

**SUN'IY INTELLEKT ASOSIDA SO'ZLOVCHINI AUTENTIFIKATSIYA QILISH
TIZIMINI ISHLAB CHIQISH VA BAHOLASH****Xoldorov Sh.I****Xasanov U.K**

*Muhammad al-Xorazmiy nomidagi TATU, robototexnika va intellektual tizimlar
kafedrası, Toshkent, O'zbekiston Email: umidjon0923@gmail.com*

Annotatsiya: *Ushbu maqolada chuqur neyron tarmoqlarga asoslangan sun'iy intellekt tizimi yordamida inson ovozi orqali foydalanuvchini ishonchli tarzda aniqlovchi va tekshiruvchi so'zlovchini autentifikatsiya qilish tizimi ishlab chiqish usuli va algoritmlari tahlil qilindi. Taklif etilayotgan tizimda signalga dastlabki ishlov berish, akustik xususiyatlarni ajratish va neyron tarmoq yordamida so'zlovchilarni ifodalovchi xususiyatlar vektorlarini shakllantirish jarayonlarini integratsiyalashgan tizimda birlashtiradi. Dastlabki ishlov berish jarayonlarida signallarni filtrlash algoritmlari bilan filtrlanadi va signal normallashtiriladi. Xususiyatlarni ajratish bosqichida log-Mel energiya koeffisientlari ajratib olinadi va uni ECAPA-TDNN modeli uchun kirish parametrlari sifatida olinadi. ECAPA-TDNN modelidan so'zlovchini ifodalovchi xususiyatlar vektori shakllantiriladi. Oxirgi bosqichda solishtirish algoritmlari yordamida xususiyatlar vektorlari solishtiriladi. Agar so'zlovchilar xususiyatlari mos bo'lsa tizim kirish uchun ruxsat beradi.*

Kalit so'zlar: *Nutq signali, nutq signallariga ishlov berish, so'zlovchilarni aniqlash, so'zlovchilarni tekshirish, so'zlovchilarni autentifikatsiya qilish, filtrlash, log-Mel energiya, ECAPA-TDNN*

Abstract: *This article analyzes the methods and algorithms for developing a speaker authentication system that reliably identifies and verifies a user through a human voice using an artificial intelligence system based on deep neural networks. The proposed system combines the processes of signal preprocessing, acoustic feature extraction, and feature vector formation using a neural network to represent speakers in an integrated system. In the initial processing processes, the signal is filtered using signal filtering algorithms and the signal is normalized. At the feature extraction stage, log-Mel energy coefficients are extracted and taken as input parameters for the ECAPA-TDNN model. A feature vector representing the speaker is formed from the ECAPA-TDNN model. At the final stage, feature vectors are compared using matching algorithms. If the speaker features match, the system allows access.*

Keywords: *Speech signal, speech signal processing, speaker identification, speaker verification, speaker authentication, filtering, log-Mel energy, ECAPA-TDNN.*

KIRISH

Zamonaviy raqamli dunyoda sun'iy intellekt texnologiyalarining jadal tushlarda rivojlanishi inson nutq signallarini tahlil qilish, ularni qayta ishlash asosida identifikatsiya hamda autentifikatsiya qilish jarayonlari uchun yangi imkoniyatlarni

ochdi. Nutq signallari inson uchun murakkab hamda takrorlanmas biometrik belgi bo'lib, uning akustik hamda lingvistik xususiyatlari shaxslarni aniqlash imkonini beradi. Bunday texnologik yechimlar bugungi kunda raqamli identifikatsiya, nutqni matnga aylantirish[1], kiberxavfsizlik, robototexnika[2] va aqlli qurilmalarning ishonchli himoya[3] mexanizmlardan biri sifatida qaraladi. Ovoz signallari orqali autentifikatsiya tabiiy hamda foydalanuvchi uchun qulay bo'lib, kontakt talab qilmaydigan va masofadan turib ishlaydigan tizimlar uchun muhim ahamiyat kasb etadi.

