Available Online at: https://www.ejal.info
https://doi.org/10.32601/ejal.11206

Eurasian Journal of Applied Linguistics, 11(2)(2025) 69-80

EJAL
Eurasian Journal of Applied Linguistics

Segmental-Prosodic Foundations of Kazakh Speech Synthesis

Bazarbayeva Zeinep Muslimovna[a], Didar Sadyk[b]*, Aisaule Amanbayeva[c], Zhanar Zhumabayeva[d], Nazgul Ospangaziyeva[e]

[a] *Institute of Linguistics named after A. Baitursunuly, Kurmangazy, 29, Almaty 050010, Kazakhstan. Email: zeynepmb@mail.ru*

[b] *Al-Farabi Kazakh National University, Almaty, 050000, Kazakhstan. Email: ames0099@ox.ac.uk*

[c] *Institute of Linguistics named after A. Baitursunuly, Kurmangazy, 29, Almaty 050010, Kazakhstan. Email: aaisaule@mail.ru*

[d] *Institute of Linguistics named after A. Baitursunuly, Kurmangazy, 29, Almaty 050010, Kazakhstan. Email: janar_tolendi@mail.ru*

[e] *Institute of Linguistics named after A. Baitursunuly, Kurmangazy, 29, Almaty 050010, Kazakhstan. Email: nb_ospangazievan@inbox.ru*

Received: 22 March 2025 | Received in Revised Form: 14 May 2025 | Accepted: 16 May 2025

APA Citation:
Muslimovna, B. Z., Sadyk, D., Amanbayeva, A., Zhumabayeva, Z., Ospangaziyeva, N. (2025). Segmental-Prosodic Foundations of Kazakh Speech Synthesis. *Eurasian Journal of Applied Linguistics, 11*(2), 69-80.
Doi: https://doi.org/10.32601/ejal.11206

Abstract

Segmental-Prosodic elements in any language's speech synthesis involves an automated procedure for blending phonetically structured meaningful units like melody, sound intensity, phoneme duration, and pause duration based. Each of these units varies in size and degree of completeness. This study aimed to examine the segmental-prosodic foundations of a database for synthesizing Kazakh speech. Using a qualitative approach, the data was collected through documentation research and content analysis of texts. The sampled texts continued elements of semantic segmentation and exemplified how an utterance can play a significant role in segmenting the speech streams. Results reveal that segmental-prosodic elements serve to unite parts of a phrase into a cohesive whole and function as meaning-differentiating tools, facilitating the clear and unambiguous understanding of text content. The boundaries of semantic-intonational and syntactic units—syntagms—can also be conveyed through syntagmatic stress. As the ultimate element of phrasing, the syntagm reflects syntactic-stylistic and lexical-phraseological segmentation of the speech stream. The boundaries of syntagms in speech are relatively fluid and are often reflected in punctuation; however, punctuation marks do not always align with syntagm boundaries. Ther study implies that speech signals from printed texts based on segmental-prosodic means can be one of the key conditions for developing speech synthesis systems.

Keywords: Prosodic Means, Syntagm, Accent Group, Speech Synthesis, Text Division, Punctuation, Intoneme, Tone, Tempo, Intensity.

*Corresponding Author
Email: ames0099@ox.ac.uk
DOI: https://doi.org/10.32601/ejal.11206

Introduction

The term 'segmentals' is associated with individual vowel and consonant sounds of a language, which are the elementary components of oral communication, and often carry semantic significance (Deterding, 2015). Phonemes change the meaning of a word by a simple change in the spelling of a word. For instance, changing the "k" sound in "cut" to a "p" sound makes "put," results in a different word. On the other hand, prosodic or suprasegmental features represent stress, rhythm, intonation, pitch, length etc. (Wang, 2022). A lot of foundational research is conducted on the segmental-prosodic synthesis and recognition, which refined the methodology of structural analysis in phonetics (Kanter, 1988; Nikolaeva, 1978; Potapova & Potapov, 2012; Svetozarova, 1993; Torsueva, 1979). Moreover, speech synthesis involves an automated procedure for forming the current contours of melody, sound intensity, phoneme duration, and pause duration based on the analysis of certain properties of the input text and its prosodic markup. The latest developments in psycholinguistics, speech activity theory, neurolinguistics, and computational linguistics are related to the creation of such Artificial Intelligence enabled programs that convert spoken voice into text and vice versa, with the press of a single button.

Contextually, speech flow is divided into individual segmentals or suprasegmentals of varying size and degree of completeness, which represent phonetically structured meaningful units: accentual units and syntagms (Hall, 2007). Accentual units are the minimal meaningful group of words united by rhythmic stress, which falls on the last syllable of the group. Accentual units are sometimes separated from one another by a pause and are marked by rhythmic stress, although the presence of a pause is not a necessary condition for distinguishing accentual units. Accentual units often coincide with syntagms; however, a syntagm may sometimes consist of several accentual units, in which case the boundary signals between them are formed by changes in melodic contour, intensity, and duration. If an accentual unit corresponds to a syntagm, in addition to pitch interval shifts and other intonational parameters, a pause is often observed at the junctions.

A syntagm is defined as independent in intonational sense, corresponding to a part of a sentence or an entire sentence, depending on the speech situation and context, and can take the form of a word, a phrase, or a sentence, expressing a complex semantic unity (Huang, 1998). The syntagm is the minimal unit of a speech segment in which syntactic and intonational meaning is manifested. It is a universal phenomenon found in all languages, and its formation in the flow of speech is not governed by rigid rules but is determined by various semantic connections within the text, as well as by the speaker's train of thought, perception, and understanding of the situation. A syntagm is understood as a universal meaningful unit characteristic of all languages, emerging in the segmentation of an utterance or text.

In the context of the Kazakh speech, particularly, accentual units and syntagms act as minimal phonetic and semantic-intonational units to determine the semantic-intonational segmentation of the Kazakh speech (Abilbekov et al., ,2024). Speech synthesis presents unique challenges due to the language's distinctive prosodic features, such as rhythmic-syntagmatic stress placement and the complex interaction between phonetic and semantic units, making accurate and natural-sounding speech particularly difficult to achieve (Karnevskaya, 2005). As is well known, in Kazakh utterances, each word does not carry separate stress; instead, it is replaced by rhythmic-syntagmatic stress localized at the end of the accentual group (Kaliyev & Rybin, 2019). The division into accentual units is determined by the meaning of the utterance. For example: *Қазақ публицистикалық стилінің // алғашқы кезеңі // синтаксис саласында // біраз ерекшеліктерді көрсетеді. Бұл силь – // қазақ тілінде // проза жанрының // алғашқы үлгілердің бірі. Мұнда проза // синтаксисінің заңдылықтары // бірден қалыптасып // кете алған жоқ [The first stage of the Kazakh public style demonstrates some features in the sphere of syntax. The tendencies of the prose syntax did not form in an instant].*

