“JOURNAL OF SCIENCE-INNOVATIVE RESEARCH IN UZBEKISTAN” JURNALI
VOLUME 04, ISSUE 11, 2026. NOVEMBER
ResearchBib Impact Factor: 10/54/ 2025 ISSN 2992-8869

# O‘ZBEK TILIDA NUTQNI AVTOMATIK ANIQLASH TIZIMLARINI YARATISHDA OVOZLI AUDIOLARNING O‘RNI: MAVJUD RESURSLARNING QIYOSIY TAHLILI

***Kurbonov Akmal Azamatovich***
*Navoiy, Sharq universiteti “Axborot-Texnologiyalar” fakulteti O‘qituvchisi*
***Tel:*** *+998973226755*
***ORCID:*** *0009-0002-8232-6700*
***Email:*** *qurbonovakmalazamatovich@gmail.com*
***Aslonov Amirxon Utkir o‘g‘li***
*Navoiy, Sharq universiteti “Axborot-Texnologiyalar” fakulteti talabasi*
***Tel:*** *+998975801481*
***Email:*** *amirxonaslonov0@gmail.com*

***Annotatsiya.*** *Ushbu maqolada o‘zbek tilida nutqni avtomatik aniqlash (ASR) tizimlarini yaratishda ovozli korpuslarning o‘rni ochiq manbalarda e’lon qilingan ma’lumotlar asosida qiyosiy tahlil etilgan. USC, Common Voice (25.0-versiya), FeruzaSpeech va boshqa resurslar hajmi, spikerlar soni, nutq turi, yozuv standarti hamda foydalanish shartlari bo‘yicha taqqoslangan; korpuslararo umumlashtirish esa e’lon qilingan tajribalar natijalari asosida baholangan. Aniqlanishicha, Common Voice o‘zbek segmentida yozib olingan 265,75 soatning atigi 101,01 soati tasdiqlangan; bitta spikerning 52,09 soatlik yozuvlarida o‘qitilgan model boshqa korpuslarda 76 foizdan yuqori WER (so‘zlar bo‘yicha xatolik darajasi) bergan; o‘quv to‘plamiga USC qo‘shilishi esa usc-test bo‘yicha WER qiymatini kamida 2,5 baravar kamaytirgan. Bosh harf va tinish belgilarini hisobga olish ayrim hollarda WER qiymatini 2 baravardan ortiq o‘zgartirishi ko‘rsatilgan. Korpuslarni tayyorlash va e’lon qilish bo‘yicha amaliy tavsiyalar ishlab chiqilgan.*

***Kalit so‘zlar:*** *Nutqni avtomatik aniqlash, ASR, ovozli korpus, o‘zbek tili, kam resursli tillar, Common Voice, so‘zlar bo‘yicha xatolik darajasi (WER), korpuslararo umumlashtirish.*

***Abstract.*** *This article presents a comparative analysis of the role of speech corpora in building automatic speech recognition (ASR) systems for the Uzbek language, based on data published in open sources. USC, Common Voice (version 25.0), FeruzaSpeech and other resources are compared by size, number of speakers,*

*speech type, transcription standard and terms of use, and cross-corpus generalization is assessed from published experimental results. It is found that only 101.01 of the 265.75 hours recorded in the Uzbek segment of Common Voice are validated; a model trained on 52.09 hours from a single speaker showed a word error rate (WER) above 76% on the other corpora; adding USC to the training set reduced the WER on the USC test set by a factor of at least 2.5. Taking casing and punctuation into account was shown to change the WER by more than a factor of two in some cases. Practical recommendations for preparing and publishing speech corpora are proposed.*

***Keywords:*** *Automatic speech recognition, ASR, speech corpus, Uzbek language, low-resource languages, Common Voice, word error rate (WER), cross-corpus generalization.*

***Аннотация.*** *В статье на основе данных из открытых источников проведён сравнительный анализ роли речевых корпусов при создании систем автоматического распознавания речи (ASR) для узбекского языка. Корпуса USC, Common Voice (версия 25.0), FeruzaSpeech и другие ресурсы сопоставлены по объёму, числу дикторов, типу речи, стандарту транскрипции и условиям использования, а межкорпусное обобщение оценено по опубликованным результатам экспериментов. Установлено, что из 265,75 часа, записанных в узбекском сегменте Common Voice, валидацию прошли лишь 101,01 часа; модель, обученная на 52,09 часа речи одного диктора, показала на других корпусах долю ошибок по словам (WER) выше 76 %; добавление USC в обучающую выборку снизило WER на тестовой выборке USC не менее чем в 2,5 раза. Показано, что учёт регистра и знаков препинания в ряде случаев изменяет WER более чем вдвое. Предложены практические рекомендации по подготовке и публикации речевых корпусов.*

***Ключевые слова:*** *Автоматическое распознавание речи, ASR, речевой корпус, узбекский язык, малоресурсные языки, Common Voice, доля ошибок по словам (WER), межкорпусное обобщение.*

## Kirish (Introduction)

Nutqni avtomatik aniqlash (ASR — Automatic Speech Recognition) tizimlari ovozli signalni matnga aylantirib, subtitrlash, ovozli yordamchilar, qo‘ng‘iroq markazlari va hujjatlashtirish kabi ko‘plab sohalarda qo‘llanilmoqda. Bunday

tizimlarning sifati nafaqat tanlangan arxitekturaga, balki o‘qitish va baholash uchun foydalaniladigan transkripsiyalangan ovozli ma’lumotlarning hajmi hamda xilma-xilligiga ham bog‘liq. Masalan, Whisper modeli 680 000 soatlik zaif nazorat ostidagi (weakly supervised) ma’lumotlarda o‘qitilgan bo‘lib, shundan 117 000 soati ingliz tilidan boshqa 96 ta tilga to‘g‘ri keladi, aksariyat tillar uchun esa ma’lumotlar hajmi 1000 soatdan kam [1]. Resursi cheklangan (under-resourced) tillar uchun ASR tizimlarini yaratishda ma’lumotlar yetishmasligi alohida muammo sifatida o‘rganilgan [2].

O‘zbek tili turkiy tillar oilasiga mansub bo‘lib, unda so‘zlashuvchilar soni turli manbalarda taxminan 31 mln dan 57 mln gacha deb baholanadi [3–5]. Shunga qaramay, o‘zbek tili ASR uchun resursi cheklangan til hisoblanadi: hajmi sezilarli bo‘lgan ochiq ovozli korpuslar faqat so‘nggi yillarda e’lon qilindi va ular tarkibi jihatidan bir-biridan farq qiladi. Tilning agglyutinativ tuzilishi so‘z shakllari sonining keskin ortishiga olib keladi, dialektlar uchun esa transkripsiyalangan ma’lumotlar cheklangan va yagona imlo me’yori mavjud emas [6]. Bundan tashqari, o‘zbek tilida lotin va kirill yozuvlari birga qo‘llanadi, tutuq belgisi esa turli ko‘rinishlarda yoziladi; bu omillar transkripsiya sifatiga va ASR tizimlarini baholashga bevosita ta’sir qiladi.

