Phonological and Acoustic Analysis of Consonant Phonemes in the Uzbek Language

Shahlo Khamroeva
Computational linguistic and digital technology
Tashkent State University of Uzbek Language and Literature
Tashkent, Uzbekistan
shaxlo.xamrayeva@navoiy-uni.uz

Gulshakhnoz Makhmudjonova
Computational linguistic and digital technology
Tashkent State University of Uzbek Language and Literature
Tashkent, Uzbekistan
gulshaxnozmahmudjonova@gmail.com

Nigmatova Lolakhon Khamidovna
Russian Language and Literature
Bukhara State University
Bukhara, Uzbekistan
nigmatovalolaxon@gmail.com

Khusanboeva Mukhlisa
Department of Uzbek Literature and Languages
Andijan State Medical Institute
Andijan, Uzbekistan
husanboevamuhlisa500@gmail.com

2025 10th International Conference on Computer Science and Engineering (UBMK) | 979-8-3315-9975-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/UBMK67458.2025.11206880

***Abstract*—This study presents a comprehensive phonological and acoustic analysis of Uzbek consonant phonemes, with the goal of proposing a structured consonant inventory suitable for grapheme-to-phoneme (G2P) conversion and text-to-speech (TTS) systems. Consonants are examined based on articulatory features (place and manner of articulation), voicing, and positional occurrence (anlaut, inlaut, auslaut). The analysis incorporates duration measurements of phoneme realizations across different phonotactic contexts, revealing patterns of allophonic variation and segmental prominence. Historical phonological observations and native linguistic descriptions are integrated with acoustic data to ensure both theoretical and practical reliability. The resulting consonant inventory offers a standardized and linguistically grounded foundation for computational modeling of Uzbek phonology, particularly in low-resource language processing tasks such as automatic phoneme annotation, G2P alignment, and high-quality TTS synthesis.**

Keywords—Uzbek consonants, phonological analysis, acoustic phonetics, grapheme-to-phoneme (G2P), text-to-speech (TTS), phoneme inventory, allophonic variation, speech synthesis.

I. Introduction

Language is the most essential means of human communication, and its complex structure is deeply studied through the disciplines of phonetics and phonology. The Uzbek language stands out for its rich phonetic and phonological system. This article aims to provide a comprehensive analysis of the phonological and acoustic features of Uzbek consonant phonemes, with a particular focus on their specific behaviors in the speech production process.

Phonological analysis investigates the distinctive function of sounds within the language system, specifically their role in differentiating meaning. Acoustic analysis, on the other hand, explores the physical properties of sounds — such as frequency, intensity, and duration — through scientific and experimental methods. These two approaches complement each other, providing a more complete and accurate understanding of speech sounds.

In addition to the phonological description of consonant phonemes in Uzbek, this study seeks to examine their acoustic properties based on parameters such as duration, resonator tone, noise component, and spectral features. During the research process, the articulatory characteristics of each consonant phoneme are analyzed, along with their realization in lexical and morphophonological contexts, and their acoustic behavior across phonetic positions (initial, medial, and final).

This analysis not only contributes to the theoretical advancement of phonology but also holds practical value for the development of speech technologies such as Automatic Speech Recognition (ASR) and Text-to-Speech synthesis (TTS) in the Uzbek language.

II. Literature Review

In phonetic studies of the Uzbek language, researchers have reported varying counts of consonant phonemes. A.F. Sjoberg [1] identified 26 consonants, stating that with their allophones, the total number of consonantal sounds reaches 33. Glynis MacMillan [2] reports 28 consonants and 37 allophones. Similarly, Mirtojiyev [3] describes 31 consonants as independent phonemes in Uzbek. In contrast, Ido [4] proposes a more compact classification, noting the presence of 24 consonants and 32 allophones.

As in other Turkic languages, Uzbek also features two phonetic realizations of voiceless plosive phonemes /p/, /t/, and /k/: aspirated ([ph], [th], [kh]) and unaspirated ([p], [t], [k]) allophones. According to Abduazizov [5], aspirated allophones generally occur in positions preceding vowels, such as in the words phul (money), tho‘k (electricity), and kho‘l (lake).