Despite the progress made, Kazakh speech synthesis systems currently face significant challenges related to accurately reproducing the language's unique prosodic structures. Therefore, a dire need was felt to develop a segment-prosodic model tailored to Kazakh speech synthesis, thereby addressing existing gaps. The rationale for this research lies in addressing these unresolved issues by developing an integrated segment-prosodic framework. Unlike previous research, which often treats prosodic features separately, this study explicitly focuses on the interconnectedness of segmental and suprasegmental elements and integrates them comprehensively into a unique modeling process. The primary objective of this study was, therefore, to investigate the prosodic foundations of Kazakh speech synthesis, explicitly focusing on segmental and suprasegmental features. The study also attempted to identify the segmentation boundaries of spoken discourse divided into syntagms and phonetic words, and how these supra phrasal units changed into larger speech segments, and their tonal levels got influenced by syntagmatic and logical stress in the Kazakh language.

Literature Review and Theoretical Framework

The transmission of intonational characteristics in segmental-prosodic synthesis gets affected by establishment of syntagm boundaries as well as semantic content. Studies recommend that, when dividing a text into syntagms, it is crucial to avoid placing a syntagm boundary where it may disrupt the semantic perception of speech, particularly between an object and its attribute. Specific rules of syntagmatic segmentation are used to determine syntagm boundaries, based on punctuation, morphological and syntactic analysis of the text, as well as statistical analysis of syntagmatic segmentation in natural speech (Ziyadullayeva, 2023). Pauses between syntagms play a role in conveying certain syntactic and semantic relationships. A syntagm boundary can be marked not only by a physical break in the speech signal but also by a sharp change in pitch and/or other prosodic characteristics that are perceived as a disruption of the smooth flow of speech. Thus, syntagm is one of the essential factors in the accurate transmission of logical meaning.

As the smallest significant part of an utterance, the syntagm is structured through prosodic means and serves as the semantic-intonational and syntactic unit in which the intonemes of spoken speech are realized (Rza & Iraj, 2022). Each intoneme is characterized by its own prosodic features. As a result of a linguistic-acoustic analysis of various types of utterances, the differential features of intonemes have been identified (Overath & Paik, 2021). For this purpose, not the absolute values of fundamental frequency, intensity, and duration — which are the primary parameters of intonation—are considered, but rather the relative changes in these primary parameters, such as higher-lower, faster-slower, wider-narrower, and so on. At the same time, linguistically relevant properties of intonation, such as ranges, intervals, and levels, were interpreted about their semantic and functional aspects. In other words, the analysis focuses on changes in acoustic parameters that carry semantic meaning and influence the content of the utterance.

The methodological basis of this study is founded on the theoretical concepts of Russian and Kazakh scholars concerning the investigation of prosodic systems in various languages. Early works in this domain laid the foundation by systematically analyzing the roles of rhythm, intonation, and stress in natural speech. Significant contributions to the study of prosody have been made by Bondarko, Verbitskaya, & Gordina (2004), who established key frameworks for understanding the interaction between segmental and suprasegmental features. These studies have highlighted the importance of prosodic elements as integral to speech comprehension and production, suggesting that prosody plays a crucial role in shaping meaning and communicative intent. Past researches on prosodic structures (Antipova, 1986; Krivnova, 1998) explore various aspects of prosodic systems – such as the organization of speech rhythm, the patterns of intonation, and the distribution of stress within utterances. Their investigations provide detailed descriptions of how prosody interacts with phonetic and morphological structures in language, offering insight into both language-specific and universal prosodic phenomena. This body of work establishes a comprehensive theoretical backdrop for understanding how prosodic features contribute to the naturalness and intelligibility of speech.

Closely aligned to these foundational studies, there are researches on speech synthesis and prosodic modeling (Dubois, 1993; Kuznetsov, 2007; Lobanov & Tsirulnik, 2008), which have extended prosodic theory into the applied realm of text-to-speech (TTS) synthesis. These studies address the challenges involved in replicating natural prosodic patterns in synthetic speech. They proposed models that integrate both segmental (phoneme-level) and suprasegmental (prosodic) features, demonstrating that explicit modeling of prosodic structures can significantly enhance the intelligibility and expressiveness of TTS systems. Further advancements in theoretical and practical approaches to prosody and speech synthesis are presented in the works like (Flanagan, 1965; Taylor, 2009; van Santen, 1994), which have introduced innovative methods for analyzing and replicating prosodic patterns in synthetic speech, emphasizing the importance of dynamic prosodic variation over static, rule-based approaches. Their work contributes to a deeper understanding of how prosodic cues—such as pitch, duration, and amplitude variations—are interwoven with linguistic information to produce speech that is both natural and contextually appropriate.

A few recent studies have also contributed significantly to the understanding of prosodic features in speech synthesis. For instance, Chen et al. (2025) investigate the use of prosodic sketches to control expressive speech synthesis. Their study not only highlights the challenges of incorporating prosodic features into expressive TTS systems but also proposes innovative solutions aimed at enhancing the emotional quality and naturalness of synthesized speech. Additionally, Bellinghausen et al. (2024) examine how prosodic cues of uncertainty are perceived differently by individuals with autism spectrum disorder compared to neurotypical adults, providing valuable insights into how prosodic features can be manipulated to express different levels of uncertainty in speech. Moreover, Goldrick & Cole (2023) explore exemplary models of speech production with a focus on how prosodic features are represented and processed in speech synthesis. Their research bridges theoretical and practical perspectives by demonstrating that when prosodic elements are accurately modeled, the naturalness of synthetic speech can be significantly enhanced. This further reinforces the argument that a deep theoretical understanding of prosody is essential for developing advanced TTS systems.

Collectively, these studies form a robust theoretical framework that underpins the present research. These studies not only validate the importance of prosody in natural speech but also demonstrate the practical benefits of integrating detailed prosodic models into speech synthesis systems. The reviewed literature emphasizes that a successful TTS system must account for both the segmental and suprasegmental aspects of speech. In the context of Kazakh, this integration becomes even more critical given the language's distinct prosodic characteristics. The current study aims to extend these theoretical insights by developing a segment-prosodic model tailored to Kazakh speech synthesis, thereby addressing existing gaps in both theoretical understanding and practical implementation.

Methodology

Research Design

The study employed several methods of experimental research to encompass auditory, prosodic, and phonological analysis, as well as perceptual analysis and computer-based data processing. In modern experimental-phonetic research, a structural method is usually applied, which involves a comprehensive analysis of all interrelated elements such as physical characteristics (fundamental frequency, intensity, and duration), and prosodic features, emphasizing the interconnectedness of segmental and prosodic phenomena.

