

DOI: <https://doi.org/10.5281/zenodo.17590662>

O‘ZBEK NUTQINI MATNGA AYLANTIRISH UCHUN MOSLASHTIRILGAN NUTQ TANISH MODELINI ISHLAB CHIQUISH

Ismatillayev Muzaffar

Toshkent Kimyo xalqaro universiteti, 2-kurs magistr; ismatilaevmuzaffar6@gmail.com

Annotatsiya - Mazkur maqolada o‘zbek tilidagi nutqni avtomatik tarzda matnga aylantirish tizimi ishlab chiqildi, sinovdan o‘tkazildi va tahlil qilindi. Tadqiqotda ikki manba bir ovozli o‘qib eshittirish va ko‘p so‘zlovchili kundalik nutqlar birlashtirilib, matnlar me‘yoriy lotin yozuviga keltirildi, audio yozuvlar 16 ming gerts namuna tezligiga moslashtirildi, uzunligi 30 soniya chegarasida boshqarildi. Ma‘lumotlar mashq (119 831 yozuv), tekshiruv (648 yozuv) va sinov (899 yozuv) to‘plamlariga qat‘iy ajratildi. Model ko‘p tilli yirik andozadan (Whisper-small) o‘zbek tili uchun moslashtirib o‘qitildi; isitib borish, og‘irlikni kamaytirish, gradientni cheklash va erta to‘xtatish choralaridan foydalanildi, aralash aniqlikdagi hisoblash orqali tezlik oshirildi. Baholashda so‘zdagi xatolik ulushi ko‘rsatkichi qo‘llanib, sinov to‘plamida 13,78 foiz natija olindi; sinov yo‘qotishi 0,1487, tekshiruv yo‘qotishi yakunda taxminan 0,15 bo‘ldi. Natijalar turli manbalarni birlashtirish, kiritmalarni yagona andozaga keltirish va ehtiyotkor moslashtirib o‘qitish orqali o‘zbek nutqini turli sharoitlarda barqaror tanish mumkinligini ko‘rsatadi. Ushbu yondashuv ta‘lim, tibbiyot, raqamli xizmatlar va ovozli muloqot talab qiluvchi dasturlarda amaliy qo‘llash imkonini kengaytiradi.

Kalit so‘zlar: avtomatik nutqni tanish; o‘zbek tili; ma‘lumotlarni tozalash; mel-spektr; moslashtirib o‘qitish; isitib borish; gradientni cheklash; erta to‘xtatish; aralash aniqlik; so‘zdagi xatolik ulushi

KIRISH

Hozirgi kunda sun‘iy intellekt va ovozli texnologiyalar sohasi jadal rivojlanib, inson-nazorat tizimlari o‘rniga ovoz orqali boshqaruv, avtomatik tarjima, yordamchi agentlar va raqamli assistentlar keng qo‘llanilmoqda. Nutqni avtomatik tarzda matnga aylantirish texnologiyasi ana shu jarayonlarning markazida turadi. Dunyo miqyosida bu yo‘nalishda Whisper, DeepSpeech, Wav2Vec 2.0 kabi yirik neyron arxitekturalar ishlab chiqilib, ular ingliz, xitoy, arab va koreys kabi tillarda yuqori aniqlikka erishgan. So‘nggi yillarda ko‘p tilli modellar ham keng rivojlanmoqda, ular kam resursli tillar uchun ham nutqni tanish imkonini kengaytirmoqda.

Shunga qaramay, o‘zbek tili uchun yaratilgan avtomatik nutqni tanish tizimlari hali ham son jihatdan cheklangan. O‘zbek tili fonetik va morfologik jihatdan boy, urg‘u tizimi va talaffuzdagi lahjalar xilma-xilligi sababli, mavjud xalqaro modellar to‘liq moslashmaydi. Shuningdek, o‘zbekcha nutqni o‘qitish uchun yetarli hajmdagi ma‘lumot to‘plamlari mavjud emas. Shu sababli, o‘zbek tili uchun alohida fonetik, leksik va morfologik xususiyatlarni hisobga olgan, aniq va barqaror ishlaydigan

avtomatik nutq tanish tizimini yaratish bugungi kunda dolzarb ilmiy vazifalardan biridir.