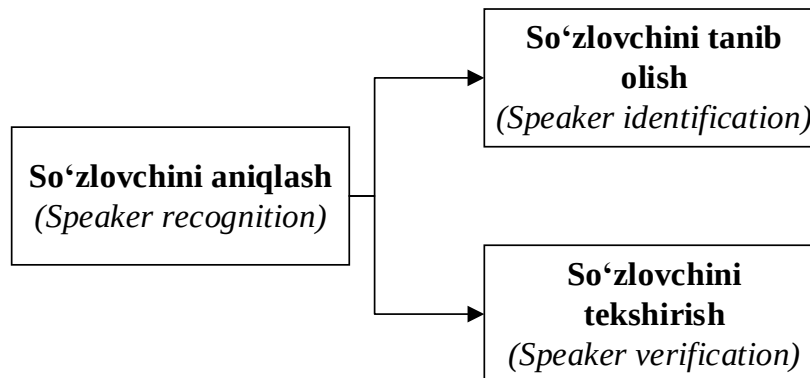
So'zlovchilarni aniqlash insonning ovozidan kim ekanligini aniqlash vazifasini bajaradi[4]. U asosan ikki turga bo'linadi:

A) So'zlovchilarni tanib olish (ing: Speaker identification)

B) So'zlovchilarni tekshirish (ing: Speaker verification)

So'zlovchilarni tanib olish noma'lum ovozni avvalda ro'yxatdan o'tgan barcha so'zlovchilar orasidan tanib olish jarayonidir. Ushbu modelda so'zlovchilarni ifodalovchi xususiyatlar vektorlarini shakllantiriladi, uni bazadagi vektorlar bilan taqqoslaydi va uni eng yaqinini tanlaydi. So'zlovchilarni tekshirish berilgan ovoz namunasi belgilangan da'vogar so'zlovchiga tegishli yoki yo'qligini tekshirish jarayonidir. Bu jarayonda asosan ikki sinfli qaror chiqarish masalasi ko'riladi[5].

So'zlovchini autentifikatsiyasi yuqoridagi ikki jarayonning amaliy integratsiyalashgan shakli bo'lib, foydalanuvchini nutqi orqali shaxs sifatida tekshirish jarayoni. U so'zlovchini tekshirish texnologiyasiga asoslanadi.



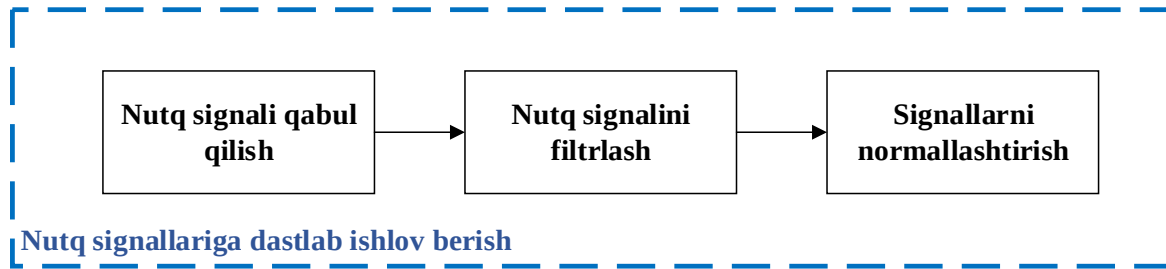
1-rasm. So'zlovchilarni aniqlash turlari

1. Metodologiya

Bugungi kunda so'zlovchilarni autentifikatsiyasi turli arxitekturali tizimlar orqali amalga oshiriladi. Eng dastlabki tizimlarda statistik xususiyatlarga asoslangan GMM-UBM[6] va i-vector[7] usullaridan foydalanilgan bo'lsa, hozirda DNN, x-vector[8] va ECAPA-TDNN[9] kabi chuqur neyron tarmoqli usullardan foydalaniladi.

Taklif etiladigan foydalanuvchini ovozi orqali autentifikatsiya qilish signalga dastlabki ishlov berish, xususiyatlarni ajratish, so'zlovchini ifodalovchi xususiyatlar vektorini hosil qilish va tasdiqlash bosqichlarini o'z ichiga oladi.

Dastlabki ishlov berish bosqichida signallarni sifatini yaxshilash uchun nutq signallari avvalo shovqin va reverbaratsiyalardan tozalanadi. Bu jarayonlarda turli filtrlash algoritmlaridan, gibrid arxitekturasida algoritmlardan foydalaniladi. Har bir bosqichning sifati umumiy tizim samaradorligiga ta'sir qiladi.

