

SHAXSNI OVOZI ASOSIDA TANIB OLISH USULLARI

Mamatov Narzullo Solidjonovich¹, Dusanov Xurshid Toshpo'latovich², Pulatov G'iyos Gofurjonovich³

¹“Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti” Milliy tadqiqot universiteti, Raqamli texnologiya va sun'iy intellekt kafedrası, t.f.d, professor, E-mail: m_narzullo@mail.ru

²O'zMU Jizzax filiali, tayanch doktorant, E-mail: xurshid3868@mail.ru

³“Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti” Milliy tadqiqot universiteti, Raqamli texnologiya va sun'iy intellekt kafedrası, katta o'qituvchi, E-mail: giyospulatov1987@gamil.com

KALIT SO'ZLAR

belgi shakllantirish, diskret kosinus, furye almashirish, identifikatsiyalash, shaxs ovozi, mel-chastota, nutqni tanib olish.

ANNOTATSIYA

Turli tillar uchun shaxsni ovozi asosida tanib olish tadqiqotchilar uchun hali ham katta muammo bo'lib qolmoqda. Agar nutq namunasining talaffuzi kamroq bo'lsa, identifikatsiya tezligining aniqligi katta muammodir. Ushbu maqola mel-chastotali kepral koefitsientlar (MFCC) belgilar to'plamini ajratish algoritmi va Gauss aralashma modeli (GMM) modellashtirish algoritmi yordamida shaxsni ovozi asosida tanib olishga qaratilgan.

I. Kirish

Shaxsni ovozi asosida identifikatsiyalash – uni ovozi xususiyatlarini mavjud nutq namunalari xususiyatlari bilan solishtirish va taqqoslash yo'li orqali tanib olish masalasidir. Nutqni tanib olish sohasida tadqiqot olib borish akustik fonetika va xususiyatlarni taqqoslash usullari haqidagi bilimlarga ega bo'lishni talab etadi [1]. Shaxsni ovozi asosida identifikatsiyalash masalasi yuzni tanib olish, nutq tilini aniqlash va nutqni tanib olish kabi ko'p tarmoqli masalalar sirasiga kiradi. Ma'ruzachini aniqlash vazifalarining aksariyati qisqa nutq yoki odamlarning ko'rsatmalariga bog'liq. Odatda shaxsni ovozi asosida identifikatsiyalash matnga bog'liq bo'lgan yoki bog'liq bo'lmagan holda amalga oshiriladi. Identifikatsiyalash matnga bog'liq bo'lganda oldindan tayyorlangan yoki tizim tomonidan hosil qilingan matn asosida, matnga bog'liq bo'lmaganda esa, ixtiyoriy matn asosida bajariladi. Shaxsni ovozi orqali identifikatsiyalash tizimiga ta'sir qiluvchi bir qancha omillar mavjud [4,5]. So'zlar sonining kam bo'lishi shaxsni tanib olishni qiyinlashtiradi [2].

Bugungi kunga kelib ovoz asosida insonning jinsini aniqlash masalalari uchun aniq yechimlar ishlab chiqilgan, biroq nutq egasining shaxsiyatini aniqlash tizimlarini ishlab chiqish masalasi bir qator kamchiliklarga ega bo'lib qolmoqda.

Ishonchli tanib olish tizimi real vaqt rejimida ishlashi uchun tezkor va qimmat uskunalarni talab qiladi. Shaxsni ovozi asosida tanib olish vazifasi odatda ikki bosqichli ishlov berish orqali amalga oshiriladi: o'qitish va testlash. O'qitish bosqichi nutq oqimidan shaxsga xos xususiyat parametrlarini hisoblab chiqadi. Xususiyatlar turli shaxslarning statistik modellarini hosil qilish uchun ishlatiladi. Testlash bosqichida noma'lum shaxslarning nutq namunalari modellar va tasniflash usullari yordamida taqqoslanadi. Tajribalarning samaradorligi va ishonchliligi identifikatsiya tezligi, aniqlik va xatolik darajasi bo'yicha baholanadi [3].

