

O'ZBEK TILIDA MATNLARNI TALAFFUZ QILISHDA PROSODIK MODELLAR

Kaxorov Azamat Abdulloyevich

Toshkent davlat iqtisodiyot universiteti katta o'qituvchisi

Annotatsiya: Ushbu maqolada o'zbek tilida matndan nutqqa aylantirish (TTS — Text-to-Speech) tizimlarida qo'llaniladigan prosodik modellar ko'rib chiqiladi. Prosodiya — nutqning ritmik va melodiyaviy tuzilishi bo'lib, u urg'u, intonatsiya, pauza va temp kabi hodisalarni o'z ichiga oladi. Maqolada prosodik modellarning turlari, ularning o'zbek tili fonetikasiga moslashuvi, zamonaviy chuqur o'qitish yondashuvlari hamda amaliy qo'llanilishi muhokama etiladi. O'zbek tilining agglutinatив xususiyati va o'ziga xos fonetik tuzilishi prosodik modellashtirish uchun alohida muammolar keltirib chiqaradi, bu esa maxsus yechimlarni talab etadi.

Kalit so'zlar: prosodiya, TTS, o'zbek tili, intonatsiya, urg'u, nutq sintezi, chuqur o'qitish, Tacotron, FastSpeech.

Abstract: This article reviews prosodic models used in Uzbek text-to-speech (TTS) systems. Prosody is the rhythmic and melodic structure of speech, which includes phenomena such as stress, intonation, pause, and tempo. The article discusses the types of prosodic models, their adaptation to the phonetics of the Uzbek language, modern deep learning approaches, and practical applications. The agglutinative nature of the Uzbek language and its unique phonetic structure pose special problems for prosodic modeling, which require special solutions.

Keywords: prosody, TTS, Uzbek language, intonation, stress, speech synthesis, deep learning, Tacotron, FastSpeech.

TTS tizimlari sun'iy intellektning eng muhim amaliy yo'nalishlaridan biridir. Ushbu tizimlar yozma matnni tabiiy eshitiluvchi nutqqa aylantiradi va avtomatik ma'lumot xizmatlari, virtual yordamchilar hamda ta'lim platformalarida keng qo'llaniladi.

O'zbek tili — turkiy tillar oilasiga mansub bo'lib, 35 milliondan ziyod kishi tomonidan so'zlashiladi. O'zbek TTS tizimlarining rivojlanishi so'nggi yillarda sezilarli tezlashdi, biroq prosodik modellashtirish sohasida hali bir qancha hal etilmagan muammolar mavjud. Prosodiya nutqning tabiiy va tushunarli bo'lishida asosiy rol o'ynaydi: bir xil so'zlar turli prosodik modellar bilan aytilganda butunlay boshqacha ma'no kasb etishi mumkin.

O'zbek tilining agglutinatив tuzilishi, ya'ni so'zga ko'plab qo'shimchalar biriktirish orqali yangi ma'nolar hosil qilish xususiyati, prosodik modellar uchun qator muammolar keltirib chiqaradi: so'z urg'usining o'zgaruvchanligi, gapning grammatik tuzilishiga bog'liq intonatsion konturlar va pauzalar o'rniga alohida e'tibor talab etadi.

Prosodiya tilshunoslikda nutqning suprasegmental hodisalari majmuini anglatadi. Bu hodisalar individual tovushlar (fonema) darajasida emas, balki bo'g'in, so'z va gap darajasida namoyon bo'ladi. Nutq sintezi uchun eng muhim prosodik parametrlar quyidagilardan iborat:

- Fundamental chastota (F0) — tovush paychalarining tebranish chastotasi va u intonatsiyaning asosini tashkil etadi.
- Davomiylik (duration) — individual tovushlar va bo'g'inlarning vaqt uzunligi.
- Amplituda (energiya) — nutq to'liqining kuchliligi, so'zlovchining ovoz balandligi bilan bog'liq.
- Pauzalar — nutq zanjirida so'z va gap orasidagi sukut davrlari.