So‘nggi yillarda ASR tizimlari an’anaviy DNN-HMM yondashuvidan to‘liq neyron (end-to-end) arxitekturalarga — Transformer [7] va Conformer [8] kabi modellarga o‘tdi. Nazoratsiz (self-supervised) oldindan o‘qitish usullari belgilangan ma’lumotlarga bo‘lgan ehtiyojni keskin kamaytirdi: wav2vec 2.0 modeli inglizcha LibriSpeech to‘plamida atigi 10 daqiqalik transkripsiyalangan nutq va 53 ming soatlik belgilanmagan audio yordamida toza va murakkab test qismlarida mos ravishda 4,8 va 8,2 % so‘zlar bo‘yicha xatolik darajasiga (WER — Word Error Rate) erishgan [9]. Biroq bu natija inglizcha to‘plamda olingan; o‘zbek tili uchun modelni sozlash (fine-tuning) ham, natijani ishonchli baholash ham sifatli transkripsiyalangan ovozli korpuslarni talab qiladi.

Muammoning dolzarbligi shundaki, o‘zbek tilidagi ovozli resurslar turli nashrlar va platformalar bo‘ylab tarqoq holda e’lon qilingan, ularning tavsifi (hajm, spikerlar, yozuv standarti, litsenziya) yagona shaklda berilmagan, e’lon qilingan xatolik ko‘rsatkichlari esa turli test to‘plamlarida va turlicha hisoblash qoidalari bilan olingan. Natijada tadqiqotchi uchun qaysi korpusni qaysi maqsadda tanlash,

ularni qanday birga ishlatish va natijalarni qanday solishtirish haqida aniq xulosa chiqarish qiyin. Mazkur maqola shu bo‘shliqni to‘ldirishga qaratilgan.

**Jumlalarni analitik tahlili**

O‘zbek tili uchun ASR tadqiqotlari uzoq vaqt kichik hajmli ma’lumotlarga tayanib kelgan: USC mualliflari avvalgi ishlarda geografik obyekt nomlaridan iborat 3500 ta yozuv va 10 soatlik o‘qilgan nutq kabi cheklangan to‘plamlar qo‘llanilganini, Speechocean korpusi esa juda qimmat yoki ommaviy foydalanish uchun mavjud emasligini qayd etgan [3]. 2021-yilda e’lon qilingan USC (Uzbek Speech Corpus) esa avvalgilariga nisbatan ancha yirik ochiq resurs bo‘ldi: u 104,9 soat va 958 spikerdan iborat bo‘lib, CC BY 4.0 litsenziyasi ostida akademik hamda tijoriy maqsadlarda foydalanishga ruxsat beradi. Mualliflar DNN-HMM hamda uchta end-to-end arxitekturani (LSTM, Transformer, Conformer) til modeli, tezlik buzilishi (speed perturbation) [10] va SpecAugment [11] usullari bilan sinab, test to‘plamida eng yaxshi natija sifatida 17,4 % WER ko‘rsatkichiga erishgan [3].

Common Voice — Mozilla tomonidan ommaviy yozuv (crowdsourcing) asosida shakllantiriladigan ko‘p tilli ochiq loyiha bo‘lib, unda foydalanuvchilar jumlalarni o‘qiydi, boshqa foydalanuvchilar esa yozuvlarni tinglab, ovoz berish orqali tasdiqlaydi [12]. Mukhamadiyev va hammualliflar Common Voice 8.0 versiyasida o‘zbek tili uchun 227 soatdan ortiq yozilgan va 80 soat tasdiqlangan material borligini ko‘rsatgan [6]; 23.0 va 25.0-versiyalarda esa yozib olingan material hajmi mos ravishda taxminan 266 va 265,75 soatni, tasdiqlangan qismi esa taxminan 100 va 101,01 soatni tashkil etadi [13, 14].

Baholash resurslari orasida FLEURS alohida o‘rin tutadi: u 102 tilda bir xil jumlalarning o‘qilgan nutqidan iborat bo‘lib, har bir til uchun o‘rtacha taxminan 12 soat nazorat ostidagi ma’lumotni o‘z ichiga oladi, o‘zbek tili ham shu tillar qatoriga kiradi [5]. FLEURS mualliflari o‘zbek tilini w2v-bert-51 va mSLAM modellarining oldindan o‘qitishida ishtirok etmagan (“unseen”) tillar qatoriga kiritgan [5].

Mukhamadiyev va hammualliflar 1281 spikerdan iborat, hajmi taxminan 207 soat bo‘lgan o‘z korpusini yaratib, unda CTC va diqqat mexanizmiga asoslangan transformer modeli 14,3 % WER ko‘rsatkichiga erishganini, DNN-HMM tizimi esa 17,9 % natija ko‘rsatganini, end-to-end model taxminan 60 soatdan keyin DNN-HMM tizimidan ustun kela boshlashini e’lon qilgan [6]. Ularning keyingi ishida o‘zbek tili uchun UzLM til modeli taklif etilgan; annotatsiyaga ko‘ra, turli manbalardagi 80 mln so‘zdan foydalanilgan va neyron tarmoqqa asoslangan til

modellari belgilar bo‘yicha xatolik darajasini (CER — Character Error Rate) 5,26 foizgacha pasaytirgan [15]. FeruzaSpeech esa bitta ayol spikerning 59,61 soatlik yozuvlaridan iborat bo‘lib, lotin va kirill yozuvidagi transkriptlarni bosh harf hamda tinish belgilari bilan saqlagan; mualliflar uni Common Voice va USC bilan birga ishlatib, korpuslararo tajribalarni o‘tkazgan [4].

Ma’lumotlar hajmi bilan xatolik darajasi o‘rtasidagi bog‘liqlik ko‘p tilli sharoitda ham o‘rganilgan: Whisper mualliflari FLEURS bo‘yicha WER logarifmi bilan tilga ajratilgan o‘qitish soatlari logarifmi o‘rtasida kuchli chiziqli bog‘liqlik ($r^2 = 0,83$) topgan va ma’lumotlar hajmi har 16 barobar oshganda WER taxminan ikki barobar kamayishini qayd etgan [1]. Qozoq tili uchun ochiq KSC korpusi taxminan 332 soat transkripsiyalangan nutqdan iborat [16], bu USC korpusidan taxminan 3 barobar katta (hisoblangan).

Shuningdek, Uzbekvoice.ai loyihasi 2023-yil mart oyida taxminan 1400 soatlik, matnlari bilan birga sifatli audio to‘plangani va ularning barchasi ommaviy foydalanish uchun ochiq ekanini e’lon qilgan [17]; biroq bu ma’lumot loyihaning o‘z e’loniga asoslangan bo‘lib, sifat, litsenziya va transkripsiya jihatidan mustaqil tasdiqlanmagan. Shunday qilib, mavjud adabiyotlarda o‘zbek tilidagi ovozli resurslar asosan alohida-alohida tavsiflangan; ularni yagona parametrlar bo‘yicha taqqoslash, tasdiqlangan ma’lumotlar ulushi va spiker xilma-xilligini baholash hamda korpuslararo umumlashtirish natijalarini tizimlashtirish masalasi yetarli yoritilmagan.