II. Usullar

Asosiy komponentlar

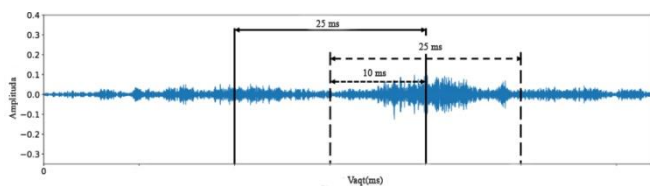
Maqolaning ushbu qismida taklif qilingan yondashuvda qo'llaniladigan algoritm tavsifi mavjud. Shuningdek, bu ularning inson nutqini aniqlash tizimiga nisbatan foydaliligi va ishonchliligini ochib beradi. Bizning ishimizda MFCC, VQ va GMM algoritmlari qo'llaniladi.

Mel-chastotali kepral koefitsientlari (MFCC-Mel Frequency Cepstral Coefficient) nutq signalini tavsiflash uchun foydalaniladigan samarali belgilar to'plami shakllantirish usuli

hisoblanadi. MFCC quyidagi yetti bosqichni o'z ichiga oladi:

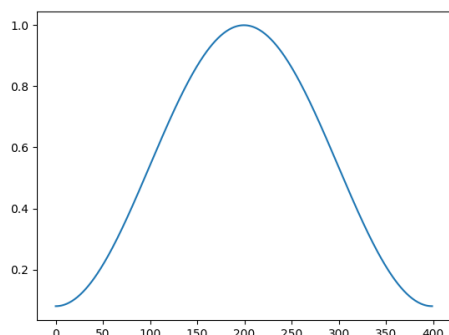
Diskretlash. Ushbu bosqichda kirish signalini analog signal shaklida raqamli signal shakliga 16 kHz namuna olish chastotasi bilan almashtirishni amalga oshiriladi.

Freym'larga ajratish. Bunda nutq signalini freym'larga ajratish uchun qo'llanilishi mumkin bo'lgan audio signalni 16 kHz chastotada yozilgan 1 sekund misolida ko'rib chiqiladi. Odatda 25 ms(millisekund) davomiylikdagi audio signal 400 ta namunadan tashkil topadi.



1-rasm. Nutq signalini freym'larga ajratish

Xemming oynasi. Birinchi 400 ta namuna 0 dan boshlanganligi sababli, keyingi 400 ta namuna 160 dan boshlanadi, ya'ni 240 ta namuna bir-biriga mos keladi.



2-rasm. Xemming oynasi

Odatda Xemming oynasidan nutq bilan bog'liq masalalarni yechishda foydalaniladi. Xemming oynasi yordamida signal quyidagi formula orqali almashtiriladi:

$$w[t] = 0.54 - 0.46 \cos\left(\frac{2\pi t}{N-1}\right), \quad 0 \leq t \leq N-1 \quad (1)$$

Diskret Furiye almashtirish (DFA): Diskret Furiye almashtirishni qo'llash orqali signal vaqtli sohadan chastotali sohaga o'tkaziladi. Nutq signallari uchun chastotali sohada tahlil qilish vaqtli soha ga nisbatan qulaydir. Vaqt signali uchun Furiye almashtirish quyidagi formula orqali amalga oshiriladi:

$$H(n, k) = \sum_{n=1}^N x(n)w(n)e^{\frac{-2\pi i k n}{N}} \quad 0 \leq k \leq K \quad (2)$$

