TURKOLOGIYADA YARATILGAN NUTQ SINTEZATORLARI TADQIQI

Ro‘za Toliboyevna O‘rinboyeva
O‘zMU magistranti
orinboyevaroza@gmail.com

ANNOTATSIYA

Maqolada o‘zbek tiliga alohida e’tibor qaratilgan holda turkiy tillar uchun nutq sintezatorlarini yaratish bo‘yicha mavjud tadqiqotlar tahlil qilinadi, shuningdek, ushbu sohaning asosiy yo‘nalishi va rivojlanish istiqboli haqida so‘z boradi.

Nutq sintezatorining tarkibiy qismlari: audio materiallar, ularni transkripsiyalash hamda teglash jarayonlari qozoq, qirg‘iz, turk, tatar hamda o‘zbek tillarida yaratilgan lingvistik ta’minotlar misolida tahlil etilgan.

Tayanch so‘zlar: tabiiy til, nutq, sintezator, sintez, Common Voice, transkripsiya, nutq korpusi.

ABSTRACT

The article analyzes existing research on the creation of speech synthesizers for Turkic languages, with a special focus on the Uzbek language, and also discusses the main directions and development prospects of this area.

The components of a speech synthesizer: audio materials, their transcription and tagging processes are analyzed using the example of linguistic resources created in the Kazakh, Kyrgyz, Turkish, Tatar and Uzbek languages.

Keywords: natural language, speech, synthesizer, synthesis, Common Voice, transcription, speech corpus.

KIRISH

Ijtimoiy faoliyat sohalari raqamlashtirilayotgan bugungi kunda inson va kompyuter o‘rtasidagi muloqotni ta’minlash uchun tabiiy tilni qayta ishlash muhim ahamiyat kasb etmoqda. Muayyan tilga ishlov berish, uni grammatik qoidalarga asoslanib tizimga solish, matematik algoritmlar bilan moslashtirish, kodlash talab etiladi. Tabiiy tilni qayta ishlashda nutqni tanib olish qurilmasi, ya’ni nutq sintezatorlarini yaratish ahamiyatga molikdir.

Nutq sintezatori ovozli xabarni matn ko‘rinishiga o‘tkazishga mo‘ljallangan dastur hisoblanadi. Dasturlar yordamida turli sohalarda, xususan, ko‘zi ojizlar uchun matnni o‘qishda, o‘yin yoki ilovalarda ovozli interfeyslarni yaratishda, avtomobillardagi navigatsiya tizimlarida yo‘l

ko‘rsatkichlarini berish hamda video-kuzatuv kameralaridagi ovozlarni matnga aylantirishda foydalanish mumkin.

ADABIYOTLAR TAHLILI VA METODOLOGIYA

Jahon kompyuter lingvistikasida nutq sintezatori yaratish masalasi 18-asrdan o‘rganila boshlangan. Rus tilshunosligida K.Kratzenshteyn (1781), nemis tilshunosligida V. fon Kempelen (1781), ingliz tilshunosligida Ch.Uitston (1800), yapon tilshunosligida Obata va Teshima (1932), qozoq tilshunosligida D.Abilhayer (2016), turk tilida D.Dalva (2000), o‘zbek tilida M.Musayev (2021), M.Qurbonova (2021), A.Narzullayev (2024) kabi olimlar dastlabki tadqiqotchilar hisoblanadi.

Turkologiyada nutq sintezatori uchun lingvistik ta’minot yaratish o‘tgan asr oxirlarida uchratish boshlangan bo‘lib, turk tilida olib borilgan izlanishlar ahamiyatlidir. Turkiyadagi bir qator universitetlar tabiiy tilni qayta ishlash (NLP) hamda nutq texnologiyasiga oid tadqiqotlar uchun yo‘naltirilgan. Ayni paytda Istanbul texnika universiteti, Bilkent universiteti hamda Yaqin Sharq texnika universitetlarida tadqiqot guruhlari mavjud.