Mazkur maqolada o‘zbek nutqini avtomatik tarzda matnga aylantirish tizimini yaratish, sinovdan o‘tkazish va takomillashtirish jarayoni yoritiladi. Hozirgi kunda axborot texnologiyalari tez sur‘atlarda rivojlanib, ovozli boshqaruv va nutqni tanish tizimlariga bo‘lgan ehtiyoj ortib bormoqda. Biroq o‘zbek tili uchun bu yo‘nalishda yaratilgan modellar soni cheklangan bo‘lib, mavjud tizimlar ko‘pincha talaffuzdagi farqlar, lahjalar va nutqdagi tabiiy ohanglarni to‘liq hisobga olmaydi. Shu sababli ushbu ishning asosiy maqsadi o‘zbek tilining fonetik va leksik xususiyatlariga mos, aniq va barqaror ishlaydigan avtomatik nutq tanish modelini ishlab chiqishdan iboratdir.

ADABIYOTLAR SHARHI

So‘nggi yillarda nutqni avtomatik tarzda matnga aylantirish texnologiyalari jadal rivojlanib, ularning imkoniyatlari turli tillarda, jumladan o‘zbek tilida ham keng qo‘llanilmoqda. Ushbu bo‘limda o‘zbek tilida nutqni tanish sohasida olib borilgan dolzarb tadqiqotlar hamda ilg‘or xalqaro yondashuvlar tahlil qilinadi. O‘zbek tilida birinchi yirik ochiq ma‘lumotlar to‘plamlaridan biri sifatida FeruzaSpeech majmuasi taqdim etilgan bo‘lib, unda so‘zlovchining ovoz yozuvlari, tinish belgilari va yozuv me‘yorlariga mos matnlar jamlangan. Mazkur to‘plam o‘zbek nutqini aniqlik bilan o‘qitish uchun muhim manba bo‘lib xizmat qiladi [1]. Shuningdek, Uzbek Speech Corpus nomli yana bir ochiq to‘plam e‘lon qilingan bo‘lib, unda yuzlab so‘zlovchilar ovozi asosida dastlabki tajribalar o‘tkazilgan va o‘zbek tili uchun avtomatik nutqni tanishning amaliy asoslari yaratilgan [2]. Xalqaro miqyosda Whisper modeli [3] nutqni aniqlashda katta yutuqlarga erishdi. U yuzdan ortiq tillarda o‘qitilgan bo‘lib, shovqinli muhitda ham yuqori aniqlik bilan ishlashga qodir. Ushbu modelning kam resursli tillarga moslashtirilgan yengil variantlari, jumladan DistilWhisper [4] va LoRA-Whisper [5], o‘qitish samaradorligini oshirgan hamda parametrlar sonini kamaytirgan holda yuqori aniqlikni saqlab qolgan.

Kam resursli tillar uchun ko‘p tilli o‘qitish muammolari bo‘yicha olib borilgan tadqiqotlarda og‘irlikli xoch-entropiya (Weighted Cross-Entropy) yondashuvi tavsiya etilgan bo‘lib, u til resurslari cheklangan sharoitda modelning aniqligini sezilarli darajada oshirishi isbotlangan [6]. Bundan tashqari, Whisper modelidan avtomatik tarzda belgilar yaratib, cheklangan belgilangan ma‘lumotlar bilan ko‘rish asosidagi nutqni tanish tizimlarini o‘qitish samarali natija bergani ko‘rsatilgan [7]. Arab tiliga oid tajribalar ham kam resursli tillar uchun muhim qiyosiy natijalarni taqdim etdi. Xususan, N-Shot Benchmarking tadqiqotida [8] modelning turli sinov hajmlaridagi barqarorligi sinovdan o‘tkazilgan va past resursli tillarga umumlashtirish imkoniyati tasdiqlangan. O‘zbek tili uchun doimiy nutqni tanish tizimlari yaratish borasida ham ilmiy ishlar olib borilgan. Mazkur tadqiqotlarda til modellari, fonetik tahlil va so‘z segmentatsiyasi orqali aniqlikni oshirishga erishilgan [9]. Shuningdek, ko‘p tilli nutqni tanish tizimlarini o‘zbek tili uchun moslashtirish yo‘nalishida olib borilgan so‘nggi izlanishlar [10] Whisper modelini qayta o‘qitish orqali ancha yuqori aniqlikka erishish