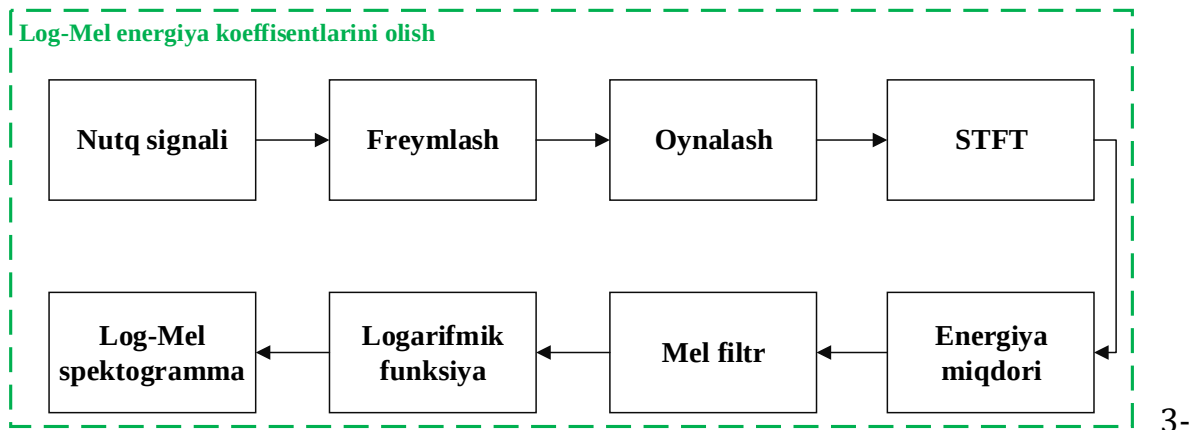


2-rasm. Nutq signallariga dastlabki ishlov berish jarayoni

Dastlabki ishlov berish jarayonidan so'ng signal freymlarga ajratiladi va har bir freym uchun log-Mel energiya koeffisientlari hisoblanadi.

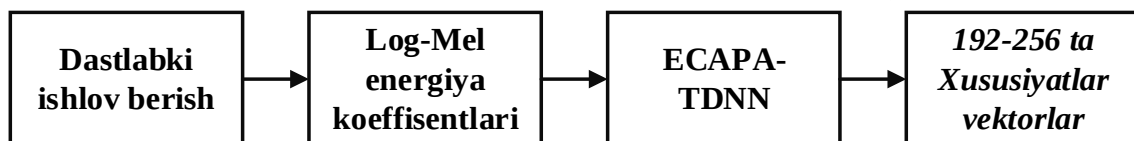
$$E_m = \log \left(\sum_{k=f_m-1}^{f_m} |X(k)|^2 \cdot H_m(k) \right)$$

bu yerda $H_m(k)$ – Mel filtri, $X(k)$ – signalning Furje spektri. Olingan log-Mel energiya koeffisientlari so'zlovchilarni ifodalovchi xususiyatlarni ifodalovchi vektorlar olish uchun neyron tarmoqqa kiruvchi parametrlar bo'lib xizmat qiladi. Bu xususiyatlar so'zlovchi haqidagi tembr, ton va formant ma'lumotlarini o'zida saqlaydi.



3-rasm. Log-Mel energiya koeffisientlarini olish bosqichlari

ECAPA-TDNN modelidan foydalanilgan holda har bir nutq namunasi uchun 192-256 ta o'lchamli xususiyatlar vektorlari olinadi. Bu modelning SE-block (ing: Squeeze-Excitation) mexanizmi kanal e'tibori asosida akustik belgilarni kuchaytiradi.



4-rasm. ECAPA-TDNN asosida xususiyatlar vektorlarini shakllantirish

So'zlovchining test xususiyatlar vektori E_{test} bazada saqlangan E_{ref} bilan kosinus o'xshashlik yoki PLDA algoritmlari orqali taqqoslanadi.

$$\text{Score}(E_{test}, E_{ref}) = \frac{E_{test} \cdot E_{ref}}{\|E_{test}\| \|E_{ref}\|}$$

Agar $\text{Score}(E_{test}, E_{ref}) > \tau$ bo'sag'a qiymatidan katta bo'lsa, foydalanuvchiga kirish uchun ruxsat beriladi. Tizimning aniqligi esa EER, FAR, FRR ko'rsatkichlari bilan baholanadi.