Voiceless plosive consonants are typically represented by two main allophonic variants [2]. This distinction is explained through two phonetic processes: palatalization, which involves the raising of the tongue's mid-section toward the hard palate, giving the sound a soft articulation; and aspiration, which refers to a noticeable burst of air accompanying the release of the consonant.

A similar system is observed in Turkish, which has 32 graphemes, although its alphabet consists of only 29 letters. The additional graphemes /â/, /î/, and /û/ are formed using diacritics. Koşaner et al. [6] report 32 phonemes in Turkish, based on a unified classification of prior linguistic sources. Importantly, Turkish does not exhibit a strict one-to-one graph-

eme-to-phoneme correspondence, as some graphemes can represent multiple phonemes. Their study presents a structured grapheme-to-phoneme mapping that can serve as a useful model for Turkic languages, especially in rule-based or hybrid G2P systems.

The Kazakh language features 20 core consonant phonemes, many of which undergo allophonic variation depending on the backness and rounding of adjacent vowels. Voiceless plosives such as /p/, /t/, /k/, and /q/ are typically aspirated in word-initial position, while voiced plosives like /b/, /d̪/, and /g/ are prevoiced. Velar and uvular obstruents alternate based on vowel harmony, with velars occurring before front vowels and uvulars before back vowels. Additional processes such as intervocalic spirantization, final devoicing, suffix desonorization, and nasal harmony are also attested. These features reflect a complex morphophonemic system relevant for accurate G2P modeling and pronunciation synthesis [7].

III. Results And Analysis

Consonant phonemes in the Uzbek language are classified based on three primary articulatory features: (1) place of articulation, (2) manner of articulation, and (3) presence or absence of vocal fold vibration (voicing). In addition, whether a consonant is produced through the oral or nasal cavity also plays an important role in its phonetic characterization [8].

1.1 Place of articulation

Bilabial - use of both lips: /p, b, m/.

Labio-dental - upper teeth and lower lip: /f/ and /v/.

Alveolar - tip of tongue touches ridge behind teeth: /t/, d, s, z, ch, j, sh, j, r, n/.

Palatal - tongue touches the middle of the palate: /y/.

Lateral – produced by airflow around the sides of the tongue: /l/.

Velar - back part of tongue touches soft palate: /k/ and /g/, /ng/.

Uvular – produced by the back of the tongue against or near the uvula (deep articulation): /q, x, g'/

Glottal - opening in vocal tract when vocal folds are apart: /h/

1.2 Manner of articulation

Plosives (Stops) – The airstream is completely blocked at a certain point in the vocal tract and then released, creating a burst of sound: /p, b, t, d, k, g, q/. Though typically oral, nasal plosives also exist in Uzbek: /m, n, ng/ (*nasal stops*).

Fricatives – The airstream passes through a narrow constriction, creating turbulent airflow or a hissing noise: /f, v, s, z, sh, j, x, g', h, y/.

Affricates – These are complex sounds that begin as plosives and release into fricatives, functioning as a single phoneme: /ch, j/.

Nasals – The velum is lowered, allowing air to flow through the nasal cavity instead of the oral tract: /m/, n, and ng/

Lateral Approximant – The airstream flows along the sides of the tongue while the central part is blocked: /l/

Trill – The tip of the tongue vibrates rapidly against the alveolar ridge, producing a series of quick taps: /r/.

1.3 Voicing

Based on the activity of the vocal folds during articulation and the presence of noise and resonance components, consonants are classified as:

- **Voiceless consonants:** /p, f, t, s, sh, ch, k, q, x, h/. These contain only turbulent noise and lack resonance from the vocal folds.
- **Voiced consonants:** /b, v, d, z, g, g', j, r, l, m, n, ng, y/. These involve vocal fold vibration and exhibit resonance characteristics.
- **Sonorants:** /m, n, ng, l, r, y/. These consonants are characterized by continuous resonance and are phonetically close to vowels in their acoustic behavior within the speech system.