Turk tilida nutqni sintez qilish muammolari 1980-yillardan o‘rganilgan bo‘lsa [1], 2000-yildan keyin nutqni avtomatik aniqlash (ASR) tizimlari ishlab chiqila boshlandi. 2019-yilda Asefisaray tomonidan shu mavzuda doktorlik dissertatsiyasi himoya qilingan [2]. Uning izlanishlari natijasida vergullarning 63,4 foizi, nuqtalarning 68,9 foizi va savol belgilarining 82,5 foizi tuzatildi. 2018-yilda Arslon atrof-muhit nutqini aniqlash bo‘yicha tadqiqot olib borishni taklif qilgan [3]. Uning izlanishlari natijasida o‘q ovozlarini aniqlashda 99,9%, qichqiriq va avtohalokat tovushlarini aniqlashda 98,4% aniqlikka erishilgan. Akin tomonidan tayyorlangan magistrlik dissertatsiyasida vizual va eshitish jihatdan ajratish qiyin bo‘lgan egizak ovozlarni tahlil qilish orqali biometrik tanib olishga asoslangan modelni ishlab chiqish maqsad qilingan [4]. 78 ta egizak ovozini o‘z ichiga olgan test to‘plami bilan o‘tkazilgan tajribada ularning 91 foizi to‘g‘ri aniqlangan. Ushbu tadqiqotlar natijasi nutq sintezatorini yaratish uchun dastlabki qadam bo‘lib xizmat qiladi.

Turkiyaning Ishik universiteti ilmiy xodimi Dog‘an Dalva o‘zining “Automatic speech recognition system for turkish spoken language” nomli kitobida turk tilining fonetik jihatlarini, nutq azolarining o‘ziga xos xususiyatlarini hamda ularning nutqni aniqlashdagi ahamiyatini batafsil yoritib beradi [5]. Olimning ta’kidlashicha, nutqni aniqlashda uning tuzilishiga ko‘ra alohidalangan so‘zni aniqlash (IWR) va doimiy nutqni aniqlash (CSR) kabi modellar qo‘llaniladi. Bundan tashqari, muallif ba’zi fonetik birliklarni turli talaffuz qilish mumkinligini, bunda fonemalarning talaffuzi yonma-yon kelgan tovushlarga bog‘liq ekanligini ta’kidlaydi. Ushbu jarayonda yana ikkita alohida so‘zni

birlashtirish ham mumkin. Masalan, turk tilida “sil” ikki so‘z orasidagi sukutni bildiruvchi “geçen-sil-aksham” so‘zi o‘rniga “geçenaksham” deb aytilishi mumkin. Bundan ko‘rinadiki, sintezator akustik signal, koartikulyatsiya effektlari va egiluvchan artikulyatsiyada kelib chiquvchi noma’lum muammolarni ham hal qilshi kerak bo‘ladi.