Tadqiqotning maqsadi — o‘zbek tilida nutqni avtomatik aniqlash uchun mavjud ochiq ovozli korpuslarni hajmi, tarkibi, yozuv standarti va foydalanish shartlari bo‘yicha qiyosiy tahlil qilish hamda ularning ASR natijalariga ta’sirini e’lon qilingan tajribalar asosida baholashdir. Maqsadga erishish uchun quyidagi vazifalar belgilandi:

- mavjud ovozli resurslarning asosiy xususiyatlarini (hajm, spikerlar, nutq turi, yozuv, litsenziya) aniqlash va taqqoslash;
- korpuslardagi tasdiqlangan ma’lumotlar ulushi va spiker xilma-xilligini baholash;
- e’lon qilingan korpuslararo tajribalar asosida korpus tarkibining ASR xatolik ko‘rsatkichlariga ta’sirini tahlil qilish;
- o‘zbek tilidagi ovozli korpuslarni tayyorlash va e’lon qilish bo‘yicha amaliy tavsiyalar ishlab chiqish.

## Material va metodlar (Methods)

### Ma’lumot manbalari

Tadqiqot hujjatlar tahlili (document analysis) usulida o‘tkazildi: yangi tajribalar o‘tkazilmadi, tahlil faqat ochiq manbalarda e’lon qilingan ma’lumotlarga asoslandi. Manbalar to‘rt guruhga ajratildi: (1) o‘zbek tilidagi ovozli korpuslarni tavsiflovchi ilmiy nashrlar (USC, FeruzaSpeech va Mukhamadiyev va hammualliflarning ishlari); (2) Mozilla Data Collective platformasidagi Common Voice o‘zbek to‘plamining ma’lumotlar varaqlari (datasheet) — 23.0 va 25.0-versiyalar; (3) ko‘p tilli benchmark va modellarga bag‘ishlangan ishlar (FLEURS, Whisper, wav2vec 2.0); (4) loyiha e’lonlari, xususan Uzbekvoice.ai loyihasining forumdagi xabari — bu manba tekshirilmagan ma’lumot sifatida alohida belgilandi. Qidiruv ilmiy nashrlar, nashriyotlar sahifalari va korpuslar joylashtirilgan platformalar bo‘yicha o‘tkazildi; manbalarga murojaat sanasi — 2026-yil 24–26-oktabr. Common Voice bo‘yicha ko‘rsatkichlar, agar boshqacha ko‘rsatilmasa, 25.0-versiyaga (cv-corpus-25.0-2026-03-09) tegishli.

Tahlilga o‘zbek tilidagi nutq va uning matnli transkripsiyasini o‘z ichiga olgan, tavsifi ochiq e’lon qilingan resurslar kiritildi. Tijoriy yoki ommaviy foydalanish uchun mavjud bo‘lmagan korpuslar (masalan, USC mualliflari juda qimmat yoki ommaviy foydalanish uchun mavjud emas deb ta’riflagan Speechocean korpusi) qiyosiy jadvalga kiritilmadi; ochiqligi aniqlanmagan resurslar esa jadvalda shu holat ko‘rsatilgan holda berildi. Manbalar bir-birini to‘ldiradi: ilmiy nashrlar korpuslarni yaratish usullari va ASR natijalarini, ma’lumotlar varaqlari esa joriy statistik ko‘rsatkichlarni beradi.

### Tahlil bosqichlari

Tahlil davomida har bir korpus bir xil parametrlar to‘plami bo‘yicha tavsiflandi, bu esa turli manbalardagi ma’lumotlarni solishtirish imkonini beradi. Tahlilning barcha bosqichlari 1-jadvalda umumlashtirilgan.

Research Science and Innovation House

**1-jadval. Tahlil bosqichlari**

| | Bosqich | Mazmuni | Maqsadi |
|---|---|---|---|
| | Manbalarni qidirish | Ilmiy nashrlar, ma’lumotlar varaqlari va loyiha sahifalaridan o‘zbek tilidagi ovozli resurslarni qidirish | O‘zbek tili ASR resurslarining dastlabki to‘liq ro‘yxatini shakllantirish |
| | Tanlash | Ochiq e’lon qilingan va tavsifi mavjud korpuslarni tijoriy hamda tavsifsiz resurslardan ajratish | Tahlil obyektini aniqlash |
| | Xususiyatlarni ajratib olish | Hajm, spikerlar, nutq turi, yozuv va litsenziya ma’lumotlarini yagona jadvalga yig‘ish | Korpuslarni bir xil parametrlar bo‘yicha taqqoslash |
| | Hisoblash | O‘rtacha yozuv uzunligi, tasdiqlangan ulush va nisbiy o‘zgarishlarni e’lon qilingan sonlardan hisoblash | Manbada bevosita berilmagan ko‘rsatkichlarni aniqlash |
| | Natijalarni qiyoslash | E’lon qilingan CER va WER qiymatlarini test to‘plami hamda hisoblash usuli bilan birga qayd etish | Taqqoslash sharoitlarining mosligini baholash |
| | Korpuslararo tahlil | Modellar qaysi korpusda o‘qitilib, qaysilarida sinalganini va hisoblash usulini solishtirish | Korpus tarkibining umumlashtirishga ta’sirini aniqlash |
| | Tavsiyalar ishlab chiqish | Aniqlangan muammolardan korpus yaratish va e’lon qilish bo‘yicha ko‘rsatmalar chiqarish | Amaliy ko‘rsatmalar shakllantirish |

Hisoblash bosqichi alohida e’tiborga loyiq. Manbalarda ba’zi ko‘rsatkichlar bevosita keltirilmagan (masalan, o‘rtacha yozuv uzunligi, tasdiqlangan ulush, WER qiymatining nisbiy o‘zgarishi). Bunday qiymatlar e’lon qilingan sonlardan hisoblab chiqildi va matnda “hisoblangan” deb belgilandi; manbalarda keltirilgan WER va

CER qiymatlari esa qayta hisoblanmadi va o‘zgartirilmadi. Manbalar o‘rtasidagi nomuvofiqliklar (masalan, FeruzaSpeech umumiy hajmi 59,61 soat deb ko‘rsatilgan bo‘lsa-da, o‘qitish, ishlab chiqish va test qismlari yig‘indisi 59,10 soatni tashkil etadi) yashirilmadi, balki natijalar bayonida alohida qayd etildi.

**Baholash mezonlari va tahlil usullari**

Korpuslar va ularning ASR tizimlariga ta’siri quyidagi mezonlar bo‘yicha baholandi:

• hajm va tarkib: yozib olingan va tasdiqlangan soatlar, yozuvlar soni, o‘rtacha yozuv uzunligi, lug‘at boyligi;

• spiker xilma-xilligi: spikerlar soni, jins va yosh bo‘yicha ma’lumotlarning to‘liqligi, o‘qitish va sinov to‘plamlarida spikerlarning kesishmasligi;

• transkripsiya va yozuv standarti: alifbo (lotin yoki kirill), bosh harf va tinish belgilarining saqlanishi, tutuq belgisining ishlatilishi;

• foydalanish shartlari: litsenziya va qayta tarqatishdagi cheklovlar.