2. Natijalar

Tizim aniqligini va samaradorligini aniqlash uchun ovoz namunalaridan foydalanilgan holda so'zlovchilarni ifodalovchi xususiyatlar vektorlari shakllantirib olinadi. Ma'lumotlat to'plami uchun 16 kHz diskretlash chastotasi, kvantlanish 16 bit bo'lgan bir so'zlovchining turli xil uzunligidagi audio signallari yig'iladi. Turli metodlar yordami so'zlovchini autentifikatsiya tizimini ishlab chiqiladi va uning samaradorliklari baholanadi.

1-jadval. So'zlovchi autentifikatsiyani usullar samaradorligi taqqoslash jadvali

Metod	EER	FAR	FRR	ROC-AUC
GMM-UBM	9.48	7.98	8.89	0.910
x-Vector + Kosinus o'xshashlik	9.23	7.87	7.82	0.921
ECAPA-TDNN +Kosinus o'xshashlik	7.55	7.29	7.81	0.9624

Xulosa

Tadqiqot natijalariga ko'ra so'zlovchini autentifikatsiya qilishda qo'llaniladigan algoritmlar samaradorliklari bilan farqlanadi. Jadvalda keltirilganidek, GMM-UBM, x-Vector, va ECAPA-TDNN modellarining natijalari orasida sezilarli farqlar kuzatildi. Klassik GMM-UBM modeli nutq signalining statistik xususiyatlariga asoslangan bo'lib, EER = 9.48%, FAR = 7.98%, va FRR = 8.89% qiymatlarini ko'rsatdi.

Ushbu natijalar modelning shovqinli va akustik jihatdan o'zgaruvchan muhitlarga nisbatan sezgirlikini bildiradi. Eng yuqori natija ECAPA-TDNN modelida qayd etildi.

Ushbu modelda EER = 7.55%, FAR = 7.29%, FRR = 7.81%, va ROC-AUC = 0.9624 bo'lib, oldingi modellarga nisbatan ancha yuqori aniqlik ko'rsatdi. ECAPA-TDNN modeli so'zlovchini autentifikatsiya qilishda klassik statistik va bazaviy neyron yondashuvlarga nisbatan 20–25% gacha yaxshilanish beradi hamda ROC-AUC ko'rsatkichi orqali tizimning ishonchlilik darajasi yuqori ekanini tasdiqlaydi.

FOYDALANILGAN ADABIYOTLAR:

1. Abrorovich, Q. A. L., O'G'Li, X. S. I., & O'G'Li, X. U. K. (2024). SO'ZLOVCHINI TANIB OLISHDA TANIB OLISH USULLARI TAHLILI. Eurasian Journal of Technology and Innovation, 2(6), 60-67.

2. Shukurov, K., Khasanov, U., Turaev, B., & Kakhkharov, A. L. (2023). The Effectiveness of the Implementation of Speech Command Recognition Algorithms in Embedded Systems. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23, 220-224.

3. Elbobo o'g'li, S. K., & Komiljon o'g'li, X. U. (2025, February). TIBBIY DIAGNOSTIKADA NUTQ SIGNALLARINING PARALINGVISTIK XUSUSIYATLARINI QO'LLASH. In INTERNATIONAL SCIENTIFIC RESEARCH CONFERENCE (Vol. 3, No. 31, pp. 118-121).

4. Kozlov, Alexander & Kudashev, Oleg & Matveev, Yu & Pekhovsky, Timur & Simonchik, Konstantin & Shulipa, Andrey. (2013). Speaker recognition system for the NIST SRE 2012. SPIIRAS Proceedings. 25. 350-370. 10.15622/sp.25.18.

5. Shukurov, K. (2021). Analysis of algorithms and implementation of real time speaker identification system. Bulletin of TUIT: Management and Communication Technologies, 4(4), 2.
6. Grozdić, Đorđe & Jovičić, Slobodan & Saric, Zoran & Subotić, Irina. (2015). Comparison of GMM/UBM and i-vector based speaker recognition systems
7. Kanagasundaram, Ahilan. (2014). Speaker Verification using I-vector Features. 10.13140/2.1.2933.7283.
8. Xu, Longting & Das, Rohan & Yilmaz, Emre & Yang, Jichen & Li, Haizhou. (2018). Generative x-vectors for text-independent speaker verification. 10.48550/arXiv.1809.06798.
9. Dawalatabad, Nauman & Ravanelli, Mirco & Grondin, François & Thienpondt, Jenthe & Desplanques, Brecht & Na, Hwidong. (2021). ECAPA-TDNN Embeddings for Speaker Diarization. 3560-3564. 10.21437/Interspeech.2021-941.