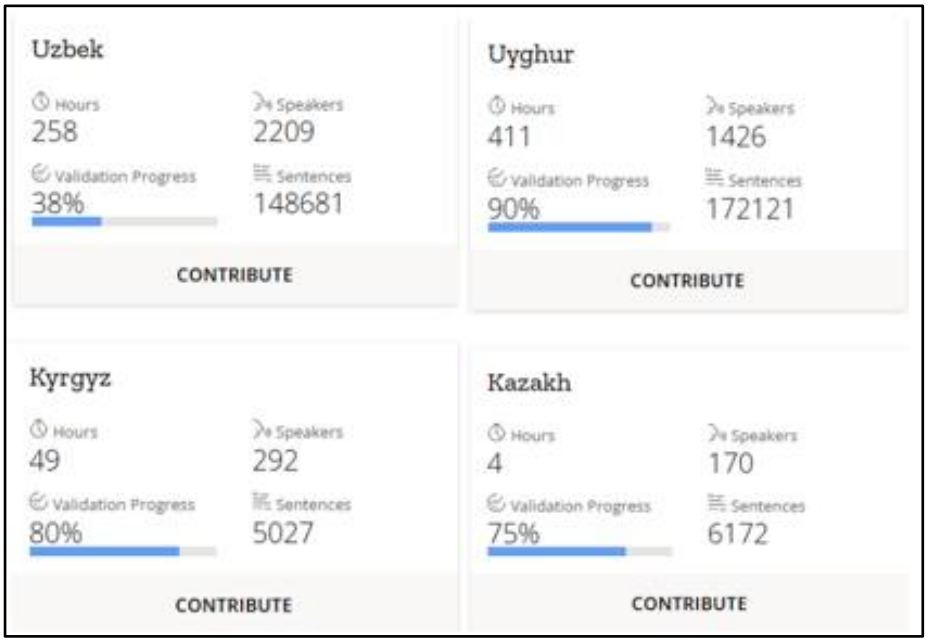
MUHOKAMA VA NATIJALAR

Og‘zaki nutqda ba’zi talaffuzi o‘xshash so‘zlar mavjud bo‘lib, so‘z qismlari aro o‘xshashlik, bu holatdagi ayrim chalkashliklar yoki turli so‘zlarning bir xil talaffuzga ega ekanligi akustik jihatdan noaniq so‘zlar deb ataladi. Masalan, “gelmek” (kelish) va “gitmek” (ketish) so‘zlarining oxiri o‘xshash bo‘lib, ular chalkashishi mumkin. So‘z boshidagi tovush“gul” (gul) va “kul” (kul) so‘zlari akustik jihatdan noaniq so‘zlardir. Shu kabi misollarning kontekst uchun mos keladigan ma’nosi xabarning taxminiy ma’nosiga ko‘ra til modelining pragmatik darajasida ko‘rib chiqiladi.

Turk tili uchun yaratilgan nutq sintezatoriga audio material sifatida turkcha chegaralangan so‘z va jumlalar olingan. Ularning har biri yuqoridagi kabi bosqichlardan o‘tkazilib, tahlil qilinadi.

Turkologiyada nutq sintezatorini yaratish bo‘yicha olib borilgan tadqiqotlardan yana biri Wenqiang Du, Yikeremu Maimaitiyiming, Mewlude Nijat, Lantian Li, Asqar Hamdulla, Dong Vanglarga tegishli bo‘lib, ularning “Automatic Speech Recognition for Uyghur, Kazakh, and Kyrgyz: An Overview” nomli maqolasida keng yoritib berilgan [6]. Unda qirg‘iz, qozoq hamda uyg‘ur tillarida nutq sintezatori yaratishga doir ilmiy izlanishlar va amaliy natijalar bilan tanishish mumkin. Maqolada har bir til uchun ma’lumotlar bazasi yaratishda Common Voicedan foydalanish samarali natija berishi qayd etiladi.

Common Voice — Mozilla va boshqa kompaniyalar, tadqiqot institutlari tashabbusi bilan yaratilgan bepul va ochiq kraudsorsingli ma’lumotlarni yig‘ish loyihasi [7]. Bunda nutq to‘plashda ko‘ngillilardan foydalaniladi. Istalgan odam bulutli serverga nutq yuklaydi hamda shu platforma orqali o‘z nutqini izohlashi mumkin. 2019-yilda ushbu loyihaning dastlabki natijasi — **Common Voice 1.0** chiqarildi. Unda jami 19 tildagi 1000 soat nutq mavjud. Ayni paytda **Common Voice 11.0** (2022-yil 21-sentabr) versiyasiga yangilangan bo‘lib, u 100 ta tilni o‘z ichiga qamrab olgan [8]. **Common Voice** — bu kraudsorsing tomonidan yaratilgan birinchi ko‘p tilli ma'lumotlar resursi. U doimiy ravishda ma'lumotlarni to‘playdi. Bu loyiha doirasida turkiy tillar uchun ham muayyan darajada ma’lumotlar bazasi to‘plangan. Quyida ularning ayrimlariga to‘xtalib o‘tamiz.



Maqolada Common Voicedan foydalanib, uyg‘ur, qozoq va qirg‘iz tillarida yaratilgan nutq sintezatorlarining shakllanish bosqichi, ishlash prinsipi, yangi versiyalari hamda ular uchun ma’lumotlar bazasi to‘plash jarayonlari haqida batafsil yoritilgan. Masalan, qirg‘iz tili uchun 2018-yilda yaratilgan CNN ning dastlabki modeli keltiriladi. 2019-yilda Common Voice orqali ma’lumotlar bazasi to‘plana boshlangan, keyingi tadqiqotlar natijasida **CPC, Wave2Vec2.0** modellari yaratiladi.

