдылықтарды сөз синтезі жүйесінде қолданудың ғылыми негіздері талданады. Әсіресе жазба мәтінді ауызша сөзге айналдыру барысында орфоэпиялық норманың сақталуының маңызы айқындалады. Сөз синтезі жүйесінің сапасы мәтіннің грамматикалық тұрғыдан дұрыс өңделуімен ғана емес, оның табиғи әрі әдеби нормаға сай дыбысталуымен де анықталады. Сондықтан синтезделген сөзде дыбыстардың өзара ықпалы, интонациялық ерекшеліктің дұрыс берілуі маңызды. Орфоэпиялық нормаларды ескермеу сөйлеудің табиғилығын төмендетіп, тыңдаушының мәтінді қабылдауына кері әсер етуі мүмкін. Осы тұрғыдан алғанда қазақ тілінің дыбыстық заңдылықтарын сөз синтезі жүйесіне енгізу өзекті ғылыми мәселелердің бірі болып табылады. Мақалада қазақ тіл біліміндегі орфоэпия мәселесінің зерттелу тарихы мен қазіргі ғылыми бағыттары сараланып, отандық және шетелдік зерттеулерге шолу жасалады. Жүргізілген зерттеулердің негізінде сөз синтезінің сапасын арттыруда мәтіннің фонетика-фонологиялық және просодикалық ерекшеліктерін жүйелі түрде модельдеу қажеттілігі тұжырымдалады. Зерттеу барысында қазақ орфоэпиясына қатысты І. Кеңесбаев, М. Балақаев, С. Мырзабеков, Р. Сыздық, Ә. Жүнісбек, З. Базарбаева сияқты ғалымдардың еңбектері негізге алынып, сөз синтезінің теориялық мәселелері шетелдік зерттеулермен сабақтастыра талданады. Сонымен қатар қазіргі қазақ тіліндегі сөз синтезі жүйелерінің (Kazakh TTS2, Kazakh ASR және басқа да жобалар) ерекшелігі қарастырылып, олардың орфоэпиялық норманы қолдану деңгейі сипатталады. Мақалада сөз синтезін жүзеге асыруда қолданылатын компилятивтік, параметрлік, форманттық және нейрондық желілерге негізделген әдістердің ерекшеліктері талданып, олардың тілдік материалды модельдеудегі қызметі көрсетіледі. Жүргізілген зерттеулердің негізінде қазақ тіліндегі сөз синтезі жүйесін жетілдіруде орфоэпиялық алгоритм мен просодикалық белгіленімдерді жүйелі түрде қолдану ұсынылады. Сондай-ақ кірме сөздердің орфоэпиялық тұрғыдан игерілу ерекшеліктері талданып, олардың дыбысталуының қазақ тілінің фонетикалық заңдылықтарына бейімделу жолдары көрсетіледі. Зерттеу нәтижесінде орфоэпиялық норма мен просодикалық ерекшелікті жүйелі түрде модельдеу синтезделген сөздің табиғилығын арттырып, тыңдаушының мәтінді қабылдауын жеңілдететіні дәлелденеді. Зерттеу қорытындылары қазақ тілінің дыбыстық және просодикалық ерекшеліктерін цифрлық ортада дәл модельдеуге мүмкіндік береді және сөз синтезі технологияларын жетілдірудің теориялық әрі қолданбалы негіздерін толықтыра түседі. Алынған нәтижелер болашақ ғылыми зерттеулерде және практикалық әзірлемелерде пайдалануға бағытталған. Сондай-ақ ұсынылған тұжырымдар қазақ тілінің орфоэпиялық қорын цифрландыру және сөйлеу технологияларының сапасын арттыру міндеттерін шешуге ықпал етеді.

Кілт сөздер: сөз синтезі, орфоэпия, просодика, модель, орфоэпиялық норма, фонетикалық заңдылықтар, кірме сөз, ырғақты том, үндесім заңы, редукция, фонема.

Ж.Т. Жумабаева[a]
[a]Институт языкознания имени Ахмета Байтурсынулы
Алматы, Республика Казахстан
ORCID: 0000-0002-0025-5201
(E-mail: zhanar.zhumabayeva.80@bk.ru)

Н. Оспангазиева[b]
[b]Институт языкознания имени Ахмета Байтурсынулы
Алматы, Республика Казахстан
ORCID: 0000-0001-8875-5747
(E-mail: nb_ospangazievan@inbox.ru)

***А.М. Жалалова**[c]
[c]Казахская национальная академия искусств имени Т. Жургенова
Алматы, Республика Казахстан
ORCID: 0000-0003-1216-3889
(E-mail: akshaim01@mail.ru)
*Автор для корреспонденции: akshaim01@mail.ru

Передача орфоэпической нормы в системе синтеза речи казахского языка

Аннотация. В данной статье рассматриваются вопросы моделирования системы синтеза речи казахского языка на основе орфоэпической нормы. В условиях стремительного развития цифровых технологий возрастает значение систем синтеза речи, направленных на естественное озвучивание текстов. В связи с этим в статье анализируются научные основы применения фонетических, орфоэпических и просодических закономерностей казахского языка в системе синтеза речи. Особое внимание уделяется значимости соблюдения орфоэпической нормы при преобразовании письменного текста в устную речь. Качество системы синтеза речи определяется не только корректной грамматической обработкой текста, но и его естественным произношением в соответствии с литературной нормой. Поэтому важное значение имеет правильная передача в синтезированной речи взаимовлияния звуков и интонационных особенностей. Игнорирование орфоэпических норм может снизить естественность звучания речи и затруднить восприятие текста слушателем. В этом контексте внедрение звуковых закономерностей казахского языка в систему синтеза речи является одной из актуальных научных задач. В статье рассматриваются история изучения орфоэпии в казахском языкознании и современные научные направления в данной области, а также представлен обзор отечественных и зарубежных исследований. На основе проведённого анализа обосновывается необходимость системного моделирования фонетико-фонологических и просодических особенностей текста для повышения качества синтеза речи. В ходе исследования использованы труды казахских учёных И. Кенесбаева, М. Балакаева, С. Мырзабекова, Р. Сыздык, А. Жунисбека, З. Базарбаевой, посвящённые вопросам орфоэпии, а теоретические аспекты синтеза речи рассмотрены в сопоставлении с зарубежными исследованиями. Кроме того, анализируются особенности современных

систем синтеза речи казахского языка (Kazakh TTS2, Kazakh ASR и другие проекты), а также характеризуется степень использования в них орфоэпических норм. В статье рассматриваются особенности компилятивных, параметрических, формантных методов и методов, основанных на нейронных сетях, применяемых при реализации синтеза речи, а также их роль в моделировании языкового материала. На основе проведённых исследований предлагается системное использование орфоэпических алгоритмов и просодической разметки для совершенствования систем синтеза речи казахского языка. Также анализируются особенности орфоэпического освоения заимствованных слов и пути адаптации их произношения к фонетическим закономерностям казахского языка. В результате исследования доказано, что системное моделирование орфоэпических норм и просодических особенностей повышает естественность синтезированной речи и облегчает восприятие текста слушателем. Результаты исследования позволяют более точно моделировать звуковые и просодические особенности казахского языка в цифровой среде и расширяют теоретические и прикладные основы совершенствования технологий синтеза речи.

Ключевые слова: синтез речи, орфоэпия, просодика, модель, орфоэпическая норма, фонетические закономерности, заимствованное слово, ритмическая группа, закон сингармонизма, редукция, фонема.