Sampling

Documentation research requires sampling of texts to meet the study prerequisites, with each text segmented into semantic-intonational and syntactic units. The eligibility criteria for sampling the texts were (a) a text must refer to methods developing speech synthesis systems based on prosodic means; (b) it must refer to typological properties of semantic segmentation, aligned with the orthoepic norms of the Kazakh language (c) it must relate to an educational platform. The speech samples, selected for auditory analysis, were carefully chosen to represent a variety of communicative contexts and prosodic patterns, ensuring the validity and comprehensiveness of subsequent instrumental analyses.

Data Collection

The secondary data was collected from the sampled texts through examples that segment the text into semantic-intonational and syntactic units, identified to further develop speech synthesis. These examples displayed the typological properties of semantic segmentation, directly relating to the development of speech synthesis systems based on prosodic means and aligned with the orthoepic norms of the Kazakh language. These examples were used to describe tonal, temporal, and dynamic characteristics of each text to examine how to transform textual data into speech signals.

Data Analysis

The study involved various stages of data analysis including subjective methods of auditory analysis, and objective methods like computer data processing. During the first stage of auditory analysis, a screening analysis was conducted, during which unsuccessful or "defective" linguistic material (poorly read or not conforming to the orthoepic norms of the language) was discarded. The auditory method analyzed segmental-prosodic characteristics of certain types of utterances in the synthesis of artificial speech. The auditory analysis included prosodic features such as pauses, pitch movement direction, tempo, and volume, with relative-comparative characteristics such as fast-slow, high-low, loud-quiet, etc. During auditory analysis, issues related to the localization of pauses in spoken discourse were also addressed. The next stage of computer data processing was carried out using specialized software, to generate numerical data on tonal, temporal, and dynamic parameters, focusing not on the absolute values of prosodic parameters but on their relative variations. The combination of subjective and objective research methods provides a reliable foundation for obtaining accurate and valid results in the description of segmental-prosodic features for further speech synthesis. This structured and phased approach ensures consistency in data collection and reliability in the identification of relevant prosodic characteristics.

Results and Findings

There are several revelations of this study. First, it was observed that the transmission of intonational characteristics in segmental-prosodic synthesis gets affected by establishment of syntagm boundaries as well as the semantic content. Specific rules of syntagmatic segmentation are used to determine syntagm boundaries, based on punctuation, morphological and syntactic analysis of the text, as well as statistical analysis of syntagmatic segmentation in natural speech. For the development of speech synthesis programs, the most relevant parameters are tonal, temporal, and dynamic characteristics, as they perform a meaning-distinguishing function. Intonation components, or prosodic features, as they are often called, function in spoken language in close interconnection and unity, with various combinations of these elements serving to structure an utterance or text. The study also found that coherent spoken speech is inconceivable without a range of inherent properties, including changes in melody (fundamental pitch), duration, tempo, pauses, intensity, and timbre. However, not all these constant features are equally well perceived by humans, nor do they all play the same role in the organization of speech flow

Melody And Speech Synthesis Across Frequency Intervals

The uniqueness of each language is most prominently expressed through melody (changes in pitch), which is why the phrase "melody of a language" exists. Melody is considered the primary and universal component of intonation across various languages. Melody can serve multiple functions. Along with pauses, it acts as a means of segmenting speech while also connecting its parts. Together with lexical and grammatical means, it contributes to the conveyance of meaning. It is also the most important tool for distinguishing communicative types of utterances, such as statements, commands, questions, and exclamations. Along with syntagmatic stress and timbre, it ensures the precise transmission of the logical structure of spoken language and its emotional nuances. While melody may vary depending on the individual characteristics of a speaker's oral speech, there are fundamental melodic patterns typical of speakers of the same language.

Acoustically, melodic characteristics of speech correspond to the fundamental frequency (F0), which changes over time. During speech production, constant fluctuations in F0 values occur. The primary indicators of F0 values primarily depend on the individual pitch of the speaker's voice. The gender-based differences in vocal fold length (where men's voices, due to longer vocal folds, are generally lower than women's) cannot be considered a distinctive feature, as intonation is determined by secondary (relative) indicators rather than absolute pitch values. Emotional arousal may cause a slight increase in F0 values, but this does not affect research outcomes. The segmental composition of an utterance has an even lesser impact on the overall melodic contour. While F0 values may change slightly depending on individual sounds and sound combinations—since speech sounds have inherent physical properties influenced by their spectral characteristics, these variations are not of primary importance to a phonetician. Differences in F0 across narrow and wide vowels or between hard and soft consonants are not crucial for phonetic research. This is particularly true for Kazakh, where, due to the law of vowel harmony, words are pronounced consistently as either entirely hard or entirely soft.

Thus, the factors mentioned above do not affect the overall melodic contour of an utterance. In Kazakh, the tonal contour may change under the influence of syntagmatic and logical stress, that is, the most semantically significant and emphasized word. A distinguishing feature of an utterance is often the direction of the fundamental frequency (F0) contour, which can be rising, falling, rising-falling, falling, or level. In addition to the overall direction of the fundamental tone, other relevant acoustic (secondary) parameters include melodic (tonal) intervals, ranges, and registers (levels). In linguistic literature, an interval is understood as the magnitude of pitch rise or fall within a syllable or between syllables, determined by the ratio between the maximum and minimum F0 values within a given segment. An interval is characterized by its direction, which can be rising, falling, or level. For example, while the intonation of declarative sentences across different languages may share a similar F0 movement direction, it differs from the intonation of imperatives or exclamations due to subtle tonal interval variations. Positive and negative frequency intervals are distinguished in speech analysis. A frequency interval is considered positive if the first measured fundamental frequency (F0) value is lower than the second. Conversely, if the first value is higher than the second, the frequency interval is classified as negative. For instance, experimental-phonetic research has demonstrated that the intonation of inter-positional parenthetical units in Kazakh and French is characterized by negative tonal intervals at both junctions with the main sentence. In Russian, in most cases, the first junction is marked by negative tonal intervals, while the second is marked by positive ones.

Frequency Ranges and Intonational Characteristics

Frequency ranges play a significant role in determining the importance of syntagms. A range is defined as the difference between the maximum and minimum F0 values in the analyzed segment. There are three types of frequency ranges: average, expanded, and narrowed. Emotionally charged utterances are typically characterized by expanded frequency ranges. For example, in many languages, including Kazakh, exclamatory sentences and imperative constructions with a commanding meaning exhibit more expanded tonal ranges compared to other communicative types. An example from M. Auezov: *Енді қаптау керек, қапта!* (Now it must be packed, pack it!). Depending on their magnitude, frequency ranges can indicate the main or secondary idea in sentences containing parenthetical elements. Regardless of their position, parenthetical constructions across different languages are characterized by narrower tonal ranges compared to the main body of the sentence. For example: *Жәнібек бұл кезде, өлі түлкінің тілін суырып ап, кәрінің алғыр қыранын қызылдатып отыр екен (*Meanwhile Zhanibek, having taken out the dead fox's tongue, was feeding the old agile eagle) (Mashakova, 2021).