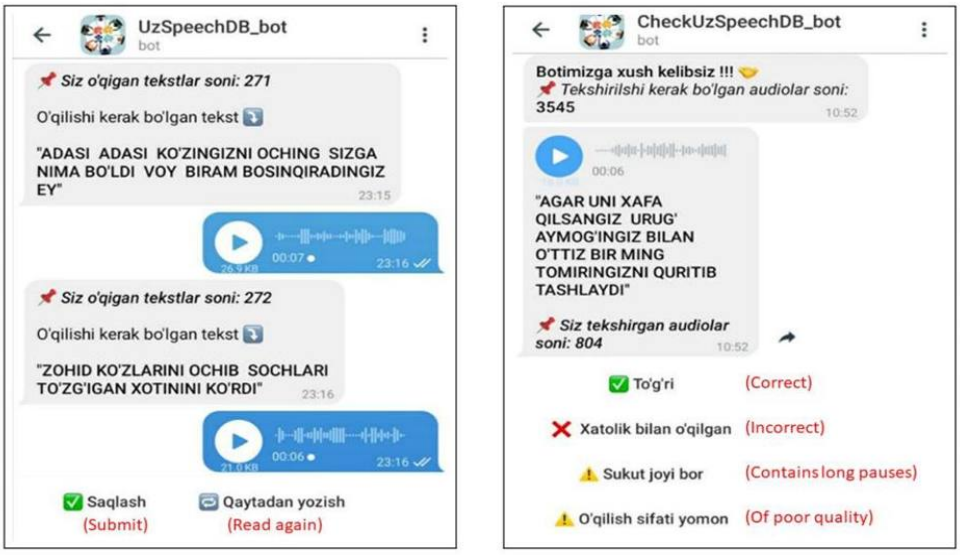
Uyg‘ur tilida bu boradagi tadqiqotlar 2016-yilda Tsinghua, Shinjon va Shimoli-g‘arbiy Minzu universitetlari hamkorligida tuzilgan **“Multilingual Minorlingual Automatic Speech Recognition (M2ASR)”** nomli qo‘shma loyiha doirasida boshlangan. 2017-yilda amaliy natija sifatida **“THUYG-20”** nomli ma’lumotlar bazasi yaratilgan [9].

Qozoq tilshunosligida 2015-yilda Xomitsevich guruhi tomonidan qozoq-rus ikki tilli nutqni aniqlash tizimi o‘rganila boshlangan [10]. 2016-yilda Davel Abilhayer hamda boshqa bir guruh olimlar “On language model construction for LVCSR in Kazakh” nomli maqolasida GMM-HMM asosida uzluksiz nutqni aniqlash tizimini yaratish masalasini yoritilgan [11]. Keyingi yillar davomida qozoq tili uchun DNN-HMM tizimi [12], BLSTM-CTC end-to-end tizimi [13], Transformer CTC/attention tizimi [14], LSTM-CTC end-to-end Qozoq ASR tizimi [15], Nazarboyev universiteti tadqiqotchilari tomonidan “Kazakh speech corpus” (KSC) [16], (KazakhTTS) [17], KSC va KazakhTTS2 ga qo‘shimcha ma’lumotlar kiritilgan holda **KSC2** nomli yangi versiyasi yaratilgan [18].

2021-yilda **M2ASR** loyihasi doirasida nutqiy ma’lumotlar bazalari (uyg‘ur tilida 277 soat, qozoq tilida 250 soat va qirg‘iz tilida 166 soat), matn korpusi (uyg‘ur tilida 88 M, qozoq tilida 100 M, qirg‘iz tilida 6 M) va orfoepik lug‘atlar (uyg‘ur tilida 45 000 ta, qozoq

tilida 100 000, qirg‘iz tilida 15 000)dan iborat resurslar to‘plangan.