Introduction

In recent years, digital technologies have increasingly permeated the field of linguistics, providing a foundation for the development of new language technologies and digital platforms. One important area in this regard is speech synthesis. The development of a Kazakh speech synthesis system requires consideration of the orthoepic, prosodic, and phonetic characteristics of the Kazakh language. Particular attention should also be paid to the natural pronunciation of loanwords within speech synthesis systems. This is important because the pronunciation of loanwords incorporated into a speech synthesis system should, as far as possible, conform to the orthoepic norms of the Kazakh language and thereby contribute to the naturalness and intelligibility of synthesized speech. To achieve this objective, loanwords need to be systematically analyzed and modeled according to the phonetic and orthoepic patterns of Kazakh. In this context, the present study examines and models the pronunciation of words in accordance with established orthoepic norms, with particular emphasis on the pronunciation and phonetic adaptation of loanwords. Such modeling contributes to the naturalness of synthesized speech and can serve as a computational model for reproducing features of natural spoken language. Modern speech synthesis systems provide a range of functionalities, including text-to-speech (TTS) conversion, playback of generated audio, and the generation or downloading of audio files in formats such as MP3. The development of such technologies is particularly relevant for languages with rapidly expanding digital resources and speech technology infrastructure. For comparison, speech synthesis technologies developed for the Uzbek language demonstrate the potential for incorporating language-specific phonetic and articulatory characteristics into synthesized speech. Some Uzbek speech synthesis systems reproduce the articulatory properties of consonants and the duration of vowels, thereby

contributing to more natural-sounding speech. Uzbek speech synthesis technologies are also being applied in various domains, including the development of tourist audio guides, the production of educational and language-learning materials, public-address announcements, multimedia presentations, and advertising content. These applications demonstrate the practical potential of speech synthesis technologies and provide a useful comparative basis for the further development of Kazakh speech synthesis systems.

When analyzing domestic and foreign research, it can be seen that speech synthesis systems are considered in two directions. One is an engineering approach based on technical modeling and relying on phonemic transcription and acoustic parameters (Taylor, 2009); the other is a linguistic approach based on the the internal structural patterns of a language, orthoepic norms, positional, and combinatorial changes. In the modern speech synthesis system, these two directions are used in a complex way, combining (Bazarbayeva et al., 2025). In the practice of foreign speech synthesis, the modeling and reproduction of vocabulary stock is mainly carried out using data-driven neural approaches. This approach enables learning of speech units from large amounts of text data. In Kazakh linguistics, speech synthesis systems are also often developed based on neural models. Orthoepy refers to the correct pronunciation of words and phrases. In general, the orthoepy system of each language is based on its internal laws. The basic pillar of phonology in the Kazakh language is the law of harmony. It is known that the law of harmony is divided into two: palatal (tongue) harmony and labial (lip) harmony. In this regard, in the speech synthesis system, the orthoepic norm is taken as a basis, and the law of harmony of sounds is observed. This is because non-compliance with the orthoepic norm and prosodic specifics in the speech synthesis system violates the word's natural sound, making it difficult to perceive. In this regard, transferring the orthoepic norm to the model is an important step toward improving the Kazakh speech synthesis system. Although loanwords used in various fields are written according to the orthographic norm, in spoken language, some are pronounced in their adopted form, while others are pronounced in their original form. In this way, it is noticeable that the pronunciation of loanwords in spoken language is not consistent. If they are used in different versions that are not regulated in the speech synthesis system, their pronunciation will also be unclear. Therefore, modeling of borrowings in a single system is important both theoretically and practically. Consequently, maintaining orthoepic norms and regulating the ways of modeling loanwords for the Kazakh speech synthesis system are scientifically relevant. So, the article considers the modeling of Kazakh and loanwords according to orthoepic standards. This increases the naturalness of the Kazakh speech synthesis system.

Materials and research methods

The research material is based on scientific studies that address theoretical problems in Kazakh orthoepy and speech synthesis. In particular, the studies of I. Kenesbayev's "Qazaq til biliminin maseleleri" (Kenesbayev, 2014), M. Balaqayev's "Qazaq adebi tili zhane onyn normalary" (Balaqayev , 1984), S. Myrzabekov's "Qazaq tilinin fonetikasy" (Myrzabekov, 2004), R. Syzdyq's "Qazaq tilinin anyqtagyshy (emle, tynys belgileri, soz sazy)" (Syzdyq, 2000), A. Zhunisbek's "Qazaq til biliminin maseleleri" (Zhunisbek, 2009), Z.M. Bazarbayeva's "Intonologia" (Bazarbayeva, 2022) served as the basis for writing the issue of orthoepy.

Furthermore, to substantiate the scientific findings on speech synthesis, the following foundational studies were utilized: T. Dutoit's "An Introduction to Text-to-Speech Synthesis" (Dutoit, 1997), P. Taylor's "Text-to-Speech Synthesis" (Taylor, 2009), J.C. Wells' "English Intonation: An Introduction Text" (Wells, 2007), D.V. Pantyukhin's "Neironnye seti sinteza rechi golosovyh pomoshnikov i poiushih avtomatov" (Pantyukhin, 2021), L.N. Yasnitsky's "Vvedenie v iskustvenny intellekt" (Yasnitsky, 2010), A.V. Frolov and G.V. Frolov's "Sintez i raspoznavanie rechi. Sovremennye resheniya" (A.V. Frolov and G.V. Frolov, 2003), as well as B.M. Lobanov and L.I. Tsirulnik's works called "Kompjuterny sintez I klonirovanie rechi" (Lobanov and Tsirulnik, 2008). Currently, several speech synthesis systems have been developed to sound out text in the Kazakh language. One of them is the demo program "Kazakh Text-to-Speech Conversion - 2 (Kazakh TTS2)" developed by the Institute of Intelligent Systems and Artificial Intelligence (ISSAI) of Nazarbayev University. This program reads Kazakh text using 3 female and 2 male voices. A demo version is available on the institute's website. Additionally, the Kazakh ASR system developed by specialists at the Institute of Informatics and Mathematics focuses on speech recognition. This program converts spoken words into text and instantly creates a ready-made text. These issues can be seen in the speech synthesis system developed by the Til-Qazyna Scientific and Practical Center. As shown in the text, words are read immediately according to their orthography. Therefore, it is very important to synthesize Kazakh words in accordance with orthoepic norms. To develop such a speech synthesis system, all the laws of the Kazakh language, phonetics, and prosodic features must be taken into account.

The history of speech synthesis dates back to the 10th century, when Gerbert de Aurillac created a bronze human head and developed it so that it could say "yes" or "no" to a questioner. And in the mid-13th century, monk Albert von Bellstedt (Albertus Magnus) and English philosopher Roger Bacon created a model of a "talking head." This was the first step towards speech synthesis at a time when technology was still in its infancy. At the end of the 18th century, the German scientist G. Kratzenstein was the first to create a model of five long vowel sounds (а, э, и, о, у) produced during human speech. Through this model, different shapes of acoustic resonators are created depending on the vibrations of vowels. In 1788, the Austrian scientist Wolfgang von Kempelen continued the work of G. Kratzenstein, creating a model of the lips and tongue and inventing an acoustic-mechanical speaking machine that could pronounce certain sounds, speaking in the voice of a 3-to 4-year-old child.

Having evolved from these early fundamental acoustic-mechanical experiments to the digital era, global speech synthesis research laid the foundation for modern algorithms, the adaptation of which in Kazakhstan required a deep consideration of the unique phonetic and prosodic characteristics of the Kazakh language.

Speech synthesis in Kazakh linguistics originates from the research of A. Zhunisbek, Z. M. Bazarbayeva, and A. Fazylzhan. The results of the scientists' research are the basis for improving modern speech synthesis. In this regard, the article is guided by the works of A. Zhunisbek "Qazaq til biliminin maseleleri", Z.M. Bazarbayeva "Qazaq fonologiasynyn negizderi", and the research of A. Fazylzhan. The study was guided by the research of Professor A. Zhunisbek in describing the acoustic and articulatory characteristics of sounds. When considering the prosody of speech synthesis, we rely on the conclusions of Academician Z.M. Bazarbayeva. At the same time, we all know that discreteness is important in the synthesis of words, and in this regard, we can rely on the research of A. Fazylzhan.

