Content is available at: CRDEEP Journals
Journal homepage: http://www.crdeepjournal.org/category/journals/ijssah/

International Journal of Social Sciences Arts and Humanities

(ISSN: 2321-4147) (Scientific Journal Impact Factor: 6.002)
A Peer Reviewed UGC Approved Quarterly Journal

Research Paper

Analysis of Audio Text Unit Problems

Urazaliyeva Mavluda Yangiboyevna*
Independent Researcher at The National University Of Uzbekistan, Russia.

ARTICLE DETAILS

Corresponding Author:
Urazaliyeva Mavluda Yangiboyevna

Key words:
Audio units, text units, linguistic analysis, speech recognition, phonetic equivalence, natural language processing, uzbek language.

ABSTRACT

Audio and text units play a crucial role in information technology, education, communication, and culture. This article examines their linguistic and technological significance, exploring their phonetic, grammatical, syntactic, and semantic characteristics. The study highlights the integration of audio and text units in artificial intelligence-based applications, such as voice assistants and automatic transcription systems. Additionally, it discusses challenges related to contextual differences, computational power, and linguistic accuracy, particularly in the Uzbek language. Addressing these challenges through advanced linguistic and technological research will enhance multimodal systems, improve transcription accuracy, and contribute to language preservation and development.

Audio and text units have become a crucial topic in the fields of information technology, education, communication, and culture in the present era. The study of these units requires a multifaceted approach, as they encompass linguistic, technological, and psychological aspects. This article provides scientifically-based information about audio and text units and examines their linguistic and technological significance. Audio text units are based on the phonetic and acoustic characteristics of human speech. From a linguistic perspective, speech sounds are classified as phonemes and allophones. Phonemes are the fundamental units that form the meaning of speech, while allophones represent the acoustic variants of phonemes.[Ladefoged, P., Johnson, K. (2014).] Audio units in the speech process encompass not only sounds but also intonation, tempo, and pauses. These elements are crucial for enhancing the meaning of words and accurately expressing context [Boersma, P., & Weenink, D. (2021).]. For example, the into national change of the word “yes” determines whether it is used as a question, affirmation, or expression of doubt. From a technological perspective, audio units are considered a fundamental component for voice assistants, automatic speech recognition systems, and speech synthesizers. For instance, systems like Deep Speech and Kaldi are successfully used to convert speech sounds into digital data [Hannun, A., Case, C., Casper, J., et al. (2014).]. Today, artificial intelligence-based systems allow for real-time processing and analysis of audio information. Text units in linguistic analysis are examined at the levels of words, phrases, sentences and text [Crystal, D. (2008)]. Words are composed of morphemes, which are considered the smallest meaningful units of language. Texts are significant as units that express a semantically coherent content.

The technological application of text units is widespread, particularly in the field of natural language processing (NLP). Advanced models such as GPT and BERT demonstrate high efficiency in semantic and syntactic analysis of text units. These models also rely on large datasets for text analysis and translation [Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019).].The grammatical rules and stylistic patterns observable in text units are of particular importance for the Uzbek language and other languages. In-depth study of these units reveals language development processes and similarities between accents and dialects. Technologies for integrating audio and text units are opening up numerous new possibilities. For instance, the combined use of audio and text in creating multimedia content enhances user experience. Additionally, voice

*Author can be contacted at: *Independent Researcher At The National University Of Uzbekistan, Russia.*
Received: *10-02-2025*; Sent for Review on: *12-02- 2024*; Draft sent to Author for corrections: *18-02-2025*; Accepted on*: 28-02-2025*; Online Available from *07-03-2025*
DOI: 10.13140/RG.2.2.32831.62885
IJSSAH: *-9910/*

assistants and automatic translation systems serve as clear examples of the convergence of these fields [Good fellow, I., Bengio, Y., & Courville, A. (2016)]. The use of artificial intelligence algorithms in integrated systems facilitates the joint processing of text and audio data. For example, these systems play a crucial role in the creation of products such as subtitled videos and audio books [Boersma, P., & Weenink, D. (2021)]. Furthermore, search engines and voice interfaces enable users to meet their needs more quickly through these technologies. However, there are several challenges in this process. From a linguistic perspective, contextual differences between audio and text create complexity. For example, pauses and intonations in audio recordings can convey meaning, but these aspects are missing in the text. Therefore, special algorithms are required to ensure logical connections between text and audio. From a technological standpoint, computational power and the breadth of the dataset are crucial when integrating audio and text units. Particularly in the Uzbek language, there is insufficient research in this area, and more studies are needed. These issues can be addressed by creating text and audio corpora in Uzbek.