In the intonational characteristics of an analyzed segment, the tonal (mid-frequency) level or register plays a crucial role. For instance, although melodic contours may be similar, imperative sentences are realized at higher tonal levels than declarative sentences. Parenthetical units in interposition and postposition are characterized by lower tonal levels compared to the main body of the sentence. Higher mid-frequency levels indicate the semantic importance of the analyzed segments, while lower tonal levels signal a reduced semantic load. In experimental studies, among the frequency characteristics commonly used is the maximum fundamental frequency (F0 peak) of various speech segments, such as phrases, syntagms, and rhythmic groups. Additionally, the rate of F0 change, which determines the steepness of intervals on ascending or descending segments, is sometimes employed as a supplementary parameter.

For differentiating various types of utterances, the following acoustic parameters have proven to be relevant namely shape, starting and ending levels, and range of the melodic contour; maximum F0 value, tonal level of the syntagm, and tonal intervals at syntagm junctions. The shape of the melodic contour varies depending on the communicative type of utterance. For example, the melody of completion, interrogative sentences with question words, imperatives, exclamations, as well as postposed parenthetical elements and vocatives, is characterized by a lowering of the fundamental frequency at the end of syntagms. Conversely, the melodic contours of incompleteness, general questions with interrogative particles, echo questions, enumerations, as well as proposed parenthetical elements and vocatives, are characterized by a rising melodic contour.

In addition to changes in fundamental frequency (F0), an essential aspect of melodic description is the position of the melodic peak within a phrase, that is, the highest F0 value. In the melody of incompleteness and certain types of interrogative intonation, the F0 peak typically occurs at the end of the utterance. In the melody of completeness, imperatives, exclamations, and some types of questions, melodic peaks are localized in the middle or closer to the beginning of the utterance. Different communicative types of utterances also vary in terms of frequency range width (melodic contour). Narrow tonal ranges characterize the melodic contours of parenthetical elements, incompleteness, and completeness. Wide frequency ranges are typical of exclamations and imperatives. Interrogative utterances generally have medium frequency ranges. The melodic contour of most models begins at the second level (incompleteness, completeness, interrogative utterances). Imperatives and exclamations start at the fourth level or, as a rule, reach the fourth level, being realized in higher tonal registers.

Parenthetical Units in Speech Synthesis

Parenthetical units are typically realized in the lower half of the speaker’s vocal range. Tonal intervals at the junctions of parenthetical units with the main part of the sentence are formed through pitch interval shifts. The junctions of the analyzed segments are primarily characterized by negative frequency intervals. A sharp shift in melodic intervals is observed in categorical imperatives and certain types of exclamations. In such phrases, a break in the melodic contour is accompanied by a steep pitch drop within the vowel of the semantic center. One of the essential conditions for the existence of speech segments is their temporal extension, that is, duration—the specific time required for the pronunciation of different speech units such as syntagms, phrases, supra-phrasal units, and texts. Any speech segment is characterized by a certain temporal length, necessary for its articulation and subsequent perception by the listener. It has been established that the physical duration of a segment depends directly on various factors, such as the inherent duration of the sound segment, its degree of prominence, which is influenced by the presence of different types of stress (phrasal, logical, emphatic), syllable type, positional placement, and so on. The intrinsic duration of sounds is determined by articulation specifically: a higher tongue position reduces the inherent duration of a vowel, while the presence of labialization, on the contrary, increases it.

The perception of speech is significantly influenced by the ratio of syllable durations within a word, words within a syntagm, syntagms within a phrase, and phrases within supra-phrasal units, which is directly related to speech tempo. Tempo is typically defined as the ratio of the total segment duration to the number of syllables. As a relative measure, tempo carries a specific linguistic function. It can indicate the degree of importance of a speech segment for the listener: segments with higher informational load are pronounced at a slower pace compared to those carrying less significant information. This universal tendency has been observed and experimentally confirmed in many languages, where syntagms with greater semantic weight require more time for pronunciation. In differentiating communicative types based on tempo, only a general tendency can be established. Syntagms with interrogative and incomplete intonation tend to have the shortest average sound durations, indicating a faster speech tempo. A slower speech tempo is observed in imperative and exclamatory utterances. In multi-syntagmatic phrases, intonationally conditioned tempo variations reflect differences in the relative importance of certain syntagms compared to others.

Syntagmatic Segmentation in a Speech Synthesis Program

When developing a speech synthesis program, particular attention is paid to segmenting the speech stream. Proper division into syntagms is determined by the semantic content of the utterance and the degree of expansion of its primary and secondary elements. The same utterance, depending on the speaker's intention, can be delivered with varying syntagmatic segmentation. Let us view the following example taken from R. Syzdyk’s book: *Сөздерді жазылуынша оқудың өзі де себепсіз емес (Reading the words exactly as they are written is not without a reason).* This utterance may consist of two or more syntagms depending on the context: *Сөздерді жазылуынша оқудың // өзі де себепсіз емес. / Сөздерді // жазылуынша оқудың // өзі де себепсіз емес./ Сөздерді // жазылуынша оқудың // өзі де // себепсіз емес.* Such a relationship between the parts of an utterance is determined by the specific interpretation of the situation by the speaker or listener. The intonational structure of a syntagm is achieved through melodic, dynamic, and temporal means. When several accentual groups are combined into a single syntagm, the syntagmatic stress is intensified by the weakening of preceding rhythmic stresses.

A more complex level of segmentation in the speech stream is represented by sentences, which may consist of several syntagms or just one. A specific intonational contour and the presence of various types of stress ensure the grouping of words into such phonetic unity. In any complete sentence, there is a communicatively structured intonational center, also known as phrasal stress, which serves two main functions (1) Differentiating the communicative type of the utterance; (2) Separating utterances from one another, performing a delimitative function. Phrasal stress, while organizing the utterance, is localized on the last word of the phrase. In the following examples, it is shown how, in neutrally pronounced utterances, phrasal stress shifts to the end of the sentence as the number of words increases, serving as the intonational center: *Тұрдым. (Woke up); Ерте тұрдым. (Woke up early); Мен бүгін // ерте тұрдым. (I woke up early today); Әдеттегідей // мен бүгін // ерте тұрдым. (As usual, I woke up early today).* The function of segmenting an utterance into speech segments is also performed by logical stress, which highlights individual words within the utterance. Logical stress is closely connected to semantics: the placement of logical stress can alter the semantic nuances of the utterance.