In considering speech synthesis from a phonetic-phonological perspective, the work of N. Uali “Orfografia, Orfoepia, Grafica” (Uali, 2018), the study of A. Zhunisbek “Qazaq fonetikasy” (Zhunisbek, 2009), as well as the studies of Z. M. Bazarbayeva entitled “Intonologia” (Bazarbayeva, 2025), “Qazaq fonologiasynyn negizderi” (Qazaq fonologiasynyn…, 2022), the K. Kuderinova's monograph “Qazaq zhazuynyn tarihy men teoriasy” (Kuderinova, 2013), and the A. Fazylzhan's work “Soz sazy zhane Intonatcia” (Fazylzhan, 2022) were used as a basis for the theoretical substantiation of the article.

Every year, thanks to advances in computer technology, the quality of speech synthesis improves, and its realistic sound becomes increasingly noticeable. As mentioned earlier, its application in various fields has also been expanding.

When speech synthesis is used to transcribe any text, the phonetic features, phoneme position, and word orthoepy must first be considered for its prosody, as well as the combination and transformation of sounds, to come out naturally. Thanks to this, the quality of speech synthesis increases; it works faster, the harmony of sounds and melody in the oral text is preserved, and it sounds realistic in accordance with the peculiarities of the Kazakh language.

For the implementation of speech synthesis, *compilative, articulatory formant, and corpus methods* are used.

In parametric synthesis, the set of messages is small and is used in cases where they do not change frequently. The quality of parametric synthesis is very high. However, parametric synthesis cannot be used for messages that are not predefined. While the formant sound modeling method uses a model that models the acoustic characteristics of speech sounds (especially vowels) based on the pitch and timbre of the voice, the parametric synthesis method represents the speech signal as a set of a small number of continuously varying parameters.

Parametric synthesis is used where the set of messages is limited and does not change frequently. Using the method of synchronous analysis of suprasegmental units, it is possible to identify true prosodic features and adapt the model to speech synthesis by analyzing intonation units in modern Kazakh linguistics.

Compilative synthesis is based on constructing a message from a dictionary of pre-written synthesis elements. The size of the synthesis elements is at least one word. Therefore, the content of synthesized messages is limited by the vocabulary. As a rule, the number of vocabulary units does not exceed several hundred words. Compilative synthesis is widely used in various fields. For example, it is used in mobile operators' reference services to obtain information about the subscriber's account status.

Full speech synthesis according to rules.

Full rule-based speech synthesis allows for the generation of speech from previously unknown texts, controlling all parameters of the speech signal. The method is distinguished by two main features:

- Articulatory synthesis - creates speech by modeling the movements of human speech organs (tongue, lips, larynx), i.e., modeling the movements of human speech organs (tongue, lips, vocal cords);

- Formant synthesis is the computational implementation of speech formats. It models sounds based on acoustics, such as formant frequencies. It takes into account the phonetic characteristics of the language. It is a phoneme-based synthesis; that is, a word is divided into

phonemes, and its transformation is determined. For example, texts that are included in speech synthesis are divided into several parts, such as syntagma, rhythmic group, prosody, and intonation modeling. The intonation characteristics of the text are important for increasing the naturalness of speech. In this regard, the text that is included in the speech synthesis is divided into rhythmic groups and syntagmas. A rhythmic group is the smallest meaningful group of words united by a rhythmic stress on the last syllable. The smallest fragment in spoken language that demonstrates the close connection between syntactic structure and intonation is a syntagma (Bazarbayeva et all, 2025: 50).

When using the formant synthesis method, the resonators quickly arrange each sound in the correct order. This method differs from the compilation method in that it places a greater load on the processor, increases the memory volume, and stores special algorithms in the synthesis program. In both cases, after production, it is necessary to smooth out the traces of connection between the sounds of the same word. The Kazakh language's articulation, that is, the organ of speech, is not accustomed to immediately forming other vowels or consonants when pronouncing sounds. To achieve this, after generating sounds in both directions, it's necessary to combine and correct the sounds within a word. To do this, the synthesis program must be provided with an algorithm for the sounds and systematized.

Research background

Kazakh orthography is a field that studies the sound system of the language and the pronunciation patterns of words. Although orthoepy was not specifically considered at the beginning of the 20th century, examples and some information show that it is a separate issue. This issue originates from the works of scientists such as Akhmet Baitursynuly, Telzhan Shonanuly, Eldes Omarov, Khalel Dosmukhameduly, and Kudaibergen Zhubanov. For example, in Akhmet Baitursynuly's studies, it is shown that the consonant sounds қ, k are pronounced as ғ, г (Baitursynuly, 2013: 87). T. Shonanuly also paid attention to the peculiarities of pronunciation of sounds and regularities in spoken speech. M. Balaqayev divided language culture into two areas, considering the first as the language of the press and the second as the language of speech. Orthoepy was described as the main component of the culture of correct speech (Balaqaev, 1984: 101). During that period, the problems of vowel pronunciation, consonants, intonation, influence, and stress were systematically presented in the textbooks by M. Balaqayev, G. Begaliev, N. Sauranbayev, I. Kenesbayev, and other scholars. In the 1970s-90s, Kazakh orthoepy began to take shape as a science. The works of M. Duisebayeva, "Qazaq adebi tili orfoepiasynyn negizgi maseleleri" (Duisebayeva, 1966), "Qazaq tilinin qysqasha orfoepikalyq sozdigi" (Duisebayeva, 1981); K. Netalieva "Orfoepia normalary turaly' (Netalieva, 1972), "Qazaq tilinin orfoepikalyq sozdigi" (Netalieva, 1977); R. Syzdyk "Soz sazy" (Syzdyk, 1983) made a great contribution to the systematization of orthoepy norms. These studies examined phonemic changes in words, the influence of syllable and stress, and suprasegmental phenomena. Similarly, among those who contributed to the study and implementation of Kazakh orthoepy, one can mention the research of A. Zhunisbek, S. Myrzabekov, N. Uali, Z. Bazarbayeva, K. Kuderinova, A. Fazylzhan. In the works of these researchers, the problem of orthoepy was considered from segmental and suprasegmental, as well as phonological perspectives.

The problem of speech synthesis has been comprehensively examined, and its practical aspects have been implemented in practice throughout the world, particularly in Russia, the USA, Japan, France, and other countries. For example, in Russian linguistics, research related to speech synthesis was first widely discussed in 1971 in the works of M.F. Derkaya, N.G. Zagoruyko, G. Fant, and I. Lizhenkrantz. Similarly, the works of scientists such as L.V. Bondarko, S.V. Golubtsova, L.V. Zlatoustov, and V.N. Sorokina served as the basis for considering various aspects of speech synthesis.

And now there are several studies related to speech synthesis systems. These studies look at speech synthesis systems from a new perspective. For example, in the work of A.V. Frolov and G.V. Frolov, "Synthesis and speech recognition. Modern solutions" (A.V. Frolov & G.V. Frolov, 2003), linguistic units and phonetic changes, as well as the correct pronunciation of words and the function of the auditory organs, are considered. In addition, the study includes speech synthesis, text recognition, and information representation through neural systems, as well as a comprehensive analysis of punctuation marks related to speech synthesis systems in the work of I.S. Chardin and O.F. Krivnova, "Pausing in Natural and Synthesized Speech" (Krivnova & Chardin, 2014).

The work also defines the function of pauses considered at the suprasegmental level. It describes various variants of pauses during unprepared speech and reading text recorded in a computer program. The next work is called "Computer synthesis and speech cloning" (2008), written by B.M. Lobanov and L.I. Tsirulnik. This work considers the phonetic-acoustic foundations and phonetic-acoustic methods of speech synthesis through text, articulatory methods, and formant and compilation methods.

Analysis

The main goal of a speech synthesis system is to make written text sound natural and understandable. To achieve this goal, large volumes of written texts were pronounced according to orthoepic norms. For synthesized speech to be of high quality, its database must be very rich, extensive, and comprehensive (representative). This database must be analyzed based on phonetic-phonological and prosodic foundations. Synthesized speech based on a system of phonetic-phonological and prosodic knowledge becomes natural and understandable to the listener.