Psychological aspects are also important. Studying the differences in how users process audio and text will help improve their effectiveness. For instance, in education, it's possible to adapt to various learning styles through the combined use of audio and text. Linguistic and technological research of audio and textual units reveals their role and significance in modern life. These units open up new opportunities in information technology, education, and culture. The combined use of audio and text units contributes to increased efficiency through multimodal systems. However, to successfully integrate them, it is necessary to solve linguistic and technological challenges. Specifically, it is advisable to conduct more extensive scientific research on audio and text units in the Uzbek language. Research in this field further expands the possibilities of integrating language into technologies and plays a crucial role in preserving cultural heritage. When analyzing audio texts, it is crucial to study the alternatives of audio (conversation or speech) and text units, examining their compatibility, similarities, and differences in content. This process is carried out phonetically, grammatically, syntactically, and semantically. The correspondence between text and audio materials determines their linguistic accuracy and precision. The process of creating audio texts involves various stages: determining pronunciation, grammatical structures, lexical units, semantic meanings, and their compatibility with one another. In this section, an attempt is made to conduct an in-depth analysis of similarities and differences when creating audio texts in the Uzbek language. Phonetic equivalence is the correspondence between pronunciation and writing in text and audio text. Ensuring phonetic equivalence is crucial when analyzing the differences between pronunciation and spelling of words in the Uzbek language. When establishing such consistency, pronunciation errors and textual errors can lead to unnecessary ambiguities. This takes into account changes between phonetic units, such as alterations in the final consonant of a word or vowel sounds, dialectal differences in pronunciation, various representations of sound combinations, and other factors. In the Uzbek language, phonetic changes between words reflect the dynamic development of the language, dialectal variations, and individual speech styles. For example, the Uzbek word "yaxshi" (meaning "good") is written as [jɑxʃi], but in some regions, it may be pronounced as [jɑxʃe]. The following table illustrates a practical example of phonetic equivalence:

Area	Word	Recording Pronunciation	Option Explanation
Tashkent	yaxshi	[jɑxʃi]	Official pronunciation
Samarkand	yaxshi	[jɑxʃe]	Territorial difference
Khorezm	yaxshi	[jɑxʃʲi]	Dialectal transformation

In the study of phonetic alternation in the Uzbek language, changes in intonation and tempo between dialects are also taken into account. For example, analysis of the differences between the Tashkent and Khorezm dialects has revealed that intonation changes can also affect semantics [Ladefoged, P., & Johnson, K. (2014)]. This difference highlights the dialectal variations and phonetic features of the Uzbek language. It is important to consider how these pronunciation variants are reflected in audio texts and how the phonetic version of the word "yaxshi" (good) is represented in the text. This issue has been studied separately when analyzing the phonetic features of dialects in the Uzbek language. Phonetic characteristics in different regions of Uzbekistan, such as Tashkent and Samarkand speech, indicate variations in pronunciation [Juraev T. 2019.].For example:The word "odam" (person) is written as [odam], but in audio text, the pronunciation of this word might be heard as [o'dam]. This requires clarification in ensuring phonetic alternation, especially during the transcription process[Ilhomov, M. 2020]. This issue is addressed in the analysis of ensuring compatibility between pronunciation and the phonetic features of writing. Grammatical alternation refers to ensuring consistency between grammatical structures in audio and text. It is necessary to consider both simple and complex grammatical structures in the Uzbek language, particularly analyzing the correspondence of verbs to person, number, and tense. Let's examine the main approaches used to ensure grammatical equivalence in the correct transcription of audio texts. For example: In an audio text, the words "Men kitobo'qiyman" (I read a book) are pronounced as [mɛnkɪtɔbo'qɪmɑn], but in writing, these words are displayed as "Men kitobo'qiyman." The correct reflection of grammar is crucial for ensuring consistency between the audio text and written text. According to the analysis, grammatical errors are more common in transcription, especially under the influence of speech rate and certain speech styles. In some cases, official texts express words with very precise and complex grammatical structures, such as "Bundayxatti-harakatlarqonunlargaziddir" (Such actions are contrary to the laws). Audio texts, however, often use simpler, abbreviated forms: "Bundayharakatlarqonungaqarshi" (Such actions are against the law). Here, the grammatical