Let us dwell on the utterance *Жазда балалар қорада ойнайды (In summer, children play in the shed).* Depending on the communicative situation, every word of this utterance can be emphasized. *1. Жазда қорада* ***балалар*** *ойнайды (It is children who play in the shed in summer). 2. Жазда балалар* ***қорада*** *ойнайды (It is in the shed where the children play in summer). 3. Балалар қорада* ***жазда*** *ойнайды (It is in summer when the children play in the shed).* In a specific contextual aura, logical stress, as an accentual realization of a latent presuppositional category, overlays phrasal stress—the main prosodic tool organizing the utterance. In the Kazakh language, logical stress is conveyed in two ways: first, by changing the word order, i.e., inversion, where the logically emphasized word is placed at the end of the utterance near the predicate; second, without altering the word order, when the speaker or reader accentuates a semantically significant word.

Punctuation in Speech Synthesis

Experimental phonetic analysis of various texts in modern languages indicates that prosodic elements serve as a means of expressing punctuation marks in spoken language. Considering the close interconnection between intonation, semantics, and syntax, one can judge its significant role in conveying punctuation in oral speech. It is essential to note that the written form of language does not have direct correspondences in oral speech. Speech intonation serves to connect parts of a phrase into a unified whole and acts as a meaning-differentiating tool, helping to accurately and unambiguously determine punctuation placement.

In fact, punctuation fixes the logical segmentation of an utterance and conveys, in writing, the semantic nuances that in live spoken speech are explicated through specific intonational structuring of the phrase. The primary purpose of punctuation marks is to delineate sentences and their parts within a text and to establish semantic relationships between them. Punctuation used in written language interacts with the prosodic segmentation of text in oral speech. In other words, the segmentation of oral speech through prosodic means is reflected in writing via punctuation marks. Furthermore, it has been established that punctuation marks in written language can correlate with the intonemes of the Kazakh language, whose prosodic characteristics have formed the basis for developing speech synthesis. The punctuation marks that typically signal the end of a sentence are the period, exclamation mark, and question mark. The period marks the end of declarative sentences, whether complete or incomplete, simple or complex. It corresponds to a longer pause than any other punctuation mark, as it is usually placed at the end of a grammatically complete unit of expression. The duration of the pause depends on the syntactic and semantic relationship between the consecutive sentences. The period intonationally corresponds to the intoneme of completeness, with a descending or rising-descending movement of the fundamental tone. The intensity changes in direct proportion to the melody.

Additionally, there is a noticeable slowing down of the final syllables of the phrase, as seen in this example: *Ана тілінің // тарихын білу // әркімгеде қажет (Everyone needs to know the history of their mother tongue).* The *exclamation mark* is used in writing at the end of imperative and exclamatory sentences, which conclude with a prolonged final pause. Such sentences are characterized by various movements of the fundamental tone, typical of the intonemes of exclamation and command: rising, rising-falling, falling-rising, and diverse tonal nuances. Logically emphasized words are accented by the maximum values of tonal, dynamic, and temporal parameters. *Қалды-ау, // қайран туған жер! Ендеше // басты қатырма (Farewell, motherland! And thus, don't confuse me).* The *question mark* signals the end of a sentence, accompanied by a prolonged pause. Depending on the type of question, sentences are characterized by rising or falling tonal endings, corresponding to the intonemes of general and specific questions. Temporal and dynamic parameters of intonation also play a role in shaping interrogative sentences. The peak values of prosodic parameters are localized on logically emphasized words. *Менің қасыма // келдіңіз бе? Соғыста // неше жыл болдың? (Have you come to me? How many years have you spent at war?)*

The *ellipsis* also belongs to punctuation marks that serve a meaning-distinguishing role, which in spoken language is conveyed through corresponding intonation. The ellipsis is used in incomplete utterances where a thought is interrupted, depending on the context and situation. Consequently, such sentences are characterized by an unfinished thought, expressed through the intoneme of incompleteness. *Бұл істерің // басқаларға өнеге ете // білеулерің керек... (You must leave these deeds as an example to the rest).* The *semicolon*, too, as a separating punctuation mark, divides parts of a sentence that are more or less complete in meaning and structure. Depending on the context, the semicolon functions either closer to a period or a comma, which is reflected in the varying length of inter-syntagmatic pauses. The intonational design of such sentences corresponds to the intonemes of completeness, incompleteness, and insertion, with their characteristic prosodic features. This punctuation mark is frequently found in asyndetic complex sentences as well as for connecting homogeneous members. *Жолдастарыңды сыйла, // өтірік жала жауып, // олардың тынышын алма. (Respect your peers; do not bother them with defamation).*

The *colon* serves as a means of structuring explanatory relationships between parts of a complex or simple sentence. In complex asyndetic sentences, the second part following the colon reveals the content of the first part. In sentences with homogeneous members, the colon is placed after a generalizing word. The first part, preceding the colon, is separated by a slight pause and is characterized by a rising or occasionally rising-falling melodic contour. Depending on the situation and context, it conveys the intoneme of incompleteness or the intoneme of completeness, along with their dynamic and temporal indicators. *Оның өлеңдерінде // орнымен жұмсалған: тұғырдан түсу, // шепті бұзу сияқты фразалар. Ақын // екі түрлі жаңалықты // енгізген: // бірі – // жаңа перифрастикалық тіркестер жасауы, // екінші – // жаңа фразеологизмдер ұсынуы (Such phrases as leave one's post or breaking the row are used successfully in his poems. The poet introduces two novelties: first is making paraphrased collocations, and the second is creating new phraseological units). (*Balaqaev, Syzdyqova, & Janpeisov, 1968*).*

The *comma* performs both separating and emphasizing functions. In complex sentences, as well as in simple sentences with homogeneous members, it serves a separating role, while in other cases it acts as an emphasizing tool (for parenthetical elements, introductory units, direct address, etc.). The comma divides an utterance into meaningful parts with minor pauses. The emphasized segments are pronounced with a rising or rising-falling melodic contour, influenced by temporal and dynamic parameters, embodying the intoneme of incompleteness and the intoneme of insertion. *Осы кезде үзіліс бітіп, // профессор аудиторияға // қайта келді. Сөйтіп, // үш топқа бөлініп, // кейде үш топ бірлесіп, // бірігіп кетіп, // таңертеңгі сағат төрттен // тұрып алып, // тынбастан даярланып жүрдік (Then the recess ended and the professor returned to the lecture hall. So we divided into three groups, these groups sometimes uniting, woke up at four in the morning and prepared rigorously).*