Bu yerda: $x(n)$ - vaqt sohasidagi signal, N - n ta namunadan iborat oyna uzunligi, K - DFA uzunligi. Quvvat spektri namuna energiyasining formulasi quyidagicha olinadi:

$$S(n, k) = |H(n, k)|^2 \quad (3)$$

Mel-filtr bankini qo'llash. Uchburchak filtrlar to'plamlarini 20 tadan 40 (26 ta standart) tagacha qo'llash mumkin. Bunda filtr bankining energiyalarini hisoblash uchun har bir filtr banki quvvat spektriga ko'paytiriladi va koeffitsientlar hosil qilinadi. Bu amal bajarilganidan so'ng, har bir filtr bankida qancha energiya borligi haqida ma'lumot beruvchi 40 ta son olinadi. Mel-filtr banki quyidagi formula orqali hisoblanadi:

$$f_{mel} = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) \quad (4)$$

Logarifmlash. Inson past chastotadagi o'zgarishlarni yuqori chastotadagi o'zgarishlarga nisbatan aniqroq sezadi. Logarifmlash shunga o'xshash xususiyatga ega. Kirishning x past qiymatida \log funksiyasining gradienti yuqori bo'ladi, lekin kirishning yuqori qiymatida gradient qiymati kichikroq bo'ladi. Bu esa inson eshitish tizimiga mos Mel-filtr chiqishiga logarifmlashni qo'llash imkonini beradi:

$$MSFB = 20 \log_{10}(S(n, k) \cdot F_m(k)) \quad (5)$$

Diskret kosinus almashtirish. Ushbu bosqichda oldingi bosqichdagi chiqish uchun teskari o'zgartirish amalga oshiriladi. MFCC usuli diskret kosinus almashtirishni qo'llaganidan so'ng signalning 13 ta koeffitsienti hosil bo'ladi.

$$C_k = \alpha_k \sum_{i=1}^M X_i \cos \frac{\pi k (2i+1)}{2M}, \quad k = 0, 1, 2, \dots, K \quad (6)$$

Nutq signalidan belgilar to'plamini ajratib olgandan so'ng shaxs ovozi modeli quriladi. Bunda modellashtirishning quyidagi usullaridan foydalaniladi:

Vektorli kvantlash(VQ-Vector Quantization) – raqamli signalni qayta ishlash usuli bo'lib, cheklangan miqdordagi vektorlar yordamida raqamli ma'lumotlarni kodlashni o'z ichiga oladi. Ushbu usul odatda tasvir, nutqni siqishda hamda ma'lumotlarni tahlil qilish va timsollarni tanib olishda qo'llaniladi. Vektor kvantlash jarayoni odatda quyidagi bosqichlarni o'z ichiga oladi:

- Ma'lumotlarni kichikroq guruhlariga yoki klasterlarga bo'lish;
- Har bir klaster uchun kod so'zi sifatida tanilgan va klasterdagi ma'lumotlar nuqtalarining qiymatlariga yaqinroq bo'lgan vakillik qiymatini topish;
- Har bir ma'lumot nuqtasini kod so'zi unga eng yaqin bo'lgan klasterga belgilash;
- Har bir ma'lumot nuqtasini tegishli kod so'zi bilan almashtirish orqali ma'lumotlarni kodlash;

Vektor kvantlash ma'lumotlarning yo'qolishiga ham olib kelishi mumkin, chunki dastlabki ma'lumotlar nuqtalari ularning kod so'zlari bilan almashtiriladi, ular asl ma'lumotlarni to'g'ri ko'rsatmasligi mumkin.

Signal sifatini dastlabki holatini saqlab qolish bilan birga uni saqlash va uzatish talablarini kamaytirishga yordam beradi. O'xshash ma'lumotlar nuqtalarini guruhlash va ularni bitta vektor bilan ifodalash orqali vektor kvantlash signalni ifodalash uchun zarur bo'lgan bitlar sonini samarali ravishda kamaytirishi mumkin. Bu esa uni ma'lumotlarni siqish, timsollarni tanib olish uchun ishlatiladi.