A. Phonological Classification of Uzbek Consonants

In contemporary Standard Uzbek, the consonants /f/, /v/, /p/, /b/, and /m/ are recognized as labial sounds. According to Mirtojiyev, however, the set of labial consonants includes four primary phonemes: /b/, /p/, /w/, and /f/. Among these, /f/ and /v/ are labiodenta**l** consonants, produced with contact between the upper teeth and the lower lip. For instance: *fabrika (factory)*, *familiya (lastname)* (/f/); *vagon (wagon)* (/v/).

Labial fricatives in Uzbek can be further categorized based on their articulatory characteristics: bilabial fricatives: /ɸ/, /w/; labiodental fricatives: /f/, /v/. The phonemes /ɸ/ and /w/ are considered bilabial fricatives, respectively [7]. These are articulated with both lips and can be observed in words such as: /ɸ/: *ɸarq (difference)*, *ɸarmon (decree)*, *daɸtar (notebook),* /w/: *waqt (time)*, *hawo (air)*, *xowli (yard).*

The voiceless labiodental fricative /f/ [f, IPA] typically occurs in word-final (*auslaut*) position and is frequently found in interjections such as *kuf (ugh)*, *suf (suf)*, *tuf (tuf)*, and *uf* (oof) In contrast, the voiceless bilabial fricative /ɸ/ [ɸ, IPA] appears in all positions and is attested in Old Turkic manuscripts as well as in loanwords of Arabic and Persian origin: *fazl (virtue)*, *Safiya (girl's name)*) [9]. Its acoustic duration is up to 125 milliseconds in initial position, 50-70 milliseconds in medial position, and 160-170 milliseconds in final position, potentially extending to 250 milliseconds when followed by a pause.

The voiced labiodental fricative /v/ [v, IPA] is characteristic of loanwords and occurs in all positions. Its duration ranges from 80-90 milliseconds in initial position to 30-60 milliseconds in intervocalic environments. It has become increasingly common in the modern literary language, especially in international terminology: *vilka (fork)*, *revolyutsiya (revolution)*, *tramvay* (tram) [3].

The voiced bilabial fricative /w/ [β, IPA] often transcribed as /w/. It appears in all word positions, with durations ranging from 70-80 milliseconds in initial position, 30-40 milliseconds intervocalically, and up to 150 milliseconds in final position with pause. It is commonly found in expressive forms such as *woy (wow)*, *wizzilamoq (to whistle)*, *mewa (fruit)*, *palow (national food)*, *kuyow(fiance*).

*The voiced bilabial nasal /m/ [m, IPA***]** is present in all positions and co-occurs with both vowels and consonants. It has an acoustic duration of 80-110 milliseconds in initial position, 90-150 milliseconds in final position, and 45-60 milliseconds intervocalically. A hard variant of /m/ is observed in Russian loanwords, characterized by a slightly more retracted tongue position during articulation, as seen in words like *dima (Dima)* and *roman (novel)* [3].

The voiced bilabial plosive /b/ [b, IPA] occurs mainly in initial position with a duration of 150-210 milliseconds. It is

rare in medial position and generally does not appear in word-final position or in clusters with voiceless consonants.

The voiceless alveolar plosive /t/ [t, IPA] and the *voiced counterpart /d/ [d, IPA]* occur in native words and are historically represented with distinct graphemes indicating hard and soft realizations in Orkhon-Enisei inscriptions [9]. The /t/ sound often clusters with sonorants and voiced segments: *lt*, *nt*, *mt*, *tl* and modern phonetic analyses identify palatalized [tʲ] and aspirated [tʰ] variants [2]. The acoustic duration of /d/ ranges from 50-70 milliseconds in anlaut and 40-50 milliseconds in inlaut, while it is typically absent in auslaut.