Turkologiyada olib borilayotgan nutq sintezatorini yaratishga doir tadqiqotlar doirasida o‘zbek tiliga oid nazariy va amaliy tajribalar olib borilmoqda. Xususan, o‘zbek tilshunosligida nutq sintezatori yaratish bo‘yicha Toshkent axborot texnologiyalari universitetida Muhammadjon Musayev tomonidan tadqiqotlar o‘tkazilmoqda. Muhammadjon Musayev tomonidan “Development of integral model of speech recognition system for Uzbek language”, “Image Approach to Speech Recognition on CNN”, “USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments” kabi maqolalar e’lon qilingan, lingvistik hamda dasturiy ta’minotlar yaratilgan. Ular orqali muallif o‘zbek tili uchun nutq sintezatori yaratish muammosini ilgari suradi, jahon tajribalarini tahlil qiladi. Nutq sintezatori yaratish texnologiyasi, uslubi va usullari haqida batafsil ma’lumot beradi. Jumladan, “USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments” nomli maqolasida o‘zbek tili uchun yaratilgan ochiq manbali o‘zbekcha nutq korpusi haqida ma’lumot beradi [19]. Maqolada 958 ta ma’ruzachi tomonidan yozilgan 105 soatlik trankripsiyalangan audio matn o‘zbekcha nutq korpusiga kiritilganligi qayd etiladi. Ularni to‘plashda an’anaviy Common Voice loyihasidan farqli ravishda alohida telegram botdan foydalaniladi. Botda dastlab foydalanuvchi ism-familiyasi, yoshi, jinsi, turar joyi kabi meta ma’lumotlarni kiritadi. Keyingi bosqichda havola etilgan matnni o‘qib beradi. Xuddi shu tartibda o‘qilgan audiolarni tekshirish uchun ham maxsus bot yaratilgan. Bu botda foydalanuvchi audioni eshitib, undagi xatoliklarni, ovoz yozish sifatini hamda nutq orasidagi sukutlarning bor-yo‘qligini aniqlab beradi:



Bu samarali usul hisoblanib, ovozlarning xilma-xilligini ta’minlaydi, audio yig‘ish jarayonini tezlashtiradi.

Bundan tashqari nutq sintezatori yaratish borasidagi tadqiqotlar Mohirdev MCHJ tomonidan UzbekVoiceAI [20], Munavvara Qurbonova tomonidan uznutq.com [21] saytlari orqali ko‘zga tashlanadi. Ularning tadqiqotlarida o‘zbek tilining fonetik hodisalari, xususan, tovush tushishi, ortishi va almashishi kabi hodisalarga alohida e’tibor qaratilgan. o‘zbek tilining barcha jihatlarini qamrab olgan holda tadqiqotchilar tomonidan o‘zbek tili shevalari uchun nutq sintezatorlari yaratish g‘oyasi ilgari surilmoqda. Areal lingvistika doirasidagi tadqiqotlarga asoslanib o‘zbek tili shevalaridan iborat nutq korpusi yaratish mumkin. Bunda lingvistik ta’minot tarkibiga to‘plangan audio matnlar adabiy nutqda va an’anaviy transkripsiyada kiritiladi. Buning uchun to‘plangan barcha audiolar segmentlarga ajratilib teglanadi. Aniqlash uchun berilgan nutq shu segmentlar asosida qidiriladi va natija uchun jamlanadi.

XULOSA

Bu kabi tadqiqotlar o‘zbek tilshunosligida NLP sohasining dolzarbligini, til imkoniyatlarini namoyon etishdagi ahamiyatini ko‘rsatadi. Tabiiy tilni qayta ishlash sohasidagi mavjud muammolarni bartaraf etishda ko‘maklashadi.

Tabiiy tilni qayta ishlashga qaratilgan izlanishlar natijasida o‘zbek tili uchun bir nechta nutq sintezatorlari yaratishga erishilgan bo‘lib, ular foydalanuvchilar tomonidan amaliyotda qo‘llanilmoqda. Tilning o‘ziga xos murakkab xususiyatlarini inobatga olgan holda nutq sintezatorlari uchun yaratilgan ma’lumotlar bazalarini takomillashtirish tilshunos olimlar oldida turgan navbatdagi vazifalardan biridir. Yaqin kelajakda nutq sintezatorlaridan rasmiy ish faoliyatda hamda ommaviy axborot vositalarida foydalanish ko‘zda tutilgan.