structure may have changed, but the meaning remains the same. The phenomenon of grammatical simplification is found in many speech genres in social communication, as people tend to prefer shorter and simpler expressions [Xolmirzaev, S. 2021]. Some differences should be taken into account when ensuring grammatical alternation, as grammatical structures may be unique in certain speech genres or in specific regions. Syntactic equivalence is ensuring the correct and logical correspondence of syntactic units between audio and text. When transcribing audio texts, syntactic errors or misrepresentation of structuresmay negatively impact the content of the text In the Uzbek language, there are sometimes several options in the syntax of sentences, and analysis is necessary to check the compatibility between audio and text. For example: Although the syntactic structure in the sentence “The person worked well” is pronounced in the audio text as [ödɑmjɑxʃɪɪʃlɑdɪ], in the text it is written that the person worked well. In this case, the syntactic compatibility between the text and the audio is fully preserved. As many studies have shown in ensuring syntactic alternation, syntactic errors between speech and writing negatively impact mutual understanding [Tursunov B. 2018]. In the Uzbek language, there are sometimes several options in the syntax of sentences, and analysis is necessary to check the compatibility between audio and text. For example: Although the syntactic structure in the sentence “The person worked well” is pronounced in the audio text as [o'dɑmjɑxʃɪɪʃlɑdɪ], in the text it is written as "Odamyaxshiishladi" (The person worked well). In this case, the syntactic compatibility between the text and the audio is fully preserved. As many studies have shown in ensuring syntactic equivalence, syntactic errors between speech and writing negatively impact mutual understanding. For example, the sentence "U yaxshio'qiydi" (He reads well) in the Uzbek language can have different pronunciation variants. For instance, instead of "U yaxshio'qiydi," there could be variations like "U yaxshio'qiyotdi" (He is reading well) or "U yaxshio'qishadi" (They read well). However, such pronunciation variants should not change the meaning. In ensuring semantic correspondence, even if the pronunciation or form of a word has changed in some cases, its meaning must remain the same. The words in the sentence "Boshqaishlarniqiling" (Do other things) may change in pronunciation, but the meaning stays consistent [Nurmatov O. 2020]. If it is expressed in the audio text as "Boshqaishlarniqilishniboshlang" (Start doing other things), this is also semantically correct. Studies that have shown how semantic changes are reflected in the transcription process have examined ensuring a high degree of semantic correspondence between speech and writing.S tudies that have shown how semantic changes are reflected in the process of transcription have studied ensuring a high degree of semantic correspondence between speech and writing. Studies conducted on ensuring phonetic, grammatical, syntactic, and semantic equivalence in transcribing audio texts show that sometimes phonetic changes between words may occur, but they should not affect the grammatical structure. Different dialects and speech genres in the Uzbek language create various alternatives for transcribing audio texts. Therefore, it is necessary to develop more advanced methods and models to eliminate phonetic and syntactic errors in the transcription process [Baxtiyorov F. 2017].Audio units include elements of human speech in the form of sounds. These include intonation, stress, rhythm, tempo, and other acoustic features. These units are crucial in determining how language is used in communication. For example, intonation helps determine the purpose of a sentence (question, command, or statement)[Crystal, D. (2008)]. In the process of studying audio units, prosody and phonetic aspects play a special role. Prosody represents the interplay of intonation, stress, and rhythm. These elements help determine not only the content but also the emotional aspects of communication. D. Crystal studied the functional significance of intonation and provided a detailed analysis of its linguistic and paralinguistic features.

On the other hand, text units are linguistic elements expressed in written form. They include components such as words, sentences, and paragraphs. Text units ensure that information is expressed and stored in a structured form. Text units are linguistically stable and are easier to process and analyze [Ladefoged, P. (2001).]. Syntactic and semantic principles play an important role in texts, which simplifies the transmission of information. Structural and contextual aspects are of great importance in the analysis of text units. While structural aspects express the grammatical arrangement of words and syntactic relationships, contextual aspects help determine the semantic content of the text. In this regard, the principles of generative grammar developed by Chomsky serve as the main theoretical basis for the analysis of text units. The primary differences between audio and text units are evident in their nature and methods of utilization. Audio units are expressed in real-time and reflect the dynamic characteristics of the communication process. Text units, on the other hand, are static, allowing for processing and storage. These differences are studied using various methods of linguistic analysis [Boersma, P. (1993)].