The *dash* and *parentheses* are used to semantically and grammatically emphasize parts of a sentence, typically serving as additional remarks, explanations, authorial clarifications, or interpretations. These punctuation marks are also employed to separate asyndetically connected sentences. The dash and parentheses correspond to pauses and shifts in the intonational contour of the sentence. The emphasized parts of the sentence align with intonemes of incompleteness and insertion, characterized by rising, level, and low tones, often accompanied by an accelerated tempo of delivery and reduced intensity. *Етістіктің тұйық түрі // (немесе қимыл атауы) // екі түрлі тәсілмен жасалған: // бірі – // қазіргіше у журнағы арқылы, // екіншісі – // мақ журнағы арқылы (The infinitive of the verb is created in two ways: first – with the contemporary "u" suffix, the second – with the help of "mak" suffix).*

As a semiotic tool, punctuation marks are chosen by the writer in accordance with their intent and the nuances of thought they wish to convey. This establishes a connection between spoken and written language, reflected in the correspondence between the content and the form of expressing thought within the framework of the selected punctuation mark. Prosodic features in spoken language operate in close interconnection and unity, with various combinations serving to structure an utterance or text. In some cases, semantic-syntactic markers align with prosodic and punctuation features, while in others, the rhythmic-melodic organization of a sentence in spoken language becomes the sole means of expressing syntactic meaning. Thus, modern punctuation in the Kazakh language, fundamentally phonetic in nature, serves as a tool for text delimitation and expresses the semantic-syntactic and intonational-semantic segmentation of written speech. This provides a reliable foundation for developing speech synthesis programs.

Discussion

This study examined the problem of segmenting the speech stream from various perspectives and aspects, presenting significant interest for linguists studying issues related to speech science and speech synthesis. With the so-called pragmatic turn in linguistics, scholars have focused on the detailed study of real speech phenomena in all their complexity and diversity. Several questions were raised, namely What is the role of the communication context? What are the principles of speech construction? Into what parts is speech divided, how are these parts connected, and which segmentation methods are semiotically relevant? In this context, it is important to note that computer technologies have given shape to a new paradigm for studying speech communication and psycholinguistic research. The phoneticians look for new directions in linguistics to determine the role that melodic, temporal, and dynamic features of intonation play not only in the semantic segmentation of the speech stream but also in speech activity in its broadest sense, in connection with extralinguistic factors.

The modeling of the segmental-prosodic database for creating speech synthesis that ensures sound as close as possible to naturally normalized speech is the result of many years of research into spoken language conducted by experienced specialists in segmental and suprasegmental phonetics. Based on experimental phonetic analysis, it has been identified that with a comma, the predominant movement of the fundamental tone is rising and level in non-final syntagms, the pause is short, and the amplitude of intensity of the vowel in the final syllable of the first syntagm is greater than in the second. With a semicolon, the tone exhibits a descending movement in the non-final syntagm, the pause is medium in length, and the intensity of the stressed syllable in the first syntagm is greater. For a dash, the tone is characterized by a level or rising movement in the non-final syntagm, and the pause is long. A colon is marked by descending intonation in the non-final syntagm, with a long pause. The juncture between syntagms is accompanied by pauses of specific lengths: minimal with a comma, medium with a semicolon, and long with a colon or dash. A longer pause signifies a weaker connection, while a shorter pause indicates a closer relationship between the units of intonational-semantic segments being divided.

Moreover, an increase in duration is one of the most significant prosodic means of emphasizing a word, highlighting its semantic weight, and creating an intonational center. A tendency toward slower tempo at the end of an intonational unit—a syntagm—has been observed. This slowdown, in our view, is related to the fact that the most semantically significant element (rheme) and the most prosodically emphasized word are often located at the end of the intonational unit. It is relevant to recall the word order in Kazakh utterances, where the predicate, which serves as the intonational center, is positioned at the end of the phrase. The deceleration of tempo at the end of a syntagm serves to create syntagmatic cohesion and functions as a marker of segmentation within the speech flow, dividing it into meaningful units. One of the key components of intonation is the pause. A pause represents a temporary stop or breaks in sound and simultaneously functions as a semiotic device, performing a specific semiological function. Depending on their purpose, pauses can be physiological, emotional, intellectual, situational, and so on. The linguistic function of pauses lies in segmenting the text into phrases, syntagms, rhythmic groups, and words, as well as in conveying additional semantic and emotional relations in combination with other intonational components.

The absolute duration of pauses depends on the tempo of conversational speech and is generally directly proportional to it. A pause between supra-phrasal units is significantly longer than an inter-phrasal pause, while inter-phrasal pauses are longer than inter-syntagmatic ones. Moreover, the temporal intervals created by pauses allow the listener to process linguistic information, retain it, and construct the semantic structure necessary for comprehension. In natural speech, different types of pauses are distinguished: grammatical pauses, which separate intonationally structured parts of a phrase; emphatic pauses, which highlight certain elements; and hesitation pauses, which indicate uncertainty and play a crucial role in speech synthesis. A pause segments speech into intonationally and semantically meaningful units, expresses the nature of connections between parts of an utterance, and serves as a means of emphasizing a word or syntagm both semantically and emotionally.

Experts working on automatic speech synthesis systems note that the localization of intonational boundaries in the synthesized text is one of the main tasks of the accentual-intonational transcriber, which is an essential component of any automatic speech synthesis system. For each intonational boundary, the issue of marking the boundary with a physical pause is addressed, and in cases where it is deemed necessary, the pause is assigned its categorical duration (Krivnova, 1998). The practice of developing automatic speech synthesis systems for arbitrary text demonstrates that punctuation marks can serve as a basis for determining intonational boundaries during text vocalization, as they could provide a foundation for automatic syntactic analysis (Monaghan, 1996). *Intensity*, as a component of intonation, is usually considered in conjunction with other intonational features. Along with fundamental frequency and duration, intensity contributes to the prominence of a particular word within an utterance. The highlighting of specific elements in the speech chain is achieved by increasing intensity. Variations in intensity are perceived auditorily as changes in loudness. However, it should be noted that loudness depends not only on intensity but also on the pitch of the fundamental tone. At the same level of intensity, a segment with a higher pitch is perceived as louder. There is a close relationship between stress and intensity. Syllables carrying logical stress generally exhibit greater intensity, although intensity is not the only component of stress. The prominence of a syllable is also ensured by other phonetic means, such as pitch and duration.

According to experimental phonetic studies, there is a general tendency for articulatory energy to decrease toward the end of an utterance, except for certain types of questions and imperative statements, where the opposite pattern is observed. In the differentiation of communicative types of utterances, the following tendency has been identified concerning the overall level of the intensity component: the lowest intensity is found in completed syntagms, parenthetical elements in interposition and postposition, while a higher intensity level is observed in imperatives, questions, and exclamations. In the intonations of completeness, imperatives, and exclamations, the peak intensity is predominantly located at the beginning of the utterance. In cases of incompleteness, certain types of questions, and parenthetical elements, depending on their position, the peak intensity may be localized at any point within the utterance.