O'zbek tilida so'z urg'usi asosan so'zning oxirgi bo'g'iniga tushadi, biroq bu qoida mutlaq emas. Qo'shimchalar qo'shilganda urg'u siljishi kuzatiladi. Masalan, "maktab" so'zida urg'u

ikkinchi bo'g'inda bo'lsa, "maktablar" shaklida u yana oxirgi bo'g'inga o'tadi. Bu dinamik xususiyat prosodik modellar uchun qo'shimcha murakkablik yaratadi.

Gap darajasida esa mantiqiy urg'u (fokus) muhim rol o'ynaydi: so'zlovchi qaysi so'zni ta'kidlamoqchi bo'lsa, uni boshqa prosodik vositalar yordamida ajratib ko'rsatadi. Bu xususiyat TTS tizimlariga matnning pragmatik ma'nosini tushunish imkoniyatini talab etadi.

O'zbek tilida intonatsiya gapning kommunikativ turini (darak, so'roq, buyruq, undov) ifodalashning asosiy vositalaridan biridir. Darak gaplarda F0 konturi odatda pasayib boradi. So'roq gaplarda esa, ayniqsa so'roq so'zsiz gaplarda, so'nggi bo'g'inda keskin ko'tarilish kuzatiladi. Buyruq intonatsiyasida baland boshlash va keskin tushish xarakterlidir.

Nutq sintezi tizimlarida prosodik modellar evolyutsion yo'l bilan rivojlandi. Dastlabki TTS tizimlari qo'lda yozilgan fonetik va prosodik qoidalar asosida ishlagan. Ushbu yondashuvda lingvistlar urg'u, intonatsiya va pauza qoidalarini eksplitsit ravishda dasturlashgan. O'zbek tili uchun bunday qoidalar qisqacha quyidagicha tasniflanadi:

- Vergul oldida qisqa pauza, nuqta oldida uzun pauza qo'yiladi;
- Ko'makchilar va bog'lovchilar urg'usiz aytiladi;
- So'roq gapining oxirida F0 ko'tariladi.

Qoida asosidagi modellar sodda va tushuntirib berish oson bo'lsa-da, ularning tabiiy ovozdan uzoqligi, chuqur lingvistik bilim talab qilishi va yangi holatlarga moslashish qiyinligi kabi kamchiliklari mavjud.

2000-yillarda yashirin Markov modellari (HMM — Hidden Markov Model) asosidagi TTS tizimlari keng tarqaldi. HMM-TTS tizimida prosodik parametrlar (F0 va davomiylik) kontekstga bog'liq fonetik birliklar (trifonlar) uchun statistik taqsimotlar sifatida modellanadi.

O'zbek tili uchun HMM-TTS tizimlarini yaratishda asosiy muammo yorliqlangan nutq korpuslarining kamligi hisoblanadi. Biroq bu yondashuv qoidali tizimlardan sezilarli darajada tabiiyroq tovush hosil qiladi va ma'lum nutq ohangi uchun moyil qilish (adaptation) imkoniyatini beradi.

Keyinchalik, chuqur o'qitish (deep learning) inqilobi TTS tizimlarini tubdan o'zgartirdi. Hozirgi zamonaviy neyron tarmoq modellari tabiiy nutqqa juda yaqin ovoz sifatini ta'minlaydi. O'zbek tili uchun qo'llaniladigan asosiy arxitekturalar sezilarli darajada o'zbek TTS tizimlarida ijobiy o'zgarishlarni yuzaga keltirdi.

Jumladan, Google tomonidan taklif etilgan Tacotron modeli end-to-end yondashuvi bilan matnni to'g'ridan-to'g'ri mel-spektrogrammaga aylantiradi. Encoder-decoder arxitekturasi va e'tibor (attention) mexanizmi yordamida model matnning prosodik usullarini o'z-o'zicha o'rganadi. O'zbek tilida Tacotron 2 modeli 10-15 soatlik yorliqlangan ma'lumotlar bilan ham maqbul natijalar berishi ko'rsatilgan.