mumkinligini ko'rsatdi. Bu esa o'zbek nutqini avtomatik tahlil qilish, subtitrlar yaratish va ovozli interfeyslar uchun yangi imkoniyatlar ochadi.

METODOLOGIYA

Har ikki to'plamdan olingan yozuvlar tekshirilib, buzilgan fayllar chiqarildi. Matnlar lotin yozuvi me'yorlariga moslashtirildi: ortiqcha belgilar, takrorlar va imlo xatolari tozalandi. Audio yozuvlar 16 000 gerts chastotaga o'tkazildi; uzunligi 30 soniyadan ortsa qisqartirildi, juda qisqa yozuvlar esa sukut segmentlari bilan to'ldirildi. Har bir audio va unga mos matn satri yagona manifest faylida JSONL formatida jamlandi. Ushbu standartlashtirish bosqichi ma'lumot sifatini oshirish bo'yicha ilgari keltirilgan usullar bilan uyg'un ishlaydi [11].

1-jadval. Ma'lumot hajmlari

To'plam turi	Yozuvlar soni	Tavsif
Mashq to'plami	119 831	Ikki manbadan birlashtirilgan qism
Tekshiruv to'plami	648	Barqaror bo'lak
Sinov to'plami	899	Mustaqil bo'lak
Jami	121 378	To'liq ishlov berilgan yozuvlar

Kiritma audio $x(t)$ dan mel-spektr tasviri olinadi. Chastota bo'yicha M ta mel kanali va vaqt bo'yicha T bo'linma hosil qilinib, kirish tenzori $X \in \mathbb{R}^{M \times T}$ yaratiladi. Matn chiquvi belgilar ketma-ketligi $y = (y_1, \dots, y_L)$ ko'rinishida ifodalanadi. Ketma-ketlikdan ketma-ketlikka o'qitiladigan dekode-asosli model quyidagi shartli ehtimollikni baholaydi [12]:

$$p_{\theta}(y | X) = \prod_{t=1}^L p_{\theta}(y_t | y_{<t}, X).$$

Til va vazifa ko'rsatmalari dekodega majburiy kirish sifatida berildi, keraksiz maxsus belgilarni bostirish yumshatildi, tinish belgilari va imlo me'yorlari izchil saqlandi. Arxitektura ko'p tilli variantdan o'zbek tiliga moslashtirildi.

Asosiy yo'qotish funksiyasi belgilar darajasidagi kesishma entropiya [13]:

$$\mathcal{L}_{\text{asosiy}} = -\frac{1}{L} \sum_{t=1}^L \log p_{\theta}(y_t | y_{<t}, X).$$

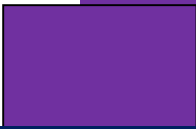
Barqarorlashtirish uchun og'irliklarni cheklash (L2) qo'shildi:

$$\mathcal{L} = \mathcal{L}_{\text{asosiy}} + \lambda \|\theta\|_2^2, \quad \lambda = 0.01.$$

Qadamdagi o'rganish tezligi $\eta(i)$ isitib borish bilan boshqarildi (chiziqli oshirish va so'ngra chiziqli pasaytirish), ikki bosqichli jadval :

$$\eta(i) = \begin{cases} \eta_{\max} \frac{i}{w}, & 0 \leq i < w, \\ \eta_{\max} \left(1 - \frac{i-w}{S-w}\right), & w \leq i < S, \end{cases}$$

bu yerda w — boshlang'ich faollashtirish qadamlar soni, S — umumiy qadamlar soni.