Thus, different communicative types of utterances are characterized by a set of prosodic features in various combinations. It has been established that the components of intonation are closely linked to semantics; in other words, prosodic features can convey subtle shades of meaning within an utterance.

Acoustic parameters such as pitch range, pitch level, tempo, duration, and intensity amplitude can express varying degrees of semantic weight in syntagms or phrases. The greatest semantic load of an utterance is conveyed through the maximum values of intonational parameters, while the minimal semantic significance is expressed through their lowest values. As a result of an experimental phonetic analysis of Kazakh texts from different genres and styles, based on the stability of relevant acoustic features, eight intonemes—basic intonational invariants of the Kazakh language—have been identified, each with several variants that account for the full diversity of Kazakh intonation. These are the following:

1. The incompleteness is characterized by a rising melodic contour and a medium pitch range, beginning at the second pitch level. Variants of incompleteness exhibit both convex and concave melodic contours, typically ending at the third and, in rare cases, the fourth pitch level. In the intoneme of incompleteness, the end of the syntagm is pronounced with syntagmatic stress, which is manifested in the lengthening of the final syllable, where the melodic peak is localized.
2. The intoneme of completeness is characterized by a convex-descending melodic contour, reaching the first pitch level, and a medium pitch range. Variants of completeness may begin at the second or third pitch level. If the melodic contour starts at the third pitch level, it sometimes rises slightly before descending. In the intonation of completeness, the end of the syntagm is pronounced at a slow tempo, with the fundamental frequency and intensity reaching their minimum values.
3. The intoneme of a general question is characterized by a rising melodic contour with a medium pitch range, occurring between the second and fourth pitch levels. Its variants may exhibit convex-rising or concave-rising melodic contours. In terms of tempo and intensity, there is a tendency for a slowed articulation and a decrease in intensity toward the end of the utterance.
4. The intoneme of a special question is characterized by a convex-descending melodic contour and a significant pitch range. This pattern begins at the third pitch level, rises to the fourth, and then descends to the first level. Variants differ only in the starting level of the melodic contour. The interrogative word serves as the logical center, where the maximum values of fundamental frequency, intensity, and duration are localized.
5. The intoneme of categorical urging is characterized by a convex-descending melodic contour, starting at the fourth pitch level and ending at the first pitch level. The melodic intervals between the beginning and end of the phrase are considerable, and the pitch range of this model is also expanded. Variants differ in the initial pitch level of the melodic contour. The intoneme of categorical urging is pronounced with an enhanced logical stress, which is conveyed through syllable lengthening and an increase in intensity.
6. The intoneme of polite urging is characterized by a convex-rising melodic contour, beginning at the second pitch level, reaching the fourth, and slightly descending. The pitch range for this type of urging is relatively small. Variants may exhibit a fully rising contour from beginning to end or a rising-falling pattern. In the intoneme of polite request, the reinforcement of logical stress is achieved through the maximum values of fundamental frequency and duration.
7. The intoneme of exclamation is characterized by a rising-falling melodic contour. The pitch range of these models is medium. Due to the diversity of emotions conveyed by intonation, its variants can be highly varied. Their melodic contours may begin at either the second or third pitch level, reach the fourth, and end at the third, second, or first pitch levels. Emotionally charged words are pronounced at the highest pitch with maximum intensity and duration. Timbre characteristics play a significant role in expressing the intoneme of exclamation.
8. The intoneme of parenthetical insertion is characterized by a convex-descending melodic contour and a narrowed pitch range. Variants of this model may exhibit rising-falling or level contours, realized within the lower range of the voice. The transition of the melodic contour at syntagm boundaries is marked by negative pitch intervals. Parenthetical insertions are typically pronounced at an accelerated tempo with low intensity. The intonemes of the Kazakh language represent the smallest functionally significant units of intonation, forming components of sufficiently long texts and capable of being extracted from them. The semantic characteristics of intonemes can depend both on changes to a single differential feature and on the melodic modification of an entire syntagm.

The qualitative and quantitative characteristics of a phrase as a whole and its individual segments are expressed comprehensively: through sound and syllabic variations, intensity levels, the pitch dynamics of the main tone frequency, the duration of segments, and the pauses between them. In the speech stream, any semantically significant word in expressive utterances can be emphasized. Logical emphasis is closely connected with the meaning of the word it focuses on. A logically emphasized word in the speech stream typically exhibits clear articulation. When located at the beginning or middle of a phrase, it is characterized by maximum values of the main tone frequency, duration, and intensity. If the emphasized word is at the end of the phrase, its intonational characteristics are realized only through temporal and dynamic peaks. Regardless of its position, emphasized words are distinguished by significant intervals at their junctions with other words in the phrase. Prosodic means, by organizing and segmenting elements of speech (or text) and expressing the degree of connection between them, dividing them into meaningful segments. In speech segmentation, rhythmic, syntagmatic, phrasal, logical, and other types of stress play a significant role, influencing the meaning of utterances.

Conclusion

Speech synthesis is an interdisciplinary field that combines knowledge from acoustics, linguistics, and computer science to develop technology capable of converting text into audible speech. The development of speech synthesis began with attempts to reproduce the sounds of the human voice using mechanical and electrical devices and continued to evolve with the use of complex mathematical models and artificial intelligence. Modern methods of speech synthesis are designed to achieve maximum naturalness, expressiveness, and flexibility in application and involve the use of deep neural networks. These methods provide more natural and expressive sound, closely resembling human speech, thanks to large volumes of data. Based on experimental-phonetic research, it has been determined that modeling the phonetic-phonological database for speech synthesis requires a detailed description of segmental and suprasegmental units. Segmental-prosodic features play a key role in the development of speech synthesis systems. These features consider context, intonation, and the emotional tone of speech, significantly improving the quality of synthesis. Such models are trained on extensive datasets that include thousands of hours of recorded human speech and their prosodic transcriptions, enabling the capture of subtle language nuances. Text-to-speech synthesis involves an automatic procedure for generating current contours of melody, sound intensity, phoneme duration, and pause duration based on the analysis of specific properties of the input text and its prosodic markup. Prosodic markup of a text involves dividing it into meaningful units, marking them as phonetic words, and labeling the intonation type of semantic units according to defined rules. Speech synthesis is of great interest to speech science specialists due to its potential for precise control over speech parameters. Tonal, temporal, and dynamic characteristics of texts serve as a foundational base for creating speech synthesis.