FastSpeech modeli paralel sintez qilish imkoniyati bilan Tacotronga nisbatan sezilarli darajada tezroq ishlaydi. Bu model alohida davomiylik va F0 prediktorlari orqali prosodik parametrlarni nazorat qiladi. O'zbek tili uchun FastSpeech 2 ni sozlash jarayonida davomiylik prediktorini o'zbek tilining agglutinativ xususiyatlariga moslashtirishga e'tibor qaratish tavsiya etiladi.

VITS (Variational Inference with adversarial learning for end-to-end Text-to-Speech) modeli variatsion autoenkoder va generativ-raqobatchi tarmoqlarni birlashtiradi. Bu model bitta bosqichda fonema kirishidan to'liq tovush signalini yaratadi va eng yuqori sifatli sintez natijalaridan birini beradi. O'zbek tili VITS modeli uchun fonema lug'atini va grafema-fonema konversiya modelini oldindan tayyorlash muhim ahamiyatga ega.

Shunga qaramay muammolar uchrab turibdi. Bunda asosan, agglutinatsiya va urg‘u o‘zgaruvchanligi, imlo va talaffuz farqlari, korpus tanqisligi ya‘ni zamonaviy neyron TTS modellari katta hajmli yorliqlangan nutq ma‘lumotlariga muhtoj. O‘zbek tili uchun ommaviy ravishda mavjud bo‘lgan ochiq litsenziyalı nutq korpuslari hali ham cheklangan. Respublikamizda bir nech marta tanlovlar o‘tkazilib, volontyorlar yordamida hamda, Mozilla Common Voice platformasi orqali o‘zbek tili korpusini to‘plash harakati davom etmoqda, ammo prosodik tadqiqotlar uchun zarur bo‘lgan fonetik yorliqlanlangan ma‘lumotlar hali ham kam. Ushbu muammoni hal qilish uchun transfer learning va low-resource TTS texnikalariga murojaat qilish maqsadga muvofiq.

Turkiy tillar oilasi (turk, qozoq, ozarbayjon, qirg‘iz va boshqalar) o‘rtasidagi fonetik va prosodik o‘xshashliklardan foydalangan holda transfer learning yondashuvlari o‘zbek TTS sifatini sezilarli oshirishi mumkin. Turk tilida o‘qitilgan model o‘zbek tiliga nisbatan yaxshi "baza" prosodik bilim beradi, chunki ikkala til ham oxirgi bo‘g‘inga urg‘u berishni afzal ko‘radi va shunga o‘xshash agglutinatıv tuzilishga ega.

Zamonaviy TTS tizimlarida foydalanuvchiga prosodik parametrlarni nazorat qilish imkoniyati berilmoqda. Style tokens yoki Global Style Token (GST) mexanizmlari yordamida mo‘ljallangan model ma‘lum nutq uslublarni (muxbir, hikoya o‘qish, suhbat) simulyatsiya qila oladi. O‘zbek tili uchun bunday nazorat mexanizmlari, jumladan SSML (Speech Synthesis Markup Language) ning o‘zbek tiliga moslashtirish, amaliy qo‘llanilishi kengaymoqda.

Yuqori sifatli o‘zbek TTS uchun alohida modullar zarur bo‘lib, ular morfologik tahlil, gap tuzilish tahlili (parser), F0 prediktorlar, davomiylik prediktorlari hisoblanadi.

TTS tizimlarining prosodik sifatini baholash murakkab vazifaligi uchun, subyektiv va obyektiv usullarni o‘z ichiga oladi:

- MOS (Mean Opinion Score) — tinglovchilar guruhi tomonidan 1-5 shkalada baholanadi; tabiiylik va tushunarlılik alohida baholanadi.
- MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) — bir necha TTS tizimini parallel solishtirishga imkon beradi.
- F0 RMSE — sintetik va tabiiy nutq orasidagi fundamental chastota farqini o‘lchaydi.
- Davomiylik farqi — sintetik va haqiqiy davomiyliklar orasidagi o‘rtacha absolyut xato.

