Website:https://worldsiencepub.com/index.php/Imc Volume:5,Issue 02,September-2026

COMPARING UZBEK GRAPHEME-TO-PHONEME CONVERTERS: EPITRAN, ESPEAK NG, AND UZG2P ON SHARED EVALUATION CRITERIA

Utkirbek Azimov
Independent researcher
ORCID: 0000-0002-8088-324X
DOI:https://doi.org/10.5281/zenodo.23027477

Abstract

We compare stock Epitran uzb-Latn, eSpeak NG's uz voice, and uzg2p, a patched Epitran wrapper, on robustness, phoneme inventory, digraphs, apostrophe canonicalization, and stress. All three complete a 356-item vocabulary without crashes or empty outputs, but their inventories differ substantially. uzg2p offers normalization, configurable o‘→/ɵ/, and stress control; eSpeak supplies a synthesis baseline. Inventory harmonization is essential for phoneme error rate (PER) comparisons. Without independently annotated pronunciations, these findings establish implementation differences, not universal phonological superiority.

1. Introduction

Uzbek phoneme-based speech pipelines need consistent text-to-phoneme conversion. Epitran supplies a built-in map; eSpeak NG provides a native Uzbek voice; uzg2p adds normalization, foreign-token flagging, stress, and map repairs. Earlier work audited 231,915 words without mechanical failures and identified standalone-okina handling as a bug (Azimov, 2026). This study compares the three converters on identical inputs rather than treating inventory choices as correctness judgments.

2. Orthographic Background

The target inventory represents o‘ as /ɵ/, g‘ as /ʁ/, and hamza as /ʔ/. Corpora interchange U+02BB, U+02BC, U+2018, U+2019, ASCII apostrophe, and backtick. The 701-hour ASR corpus contains 603k U+02BB occurrences; an independent TTS corpus showed 200:1 ASCII dominance. Canonicalization prevents this variation from contaminating phoneme sequences. Agglutinative suffixes also require digraph and boundary handling. Stress defaults to the final syllable, with exceptions.

3. Evaluation and Setup

We assess mechanical robustness, inventory consistency, digraph integrity, apostrophe canonicalization, and stress/boundary marking. The 356-item vocabulary comprises 112 core words and 244 apostrophe-variant/control items covering o‘, g‘, and hamza. All systems run single-core on the same AMD EPYC machine.

System	Version / invocation
Epitran	1.35.2; Epitran('uzb-Latn').transliterate(word)
eSpeak NG	1.51; espeak-ng -v uz -q --ipa=3 "word"
uzg2p	0.1.0; uzg2p.G2P(strict_oo=True)(word)

The harness, scripts/compare_g2p.py, records failures, empty outputs, raw and normalized strings, and per-word timing. Comparison normalization removes stress/tie-bars and applies NFC; raw outputs remain available.

Robustness checks target crashes, empty strings, and deterministic behavior. Inventory checks document consistent symbols, stress, and affricates. Digraph tests cover initial, medial, and final positions; apostrophe tests examine whether alternate glyphs leak into output. Boundary and stress checks concern alignment and exception handling. The vocabulary includes frequent nouns, verbs, numbers, and function words alongside orthographic controls.

4. Results

All systems produce zero failures and zero empty outputs. Mean seconds per word are 0.00003 for Epitran, 0.013 for eSpeak, and 0.00004 for uzg2p. eSpeak's per-word process spawning penalizes timing; batching would improve it.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.

Word	Epitran (norm)	eSpeak (norm)	uzg2p (strict)
o‘zbek	ozbek	ozbek	ɵzbek
shahar	ʃahar	ʃæhæɾ	ʃahar
osha	ɒʃa	ɑʃæ	ɒʃa
o‘sha	oʃa	oʃæ	ɵʃa

Inventories differ in vowel quality, affricate notation, and dental diacritics. Epitran/uzg2p use dental t̪/d̪ and tied affricates; eSpeak uses plain consonants and different affricate strings. uzg2p’s strict option maps o‘ to /ɵ/.

The tested sh/ch/ng/ts sequences leave no unconverted digraphs. uzg2p additionally uses an ng-split exception list for words such as singil, bolangiz, and bugungi. Stock Epitran passes standalone okina through. uzg2p reports canonicalization across all six apostrophe forms and a glottal-stop repair; eSpeak handles several forms but exhibits variant-dependent failures, notably backtick.

Representative output differences include g‘alati: ʁalat̪i in Epitran/uzg2p versus ʁælæty in eSpeak; tong: t̪ɒŋ versus tɑŋ; and tsikl: tsikl versus tsykɬ. For chiroyli, normalized Epitran/uzg2p give tʃirɒjli, while eSpeak gives tSyɾɑjły. Such differences require explicit representation policies before scoring. The wrapper’s ng-split mechanism distinguishes selected boundary sequences from ordinary ng→ŋ conversion, rather than applying one mapping indiscriminately.

In the listed ya‘ni variants, stock Epitran preserves ASCII apostrophe, U+02BB, and U+2018, while U+02BC and U+2019 produce /ʔ/. eSpeak yields jæʔny for several variants but shows problematic U+02BC and backtick outputs. uzg2p reports consistent normalization.

eSpeak emits stress by default and has internal exceptions. Epitran has no stress-marking option. uzg2p is unmarked by default but supports mark_stress=True, using final-syllable stress and a configurable dictionary of approximately 200 loanwords/proper nouns.

5. Discussion and Limitations

uzg2p’s audited map, normalization, stress control, and configurable inventory make it the recommended foundation for phoneme-based TTS and PER evaluation. eSpeak remains a valid direct-synthesis baseline; claims that Uzbek lacks a phonemizer are obsolete. Epitran remains useful for readable-map inspection.

PER can change with inventory conventions even for identical recognition errors. Reports should freeze the converter/version, stress policy, affricate segmentation, boundaries, and o‘ representation. uzg2p’s normalize_for_comparison utility is described as removing stress, expanding tie-bars, and harmonizing inventories. G2P decisions should accompany PER/CER/WER reporting.

This is a small, orthography-focused comparison, not a pronunciation-accuracy benchmark: no independent lexicon or native-speaker annotations are used. A held-out lexicon with inter-annotator agreement is needed. Coverage targets Tashkent-standard Uzbek; dialects, Cyrillic, and exhaustive loanword handling remain outside evaluation. The earlier full-vocabulary audit addresses mechanical robustness only.

6. Conclusion

All three converters are mechanically robust on the tested vocabulary. uzg2p offers configurable normalization and inventory control; eSpeak provides a credible synthesis baseline. Cross-system phoneme comparisons require explicit harmonization, not assumptions of interchangeable output.

Planned harness repository: github.com/azimov/uzbek-g2p-comparison (to be created).

References

1. Mortensen, D. R., Dalmia, S., & Littell, P. (2018). Epitran: Precision G2P for Many Languages. LREC 2018.
2. eSpeak NG contributors. eSpeak NG supported languages. GitHub.
3. Azimov, U. (2026). uzg2p: Auditing Uzbek Grapheme-to-Phoneme Conversion and a Phoneme Error Rate Benchmark for Agglutinative-Language ASR. Preprint.
4. uzinfocom-edu-ai. uzbek-asr-curated-701h. HuggingFace Datasets.
5. Canopy Labs. Orpheus-TTS. GitHub: canopyai/Orpheus-TTS.

© 2026 International Conference on Artificial Intelligence and Applications . All Rights Reserved.