Ushbu tadqiqotda quyidagi usullardan foydalanildi:

• qiyosiy tahlil — resurslar yagona parametrlar to‘plami bo‘yicha jadvalga joylashtirildi;

• miqdoriy hisob-kitoblar — o‘rtacha yozuv uzunligi (jami soat / yozuvlar soni), tasdiqlangan ulush va WER qiymatining nisbiy o‘zgarishi e’lon qilingan sonlardan hisoblandi;

• korpuslararo natijalar tahlili — modellar qaysi korpusda o‘qitilib, qaysilarida sinalgani, shuningdek hisoblash usuli (bosh harf va tinish belgilari bilan yoki ularsiz) bo‘yicha taqqoslandi;

• manbalar o‘rtasidagi nomuvofiqliklarni aniqlash — bir xil kattalik uchun turli manbalardagi qiymatlar (masalan, o‘zbek tilida so‘zlashuvchilar soni yoki FeruzaSpeech umumiy hajmi) o‘zaro solishtirildi.

Belgilangan mezonlar bir-birini to‘ldiradi. Hajm va tarkib ko‘rsatkichlari ma’lumotlarning miqdoriy yetarliligini, spiker xilma-xilligi modelning yangi ovozlarga umumlashtirish imkoniyatini, yozuv standarti transkripsiyaning izchilligi va natijalarning solishtirilishini, foydalanish shartlari esa resursdan amalda foydalanish mumkinligini aks ettiradi.

**Natijalar (Results)**

Tahlil natijasida o‘zbek tili uchun to‘rtta yirik o‘qitish resursi (USC, Common Voice, FeruzaSpeech va Mukhamadiyev va hammualliflar korpusi) hamda FLEURS

benchmarki ko'rib chiqildi. Resurslarning asosiy xususiyatlari 2-jadvalda, korpuslararo tajribalar natijalari esa 3-jadvalda keltirilgan. Natijalarni quyidagi oltita guruh bo'yicha bayon etamiz:

**1. Hajm va tasdiqlangan ulush (Data volume vs validated share):** USC jami 104,9 soatni (108 387 ta yozuv) tashkil etadi va o'qitish, validatsiya hamda test to'plamlariga mos ravishda 96,4, 4,0 va 4,5 soatdan ajratilgan; korpus 618,6 ming so'z va 63,1 ming unikal so'zdan iborat (unikal so'zlar ulushi taxminan 10,2 %, hisoblangan) [3]. Hisoblashlarga ko'ra, USC korpusida o'rtacha yozuv uzunligi taxminan 3,5 soniya (bitta yozuvda o'rtacha 5,7 so'z) [3], Common Voice to'plamida esa taxminan 4,2 soniya [14]; FeruzaSpeech korpusida o'rtacha segment uzunligi manbada 16,39 soniya (eng qisqasi 3,78, eng uzuni 50,69 soniya) deb ko'rsatilgan [4]. Common Voice 25.0 versiyasida 230 040 ta yozuv jami 265,75 soatni tashkil etadi, biroq ulardan faqat 87 442 tasi (38,0 %; 101,01 soat) tasdiqlangan, 14 158 tasi (6,2 %) rad etilgan, 128 440 tasi (55,8 %) esa hali hal qilinmagan (unresolved) [14]; foizlar yozuvlar sonidan hisoblangan. Qaror chiqarilgan yozuvlar orasida tasdiqlanganlar ulushi 86,1 foizni tashkil etadi (hisoblangan). Matn tomonida ham shunga o'xshash holat kuzatiladi: 286 626 ta jumlaning 148 738 tasi (51,9 %, hisoblangan) tasdiqlangan, 137 871 tasi ko'rib chiqilishini kutmoqda [14]. 23.0-versiyadan 25.0-versiyagacha yozuvlar soni atigi 335 taga (229 705 dan 230 040 gacha, hisoblangan) ko'paygan [13, 14]. FeruzaSpeech uchun umumiy hajm 59,61 soat deb ko'rsatilgan, biroq o'qitish (52,09), ishlab chiqish (2,93) va test (4,08 soat) qismlari yig'indisi 59,10 soatni tashkil etadi; farq sababi ko'rib chiqilgan matnda izohlanmagan [4].

**2. Spiker xilma-xilligi va metadata (Speaker diversity):** USC korpusida 958 ta spiker bor (o'qitish, validatsiya va test to'plamlarida mos ravishda 879, 41 va 38 nafar), spikerlar to'plamlar orasida kesishmaydi [3]. Common Voice 25.0 versiyasida spikerlar soni 2299 ta; erkak-maskulin toifasidagi 631 nafar spiker (27,4 %) yozuvlarning 44,2 foizini (101 757 ta) bergan [14]. 23.0-versiyada yozuvlarning 41 foizida jins va yosh ko'rsatilmagan, 44 foizi erkak, 15 foizi ayol spikerlarga tegishli; yoshi ko'rsatilgan yozuvlarning taxminan 97 foizi (yaxlitlangan foizlardan hisoblangan) o'smirlar (teens) va yigirma yoshlilar (twenties) toifalariga to'g'ri keladi [13]. FeruzaSpeech esa Toshkentlik yagona ayol spikerning ovozidan iborat [4]; Mukhamadiyev va hammualliflar korpusida 1281 spiker qatnashgan [6].

Manbalar o‘rtasidagi nomuvofiqlik o‘zbek tilida so‘zlashuvchilar soniga ham tegishli: USC maqolasida 35 mln dan ortiq, FeruzaSpeech maqolasida 31 mln dan ortiq, FLEURS maqolasida esa 57 mln deb ko‘rsatilgan [3–5].

**3. Transkripsiya va yozuv standarti (Transcription standard):** USC transkriptlari lotin alifbosida (29 harf va tutuq belgisi) bo‘lib, bosh harf va tinish belgilarisiz me’yorlashtirilgan [3]. FeruzaSpeech esa lotin va kirill transkriptlarini bosh harf hamda tinish belgilari bilan saqlagan; mualliflar kirilldan lotinga avtomatik o‘tkazishda yumshatish belgisi (ь) yo‘qolishi yoki noto‘g‘ri ravishda boshqa belgiga aylanib qolishini aniqlagan [4]. Natijada bir xil matnga turli me’yorlar qo‘llanadi, ASR tizimining chiqishi va baholash natijasi ham shunga mos ravishda farq qiladi (3-jadvalga qarang). Resurslarning umumlashtirilgan tavsifi 2-jadvalda keltirilgan.