The practical significance of speech synthesis. Speech synthesis is of great practical importance, as it is currently widely used in many areas with the rapid development of information and communication technologies. In education, text-to-speech tools are widely used by students and in educational resources. This tool is becoming an effective learning resource, especially for visually impaired students. On mobile devices and with virtual assistants, platforms like Siri, Google Assistant, and Yandex Alice can interact with the user in natural language, converting text to speech. In transportation systems, GPS navigators provide voice-activated directions, helping drivers quickly navigate their surroundings. In government agencies and the banking sector, automated voice systems (for example, in call centers) provide customer service without a human operator. Furthermore, the development of the Kazakh language speech synthesis system is an important step in strengthening the position of the state language in the digital space. Currently, the number of systems implementing speech synthesis in the

Kazakh language is limited, and most are still under development. This increases the need for high-quality speech synthesis that takes into account the orthoepic features of the Kazakh language. When creating a speech synthesis system, its quality should be based on certain requirements. These requirements should take into account the naturalness, intelligibility, and phonetic accuracy of speech. Accordingly, the main requirements for synthesized speech are systematized below.

Naturalness. Synthesized speech should not be too far removed from the natural human voice. Voice intonation, speech rate, melody, and pauses should correspond to the context of the speech.

Intelligibility. Words and sentences should be heard clearly and distinctly. To do this, it is necessary to exclude violations of the articulation of announcers in the sounding of texts obtained for speech synthesis, that is, deviations in the pronunciation of sounds and incompatible movement of the organs of speech (tongue, lips, palate), as well as distortion or merging of words. Synthesized speech should be equally understandable to users of different age groups and social groups.

Phonetic accuracy. Speech synthesis should be developed in accordance with the phonetic laws of the language. It is especially important to maintain harmony, assimilation of vowels and consonants in the Kazakh language, as well as the orthoepic norm, because demand directly affects the natural transmission of the language.

Intonation structure. Intonation must be correctly constructed to match the meaning of the sentence. For example, a question should have a rising intonation, and an exclamatory sentence should have an emotional accent. A variety of intonation patterns helps prevent monotony and enhances the naturalness and expressiveness of synthesized speech.

To improve the quality of Kazakh speech synthesis, it is important to develop both an algorithm for modeling phonetic changes within words and phrases and a system of prosodic annotation. The interaction of these two mechanisms can contribute to the naturalness of synthesized speech, its conformity with the pronunciation norms of the literary language, and its intelligibility to listeners. First, the orthoepic algorithm aims to accurately pronounce words in written text in accordance with the Kazakh language sound regularities.

In the Kazakh language, there are frequent differences in spelling and pronunciation, such as sound interaction, the softening of hard consonants, and the front-back distinction of vowels, which directly affect the actual pronunciation of a word. If these changes are not taken into account in the synthesis algorithm, the sound will be artificial and will not correspond to the literary norm. Secondly, prosodic notation takes into account such prosodic components as melody, stress, pauses, and tempo. Without this notation, synthesized speech will not sound monotone, smooth, or natural. Intonation in the Kazakh language is essential not only for conveying meaning but also for facilitating the listener's comprehension. For example, a declarative sentence and an interrogative sentence have different intonations. Prosodic notation allows us to divide sentences into phrases, syntagmas, and rhythmic groups, so that the syntagmas of the texts included in the synthesis system are correctly placed, and the intonation is formed naturally (Bazarbayeva, 2022).

While the orthographic algorithm ensures accuracy at the sound level, prosodic notation increases the expressiveness and naturalness of speech. In this regard, in accordance with the above task, the orthographic algorithm within words and within word groups was based

on the model presented in the work “Orthographic identifier of the Kazakh language (an auxiliary tool for radio and television journalists)” (Qazaq tilinin orfoepiyalyq anyqtagyshy (radio, telezhurnalisterge arnalgan komekshi qural, 2004). An example of this can be seen in the tables below (Tables 1, 2, 3).

Table 1. Reduction of the vowel sounds /ы/i/ within or between words
1-кесте. Сөз ішіндегі немесе сөз аралығындағы дауысты /ы/i/ дыбыстарының редукциялануы
Таблица 1. Редукция гласных звуков /ы/i/ внутри или между словами

The omission of the vowels /ы/i/	
Written form	Spoken form
арғы ата (ancestor)	арғ'ата
екі ұдай (ambiguous)	йек'ұдай
жеті ата (ancestry up to the seventh generation)	жет'ата
екі ара (the space between the two)	йек'ара
ертегі-аңыз (folktale and legend)	йертег'аңыз
саптыаяқ (handled wooden bowl/cup (traditional Kazakh vessel)	сапт'айақ
сарымай (butter)	сар'май
Сарыағаш (Saryagash - proper noun/place name)	Сар'ағаш
бері әкел (bring it here)	бер'әкел
ешкіемер (lizard, whippoorwill)	йешк'емер

Table 2. Reduction of the vowel sound /a/ within or between words
2-кесте. Сөз ішіндегі немесе сөз аралығындағы дауысты /a/ дыбысының редукциялануы
Таблица 2. Редукция гласного звука /a/ внутри или между словами

The omission of the vowel /a/ (a + a, a + e, a + ө, a + o)	
Written form	Spoken form
алма ағаш (apple tree)	алм'ағаш
ата-ана (parents)	ат'ана
ауаөткел (air passage, etc.)	ау'өткөл
жаңа ай (new month)	жаң'ай
дала аптабы (the heat of the steppe)	дал'аптаб̆ы
қара өлең (traditional Kazakh folk verse, Kazakh folk song-poetry)	қар'өлөң

қара ет (lean, clean meat or meat pulp)	қар'ет
қара орман (black forest, dark forest)	қар'орман

Table 3. Reduction of the vowel sound /e/ within or between words
3-кесте. Сөз ішіндегі немесе сөз аралығындағы дауысты /e/ дыбысының редукциялануы
Таблица 3. Редукция гласного звука /e/ внутри или между словами

The omission of the vowel /e/	
Written form	Spoken form
жөке ағаш (linden tree, lime tree, Tilia)	жөк'ағаш
екпе ағаш (planted forest, forest plantation)	ʲекп'ағаш
іске асу (to be implemented)	іск'асұу
телеөнім (TV product)	тел'өнүм
теңге алу (taking money (literal; context may be needed)	теңг'алұу
телеарна (TV channel)	тел'арна
бейне өнім (video product)	бейн'өнүм
келе алмады (could not come)	кел'алмады

The tables above clearly demonstrate the patterns of sound changes within and between words. Only by systematically transforming these patterns into an algorithmic basis will synthesized speech become natural and easily understandable, in accordance with the norms of literary language.

Sound changes within a word are determined by the interaction of neighboring sounds in spoken speech. For example, the omission of the intervening vowel in the words argy ata → [arg'ata], dala aptaby → [dal'aptaƀy] allows for easy and rapid pronunciation of the word rhythm. Maintaining orthoepic and prosodic norms is a basic requirement for restoring the natural sound of speech in the Kazakh language. Only when these norms are implemented algorithmically will the synthesized word meet the requirements of the literary language, be natural, and easy to understand. As an example of these laws, the changes in sounds within words and in word groups are shown in our study. Similarly, the phenomenon of reduction is also of particular importance in the speech synthesis system. Reduction (from Latin *reductio* - return, moving back, retraction) is a phenomenon related to vowels, that is, the obscuring of a vowel within a word. Reduction is mainly experienced by stressed vowels (Qazirgi qazaq tili stilistikasy men orfoepiiasy, 2025). In the Kazakh language, reduction should be understood as the weakening and being pronounced with a reduced (obscured) quality of vowels in a certain sound environment. From an orthoepic point of view, ignoring such phenomena increases the artificiality of the synthesized word, since the natural harmony of sounds is disturbed. Therefore, the synthesis system should

automatically recognize sound phenomena at the algorithm level and select the correct pronunciation option based on them. Such an automated orthographic algorithm not only ensures phonetic accuracy but also, in conjunction with prosodic features, increases the naturalness of speech. The intonation and rhythmic structures of speech also depend on these sound compactifications. That is, the interrelation between orthographic and prosodic norms is the main indicator of the quality of speech synthesis in Kazakh. Accordingly, in our article, we provide the following rules for the speech synthesis system to sound in accordance with orthoepic norms:

- when the voiceless consonants *k, қ* are placed among words, between two vowels, and after a vowel, they are pronounced as *г, ғ*, for example, *көгжійек (көкжиек), көғалма (көк алма), көгмұз (көк мұз), қазағжері (қазақ жері), құлаг͡асұу (құлақ асу), жарығдүнүйе (жарық дүние), бұлшұғет (бұлшық ет), қонағжәй (қонақжай), қоңұргүз (қоңыр күз), қора-ғоӈсұ (қора-қопсы);*

- if the first syllable of two words forming a rhythmic group ends in the consonants *з, р, л, м, н, ң, й, у,* the sounds *қ, к* at the beginning of the next syllable are pronounced as *ғ, г,* for example, *қаз-ғатар (қаз-қатар), өнөргәсіп (өнеркәсіп), балғаймақ (бал қаймақ), азаң-ғазан (азан-қазан), жылаңгөз (жылан көз), қойгөз (қой көз), қайғайдағы (қай қайдағы) ұугекіре (укекіре), тұсаугесер (тұсаукесер);*

- if the first syllable of the words forming a rhythmic group ends in one of the vowels *а, е, у, і, и, і,* the sounds *k, k* in the second pair will be pronounced as voiced sounds: *атағонұс (ата қоныс), ботагөз (бота көз), үйгөгелүу (үйге келу), қысқыгійім (қысқы киім), қылыйгөз (қыли көз), ескігійім (ескі киім), елүугаралы (елу қаралы), үйгасында (үй қасында);*

- if the first part of a compound word or phrase ends in a vowel, sonorant, or voiced consonant sound, the sounds *қ, к* in the second part will be pronounced as voiced sounds *ғ, г: алагөл (ала көл), қарағой (қара қой), қойғора (қой қора), Талдығорған (Талдықорған), қарагөк (қара көк), қарагөлеңке (қара көлеңке), көзгарас (көзқарас), жаңғалта (жаңқалта), не͡ғыласың? (не қыласың?)* (Qazaq tilinin orfoepiyalyq anyqtagyshy (radio, telezhurnalisterge arnalgan komekshi qural, 2004).

According to these rules, if the announcer reads the text while strictly adhering to orthographic rules, the quality of speech synthesis will improve, the naturalness of the pronunciation will increase, and the listener will be able to perceive the text more easily. Similarly, when the *н* sound is adjacent to the sounds *қ, к* within a word or between words, the *н* sound becomes *ң,* and the *қ, к* sounds become *ғ, г: жаңғыйар (жаңқияр), жаңғалта (жан қалта), Қазаңғап (Қазанқап), қаңғыз (қанқыз), таңғалу (таң қалу), түң͡ғатұу (түн қату), жаңгүйөр (жанкүйер), дұшпаңгөз (дұшпанкөз)* (Qazaq tilinin orfoepiyalyq anyqtagyshy (radio, telezhurnalisterge arnalgan komekshi qural, 2004).

To ensure accurate and natural pronunciation, speech synthesis should be based on the established orthoepic norms of the Kazakh language. This is because the synthesized sound must be pleasant to the ear and pronounced in accordance with orthoepic norms, keeping the phonetic laws of the Kazakh language. It is the main indicator of language culture. The fact that some vowels and consonants change during speech is not only a phonetic norm but also demonstrates the natural form, intonation, and tone of the Kazakh language, which distinguishes it from other languages. If the synthesized word preserves all of this,

we believe we can reveal the nature of the Kazakh language and create a lexical synthesis based on orthoepy. Furthermore, the selection of speakers reading the audio recording is crucial in this study. After all, it is the speaker who brings the prepared text to life. The main task of the speaker reading the text is to read according to the rules of orthoepy. This allows the speaker to develop a consistent speech style and adhere to orthoepical norms throughout their presentation. Thus, each text entered into the synthesis system is first created in accordance with orthoepic standards. Thus, a verbal synthesis created in accordance with the natural pronunciation of the Kazakh language preserves the sonic richness of the national language and paves the way for its development in a digital society (Qazaq soz sintezin qurastyrushylarga arnalgan nusqaulyq, 2025). Here is a sample of a text created according to the orthoepic norm:

ᵘ̆ендi͡ бұл͡ гүндө // сол шыңға шыққан // жалғыш͡ шынар // балғын тартып // жас͡ құуатқа толұптұ /// қәзір͡ ʸоған // қыс͡ пен͡ айаз͡ да // тіпті / тау дауұлұ͡ да // қәтер͡ болұудаң͡ ғалған͡ ᵘ̆еді /// таң͡ алдында / бірғана сағат мызғығаны͡ болмаса // аба͡й бұл͡ түндү // ұйқұсұз ʸөткүздү /// бірәқ / әлі͡ талған͡ жоқ /// әлі͡ де / кітәп͡ петін'үңұлүуде /// үлкөн үстөл үстүндө // шалқасᵘ̆нан ашылып жатқаң͡ гітәп // бір͡ ᵘ̆емес әлденеше /// әралұуан тілдерде тұуған / бұл͡ гітаптәр // бүгүн ʸосұ жерге ʸоқұс͡ келіп // қызық͡ пас͡ қосқан /// аба͡й жаᵏ̆с'ұғұнатын шағатай / түркү кітәптарымең͡ ғатар тілін // аба͡й͡ ғыйналып түсүнөтін // араб / парсы және ʸолардаң͡ гөрү͡ де // әлі / аʸұрұрақ тійетін // ʸорұстіліндегі͡ гітәптар /// аба͡йдың жәйшылықтағ'оқұʸұнан // бүгүңг'оқұуұнұң маᵏ̆саты͡ да͡ басқа /// кітәптан алатым͡ білім нұсқаның // бүгүн тіршіліктің͡ голма͡ гол͡ герег'үшүн // асығыс қәжет͡ полғам͡ біршағ'еді /// аба͡й соңғұ͡ гүндөрдө // әсірес'осұ соңғұ түндө // ʸөзүнө͡ бұрұн / дағдылы͡ болмағам͡ бірхәлде /// ʸол бір͡ ᵘ̆ескі ғалымдар͡ ма // ᵘ̆рауаилар͡ ма // солардай͡ біргүйд'отұр /// ʸосұ жәйіна / уөзү таңырқай͡ түсүп // қызығып͡ та͡ ғарайды /// парсы / түркү͡ кітәптары // бұнұ͡ біресе шійраздың гүлзарын'әкетеді /// самарқанның мазар ғыймараттарына͡ ғадалтады /// мәрүү͡ мешхеттің мійуалы // бұлбұлдұ͡ бақтарына // салқын / самал хауұздарын'үңұлтөді /// ұл'ақындар мекен ᵘ̆еткең͡ ғырат / газᵘ̆на // бағдаттың сарайларына // медреселеріне // кітәпқаналарына͡ тартады /// нелер ықылым заманда͡р бойұ // асфақаный қайқы қылыштары // тынымсыз тайталаспен // қарш ұрұсұп͡ ʸөткөн араб / парсы / түрүк / мұңғұл тарійхына͡ да // көп͡ көп͡ көңүл тартады /// (The Path of Abai).

It is known that today, many words in the Kazakh language are pronounced according to the Russian pronunciation pattern. Such a trend is especially pronounced in everyday language use. As a result, the pronunciation of loanwords has diverged somewhat from the Kazakh language phonetic laws. During the development of the speech synthesis system, the following model was created, designed to bring the pronunciation of loanwords as close as possible to the orthoepic norms of the Kazakh language. This model was created using the spelling rules for foreign words outlined in the document *“Latyn alipbiinin zhobasy negizdemesi”*. This is because the naturalness and intelligibility of synthesized speech largely depend on the accurate pronunciation of words. This allows the pronunciation of loanwords to be brought closer to the phonetic patterns of the Kazakh language (Table 4).