O‘zbek tilida matnlarni tabiiy talaffuz qilish uchun prosodik modellash tirish hal qiluvchi ahamiyatga ega. So‘nggi yillarda chuqur o‘qitishga asoslangan modellar (Tacotron 2, FastSpeech 2, VITS) o‘zbek TTS sifatini sezilarli darajada oshirdi. Biroq quyidagi muammolar hali ham muhim tadqiqot yo‘nalishlari hisoblanadi:

- O‘zbek tili uchun katta hajmli, fonetik jihatdan annotatsiyalangan nutq korpuslarini yaratish;
- Agglutinatsiya ta‘sirini hisobga oluvchi morfologik-prosodik integratsiyalangan modellar ishlab chiqish;
- Dialektal va uslubiy xilma-xillikni qamrab oluvchi ko‘p uslubli TTS tizimlari yaratish;
- Transfer learning va kam resursli o‘qitish texnikalarini o‘zbek tili uchun maqbullashtirish.

Ushbu sohadağı tadqiqotlar o‘zbek tilida sun‘iy intellekt texnologiyalarining keng qo‘llanilishiga — ta‘lim, sog‘liqni saqlash, ma‘muriyat va media sohalarida — yo‘l ochadi hamda o‘zbek tilidagi raqamli kontentga egalik darajasini oshiradi.

Foydalanilgan adabiyotlar

1. Kakhorov, A. (2023). Og‘zaki muloqot tizimlarini ishlab chiqish uchun noravshan qoidalarga asoslangan evolyutsion klassifikatorlarning qo‘llanilishi. *Digital Transformation And Artificial Intelligence*, 1(2), 108–115. Retrieved from

2. A Kakhorov, D Yodgorova, M Khayrullayev. "AI-DRIVEN ENVIRONMENTAL COMMUNICATION: TRANSLATING SUSTAINABILITY REPORTS INTO UZBEK FOR GLOBAL AWARENESS" Экономика и социум 6-1 (121), 292-301
3. Кахоров, А. "ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА (NLP) НА ОСНОВЕ СИНТАКСИЧЕСКОГО И СЕМАНТИЧЕСКОГО АНАЛИЗА ТАКЖЕ ЧАСТЕРЕЧНОЙ РАЗМЕТКИ." International Scientific and Practical Conference on Algorithms and Current Problems of Programming. 2023.
4. Wang, Yuxuan, et al. "Tacotron: Towards end-to-end speech synthesis." arXiv preprint arXiv:1703.10135 (2017).
5. Kim, Jaehyeon, Jungil Kong, and Juhee Son. "Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech." International conference on machine learning. PMLR, 2021.
6. Muradovna, Allaberganova Nasiba. "O'ZBEK TILI MATNLARI SEMANTIK O'XSHASHLIGINI ANIQLASH MODELLARINING ILMIY-NAZARIY TAHLILI." Илм-фан ва инновацион ривожланиш/Наука и инновационное развитие 9.1 (2026): 34-44.
7. Кахоров А. Анализ текста методом машинного обучение //International Scientific and Practical Conference on Algorithms and Current Problems of Programming. – 2023.
8. Кахоров А., Qalandarov J. RAQAMLI AUDIO SIGNALNI QAYTA ISHLASH //Research and implementation. – 2023.
9. Xusainova, Zilola, and Nasiba Allaberganova. "О создании мобильного приложения электронного словаря произведений Алишера Навои." COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS 1.1 (2021).
10. B.B. Muminov, N.M. Allaberganova. A SEMI-SUPERVISED FRAMEWORK NAMED AUGSBERT-UZ FOR HIGH-PERFORMANCE SEMANTIC TEXTUAL SIMILARITY IN UZBEK. Zenodo, ноябрь 2025 г., <https://doi.org/10.5281/zenodo.17693973>.