**2-jadval. O‘zbek tilidagi ovozli resurslarning qiyosiy xususiyatlari**

| Korpus | Hajm | Spikerlar | Nutq turi va yozuv | Foydalanish shartlari |
|---|---|---|---|---|
| USC [3] | 104,9 soat; 108 387 ta yozuv; 618,6 ming so‘z | 958 ta; to‘plamlar orasida kesishmaydi | O‘qilgan nutq: ommaviy yozuv (Telegram-bot) va 20 ta diktorning audiokitob parchalari; lotin yozuvi, bosh harf va tinish belgilarisiz; WAV | CC BY 4.0; akademik va tijoriy foydalanish |
| Common Voice 25.0 [14] | 265,75 soat yozilgan, 101,01 soati tasdiqlangan; 230 040 ta yozuv | 2299 ta | O‘qilgan nutq: ommaviy yozuv va ovoz berish orqali tasdiqlash; MP3 | CC0-1.0 (varaqqa ko‘ra); qayta joylashtirish, ulashish va spikerni aniqlash taqiqlangan |

| Korpus | Hajm | Spikerlar | Nutq turi va yozuv | Foydalanish shartlari |
|---|---|---|---|---|
| FeruzaSpeech [4] | 59,61 soat (qismlar yig'indisi 59,10 soat); o'rtacha segment 16,39 soniya | 1 ta (ayol, Toshkent) | Diktor ovozi: badiiy asar (21,57 soat) va BBC Uzbek matnlari (38,04 soat); lotin va kirill transkriptlari, bosh harf va tinish belgilari saqlangan; 16 kHz WAV | Akademik tadqiqotlar uchun bepul |
| Mukhamadiyev va boshq. (2022) [6] | taxminan 207 soat (+52 soat audiokitob); 180 810 ta yozuv | 1281 ta | Suhbat 63 %, hisobot (reports) 18 %, intervyu 19 % | Ommaviy foydalanish imkoniyati aniqlanmadi |

**4. Korpuslararo umumlashtirish (Cross-corpus generalization):** FeruzaSpeech mualliflari RNN-T Conformer modelini Common Voice (CV, 54,88 soat), FeruzaSpeech (FS, 52,09 soat) va USC (90,70 soat) korpuslarida hamda ularning birikmalarida o'qitib, uchala korpusning test to'plamlarida (cv-test, fs-test, usc-test) sinagan [4]; natijalar 3-jadvalda berilgan. Birinchidan, deyarli teng hajmdagi ma'lumotlar turlicha umumlashtirish bergan: faqat FeruzaSpeech korpusida o'qitilgan model o'z test to'plamida 5,50 % WER ko'rsatgan, biroq cv-test va usc-test to'plamlarida mos ravishda 86,26 % va 76,24 % xatolikka yo'l qo'ygan; faqat Common Voice korpusida o'qitilgan model esa o'z test to'plamida 20,16 %, fs-test to'plamida 19,34 %, usc-test to'plamida 34,03 % natija bergan (bosh harf va tinish belgilarisiz, modified beam search). Ikkinchidan, o'quv to'plamiga FeruzaSpeech qo'shilishi cv-test bo'yicha WER qiymatini 20,16 foizdan 18,33 foizgacha (1,83 foiz punkti, f.p.; hisoblangan), USC korpusining qo'shilishi esa uni yana 11,17 foizgacha (7,16 f.p., hisoblangan) tushirgan; usc-test bo'yicha WER 29,67 foizdan 11,67 foizgacha (18,00 f.p., hisoblangan) kamaygan. Bu yaxshilanishlarga ma'lumot hajmining taxminan 107 soatdan (54,88 + 52,09;

hisoblangan) 197,68 soatgacha oshishi ham ta’sir qilgan bo‘lishi mumkin, ya’ni tarkib va hajm ta’sirini bu tajribalardan ajratib bo‘lmaydi. Uchinchidan, USC transkriptlari bosh harf va tinish belgilarisiz me’yorlashtirilgani sababli, USC korpusini o‘z ichiga olgan modelda usc-test bo‘yicha ikkala hisoblash usuli bir xil natija (11,67 %) bergan. Shuni ham qayd etish kerakki, USC korpusining o‘qitish qismi USC maqolasida 96,4 soat deb ko‘rsatilgan bo‘lsa [3], FeruzaSpeech tajribalarida 90,70 soat sifatida berilgan; buning sababini ko‘rib chiqilgan matndan aniqlab bo‘lmadi.

**3-jadval. Korpuslararo tajribalarda WER, % (“UNP / C&P”)**

| O‘qitish to‘plami | cv-test | fs-test | usc-test |
|---|---|---|---|
| CV (54,88 soat) | **20,16 / 31,98** | 19,34 / 31,88 | 34,03 / 51,61 |
| FS (52,09 soat) | 86,26 / 89,10 | **5,50 / 11,25** | 76,24 / 85,22 |
| CV + FS (54,88 + 52,09 soat) | **18,33 / 30,47** | **4,24 / 9,85** | 29,67 / 48,6 |
| CV + FS + USC (197,68 soat) | **11,17 / 27,81** | **4,05 / 9,56** | **11,67 / 11,67** |

Izoh. Dekodlash usuli — modified beam search. Har bir katakda “UNP / C&P” tartibida berilgan: UNP — barcha harflar bosh harfga o‘tkazilib, tinish belgilarisiz hisoblangan WER; C&P — bosh harf va tinish belgilari hisobga olingan WER. Qalin shriftda modelning o‘qitish to‘plamiga kirgan korpusning test qismi (in-domain) natijalari ajratilgan. Manba: [4] (7 va 8-jadvallar).

**5. Arxitektura va augmentatsiyaning ta’siri (Model and augmentation effects):** USC korpusida test to‘plamidagi eng yaxshi natijalar DNN-HMM uchun 23,5 %, E2E-LSTM uchun 27,0 %, E2E-Transformer uchun 21,0 % va E2E-Conformer uchun 17,4 % WER bo‘lgan (Conformer uchun CER 5,8 %) [3]. Til modelini qo‘shish WER qiymatini 7,7–12,6 f.p., tezlik buzilishi (SP) 0,8–5,1 f.p., SpecAugment (SA) esa 2,0–3,8 f.p. pasaytirgan (end-to-end modellar bo‘yicha test to‘plamida, hisoblangan); DNN-HMM uchun SA yaxshilanish bermagan: WER 23,5 foizdan 24,9 foizga oshgan. Conformer modelida barcha usullar birgalikda qo‘llanganda test WER 35,1 foizdan 17,4 foizgacha, ya’ni 17,7 f.p. (taxminan 50,4 % nisbiy) kamaygan (hisoblangan).

**6. Ko‘p tilli benchmarklarda o‘zbek tili (Uzbek in multilingual benchmarks):** FLEURS benchmarkida o‘zbek tili uchun belgilar bo‘yicha xatolik darajasi (CER) w2v-bert-51 modelida 16,8, mSLAM modelida 15,5 foizni tashkil etgan; bu qiymatlar barcha tillar bo‘yicha o‘rtacha ko‘rsatkichlardan (mos ravishda 14,1 % va 14,6 %) yuqori [5]. Oldindan o‘qitishda ishtirok etgan qardosh turkiy tillarda natija ancha yaxshi: qozoq tilida 6,5 va 5,7 %, qirg‘iz tilida 8,3 va 8,0 %, turk tilida 8,5 va 9,7 %; oldindan o‘qitishda qatnashmagan ozarbayjon tilida esa 12,7 va 15,9 % [5]. Hisoblashlarga ko‘ra, o‘zbek tilining CER ko‘rsatkichi ishtirok etgan qardosh tillarnikidan taxminan 1,6–2,7 barobar yuqori.