Gradient cheklash orqali barqarorlik ta'minlandi (L2-norma asosida qayta masshtablash):

$$g \leftarrow \frac{g}{\max(1, \|g\|_2/c)}, \quad c = 0.5.$$

2-jadval. O'qitish tartibi va asosiy parametrlar

Band	Qiymat	Izoh
Davrlar soni	3	Uch marta to'liq o'qitish
O'rganish tezligi	0.000005	Isitib borish bilan
Baholash oraliqlari	500	Tekshiruvda muntazam nazorat
Og'irlikni kamaytirish	0.01	Barqarorlashtirish
Gradient cheklash mezoni	0.5	Qayta masshtablash
Paket hajmi	32	Yozuvlar bo'yicha
Hisoblash aniqligi	16-bit	Aralash aniqlik

Taxminiy umumiy qadamlar soni:

$$S \approx \frac{N}{B} \times E = \frac{121\,378}{32} \times 3 \approx 11\,387.$$

Tekshiruv yo'qotishi yoki WER ko'rsatkichi bir necha marotaba yaxshilanmasa, o'qitish to'xtatildi va eng yaxshi holatdagi parametrlar saqlandi. Bu mexanizm ortiqcha yodlab olishni oldini olishga xizmat qildi [13,14].

Yakuniy baho so'zdagi xatolik ulushi (WER) orqali olindi [14]:

$$WER = \frac{S + D + I}{N} \times 100\%,$$

bu yerda S — almashtirishlar soni, D — tushirib qoldirishlar, I — ortiqcha kiritishlar, N — etalon so'zlar soni.

TAHLIL VA NATIJALAR

Ushbu ishda yig'ilgan ma'lumotlar bir xil andozaga keltirilib, moslashtirib o'qitish uch bosqichda bajarildi va yakunda alohida sinov to'plamida baholandi. Tayyorlangan to'plamlar hajmi mashq uchun 119 831 yozuv, tekshiruv uchun 648 yozuv, sinov uchun 899 yozuvdan iborat bo'ldi. Moslashtirib o'qitish jarayonida tekshiruv bahosi muntazam kuzatildi va eng yaxshi holat parametrlarini saqlash uchun erta to'xtatish mexanizmi qo'llandi.

Model chiquvi sifatini baholashda so'zdagi xatolik ulushi ko'rsatkichi qo'llandi. Sinov to'plamida olinadigan yakuniy baho quyidagicha bo'ldi: so'zdagi xatolik ulushi 13.78%, sinov yo'qotishi 0.1487, tekshiruv yo'qotishi yakunda taxminan 0.15. Bu natija turli manbalardan yig'ilgan yozuvlarni birlashtirish, matnni me'yoriy shaklga keltirish, audio parametrlarini yagona ko'rinishga moslashtirish hamda o'qitish jadvalini ehtiyotkor boshqarish (isitib borish va gradient cheklash) bilan birga qo'llanganda barqaror samaraga erishish mumkinligini ko'rsatadi. Amaliy kuzatishlar shuni ko'rsatdiki, tinish belgilari va yozuv me'yorlariga rioya qilish natijaviy matnni silliqroq va izchilroq qiladi; talaffuz farqlari va sukut oralig'idagi o'zgarishlarga chidamlilik oshiriladi. Shuningdek, kiritmalarni yagona namuna olish tezligiga keltirish hamda uzunlikni boshqarish natijalarni sezilarli barqarorlashtiradi.



Ma'lumotlar sifati bo'yicha boshqaruv choralarining uyg'unligi tekshiruv bahosida tebranishlarni kamaytirib, chiquvdagi xatoliklarni pasaytirishga xizmat qildi .