The voiceless alveolar fricative /s/ [s, IPA] occurs in all syllabic positions and has a variable duration: 150-160 milliseconds in word-initial (anlaut), 50-70 milliseconds intervocalically, and 130-150 milliseconds in word-final (auslaut) position. *The voiced counterpart /z/ [z, IPA]* also appears in all positions, with durations of 150-170 milliseconds in anlaut/inlaut, and up to 230-250 milliseconds in auslaut when followed by a pause.

The voiceless postalveolar fricative /sh/ [ʃ, IPA] and voiced postalveolar fricative /dj/ [ʒ, IPA] are common in native words and historical loanwords, respectively. The /ʃ/ phoneme is pronounced with durations of 120-130 ms in anlaut, 80-130 ms in inlaut, and up to 250 ms in auslaut with pause. The /ʒ/ sound occurs in early Turkic loanwords such as *jafo (affliction)*, *ranj (bitternes)*, and *javob (answer)*, with durations of 120-130 ms (anlaut) and 70-90 ms (inlaut).

The voiceless postalveolar affricate /ch/ [ʧ, IPA] and voiced affricate /j/ [ʤ, IPA] occur in all positions of native words. The /tʃ/ segment shows context-dependent acoustic variation: 70-80 ms in anlaut, 50-90 ms in inlaut/intervocalic position, 30-40 ms in auslaut following consonants, and 120-150 ms in other final environments. The /dʒ/ phoneme has a duration of 70-80 ms in anlaut and 60-70 ms in inlaut.

The voiced alveolar nasal /n/ [n, IPA] is widely distributed and context-sensitive in duration. It is realized with 50-80 milliseconds in syllable-initial/final and auslaut positions, 120-190 milliseconds when followed by a pause, and up to 250 milliseconds at the end of an utterance.

The voiced alveolar tap /r/ [ɾ, IPA] appears in all positions. Its duration ranges from 25-30 milliseconds intervocalically to 40-60 milliseconds in syllable-final positions, with an extended duration of 100-150 milliseconds in imperative forms when placed in a stressed auslaut position.

The voiced palatal approximant /y/ [j, IPA] occurs in all positions of native Uzbek words and co-occurs with both hard and soft vowels and consonants: *yana (else)*, *boy* (rich). Its duration ranges from 60-90 ms in anlaut, 15-20 ms in intervocalic position, 40-50 ms in other medial contexts, and 120-150 ms in auslaut.

The voiceless velar plosive /k/, [k, IPA] occurs in all positions of native Uzbek words, particularly in those with soft articulation. It co-occurs with both voiced and voiceless consonants and sonorants, as in *turk (Turk)*, *kun (day)*, *ichki (inner)*, and *ko'l (lake)* [9]. In Uzbek, /k/ has at least two major allophonic variants: the palatalized [kʲ] and the aspirated [kʰ], also transcribed as *[kx, IPA]* [2]. As shown in Fig. 1, the spectrogram of the voiceless velar plosive [k] exhibits a silent closure followed by a brief high-frequency burst, whereas Fig. 2 displays the spectrogram of the voiceless velar affricate [kx], which combines the same closure-burst sequence with a subsequent, sustained velar frication band.

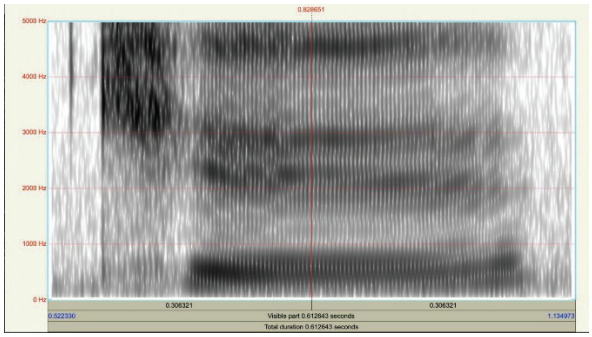

Fig. 1. Narrow-band spectrogram of an *isolated* voiceless velar plosive /k/.