Table 4. Rules of pronunciation of loanwords in the speech synthesis system (model)
4-кесте. Сөз синтезі жүйесінде кірме сөздердің дыбысталу ережесі (үлгі)
Таблица 4. Правила произношения заимствованных слов в системе синтеза речи (модель)

Transmission of loanwords	*Kazakh pronunciation*	*Additional note*
The letter “ё” is pronounced as /ө/	*бруцелөз, актөр, амөба, дерійжөр, режійсөр*	
The letter “ц” is pronounced as /ц/	*ційрк, цемент*	
The letter “э” is pronounced as /е/	*йелемент, йелектір*	
The letter “ю” is pronounced as /йу/ at the beginning of a word and after a vowel	*йуан, йуневал, кайута,*	
The letter “ю” is pronounced as /ү/ after a consonant.	*бүро, бүрократ, лүстра, рүкзак, сүжет, бүлетен, бүджет, продүсер*	
Suffixes are attached in harmony with the final syllable		
The letter “я” is pronounced as /йа/ at the beginning of a word and after a vowel	*йахта, йадро*	
The letter “я” is pronounced as /ә/ after a consonant	*бійләрд, ійталән, разрәд, калкүләтор*	
The letter “я” after the letter “и” is pronounced as /а/	*хіймійа, биологійа*	
The letter “ч” is pronounced as /ч/	*чемпійон, чат, скоч, мач*	
A syllable containing the soft sign (ь) is pronounced with front vowels	*әлбом, бәлзам, ансәмбіл, фөлклор, баталөн, корөл*	*But* If the syllable with the soft sign (ь) contains the letters “е”, “і” the front vowels are not taken into account: *премера, ателе, акварел, артикіл, барер, кегіл, лагер, релеф*
If a syllable (or word) contains the hard sign (ъ), it is dropped in any case while pronouncing	субект, обект, фелдегер, абектійв	

The letter “в” is pronounced as /в/	*авто, евро, евразия*	
The letter combination “нг” at the end of a word is pronounced as /ңк/	*боулійңк, брійфійңк, рійңк, раңк, слеңк*	
In words with the letter combination “дж”, the consonant “д” is not pronounced; only the /ж/ sound is pronounced	*менежмент, картрійж, жентелмен, котеж*	*But* When the letter combination “дж” affects its meaning, it is pronounced as *джаз, джем, бүджет, лоджій*
If the doubled letters “сс”, “мм”, “лл”, “тт”, “фф”, “нн”, “бб”, “рр”, “пп”, “гг”, “дд”, “кк”, “уу”, “оо” appear in the middle of a word, one of them will drop	*комійсар, патійсон, сійлабұус, колеж, кореспондент, кордійнат*	*But* In cases where the doubled letters affect the meaning of the word, they are pronounced as *масса, тонна, касса, доллар, брутто, нетто, ноосфера*
The final letters of words ending with the letter combinations “ст”, “вт”, “фт’, “нкт”, “ркт” are dropped during pronunciation	*фантас, фонетійс, фұутұурійс, тұурійс*	*But* It is pronounced without dropping at the end of monosyllabic words: *тест, пост, лифт*
The sound “а” at the end of loanwords is dropped if it does not affect the meaning of the word.	*кардійограм, орфрграм, капсұул, кордійнат, фонем, морфем*	*But* The sound “а” is retained in cases where it affects the meaning of the word and in two-syllable words: *математика, педагогика, статис*тика, норма, фаза
Some borrowings that start with the letter “о” are pronounced according to their version in lingua franca	*банкомат, модератор, дійректор, фонем, морфем, фонетійка, фонологійа*	
The letter “и” represents the sounds /ый/ and /ий/; this system is also preserved in loanwords	лексійкологійа, түркологійа, гендер͡патій, медійа	

According to this given model, the following examples can be given. The example is presented in a table (Table 5).

Table 5. Pronunciation of loanwords in the speech synthesis system
5-кесте. Сөз синтезі жүйесінде кірме сөздердің дыбысталуы
Таблица 5. Произношение заимствованных слов в системе синтеза речи

Loanwords pronounced with the sound /o/		
Loanwords	Pronunciation of the loanwords	Examples
остеон *(osteon; Haversian system)*	*ᵘостеᵘон*	Остейон – орталық өзекті қоршай орналасқан өте майда сүйекті табақшаларының түтікті жүйесі *(Osteon is a tubular system of very small bone plates surrounding the central nucleus)*
рентгенограмма *(a radiograph or X-ray image)*	*реҢгЕнограма*	Сүйек кесінділерінде немесе реңгЕнограмада екі құрылымды айыруға болады *(Two structures can be distinguished in bone sections or X-rays)*
The letter “э” is pronounced as the sound /e/		
Эндесмалды *(desmal (intramembranous)*	*ᵘендесмалды*	Ол ᵘендесмалды сүйек деп аталады *(It is called desmal (intramembranous) bone*
The letter “я” is pronounced as /ə/ or /a/		
Перпендикуляр *(perpendicular)*	*перпендійкуляр*	Перпендійкуляр табақша мұрын қалқасының құрамына кіреді *(The perpendicular plate is included in the nasal septum)*
Анатомия *(anatomy)*	*анатомійа*	Тізе буынында көптеген қапшық ішіндегі және қапшықтан тыс анатомійалық құрылымдар бар *(The knee joint contains many intracapsular and extracapsular anatomical structures)*
Words pronounced with the omitted soft sign (ь)		
Фронталды / фронтальды *(frontal)*	*фронталды*	Он екінші кеуде омыртқасынан басқаларының буындық өсінділері фронталды жазықтыққа бағытталған *(The articular processes of all vertebrae, except the twelfth thoracic vertebra, are directed in the frontal plane)*
Pronunciation of the soft sign (ь) with front vowels		
Монастырь *(monastery)*	*монәстыр*	Мешіт жанынан медреселер, шіркеу жанынан прійходтық, соборлық, монәстырлық мектептер ашылды *(Madrasahs were opened near mosque, and parish, cathedral, and monastic schools were opened near churches)*
Батальон *(battalion)*	*баталөн*	Орал казак-орыс атқыштар баталөнында әскери дәрігерлік қызмет атқарады *(Oral serves as a military doctor in the Kazakh-Russian Rifle Battalion)*

Pronunciation of the sound /ц/		
Висцералдық *(visceral)*	*війсцералдық*	Аталған екі доға війсцералдық болса, басқалары желбезектік деп аталады *(These two arches are called visceral, while the others are called gill arches)*
The pronunciation of two adjacent consonants		
Профессор *(professor)*	*професор*	Професор Халел Досмұхамедұлының тарих, әдебиет, тіл білімі ғылымдары саласында да бірнеше ғылыми еңбектері бар *(Professor Khalel Dosmukhamedov had also several scientific works in the fields of history, literature, and linguistics)*

Pronunciation of borrowed foreign words without changing the Russian or original version leads to a violation of the phonetic and orthoepic laws of the Kazakh language. In particular, the inconsistency of the pronunciation of loanwords with the phonetic features of the Kazakh language affects the non-compliance with the law of consonance and violation of the natural sound system of the language. Therefore, when developing the speech synthesis system, the goal was to bring the pronunciation of borrowings as close as possible to the orthoepic norms of the Kazakh language. This approach increases the naturalness of synthesized speech and makes it easier for the listener to perceive and understand. In addition, the inclusion of orthoepy for loanwords in the model and the given linguistic data provides support for the results of this study.