3-jadval Yakuniy ko'rsatkichlar

Ko'rsatkich	Qiymat
So'zdagi xatolik ulushi	13.78%
Yo'qotish (sinov)	0.1487
Yo'qotish (tekshiruv, yakun)	0.15

MUHOKAMA

Natijalar tahlili shuni ko'rsatadiki, ikki turdagi ma'lumot manbalarini uyg'unlashtirish hamda kiritmalarni yagona ko'rinishga keltirish o'zbek nutqini turli sharoitlarda yanada aniqroq tanishga yordam beradi. Ushbu yondashuv turli so'zlovchilarning nutqidagi tabiiy farqlar, talaffuzdagi o'zgarishlar va lahjaviy xususiyatlarni qisman neytrallashtiradi hamda modelni umumlashtirish qobiliyatini oshiradi. Yozuvlar orasidagi lahja farqlari, talaffuzdagi o'zgarishlar va sukut oralig'i kabi omillar mavjud bo'lsa-da, moslashtirib o'qitish usullari barqaror natija beradi va ortiqcha yodlab olishning oldini oladi. Bu holat modelning yangi, ilgari ko'rilmagan namunalar bilan ishlashda ham moslashuvchanligini ta'minlaydi. Shu orqali o'qitilgan tizim turli ovoz sifati, mikrofondan uzoqlik yoki fon shovqini mavjud vaziyatlarda ham maqbul aniqlik darajasini saqlaydi. Dekoderga til va vazifa ko'rsatmalarini majburiy berish, maxsus belgilarni ehtiyotkorlik bilan boshqarish, tinish belgilari hamda imlo me'yorlarini izchil saqlash chiquv matnining tabiiyligiga sezilarli ijobiy ta'sir ko'rsatdi. Natijada model tomonidan yaratilgan matnlar grammatik jihatdan to'g'ri, ohang va mantiqiy yaxlitlik nuqtai nazaridan esa o'qish uchun qulay shaklga ega bo'ldi. Bu, ayniqsa, nutqni matnga aylantirish tizimlarini amaliy platformalarda qo'llashda muhim ahamiyat kasb etadi.

Shuningdek, tekshiruv bahosini qadamlar bo'yicha muntazam kuzatish hamda erta to'xtatish mexanizmi o'qitish jarayonini muvozanatga keltirib, eng yaxshi holatdagi model parametrlarini saqlab qoldi. Bu usul ortiqcha o'qitilish xavfini kamaytirib, umumlashtirish qobiliyatini oshirdi.

Texnik tomondan qaralganda, aralash aniqlikdagi hisoblash (mixed precision training) va xotirani joyida band qilish (gradient checkpointing) kabi choralar hisoblash samaradorligini oshirdi hamda chiquv sifatiga salbiy ta'sir ko'rsatmasdan tezkor ishlashni ta'minladi. Bu yondashuv resurslar cheklangan muhitlarda, ayniqsa real vaqt rejimidagi tizimlarda, amaliy ahamiyatga ega. Shuni qayd etish lozimki, ma'lumotlarda raqamlar, sanalar, qisqartirilmasdan yozilgan murakkab iboralar yoki xorijiy atamalar mavjud bo'lganda ayrim chalkashishlar yuzaga kelishi mumkin. Bunday holatlar ko'pincha yozuvlardagi orfografik nomuvofiqliklar va talaffuzdagi farqlar bilan bog'liq. Shu sababli, kelajakdagi ishlar uchun ma'lumotlarni me'yoriy tozalash, yozuvlarning bir xilligini oshirish va maxsus belgilarni standartlashtirish muhim hisoblanadi. Keyingi bosqichlarda turli shovqin darajalarini sun'iy ravishda qo'shib mashq qilish, matnlarni mavzu bo'yicha muvozanatlash, shuningdek, lahjaviy o'zgarishlarni yanada kengroq qamrab olish orqali natijalarni yanada yaxshilash

kutilmoqda. Bundan tashqari, modelni ko‘p tilli ma‘lumotlar bilan qayta o‘qitish orqali turli lahjalarning kesishuvchi elementlarini aniqlash ham o‘zbek tilining nutq tanish sohasida universallik darajasini oshiradi.