**Muhokama (discussion)**

**Natijalarning talqini**

Olingan natijalar o‘zbek tilidagi ASR tizimlari uchun korpusning foydaliligi yagona ko‘rsatkich — soatlar soni bilan emas, o‘zaro bog‘liq bir necha omil majmui bilan belgilanishini ko‘rsatadi. Birinchidan, nominal hajm va foydalanish mumkin bo‘lgan hajm bir xil emas: Common Voice to‘plamida yozib olingan 265,75 soatning atigi 101,01 soati tasdiqlangan, 128 440 ta yozuv esa hali hal qilinmagan. Bu validatsiya jarayoni o‘zbek segmentida asosiy to‘siqlardan biri ekanini, ya’ni mavjud resursning katta qismi hozircha o‘qitishda ishlatilmayotganini bildiradi. 23.0-versiyadan 25.0-versiyagacha yozuvlar soni deyarli o‘zgarmagani (335 ta) yangi yozuvlar qo‘shilishi sezilarli bo‘lmaganini ko‘rsatadi, shu sababli hajmni oshirishning eng arzon yo‘li — mavjud yozuvlarni ko‘rib chiqishdir. Agar hal qilinmagan yozuvlarning ko‘rib chiqilish natijasi hozirgi ulushga (86,1 %) yaqin bo‘lsa, bu o‘qitish uchun yaroqli hajmni sezilarli oshirishi mumkin; bu faqat faraz bo‘lib, uni tekshirish alohida tadqiqotni talab qiladi.

Ikkinchidan, hajmning o‘zi umumlashtirishni kafolatlamaydi. Common Voice (54,88 soat) va FeruzaSpeech (52,09 soat) deyarli bir xil hajmdagi o‘qitish ma’lumotini bergan bo‘lsa-da, birinchisida o‘qitilgan model boshqa korpuslarda 19,34 % va 34,03 % WER, ikkinchisida o‘qitilgani esa 86,26 % va 76,24 % WER ko‘rsatgan. Bunday farq spiker xilma-xilligi (ko‘p spikerli va bitta spikerli korpus), matn mavzulari, yozuv sharoitlari va segment uzunligi (taxminan 4,2 va 16,39 soniya) bilan bog‘liq bo‘lishi mumkin; mavjud tajribalar bu omillarni bir-biridan ajratmaydi, shu sababli har birining hissasini aniqlash uchun maxsus nazoratli tajribalar kerak. Shuningdek, usc-test bo‘yicha yaxshilanishga USC

korpusining o‘qitish qismi usc-test bilan bir manbadan (in-domain) olinganligi ham ta’sir qilgan bo‘lishi mumkin.

Uchinchidan, WER qiymati faqat modelga emas, transkripsiya me’yoriga ham bog‘liq: bir xil model uchun bosh harf va tinish belgilarini hisobga olish natijani 2 baravardan ortiq o‘zgartirishi mumkin (cv-test: 27,81 % va 11,17 %). Demak, turli ishlarda e’lon qilingan raqamlarni solishtirish uchun hisoblash me’yori ham ko‘rsatilishi shart. Masalan, FeruzaSpeech mualliflari o‘zlarining usc-test bo‘yicha 11,67 % natijasini USC mualliflari e’lon qilgan 17,4 % bilan solishtirib, 5,73 f.p. yaxshilanishni qayd etgan [4]; biroq modellar turlicha (RNN-T va CTC/diqqat asosidagi tizimlar), o‘quv ma’lumotlari hajmi esa (197,68 soat) USC korpusidagi 96,4 soatdan taxminan 2 barobar katta, shu sababli bu taqqoslash to‘liq mos emas.

To‘rtinchidan, FLEURS natijalari o‘zbek tilining yirik ko‘p tilli modellarda hali yetarli darajada qamrab olinmaganini ko‘rsatadi: oldindan o‘qitishda qatnashmagan o‘zbek tilida CER (16,8 va 15,5 %) o‘rtacha qiymatlardan yuqori, ishtirok etgan qardosh turkiy tillarda esa ancha past. Bunga faqat ma’lumot hajmi emas, til oldindan o‘qitishda ishtirok etgan-etmaganligi ham ta’sir qilishi mumkin. Beshinchidan, tahlil qilingan ochiq korpuslar (USC, Common Voice, FeruzaSpeech) asosan o‘qilgan matnlarga tayanadi; tabiiy suhbat va intervyu nutqini o‘z ichiga olgan korpusning (Mukhamadiyev va hammualliflar) ochiqligi esa aniqlanmadi. Shu sababli real sharoitdagi (suhbat, shovqin, dialekt) ishlash sifati ochiq resurslarda kam baholanadi. Aniqlangan muammolar, ularning ASR tizimlariga ta’siri va taklif etilayotgan yechimlar 4-jadvalda umumlashtirilgan.

**4-jadval. Ovozli korpuslar sifatiga ta’sir etuvchi muammolar va ularning yechimlari**

| Muammo | ASR tizimlariga ta’siri | Yechim |
|---|---|---|
| Tasdiqlanmagan yozuvlarning yuqori ulushi | Common Voice to‘plamida 265,75 soatning atigi 101,01 soati o‘qitishga yaroqli; 128 440 yozuv hal qilinmagan | Validatsiyani tezlashtirish; tasdiqlangan va hal qilinmagan yozuvlar ulushini muntazam e’lon qilish |
| Spiker xilma-xilligining yetishmasligi | Bitta spikerli korpusda o‘qitilgan model boshqa korpuslarda 76 foizdan yuqori WER bergan | Bunday korpuslarni mustaqil emas, ko‘p spikerli korpuslar bilan birga ishlatish |

| Muammo | ASR tizimlariga ta’siri | Yechim |
|---|---|---|
| Spiker metadatasining to‘liq emasligi va yosh–jins bo‘yicha siljish | Yozuvlarning 41 foizida jins va yosh ko‘rsatilmagan; yoshi ko‘rsatilganlarning taxminan 97 foizi ikki yosh toifasiga (teens, twenties) to‘g‘ri keladi; guruhlar bo‘yicha sifatni baholash qiyin | Metadatani anonim holda to‘ldirishni rag‘batlantirish; yosh va jins bo‘yicha muvozanatli yig‘ish |
| Transkripsiya me’yorlarining farqi | Bir xil modelning WER qiymati 2 baravardan ortiq o‘zgarishi mumkin (27,81 % va 11,17 %) | WER qiymatini ikki usulda — xom (C&P) va me’yorlashtirilgan (UNP) — e’lon qilish; me’yorlash qoidalarini hujjatlashtirish |
| Lotin va kirill yozuvi, tutuq belgisi variantlari | Kirilldan lotinga o‘tkazishda ь belgisi yo‘qolishi yoki buzilishi; bir xil so‘zning bir necha yozilish varianti | Hujjatlashtirilgan o‘tkazish qoidalari; ikkala yozuvdagi transkriptlar; yagona tutuq belgisi |
| Standart bo‘lingishlar va yagona test to‘plamlarining yo‘qligi | Turli ishlarda e’lon qilingan natijalarni to‘g‘ridan-to‘g‘ri solishtirib bo‘lmaydi | Spikerlari kesishmaydigan yagona o‘qitish, validatsiya va test bo‘lingishlarini e’lon qilish (USC kabi) |
| O‘qilgan nutqning ustunligi | Tabiiy suhbat va dialektlar kam qamrab olingan; real sharoitdagi sifat noaniq | Spontan nutq va dialektlarni o‘z ichiga olgan ochiq korpus yaratish |

**Mavjud tadqiqotlar bilan taqqoslash**

Natijalarimiz xalqaro tajriba bilan mos keladi. Whisper mualliflari ko‘p tilli sharoitda FLEURS bo‘yicha WER va tilga ajratilgan o‘qitish soatlari orasida logarifmik shkalada kuchli chiziqli bog‘liqlik ($r^2 = 0,83$) topgan va ma’lumotlar hajmi har 16 barobar oshganda WER taxminan ikki barobar kamayishini qayd etgan [1]. O‘zbek tili uchun bu bog‘liqlik alohida tekshirilmadi: Whisper modelining

o‘zbek tiliga xos o‘qitish soatlari va natijalari ushbu tadqiqotda ko‘rib chiqilmagan. Shunga qaramay, bizning tahlil ushbu bog‘liqlikka sifat jihatini — tasdiqlangan ulush, spiker xilma-xilligi va transkripsiya izchilligini — qo‘shadi: deyarli bir xil hajm turlicha umumlashtirish bergan.