The methodology of speech synthesis developed to date allows for the correlation of the speech signal with prosodic-level units. From this perspective, prosodic features become an actual set of rules for synthesis, offering a precise procedure for transitioning through the sequence of accentual structures, syntagmas, phrases, and texts. At the present stage of the development of science and education, the widespread use of information technologies in all fields of science is among the most urgent and pressing issues. Our state has begun implementing digital technologies in all areas of science and education. At the same time, to enhance the status of the Kazakh language as the state language of the Republic of Kazakhstan, strengthen its social functions, and expand its scope of application, it is necessary to integrate it into the global information and technological space. In the era of globalization, digital technologies are widely used in all fields of science and in linguistics. In the context of societal modernization, linguo-technology is advancing ahead of other fields. The global achievements of our time (automatic translators, machine analysis and synthesis of words, text editors, electronic linguistic resources, and a vast number of other technical advancements) indicate that linguo-technology, developing at a rapid pace, is a promising direction for Kazakh linguistics.

Acknowledgments

The article was prepared within the framework of program-specific financing BR 24993111 "Development of a speech synthesis system based on the orthoepical norm of the Kazakh language".

References

Antipova, A. M. (1986). The Trends in Examining Intonation in Contemporary Linguistics. *Voprosy yazykoznaniya, 1*, 122-132.

Balaqaev, M., Syzdyqova, R., & Janpeisov, E. (1968). *Qazaq adebi tinin history: oqu quraly [History of the Kazakh literary language: textbook]*. Almaty: Mektep.

Bellinghausen, C., Schröder, B., Rauh, R., Riedel, A., Dahmen, P., Birkholz, P., et al. (2024). Processing of prosodic cues of uncertainty in autistic and non-autistic adults: a study based on articulatory speech synthesis. *Frontiers in Psychiatry, 15*, 1347913. doi: https://doi.org/10.3389/fpsyt.2024.1347913

Bondarko, L., Verbitskaya, L., & Gordina, M. (2004). *Osnovy obshchei fonetiki [Fundamentals of general phonetics]*.

Chen, W., Yang, S., Li, G., & Wu, X. (2025). DrawSpeech: Expressive Speech Synthesis Using Prosodic Sketches as Control Conditions. *arXiv preprint arXiv:2501.04256, 1*, 1-5. doi: https://doi.org/10.48550/arXiv.2501.04256

Deterding, D. (2015). Segmentals. In J. M. L. Marnie Reed (Ed.), *The Handbook of English Pronunciation* (pp. 67-84). Wiley Online Library. doi: https://doi.org/10.1002/9781118346952.ch4

Dubois, J. (1993). Research on French intonation. *Studies in Applied Linguistics, 2*, 45-61.

Flanagan, J. L. (1965). Acoustical Properties of the Vocal System. In J. L. Flanagan (Ed.), *Speech Analysis Synthesis and Perception* (pp. 21-75). Springer Berlin Heidelberg. doi: https://doi.org/10.1007/978-3-662-01562-9_3

Goldrick, M., & Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics, 99*, 101254. doi: https://doi.org/10.1016/j.wocn.2023.101254

Hall, T. A. (2007). Segmental features. In d. L. Paul (Ed.), *The Cambridge Handbook of Phonology* (Vol. 1118, pp. 311-334). Cambridge university press. Retrieved from https://www.cambridge.org/core/books/abs/cambridge-handbook-of-phonology/segmentalfeatures/521591C65A262B26B06B37034AEE68A1

Huang, F. (1998). Syntagms in development and evolution. *International Journal of Developmental Biology, 42*(3), 487-494. Retrieved from https://ijdb.ehu.eus/article/9654036

Kaliyev, A. K., & Rybin, S. V. (2019). Acoustic modeling for Kazakh speech synthesis. *Journal Scientific and Technical Of Information Technologies, Mechanics and Optics, 19*(5), 951-954. doi: https://doi.org/10.17586/2226-1494-2019-19-5-951-954

Kanter, L. A. (1988). *Systemic analysis of speech intonation*. Moscow: Higher School.

Karnevskaya, E. (2005). Current Problems of Prosodic Modelling for Speech Synthesis. *Białostockie Archiwum Językowe, 5*, 21-25. doi: https://doi.org/10.15290/baj.2005.05.02

Krivnova, O. V. (1998). Automatic synthesis of Russian speech from arbitrary text. In *Proceedings of the International Seminar on Computational Linguistics and Its Applications* (pp. 175-180). Moscow: Moscow State University.

Kuznetsov, V. (2007). *Automatic speech synthesis: "Letter-to-sound" conversion algorithms and speech segment duration control*. Moscow: Valgus.

Lobanov, B. M., & Tsirulnik, L. I. (2008). *omputer speech synthesis and cloning*. Minsk: Belarusian Science.

Mashakova, A. (2021). Perception of Literary Heritage of Mukhtar Auezov in Germany. *Deutsche Internationale Zeitschrift für zeitgenössische Wissenschaft,* (16), 46-48. doi: https://doi.org/10.24412/2701-8369-2021-16-46-48

Monaghan, A. I. (1996). Rhythm and stress shift. In *Computer Speech and Language* (pp. 157-170). Boston: Kluwer.

Nikolaeva, T. M. (1978). Text linguistics: Current state and prospects. In *New in foreign linguistics* (pp. 213-234). Moscow: Higher School.

Overath, T., & Paik, J. H. (2021). From acoustic to linguistic analysis of temporal speech structure: Acousto-linguistic transformation during speech perception using speech quilts. *NeuroImage, 235*, 117887. doi: https://doi.org/10.1016/j.neuroimage.2021.117887

Potapova, R. K., & Potapov, V. V. (2012). *Speech communication: From sound to utterance*. Moscow: Languages of Slavic Cultures.

Rza, G. A., & Iraj, I. J. (2022). Syntagmatic and paradigmatic approach to parallelism in literary texts. *Foundations and Trends in Research,* (1). Retrieved from https://ojs.scipub.de/index.php/FTR/article/view/430

Svetozarova, N. D. (1993). *Pause. In Linguistic encyclopedic dictionary*. Moscow: Encyclopedia.

Taylor, P. (2009). *Text-to-Speech Synthesis*. Cambridge university press. doi: https://doi.org/10.1017/CBO9780511816338

Torsueva, I. G. (1979). *Intonation and meaning of an utterance*. Moscow: Nauka Publishing House.

van Santen, J. P. H. (1994). Assignment of segmental duration in text-to-speech synthesis. *Computer Speech & Language, 8*(2), 95-128. doi: https://doi.org/10.1006/csla.1994.1005

Wang, X. (2022). Segmental versus Suprasegmental: Which One is More Important to Teach? *RELC Journal, 53*(1), 194-202. doi: https://doi.org/10.1177/0033688220925926

Ziyadullayeva, M. (2023). Paradigmatic and Syntagmatic Relations of Linguistic Units in Uzbek Language. *Science and Innovation, 2*(C7), 35-38. Retrieved from https://cyberleninka.ru/article/n/paradigmatic-and-syntagmatic-relations-of-linguistic-units-in-uzbek-language