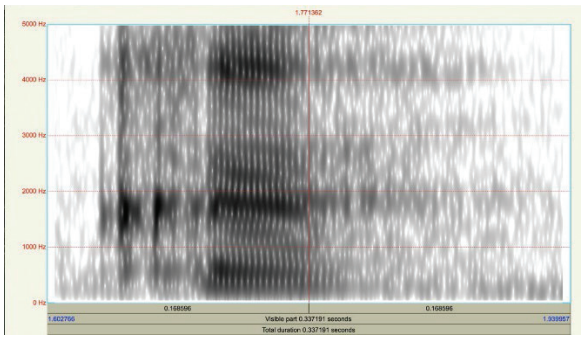

Fig. 2. Narrow-band spectrogram of an isolated voiceless velar affricate /kx/.

The voiced velar plosive /g/, [g, IPA] founds in medial and final positions of words with soft, non-palatalized articulation, /g/ co-occurs with sonorants, voiced and voiceless consonants, and front vowels. It has at least two allophonic variants in Uzbek: [gʲ] (palatalized) and [gɣ].

The voiced velar nasal /ng/, [ŋ, IPA] occurs primarily in medial and final positions of native Turkic words, but not in initial (*anlaut*) position. Its duration is 50-110 milliseconds intervocalically, and 80-100 milliseconds in syllable-initial and final (*inlaut* and *auslaut*) positions: *ko'ngil (soul)*, *singil (sister)*, *yengil(light*).

The voiceless uvular plosive /q/, [q, IPA] has a wide positional distribution in Uzbek. In word-initial (*anlaut*) position, its acoustic duration is typically 50-70 milliseconds, and may reach up to 100 milliseconds when occurring after a prosodic pause: *qachon (when)*, *qancha (how much)*, *qizil (red*). In medial (*inlaut*) and final (*auslaut*) positions, it ranges between 50-60 milliseconds: *toq (odd)*, *tiqin (plug)*, *chaqmoq(lighting*). It has no voiced counterpart in the phonemic inventory.

The voiced uvular fricative /g'/ [ʁ, IPA] occurs in the medial and final positions of Turkic-origin words with a "hard" phonetic character. It typically co-occurs with sonorants, both voiced and voiceless consonants, and back vowels: *bog' (garden)*, *yig'lamoq (to cry*).

The voiceless uvular fricative /x/ [χ, IPA] occurs in word-initial position with a duration of approximately 110 milliseconds: *xalq (nation)*, *xolos (only*), and in medial and intervocalic positions with a shorter duration of 60-80 milliseconds, as in *shaxs (individual)*, *axir (after all)*, *axborat (information)*, and *axloq (morality)*.

The voiced alveolar lateral approximant /l/, [l, IPA] does not occur in word-initial position in native Uzbek words but is

common in medial and final positions, as in *olmoq (to take)* and *el* (people). In intervocalic position, its duration ranges from 30 to 40 milliseconds; in non-final syllables of polysyllabic words, 35-45 milliseconds; and rarely up to 50 milliseconds. When occurring at the beginning of a stressed syllable or in the anlaut of plural forms, it may last 50-70 milliseconds, and in word-final (*auslaut*) position, it can extend up to 150 milliseconds. In borrowings from Russian, /l/ exhibits a secondary variant [ɬ], often referred to as the "dark l" or hard lateral. This variant is articulated further back in the oral cavity and contrasts with the soft [l]. While both have nearly identical articulation points, the dark [ɬ, IPA] involves a lower and more retracted tongue body, particularly when adjacent to back vowels such as /a/, /o‘/, /u/, and /i/ [3]. Examples include *stol (table)*, *stil (style)*, *kul* (ash), *xulq* (behavior), *film (movie)*. As illustrated in Fig 3 (*lizing*, initial /ɬ/) and Fig 4 (*lola*, initial /l/), the two Uzbek laterals exhibit sharply different acoustic signatures.