Results

For a speech synthesis system to produce natural-sounding and intelligible speech, established orthoepic norms must be taken into account. For this, the orthoepic norm was determined from two perspectives. One is the orthoepic norm of Kazakh texts, and the other is the orthoepic norm of loanwords. In this regard, all texts were based as much as possible on the orthoepic norm in modeling. The incorporation of these norms enables a speech synthesis system to reflect the language-specific phonetic and prosodic characteristics of Kazakh. Since the speech synthesis system is the pronunciation of the written text, compliance with the orthoepic norm enhances its functionality. Also, through the speech synthesis system, the originality, national color, and sound changes of Kazakh orthoepy will be preserved in the future. Through this, young people will be able to recognize the features, models, and rules of correct Kazakh pronunciation. Therefore, the article presents rules for the pronunciation of Kazakh words according to the orthoepic norm, based on linguistic data, and develops a model for the pronunciation of loanwords. The loanwords were taken from texts included in the speech synthesis system. These texts were taken from natural science textbooks. The sound characteristics of the loanwords in the textbook texts were analyzed and adapted to conform to the Kazakh language phonological rules. Their orthoepic model was proposed. The use of these models increases the naturalness of the synthesized speech. This allows the listener to easily perceive oral speech and correctly understand any information. The results

of the study are of practical value in improving speech synthesis systems. The proposed rules and models will also serve as a basis for future orthoepic pronunciation of various texts. They are especially useful in the fields of education, mass media, and digital content. Therefore, it is necessary to introduce loanwords into the speech synthesis system while maintaining the model of oral speech, which is based on the natural and sound harmony rules of the Kazakh language.

Conclusion

The article considers the issues of modeling the speech synthesis system in the Kazakh language based on orthoepic norms and analyzes its theoretical and applied aspects. During the study, it was found that the quality of speech synthesis depends primarily on the correct transmission of orthoepic and prosodic patterns. In this regard, the importance of systematically modeling the phonetic and phonological features of the Kazakh language, positional and combinatorial changes of sounds, as well as suprasegmental phenomena such as intonation, rhythm, and stress in the speech synthesis system was identified. The study showed that the use of orthoepic and prosodic notation in interrelation is one of the main conditions for ensuring the naturalness of speech synthesis. It has also been proven that a systematic description of phonetic patterns, such as sound changes within words and word groups, the phenomenon of reduction, and the interaction of sounds, allows for improving the quality of the synthesis system. It has also been proven that a systematic description of phonetic patterns, such as sound changes within words and word groups, the phenomenon of reduction, and the interaction of sounds, allows for improving the quality of the synthesis system. The article also examines the orthoepic features of loanwords in the modern Kazakh language and adapts their pronunciation to the phonetic patterns of the Kazakh language. A model was created to systematize the pronunciation of loanwords in spoken language. This is especially important when text-to-speech applications, since the model systematizes a large amount of material, which increases the quality of speech synthesis, and in the process of mastering it, there is a lack of consistency.

In general, the conducted study results indicate the need to comprehensively take into account orthoepic norms, phonetic-phonological patterns, and prosodic features in improving the speech synthesis system in the Kazakh language. The proposed orthoepic models and pronunciation rules can contribute to more natural pronunciation in Kazakh speech synthesis, thereby improving the intelligibility and overall quality of synthesized speech.

Reference

Базарбаева З.М., Садық Д.А., Жұмабаева Ж., Аманбаева А., 2025. Қазақ тіліндегі сөйлеу синтезінің фонетика-фонологиялық негіздері. Қарағанды университетінің хабаршысы. Филология сериясы. 30. 2 (118).

Базарбаева З.М., 2022. Қазақ фонологиясының негіздері. Алматы: Everest. 460 б.

Базарбаева З.М., 2022. Казахская интонация. Алматы: Everest. 380 с.

Базарбаева З.М., 2022. Интонология. Т 1. Алматы: Everest. 440 бет.

Байтұрсынұлы А., 2013. Алты томдық шығармалар жинағы. Тіл-құрал. Алматы: Ел-

шежіре. Т. III. 384 б.

Балақаев М., 1984. Қазақ әдеби тілі және оның нормалары. Алматы: Ғылым. 184 б.

Жүнісбек Ә., 2018. Қазақ тіл білімінің мәселелері. Алматы: Абзал-ай. 368 б.

Жүнісбек Ә., 2009. Қазақ фонетикасы. Алматы: Арыс. 312 б.

Кеңесбаев І., 2014. Қазақ тіл білімінің мәселелері. Алматы: Абзал-ай. 640 б.

Күдеринова Қ.Б., 2013. Қазақ жазуының тарихы мен теориясы. Алматы: Елтаным. 242 б.

Кривнова О.Ф., Чардин И.С. Паузирование в естественной и синтезированной речи [Электронный ресурс] URL: http://www.philol.msu.ru. (Accessed: 29.01.2026)

Қазақ тілінің орфоэпиялық анықтағышы (радио, тележурналистерге арналған көмекші құрал), 2004. Алматы: Арыс. 200 б.

Қазіргі қазақ тілі стилистикасы мен орфоэпиясы: ұжымдық монография, 2025. Алматы: Асыл ЖК. 222 б.

Қазақ сөз синтезін құрастырушыларға арналған нұсқаулық, 2025. Алматы: Асыл ЖК. 74 б.

Лобанов Б.М., 2008. Цирульник Л.И. Компьютерный синтез и клонирование речи. Минск: Белорусская наука. 316 с.

Мырзабеков С., 2004. Қазақ тілі фонетикасы. Алматы: Қазақ университеті. 247 б.

Пантюхин, Д.В., 2021. Нейронные сети синтеза речи голосовых помощников и поющих автоматов. Речевые технологии. № 3-4. С. 3-16.

Сыздық Р., 2000. Қазақ тілінің анықтағышы (емле, тыныс белгілері, сөз сазы). Астана: Елорда. 532 б.

Уәли Н., 2018. Графика, орфография, орфоэпия. Алматы: Сардар. 250 б.

Фазылжанова А., 2022. Сөз сазы және интонация (эксперименттi-фонетикалық зерттеу). Алматы: Алматы-Болашақ. 207 б.

Фролов А.В., Фролов Г.В., 2003. Синтез и распознавание речи. Современные решения. Москва: Связь. 208 с.

Ясницкий Л.Н., 2010. Введение в искусственный интеллект. Москва: Академия. 176 с.

Dutoit, T., 1997. An Introduction to Text-to-Speech Synthesis. Dordrecht: Kluwer Academic Publishers. 285 p.

Taylor P., 2009. Text-to-speech synthesis, Cambridge University Press. 567 p.

Wells, J.C., 2007. English intonation. An introduction [Text]. Cambridge: Cambridge University Press. 276 p.

Reference

Bazarbayeva, Z. M., Sadyq, D. A., Zhumabaeva, Zh., & Amanbaeva, A., 2025. Qazaq tilindegi soileu sintezinin fonetika-fonologiyalyq negizderi [Phonetic-phonological foundations of speech synthesis in the Kazakh language]. Qaragandy universitetinin khabarshysy. Filologiya seriiasy, 30, 2 (118). [in Kazakh].

Bazarbayeva, Z. M., 2022. Intonologia [Intonology]. Vol. 1. Almaty: Everest. 440 p. [in Kazakh].

Bazarbayeva, Z. M., 2022. Qazaq fonologiiasynyn negizderi [Fundamentals of Kazakh phonology]. Almaty: Everest. 460 p. [in Kazakh].

Bazarbayeva, Z. M., 2022. Kazakhskaia intonatsiia [Kazakh intonation]. Almaty: Everest.

380 p. [in Russian].

Baitursynuly, A., 2013. Alty tomdyq shygarmalar zhinagy. Tildik-qural [Six-volume collected works. Language tool]. Vol. III. Almaty: El-shezhire. 384 p. [in Kazakh].

Balaqayev, M., 1984. Qazaq adebi tili zhane onyn normalary [The Kazakh literary language and its norms]. Almaty: Gylym. 184 p. [in Kazakh].

Zhunisbek, A., 2009. Qazaq fonetikasy [Kazakh phonetics]. Almaty: Arys. 312 p. [in Kazakh].

Zhunisbek, A., 2018. Qazaq til biliminin maseleleri [Problems of Kazakh linguistics]. Almaty: Abzal-ai. [in Kazakh].