Gauss aralashmasi modeli (GMM – Gaussian Mixture Model) - timsollarni tanib olish vazifalarida keng qo'llaniladigan statistik modeldir. bu suhandon ovozi ehtimollik modelini taminlovchi shaxsni ovozi asosida identifikatsiyalash yondashuvi hisoblanadi. Har qanday urg'uni modellashtirishning muhim jihati – har bir urg'u va aralashmaning o'rtacha vektorining og'irligini o'quv nutqidan to'plash va topishdan iborat. Gauss aralashmasi modelining parametrlari maksimal ehtimollik – Gauss aralashmasi modeli yordamida baholanadi. Shakllarning ehtimollik zichligi funksiyasi taxminiy hisoblanadi. GMM ehtimollik zichligi funksiyasi kovariatsiya matritsasi (Σ_i), matematik

kutilma (μ_i) va aralashmaning vazn parametrlari to'plami (K_i) orqali taqdim etilishi mumkin. Ko'p o'lovchi Gauss aralashmasi modeli (7) tenglama bilan berilgan.

$$K_i = \frac{1}{\sqrt{(2\pi)^{D/2} |\Sigma_i|^{1/2}}} \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right] \quad (7)$$

Taklif etilayotgan algoritm xususiyatlarni ajratib olishdan keyingi tasniflash usullaridan iborat. Analog signal birinchi navbatda raqamli shaklga aylanadi. O'zbek tilidagi nutq signallarining xususiyatlari MFCC yordamida hisoblab chiqiladi. Keyin olingan xususiyat Gauss aralashmasi modeli yordamida tasniflanadi. Yakuniy natija maksimal logorifik ehtimollik funksiyasi yordamida hisoblanadi. MFCC-GMM usulidan foydalangan holda ma'ruzachini aniqlash algoritmi quyida keltirilgan.

Shaxsni ovozi asosida tanib olish uchun MFCC-GMM usuli quyidagi bosqichlardan iborat:

Belgilar to'lamini shakllantirish: nutq signali oldindan qayta ishlanadi va freymlarga bo'linadi. Keyin har bir freym MFCC koeffitsientlariga aylantiriladi. Koeffitsientlar soni odatda 12 dan 20 gacha.

Xususiyatlarni normallashtirish: MFCC koeffitsientlari nol o'rtacha va birlik farqiga ega bo'lish uchun normallashtiriladi. Bu qadam Shaxsni ovozi asosida tanib olish tizimining aniqligini yaxshilashga yordam beradi [7].

GMM o'qitish: GMM har bir shaxs uchun normallashtirilgan MFCC koeffitsientlari bo'yicha o'qitiladi. GMMdagi komponentlar soni o'quv ma'lumotlar to'plamining hajmiga qarab farq qilishi mumkin.

Suxondonni modellashtirish: GMM o'qitilgandan so'ng, har bir suxondon nutqini modellashtirish uchun foydalanish mumkin. Har bir suxondon uchun GMM ularning nutq signalining taqsimlanishini ifodalaydi.

Shaxsni ovozi asosida tanib olish: testlash nutq signalidan shaxsni ovozi asosida tanib olish uchun avval testlash signalining MFCC koeffitsientlari chiqariladi va normallashtiriladi.

Keyin, har bir suxondonning GMM ga tegishli bo'lgan testlash signalining ehtimolligi hisoblanadi. Eng yuqori ehtimolga ega bo'lgan suhandon testlash signalining dinamik sifatida aniqlanadi.

MFCC-GMM usulini istiqbolli natijalar bilan o'zbekcha nutq signalini shaxsni ovozi asosida tanib olishda qo'llanildi. Yondashuv aniqlik va mustahkamlik nuqtai nazaridan shaxsni ovozi asosida tanib olishning boshqa usullaridan ustun ekanligini ko'rsatdi. MFCC-GMM usulining muvaffaqiyati MFCC ning nutq signallaridan belgilarni ajratib olish qobiliyati va turli suxondonlar nutqini modellashtirishda GMM samaradorligi bilan bog'liq [6,8].