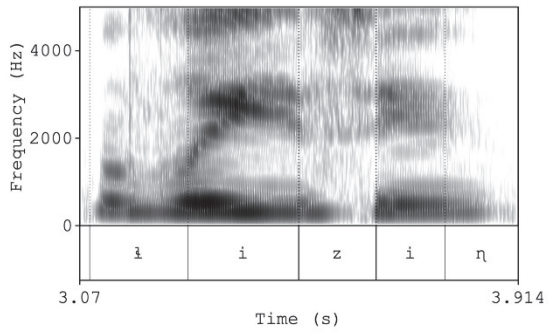

Fig. 3. Narrow-band spectrogram of the Uzbek borrowing *lizing (leasing)*

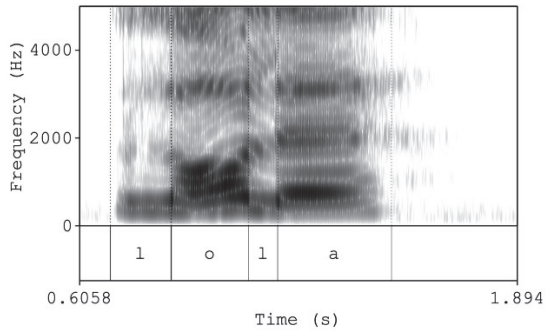

Fig. 4. Narrow-band spectrogram of the Uzbek borrowing *lola (tulip)*

The initial lateral in lizing shows an F2 that is ≈ 460 Hz lower than the one in lola (1285 Hz vs 1748 Hz). Because a velarised /ɬ/ lengthens the back cavity, its second formant drops; a non-velarised (clear) /l/ keeps the tongue body forward, so F2 rises.

Consistent with cross-linguistic findings [10], the mean F2 of the dark lateral in lizing (1285 Hz) is markedly lower than that of the clear lateral in lola (1748 Hz), confirming velarisation as the primary acoustic cue distinguishing the two Uzbek allophones.

The voiceless glottal fricative /h/, [h, IPA] occurs in both word-initial and medial positions with an acoustic duration of 70-90 milliseconds, as in *hali (yet)*, *hovli (yard)*, *hukm (judgment)*, *mehr (kindness)*, *bahor (spring)*, and *shahar (city)*. In word-final (*auslaut*) position, when not subject to reduction, its duration can reach up to 120 milliseconds.

Each Uzbek consonant is listed in its standard orthographic form (e.g., sh, g‘, ch, ng) rather than IPA symbols. The classification uses standard phonetic abbreviations: voiceless (vsl), voiced (vcd), and place of articulation labels such as bilabial (blb), alveolar (alv), velar (vel), etc. These descriptors provide a compact and linguistically accurate representation of the Uzbek consonant system as represented in the standard Uzbek orthography.

TABLE I. MANNER AND PLACE OF ARTICULATION FOR UZBEK CONSONANTS

		blb	lbd	alv	pla	vel	uvl	glt
nasal		m		n		ng		
plosive	vsl	b		d		k	q	
	vcd	p		t				
fricative	vsl	ɸ	f	s, sh			x	h
	vcd	w	v	z, j		g	g’	
affricate	vsl			ch				
	vcd			dj				
approximant				l	y			
tap/flap				r				

As shown in Table I, a phonological classification of Uzbek consonants is presented based on the native Uzbek alphabet.

Each phoneme is classified by its manner and place of articulation, voicing properties, and is accompanied by practical examples from the Uzbek language. The inclusion of detailed articulation descriptors (e.g., voiced bilabial plosive, voiceless uvular fricative, etc.) makes this chart a valuable reference for phonological and phonetic analysis, G2P modeling, and TTS system development.