Kenesbayev, I., 2014. Qazaq til biliminin maseleleri [Problems of Kazakh linguistics]. Almaty: Abzal-ai. 608 p. [in Kazakh].

Kuderinova, K. B., 2013. Qazaq zhazuynyn tarihy men teoriyasy [History and theory of Kazakh writing]. Almaty: Eltanym. 242 p. [in Kazakh].

Krivnova, O.F., & Chardin, I. S., 2014. Pauzirovanie v estestvennoi i sintezirovannoi rechi [Pausing in natural and synthesized speech]. Retrieved from http://www.philol.msu.ru (Accessed: 29.01.2026) [in Russian].

Qazaq tilinin orfoepiyalyq anyqtagyshy (radio, telezhurnalisterge arnalgan komekshi qural) [Orthoepic guide of the Kazakh language (a manual for radio and TV journalists)], 2004. Almaty: Arys. 200 p. [in Kazakh].

Qazirgi qazaq tili stilistikasy men orfoepiiasy [Stylistics and orthoepy of the modern Kazakh language]: collective monograph, 2025. Almaty: Asyl ZhK, 222 p. [in Kazakh].

Qazaq soz sintezin qurastyrushylarga arnalgan nusqaulyq [Guide for developers of Kazakh speech synthesis systems], 2025. Almaty: Asyl ZhK. 76 p. [in Kazakh].

Lobanov, B.M., & Tsirulnik, L.I., 2008. Kompyuternyi sintez i klonirovanie rechi [Computer speech synthesis and cloning]. Minsk: Belorusskaya nauka. 316 p. [in Russian].

Myrzabekov, S., 2004. Qazaq tili fonetikasy [Phonetics of the Kazakh language]. Almaty: Qazaq universiteti. 247 p. [in Kazakh].

Pantyukhin, D.V., 2021. Neironnye seti sinteza rechi golosovykh pomoshchnikov i poyushchikh avtomatov. Rechevye tekhnologii [Neural networks for speech synthesis of voice assistants and singing machines]. № 3–4, p. 3–16 [in Russian].

Syzdyq, R., 2000. Qazaq tilinin anyqtagyshy (emle, tynys belgileri, soz sazy) [Guide to the Kazakh language (spelling, punctuation, speech rhythm)]. Astana: Elorda. [in Kazakh].

Uali, N., 2018. Grafika, orfografiya, orfoepiya [Graphics, orthography, orthoepy]. Almaty. 250 p. [in Kazakh].

Fazylzhanova, A., 2022. Soz sazy zhane intonatsiya (eksperimentti-fonetikalyq zertteu) [Speech rhythm and intonation (experimental-phonetic research)]. Almaty: Almaty-Bolashak. 207 p. [in Kazakh].

Frolov, A.V., & Frolov, G.V., 2003. Sintez i raspoznavanie rechi. Sovremennye resheniya [Speech synthesis and recognition. Modern solutions]. Moscow: Svyaz. [in Russian].

Yasnitskiy, L.N., 2010. Vvedenie v iskusstvennyi intellekt [Introduction into Artificial Intelligence]. Moscow. 176 p. [in Russian].

Dutoit, T., 1997. An Introduction to Text-to-Speech Synthesis. Dordrecht: Kluwer Academic Publishers. 285 p.

Taylor, P., 2009. Text-to-speech synthesis. Cambridge: Cambridge University Press. 567

p. [in English].

Wells, J. C., 2007. English Intonation: An Introduction. Cambridge: Cambridge University Press. 276 p.

Авторлар туралы мәлімет:

Жұмабаева Жанар Төлендиевна, филология ғылымдарының кандидаты, фонетика бөлімі, Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Құрманғазы көш., 29, Алматы, Қазақстан Республикасы.

Оспанғазиева Назгүл, PhD, Фонетика бөлімі, Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Құрманғазы көш., 29, Алматы, Қазақстан Республикасы.

Жалалова Акшай Максутовна, PhD, Шет тілдер және ҚХА кафедрасы, Темірбек Жүргенов атындағы Қазақ ұлттық өнер академиясы, Панфилов көш., 127, Алматы, Қазақстан Республикасы.

Сведения об авторах:

Жумабаева Жанар Тулендиевна, кандидат филологических наук, отдел фонетики, Институт языкознания им. Ахмета Байтурсынулы, ул. Курмангазы, 29, Алматы, Республика Казахстан.

Оспанғазиева Назгуль, PhD, отдел фонетики, Институт языкознания им. Ахмета Байтурсынулы, ул. Курмангазы, 29, Алматы, Республика Казахстан.

Жалалова Акшай Максутовна, PhD, кафедра иностранных языков и АНК, Казахская национальная академия искусств имени Темирбека Жургенова, ул. Панфилова, 127, Алматы, Республика Казахстан.

Information about authors:

Zhumabayeva Zhanar Tolendievna, Candidate of Philological Sciences, Ass. Prof., Department of Phonetics, Akhmet Baitursynuly Institute of Linguistics, 29 Kurmangazy Str., Almaty, Republic of Kazakhstan.

Ospangazieva Nazgul, PhD, Department of Phonetics, Akhmet Baitursynuly Institute of Linguistics, 29 Kurmangazy Str., Almaty, Republic of Kazakhstan.

Zhalalova Akshay Maksutovna, PhD, Department of Foreign languages and APK, Temirbek Zurgenov Kazakh National Academy of Arts, 127 Panfilov Str., Almaty, Republic of Kazakhstan.

Authors contribution:

Zhumabayeva Zhanar Tolendievna developed the article structure, considered the issues of modeling the speech synthesis system in the Kazakh language based on orthoepic norms, analyzed

its theoretical and applied nature, and advanced the concept and methodology of the study.

Ospangazieva Nazgul analyzed the phenomenon of sound changes within words and at word boundaries, sound reduction based on phonetic laws, and their interaction, as well as how such an occurrence can improve the quality of the synthesis system.

Zhalalova Akshay Maksutovna analyzed the orthoepic features of loanwords in the Kazakh language, proposed ways to adapt their pronunciation to the phonetic laws of the Kazakh language, and edited the article.

Авторлардың үлесі:

Жұмабаева Жанар Төлендиевна мақаланың құрылымын жасап, орфоэпиялық нормаларға сүйене отырып, қазақ тіліндегі сөз синтезі жүйесін модельдеу мәселелерін қарастырып, оның теориялық және қолданбалы сипатын талдап, зерттеудің тұжырымдамасы мен әдіснамасын әзірледі.

Оспанғазиева Назгүл сөз ішіндегі және сөздер жігіндегі дыбыс өзгерістерін, фонетикалық заңдылықтарға негізделген дыбыстың редукциясы мен олардың өзара әрекеттесу құбылысын, сондай-ақ мұндай құбылыстың синтез жүйесінің сапасын жақсартуға мүмкіндік беретінін талдады.

Жалалова Акшай Максутовна қазақ тіліндегі кірме сөздердің орфоэпиялық ерекшеліктерін талдап, олардың дыбысталуын қазақ тілінің фонетикалық заңдылықтарына бейімдеу жолдарын ұсынды, мақаланы редакциялады.

Вклад авторов:

Жумабаева Жанар Толендиевна разработала структуру статьи, рассмотрела вопросы моделирования системы синтеза слов в казахском языке на основе орфоэпических норм, проанализировала ее теоретическую и прикладную природу, а также разработала концепцию и методологию исследования.

Оспангазиева Назгуль проанализировала феномен звуковых изменений внутри слов и на стыке слов, редукцию звука на основе фонетических законов и их взаимодействие, а также то, как такое явление может улучшить качество системы синтеза.

Жалалова Акшай Максутовна проанализировала орфоэпические особенности заимствованных слов в казахском языке, предложила способы адаптации их произношения к фонетическим законам казахского языка, отредактировала статью.

Conflict of Interest.

onflict of interest related to this article.

…ақтығысы.

Мақалаға байланысты мүдде қақтығысы жоқ.

Конфликт интересов.

Нет конфликта интересов, связанного со статьей.