USC bo‘yicha va Mukhamadiyev hamda hammualliflarning natijalari ham o‘zaro mos: USC korpusida end-to-end Transformer (21,0 %) va Conformer (17,4 %) DNN-HMM tizimidan (23,5 %) yaxshi natija bergan, E2E-LSTM (27,0 %) esa undan past bo‘lgan [3]; Mukhamadiyev va hammualliflar end-to-end modelning DNN-HMM tizimidan ustunligi taxminan 60 soatdan keyin namoyon bo‘lishini qayd etgan [6]. Ikkala ishda ham eng yaxshi end-to-end modellar DNN-HMM tizimidan yuqori natija bergan; biroq bu ishlar turli test to‘plamlarida va turli tarkibdagi korpuslarda o‘tkazilgani sababli ularning WER qiymatlari (17,4 va 14,3 %) to‘g‘ridan-to‘g‘ri solishtirilmaydi.

Qardosh qozoq tili bilan solishtirish ham ibratli. Qozoq tili uchun ochiq KSC korpusi taxminan 332 soat bo‘lib [16], USC korpusidan taxminan 3 barobar katta (hisoblangan); FLEURS benchmarkida esa qozoq tilining CER ko‘rsatkichi o‘zbek tilinikidan taxminan 2,6–2,7 barobar past (6,5 va 5,7 foizga nisbatan 16,8 va 15,5 foiz; hisoblangan) [5]. Bu farq faqat ochiq korpus hajmi bilan izohlanmaydi — qozoq tili oldindan o‘qitish tillari qatoriga kirgan, o‘zbek tili esa kirmagan — lekin u o‘zbek tili uchun ochiq va yirik resurslarni kengaytirish dolzarbligini ko‘rsatadi.

Nazoratsiz oldindan o‘qitish yondashuvlari kam belgilangan ma’lumot sharoitida istiqbolli ekani wav2vec 2.0 misolida ko‘rsatilgan: inglizcha LibriSpeech to‘plamida atigi 10 daqiqalik transkripsiyalangan nutq bilan 4,8 va 8,2 % WER olingan [9]. O‘zbek tili uchun bunday yondashuvlarning natijalari ushbu tahlil doirasida o‘rganilmadi; ular o‘zbek tilida mavjud, ammo tasdiqlanmagan yoki belgilanmagan audio materiallardan foydalanish imkonini berishi mumkin, bu esa kelgusi tadqiqotlar uchun yo‘nalish hisoblanadi.

**Tadqiqotning cheklovlari**

Tadqiqot natijalarini sharhlashda quyidagi cheklovlarni hisobga olish lozim. Birinchidan, tahlil ikkilamchi ma’lumotlarga — e’lon qilingan maqolalar va ma’lumotlar varaqlariga — asoslangan; o‘z tajribalarimiz o‘tkazilmagan, shu sababli korpuslarning ASR sifatiga ta’siri haqidagi xulosalar manbalardagi tajribalar bilan cheklangan. Ikkinchidan, turli ishlardagi test to‘plamlari, modellar va hisoblash me’yorlari bir xil emas; shu sababli ularning WER qiymatlari bir-biri bilan

to‘g‘ridan-to‘g‘ri solishtirilmadi va korpuslararo xulosalar asosan bitta ishdagi (FeruzaSpeech) nazoratli sharoitga tayanadi. Uchinchidan, Uzbekvoice.ai loyihasining 1400 soatlik ma’lumoti loyihaning o‘z e’loniga asoslangan va mustaqil tasdiqlanmagan, Mukhamadiyev va hammualliflar korpusining ommaviy mavjudligi esa aniqlanmadi; bu resurslar umumiy xulosalarga kiritilmadi. To‘rtinchidan, Common Voice ko‘rsatkichlari ma’lumotlar varaqlarining 2026-yil 24-oktabrdagi nusxasiga (25.0 va 23.0-versiyalar) asoslangan; loyihaning yangi versiyalari muntazam chiqarilgani uchun bu qiymatlar o‘zgargan bo‘lishi mumkin. Beshinchidan, Whisper kabi yirik ko‘p tilli modellarning o‘zbek tiliga xos o‘qitish soatlari va natijalari ushbu tadqiqotda ko‘rib chiqilmadi. Oltinchidan, qidiruv asosan ingliz tilidagi manbalarda o‘tkazilgan, shu sababli mahalliy nashrlar to‘liq qamrab olinmagan bo‘lishi mumkin.

**Amaliy tavsiyalar**

O‘tkazilgan tahlil asosida o‘zbek tilidagi ovozli korpuslarni tayyorlash va e’lon qilish bo‘yicha quyidagi tavsiyalar ishlab chiqildi (ular mualliflarning takliflari bo‘lib, alohida tajribada sinalmagan):

- har bir korpus uchun spikerlari kesishmaydigan yagona o‘qitish, validatsiya va test bo‘lingishlarini e’lon qilish;
- transkriptlarni lotin va kirill yozuvlarida, kirilldan lotinga o‘tkazish qoidalarini hujjatlashtirgan holda taqdim etish; tutuq belgisi uchun yagona belgini qo‘llash;
- WER qiymatini ikki usulda — bosh harf va tinish belgilari bilan (xom) hamda ularsiz (me’yorlashtirilgan) — e’lon qilish va me’yorlash qoidalarini ko‘rsatish;
- Common Voice to‘plamida yozuvlarni tasdiqlash jarayonini tezlashtirish, tasdiqlangan va hal qilinmagan yozuvlar ulushini muntazam e’lon qilish;
- spiker metadatasini (yosh guruhi, jins, dialekt) anonim holda to‘ldirishni rag‘batlantirish va yosh hamda jins bo‘yicha muvozanatni ta’minlash;
- o‘qilgan nutq bilan birga spontan nutq va dialektlarni qamrab oluvchi korpuslar yaratish;
- bitta spikerli korpuslardan mustaqil emas, boshqa ko‘p spikerli korpuslar bilan birga foydalanish;
- har bir korpus uchun manba matnlari, yozib olish sharoitlari, litsenziya va versiya haqidagi ma’lumotlarni hujjatlashtirish.