TABLE II. PROPOSED UZBEK CONSONANTS PHONEME INVENTORY FOR TTS AND G2P SYSTEMS

IPA	Description	Example
b	voiced bilabial plosive	bola, obod, bormoq
p	voiceless bilabial plosive	paxta, po’lat, opa
d	voiced alveolar plosive	daraxt, dono, odil
t	voiceless alveolar plosives	taxta, ota, botir
g	voiced velar plosive	gul, gap, agar
k	voiceless velar plosive	kasb, akmal
kx	voiceless velar affricate	kompaniya, vokal
q	voiceless uvular plosive	qozon, qiz, aql
ʁ	voiced uvular fricative	g’uncha, g’aflat
χ	voiceless uvular fricative	xayol, xat, axborat
h	voiceless glottal fricative	hafta, hamma, daho
ʤ	voiced postalalveolar affricate	jiyda, jamoa, ajdod
ʒ	voiced postalveolar fricative	jurafa,
ʧ	voiceless postalveolar affricate	choy, ochmoq, qilich
ʃ	voiceless postalveolar fricative	shaxmat, shahar, oshiq
v	voiced labiodental fricative	vagon, tramvay, revizor
β	voiced bilabial fricative	ovoz, suv, voydod
f	voiceless labiodental fricative	fabrika familiya
ɸ	voiceless bilabial fricative	tuɸ, lutɸ, qulɸ
z	voiced alveolar fricative	zarar, yozmoq
s	voiceless alveolar fricative	osmon, salom, soyabon
m	voiced bilabial nasal	maqsad, amma, men
n	voiced alveolar nasal	non, tanimoq, tana
ŋ	voiced velar nasal	siŋil, toŋ, meniŋ
ɾ	voiced alveolar tap	ɾang, ɾahmat, qoɾa
l	voiced alveolar, lateral approximent	lola, elak

ɬ	voiced retroflex lateral approx-imant	ɬabaratoriya, stoɬ, stiɬ

Table II provides a comprehensive inventory of Uzbek consonant phonemes, transcribed in the International Phonetic Alphabet (IPA).

IV. Implications for Neural Text-to-Speech (TTS) Modeling

The detailed classification and acoustic profiling of Uzbek consonant phonemes presented in this study offer substantial benefits for the development of neural text-to-speech (TTS) systems. State-of-the-art architectures such as FastSpeech2 and VITS rely on accurate phoneme alignment, prosodic consistency, and temporal modeling to achieve natural-sounding synthesis. The empirical findings particularly regarding stop consonants, aspiration contrasts, and segmental duration measurements can be leveraged in multiple ways:

- Enhancing grapheme-to-phoneme (G2P) modules by refining Uzbek-specific phoneme mapping rules;
- Constructing more precise phoneme embeddings that capture articulatory distinctions (e.g., voicing, place, and manner of articulation);
- Distinguishing phonetic contrasts such as /l/ vs. /ɬ/ and /k/ vs. /kx/, which are relevant in both native and loanword pronunciation;
- Accurately modeling labiodental consonants /v/ and /f/, and contrasting them with labialized variants such as /w/, [β] (voiced bilabial fricative), and /ɸ/ (voiceless bilabial fricative), where appropriate;
- Fine-tuning duration prediction models using the mean segmental durations derived from native speaker recordings.

Importantly, the phonetic distinction between aspirated and unaspirated plosives in suffix position such as in /*k*/ , /*kx*/ is likely to contribute to improved naturalness and intelligibility in synthesized Uzbek speech.

Currently, one of the few existing Uzbek TTS systems is Muxlisa AI [12], which generates synthetic speech using neural network architectures. While it demonstrates notable progress in producing intelligible Uzbek output, the system still lacks detailed phonetic modeling particularly in capturing contrastive features such as aspiration and consonant velarization. For example, it tends to neutralize the distinction between unaspirated and aspirated voiceless stops, rendering both “kitob” (/k/) and “kompyuter” (/kx/) with a uniform unaspirated /k/, thereby diminishing the perceptual distinctiveness of foreign-origin words. Similarly, the system fails to reflect the lateral contrast between the clear /l/ in “lola”, a native Uzbek word, and the velarized /ɬ/ in “laboratoriya”, a borrowed term. By pronouncing both with a generic /l/, Muxlisa AI overlooks key phonological distinctions necessary for accurate and natural-sounding synthesis. These limitations underscore the importance of integrating fine-grained phoneme-level contrasts in future Uzbek TTS models.