REFERENCES

1. Arslan, Recep S. and Barişçi N. A detailed survey of Turkish automatic speech recognition//Turkish Journal of Electrical Engineering and Computer Sciences: Vol. 28: No. 6, Article 12, 2020. https://doi.org/10.3906/elk-2001-38
2. Asefisaray B. End-to-end speech recognition model: experiments in Turkish. PhD, Hacettepe University, Ankara, Turkey, 2018.
3. Arslan Y. Detection and recognition of sounds from hazardous events for surveillance applications. PhD, Yıldırım Beyazıt University, Ankara, Turkey, 2018.
4. Akın C. Voice recognition system with score level fusion methods and embedded system design. MSc, İstanbul Technical University, İstanbul, Turkey, 2019.
5. Dalva_MSc_320753-libre.pdf (d1wqtxts1xzle7.cloudfront.net)

6. Wenqiang D., Yikeremu M., Mewlude N., Lantian L., Hamdulla A., Dong V. Automatic Speech Recognition for Uyghur, Kazakh, and Kyrgyz: An Overview, Appl. Sci. 2023. https://doi.org/10.3390/app13010326
7. Ardila R., Branson M., Davis K., Henretty M., Kohler M., Meyer J., Morais R., Saunders L., Tyers F.M., Weber G. Common voice: A massively-multilingual speech corpus. arXiv 2019. https://commonvoice.mozilla.org/en
8. https://commonvoice.mozilla.org/en
9. Rouzi A., Shi Y., Zhiyong Z., Dong W., Hamdulla A., Fang Z. THUYG-20: A free Uyghur speech database// Science Technology, 2017. pp. 182–187. [Google Scholar]
10. Khomitsevich O., Mendelev V., Tomashenko N., Rybin S., Medennikov I., Kudubayeva S. A bilingual Kazakh-Russian system for automatic speech recognition and synthesis//In Proceedings of the International Conference on Speech and Computer, 2015. pp. 25–33. [Google Scholar]
11. Dawel A., Nurmemet Y., Liu Y. On language model construction for LVCSR in Kazakh. 2016. 178–181. [Google Scholar]
12. Mamyrbayev O., Turdalyuly M., Mekebayev N., Mukhsina K., Keylan A., Baba A.B., Nabieva G., Duisenbayeva A., Akhmetov B. Continuous speech recognition of Kazakh language//In Proceedings of the ITM Web of Conferences, 2019. [Google Scholar]
13. Mamyrbayev O., Alimhan K., Zhumazhanov B., Turdalykyzy T., Gusmanova F. End-to-end speech recognition in agglutinative languages//In Proceedings of the Asian Conference on Intelligent Information and Database Systems, 2020. pp. 391–401. [Google Scholar]
14. Orken M., Dina O., Keylan A., Tolganay T., Mohamed O. A study of transformer-based end-to-end speech recognition system for Kazakh language. 2022. [PubMed]
15. Beibut A., Darkhan K., Olimzhan B., Madina K. Development of Automatic speech recognition for Kazakh language using transfer learning//Int. J. Adv. Trends Comput. Sci. Eng. 2020. [Google Scholar] [CrossRef]
16. Khassanov Y., Mussakhojayeva S., Mirzakhmetov A., Adiyev A., Nurpeiissov M., Varol H.A. A crowdsourced open-source Kazakh speech corpus and initial speech recognition baseline. arXiv 2020. [Google Scholar]
17. Mussakhojayeva S., Janaliyeva A., Mirzakhmetov A., Khassanov Y., Varol H.A. KazakhTTS: An open-source Kazakh text-to-speech synthesis dataset. arXiv 2021. [Google Scholar]
18. Mussakhojayeva S., Khassanov Y., Varol H.A. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus//In

Proceedings of the INTERSPEECH, 2015. pp. 1367–1371. [Google Scholar]
19. Musayev M., Mussakhojayeva S., Khujayorov I., Khassanov Y., Ochilov M., Atakan H., Varol. USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments, (PDF) USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments
20. UzbekVoiceAI
21. uznutq.com