**Kelgusi tadqiqot yo‘nalishlari**

Kelgusi tadqiqotlarda bir xil arxitektura va bir xil hisoblash qoidalari bilan har bir korpusda o‘qitilgan modellarni umumiy test to‘plamlarida baholash, o‘qitish ma’lumotlarining hajmi va spiker xilma-xilligi ta’sirini alohida ajratib o‘rganish, Whisper kabi ko‘p tilli modellar va nazoratsiz oldindan o‘qitilgan modellarning o‘zbek tilidagi natijalarini yagona benchmarkda solishtirish hamda spontan nutq va dialektlar uchun ochiq korpuslar yaratish maqsadga muvofiqdir.

**Xulosa (Conclusion)**

Xulosa o‘rnida ta’kidlash joizki, o‘zbek tilida nutqni avtomatik aniqlash tizimlarining sifati faqat model arxitekturasiga emas, balki ovozli korpuslarning hajmi, tasdiqlangan ulushi, spiker xilma-xilligi va transkripsiya izchilligiga ham bog‘liq. Mavjud ochiq resurslar bir-birini to‘ldiradi: USC — spikerlari kesishmaydigan standart bo‘lingishli ko‘p spikerli to‘plam, Common Voice — spikerlari eng ko‘p, ammo tasdiqlanmagan yozuvlari ham ko‘p to‘plam, FeruzaSpeech esa tarkibi bir xil, lekin bitta spikerli korpus. Ularni birga ishlatish umumlashtirishni sezilarli oshirishi mumkin. Muxtasar qilib aytganda, o‘zbek tili uchun “ko‘proq soat” tamoyilidan ko‘ra “tasdiqlangan, xilma-xil va izchil transkripsiyalangan soat” tamoyili samaraliroq yo‘l bo‘lib ko‘rinadi, bu esa nafaqat ASR tizimlarining, balki o‘zbek tilining raqamli dunyodagi o‘rnini mustahkamlashga ham xizmat qiladi.

Tadqiqotning asosiy natijalari quyidagilardan iborat:

• ochiq o‘zbek korpuslarining hajmi o‘nlab–yuzlab soat bilan o‘lchanadi (USC — 104,9 soat; Common Voice 25.0 versiyasida 101,01 soat tasdiqlangan; FeruzaSpeech — 59,61 soat), ularning spiker tarkibi, nutq turi va yozuv standarti bir-biridan farq qiladi;

• Common Voice o‘zbek segmentida yozuvlarning 55,8 foizi hali hal qilinmagan, shu sababli validatsiya foydalanish mumkin bo‘lgan hajmni oshirishdagi asosiy zaxira hisoblanadi;

• deyarli teng soatdagi ma’lumotlar (54,88 va 52,09 soat) turlicha umumlashtirish bergan: bitta spikerli korpusda o‘qitilgan model boshqa korpuslarda 76 foizdan yuqori WER ko‘rsatgan, USC korpusining qo‘shilishi esa usc-test bo‘yicha WER qiymatini 29,67 foizdan 11,67 foizgacha tushirgan;

• bosh harf va tinish belgilarini hisobga olish WER qiymatini 2 baravardan ortiq o‘zgartirishi mumkin, shu sababli natijalarni solishtirish uchun standart bo‘lingishlar va yagona hisoblash me’yorlari zarur.

**Foydalanilgan adabiyotlar ro‘yxati:**

**1.** Radford A., Kim J. W., Xu T., et al. Robust Speech Recognition via Large-Scale Weak Supervision // Proceedings of the 40th International Conference on Machine Learning (ICML 2023). - 2023. - arXiv:2212.04356.

**2.** Besacier L., Barnard E., Karpov A., Schultz T. Automatic speech recognition for under-resourced languages: A survey // Speech Communication. - 2014. - Vol. 56. - P. 85–100.

**3.** Musaev M., Mussakhojayeva S., Khujayorov I., et al. USC: An Open-Source Uzbek Speech Corpus and Initial Speech Recognition Experiments // Speech and Computer: 23rd International Conference, SPECOM 2021. - 2021. - P. 437–447. - DOI: 10.1007/978-3-030-87802-3_40.

**4.** Povey A., Povey K. FeruzaSpeech: A 60 Hour Uzbek Read Speech Corpus with Punctuation, Casing, and Context // Proceedings of the International Conference on Natural Language and Speech Processing (ICNLSP 2024). - 2024. - P. 360–364. - arXiv:2410.00035.

**5.** Conneau A., Ma M., Khanuja S., et al. FLEURS: Few-shot Learning Evaluation of Universal Representations of Speech // 2022 IEEE Spoken Language Technology Workshop (SLT). - 2022. - arXiv:2205.12446.

**6.** Mukhamadiyev A., Khujayarov I., Djuraev O., Cho J. Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language // Sensors. - 2022. - Vol. 22, No. 10. - Art. 3683. - DOI: 10.3390/s22103683.

**7.** Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need // Advances in Neural Information Processing Systems 30 (NIPS 2017). - 2017. - P. 5998–6008.

**8.** Gulati A., Qin J., Chiu C.-C., et al. Conformer: Convolution-augmented Transformer for Speech Recognition // Proceedings of Interspeech 2020. - 2020. - P. 5036–5040.

**9.** Baevski A., Zhou Y., Mohamed A., Auli M. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // Advances in

Neural Information Processing Systems 33 (NeurIPS 2020). - 2020. - P. 12449–12460.

**10.** Ko T., Peddinti V., Povey D., Khudanpur S. Audio augmentation for speech recognition // Proceedings of Interspeech 2015. - 2015. - P. 3586–3589.

**11.** Park D. S., Chan W., Zhang Y., et al. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition // Proceedings of Interspeech 2019. - 2019. - P. 2613–2617.

**12.** Ardila R., Branson M., Davis K., et al. Common Voice: A Massively-Multilingual Speech Corpus // Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020). - 2020. - P. 4218–4222.

**13.** Mozilla Data Collective. Common Voice Scripted Speech 23.0 – Uzbek [Elektron resurs]. - 2025. - URL: https://datacollective.mozillafoundation.org/datasets/cmflnuzw7fd2f0vzknodytok4 (murojaat sanasi: 24.10.2026).

**14.** Mozilla Data Collective. Common Voice Scripted Speech 25.0 – Uzbek (cv-corpus-25.0-2026-03-09) [Elektron resurs]. - 2026. - URL: https://datacollective.mozillafoundation.org/datasets/cmn29fa4y0150o1073fqww73p (murojaat sanasi: 24.10.2026).

**15.** Mukhamadiyev A., Mukhiddinov M., Khujayarov I., Ochilov M., Cho J. Development of Language Models for Continuous Uzbek Speech Recognition System // Sensors. - 2023. - Vol. 23, No. 3. - Art. 1145. - DOI: 10.3390/s23031145.

**16.** Khassanov Y., Mussakhojayeva S., Mirzakhmetov A., et al. A Crowdsourced Open-Source Kazakh Speech Corpus and Initial Speech Recognition Baseline // Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021). - 2021. - P. 697–706. - arXiv:2009.10334.

**17.** gina. Creating Open and Accessible Resources for Uzbek: The Uzbekvoice.ai Project [Elektron resurs] // Mozilla Discourse. - 27.03.2023. - URL: https://discourse.mozilla.org/t/creating-open-and-accessible-resources-for-uzbek-the-uzbekvoice-ai-project/112634 (murojaat sanasi: 26.10.2026).