V. Conclusion

As a result, the study carried out a detailed phonological and acoustic analysis of consonant phonemes in the Uzbek language, focusing on their articulatory classification in various phonotactic contexts, sound properties and tense characteristics. The place and style of articulation, sound differences and positional variability (anlaut, inlaut, auslaut) are systematically studied, the study identifies both standard phonemes and their allophonic variants, thereby capturing the complexity and richness of the Uzbek consonant system.

In addition to the theoretical classification, acoustic data were empirically collected and analyzed using Praat [13], a widely adopted tool in phonetic research. The analysis was based on audio recordings of four native Uzbek speakers: two males aged 20-22 and two females aged 21-27.

VI. Future Work

The current study is based on a limited set of audio recordings obtained from native speakers representing the Tashkent and Andijan regions, analyzed using Praat. In future work, a more extensive and balanced dataset will be collected, involving recordings from 20 to 30 participants, which will allow for statistically robust acoustic analysis.

To ensure greater variability and precision in phonetic modeling, separate recordings will be collected and analyzed by gender, distinguishing between male and female speakers. Additionally, the study will be expanded to include regional dialects of Uzbek, enabling a comparative acoustic analysis of consonantal features across dialectal varieties. This will contribute to the creation of a more representative phonetic corpus for Uzbek, which is essential for dialect-aware G2P modeling, speech synthesis, and recognition applications.

References

[1] A. F. Sjoberg, Uzbek Structural Grammar. New York, USA: American Council of Learned Societies, 1962.

[2] D. MacMillan, Descriptive Phonetics of Uzbek. https://www.yumpu.com/en/document/view/33134773/descriptive-phonetics-of-uzbek

[3] M. M. Mirtojiyev, “O‘zbek tili fonetikasi”. Tashkent: Fan, 2013.

[4] S. Ido, “Uzbek,” Journal of the International Phonetic Association, pp. 1–17, 2025. doi:10.1017/S0025100324000148

[5] A. Abduazizov, O‘zbek Tili Fonologiyasi va Morfonologiyasi, 2nd ed. Tashkent, Uzbekistan: Universitet, 2010.

[6] Ö. Koşaner, Ç. C. Bırant, and Ö. Aktaş, “Improving Turkish language training materials: Grapheme-to-phoneme conversion for adding phonemic transcription into dictionary entries and course books,” Procedia - Social and Behavioral Sciences, vol. 103, pp. 473–484, 2013, doi: 10.1016/j.sbspro.2013.10.359.

[7] G. A. Fielder and B. Khassenov, “Kazakh,” Journal of the International Phonetic Association, vol. 51, no. 1, pp. 111–132, 2021, doi: 10.1017/S0025100319000269.

[8] S. Otamirzayeva and M. Yusupova, “O‘zbek tili (Fonetika, grafika, orfoepiya va orfografiya)”. Tashkent: O‘zbekiston Milliy Ensiklopediyasi, 2004. – Б. 17.

[9] F.Kamal, Hozirgi zamon o‘zbek tili: Leksikologiya, fonetika, grafika, orfografiya, morfologiya. Tashkent: Fanlar Akademiyasi nashriyoti, 1957.

[10] Q. Mahmudov, “O‘zbek tilining tarixiy fonetikasi”. Tashkent: Ixtiyor, 2006.

[11] D. Recasens and A. Espinosa, “Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /ɬ/: evidence from two Catalan dialects,” Journal of the International Phonetic Association, vol. 35, no. 1, pp. 1–25, Jun. 2005, doi: 10.1017/S0025100305001878.

[12] https://muxlisa.uz

[13] P. Boersma and D. Weenink, Praat: Doing phonetics by computer [Computer program]. Version 6.4.10, 2025. https://www.fon.hum.uva.nl/praat/