The following table compares the main features of audio and text units:

Properties	**Audio units**	**Tex units**
Form of expression	Intonation, rhythm, tempo	Word, sentence, paragraph
Time dependence	Real-time	Can be processed and saved
Dynamic characteristics	Interactive and variable	Stable and structured
Ability to analyze	Based on acoustic and prosodic aspects	Based on grammatical and semantic aspects

The above definitions and descriptions demonstrate how linguistic analysis aids in understanding the relationship between audio and textual units. The study of the connection between audio and text units has been conducted by numerous linguists. Notably, N. Chomsky developed general principles of syntactic units in his work “Syntactic Structures” [Chomsky, N. (2002)]. These principles play a crucial role in comprehending the grammatical correspondence between written and spoken language.In the field of phonetic analysis, P. Boersma conducted advanced work on determining sound frequencies and

studying the acoustic properties of speech [Boersma, P. (1993)]. He developed a short-term spectral analysis of speech and proposed a method for identifying the main characteristics of the sound signal. This method serves as the basis for audio analysis.

Conclusion

The study of audio and text units reveals their fundamental role in modern communication, information processing, and artificial intelligence applications. These units differ in their structure, dynamic properties, and methods of analysis, yet their integration offers significant advancements in speech technology and language processing. Ensuring phonetic, grammatical, syntactic, and semantic equivalence is essential for accurate transcription and interpretation, especially in languages with dialectal variations like Uzbek. While technological advancements have enabled efficient processing of audio and text, further research is required to refine transcription models, expand linguistic datasets, and improve multilingual AI systems. By addressing these challenges, the integration of audio and text units will continue to enhance human-computer interaction, education, and cultural preservation.

References:

1. Abduraxmonova, N. Z. (2018). Linguistic support of the program for translating English texts into Uzbek (on the example of simple sentences): Doctor of Philosophy (PhD) il dis. aftoref.
2. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. *IzvestiyaKyrgyzskogogosudarstvennogotekhnicheskogouniversiteta*, *2*(38), 12-7.
3. Suleymanov, D., Nevzorova, O., Gatiatullin, A., Gilmullin, R., &Khakimov, B. (2013). National corpus of the Tatar language "Tugan Tel": grammatical annotation and implementation. *Procedia-SocialandBehavioralSciences*, *95*, 68-74.
4. Baxtiyorov F. Matnva audio transkriptsiyasi: semantikjihatlar. Toshkent. Yangiakademiya. – 2017. – 126b.
5. Boersma, P. (1993). Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound. Proceedings of the Institute of Phonetic Sciences
6. Boersma, P., &Weenink, D. (2021). Praat: Doing phonetics by computer. [Praat software documentation]
7. Chomsky, N. (2002). Syntactic Structures. The Hague: Mouton.
8. Crystal, D. (2008). A Dictionary of Linguistics and Phonetics. Blackwell Publishing.
9. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding.
10. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
11. Hannun, A., Case, C., Casper, J., et al. (2014). DeepSpeech: Scaling up end-to-end speech recognition.
12. Ilhomov, M. O‘zbekfonetikasivatranskriptsiya: fonetiko‘zgarishlar. Toshkent. O‘qituvchi. – 2020. – 98b.
13. Juraev T. O‘zbektilidagidialektalfarqlar: fonetikvagrammatiktahlil. Toshkent. Fan. – 2019. – 110b.
14. Ladefoged, P. (2001). A Course in Phonetics. Boston: Heinle.
15. Ladefoged, P., Johnson, K. (2014). A Course in Phonetics. Cengage Learning
16. Nurmatov O. Semantiko‘zgarishlarvaularningtranskriptsiyadagio‘rni. Toshkent. Yanginashr. – 2020. – 125b.
17. Tursunov B. Sintaksisvasemantika: tahlilvatadqiqotlar. Samarkand. Sharq. – 2018. – 145b.
18. Xolmirzaev, S. Grammatik muqobillikvanutq: soddalashtirishuslubi. Buxoro. O‘zbekiston. – 2021 – 126b.

1.