XULOSA

Tadqiqot jarayonida biz FeruzaSpeech va Uzbek Speech Corpus ma‘lumot to‘plamlarini birlashtirdik hamda ularni tozalash, muvozanatlashtirish va yagona formatga keltirish orqali sifatli o‘quv to‘plamini yaratdik. Ushbu ma‘lumotlar asosida “Whisper-small” modeli tanlanib, uni o‘zbek tiliga moslashtirish uchun chuqur o‘qitish usuli qo‘llandi. Modelni qayta o‘qitish jarayonida tovush sifati, uzunligi va talaffuzdagi tafovutlar hisobga olindi. Natijada model o‘zbekcha nutqni aniqlik bilan matnga aylantira oladigan darajaga yetkazildi. Tajriba natijalari shuni ko‘rsatdiki, yaratilgan tizim sinov ma‘lumotlarida yuqori aniqlikka erishdi va xatolik darajasi past bo‘ldi. Bu esa o‘zbek tilidagi nutqni avtomatik tahlil qilish va amaliy dasturlarda qo‘llash imkoniyatlarini kengaytiradi. Mazkur ish natijalari ta‘lim, tibbiyot, raqamli xizmatlar, ommaviy axborot vositalari va foydalanuvchi bilan ovozi muloqot talab etiladigan ko‘plab sohalarda samarali qo‘llanishi mumkin. Shu tariqa, bizning tadqiqotimiz o‘zbek tili uchun moslashgan, samarali va yuqori sifatli avtomatik nutq tanish tizimini yaratishga qaratilgan bo‘lib, bu yo‘nalishda keyingi ilmiy izlanishlar uchun mustahkam poydevor yaratadi.

FOYDALANILGAN ADABIYOTLAR

1. Radford, A. va boshq. Whisper: Robust Speech Recognition via Large-Scale Weak Supervision. OpenAI texnik hisobot, 2022.
2. Uzbek Speech Corpus va Avtomatik Nutqni Tanish. ISSAI loyiha sahifasi, 2021.
3. Uzbek Speech Corpus — dataset card. Hugging Face, 2021.
4. Povey, A.; Povey, K. FeruzaSpeech: A 60-Hour Uzbek Read Speech Corpus with Punctuation, Casing, and Context. ICNLSP 2024 materiallari, 2024.
5. Mozilla Common Voice — umumiy portal. Ma‘lumotlar platformasi, 2017.
6. Common Voice Scripted Speech 23.0 — Uzbek datasheet. Mozilla Data Collective, 2024.
7. Mukhamadiyev, A. va boshq. Development of Language Models for Continuous Uzbek Speech Recognition System. *Sensors* (MDPI), 23(3):1145, 2023. (PMC to‘liq matn).
8. Li, J. va boshq. Improving Whisper’s Recognition Performance for Under-Represented Language (Kazakh) Leveraging Unpaired Speech and Text. Interspeech 2024 (ISCA Archive), 2024.
9. NVIDIA NeMo Framework User Guide — Checkpoints/Training results, 2024.
10. NVIDIA FastConformer-Hybrid Large (Uzbek) — model sahifasi. Hugging Face, 2024.
11. Whisper models in Amazon SageMaker JumpStart. AWS Machine Learning Blog, 10 Oct 2023.
12. Nguyen, M. T. va boshq. Improving Speech Recognition with Jargon Injection. SIGDIAL 2024 materiallari (ACL Anthology), 2024.

13. Peng, Y. va boshq. OWSM v4: Improving Open Whisper-Style Speech Models via Data Scaling and Cleaning. Interspeech 2025 (ISCA Archive), 2025.
14. Khamidov, A., Ismatov, M., Karimova, F. (2024). Integration of Whisper Models for Uzbek ASR in Noisy Environments. Journal of Modern Information Technologies and AI Research, 2(1), 58-66.