MFCC-GMM usuli o'zbek tilidagi nutq signalini shaxsni ovozi asosida tanib olishning samarali usuli ekanligi ko'rsatilgan. Ushbu

yondashuvning muvaffaqiyati shaxsni ovozi asosida tanib olish vazifalarida belgilarni ajratib olish va modellashtirish usullarining muhim hisoblanadi.

III.Natijalar

Maqolada 5 erkak (M) va 5 ayol (F) dan iborat. Har bir suxondon uchun 15 ta turli xil muhitdagi ma'lumotlar to'plamida amalga oshiriladi. Jadvallar o'zbek tilida nutq uchun MFCC-VQ usuli va MFCC-GMM usuli yordamida tanib olish natijasini berilgan. Suxondon uchun mos ravishda 1-jadval va 2-jadval mos ravishda matndan mustaqil va matnga bog'liq shaxsni ovozi asosida tanib olish natijalarini ko'rsatdi. Matnga bog'liq holda ma'ruzachini aniqlash uchun grafiklar 3-rasmda ko'rsatilgan. 4-rasmda matnga bog'liq holda tanib olish uchun grafik ko'rsatilgan. Ikkala usulning umumiy aniqligi 5-rasmda ko'rsatilgan.

1-jadval.

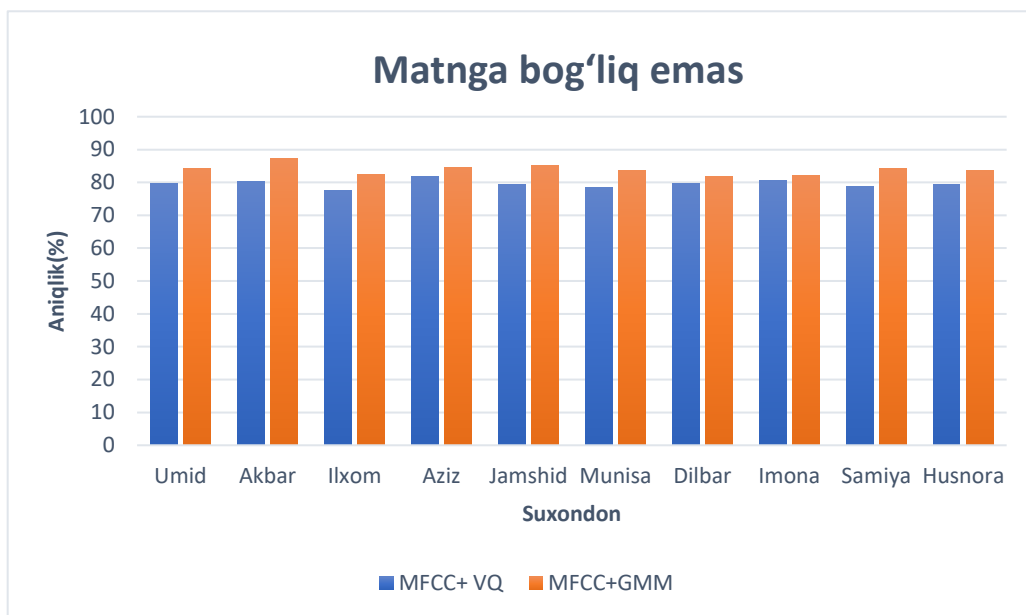
Matnga bo'g'liq bo'lmagan shaxsni ovozi asosida tanib olish.

Suxan-don	Algoritm	Umid	Akbar	Ilxom	Aziz	Jamshid	Munisa	Dilbar	Imona	Samiya	Husnora	Natija
Matnga bo'g'liq bo'lmagan	MFCC+VQ	79,6	80,3	77,6	81,7	79,3	78,5	79,6	80,6	78,7	79,3	79,5
	MFCC+GMM	84,3	87,4	82,4	84,6	85,1	83,7	81,9	82,2	84,2	83,7	83,9

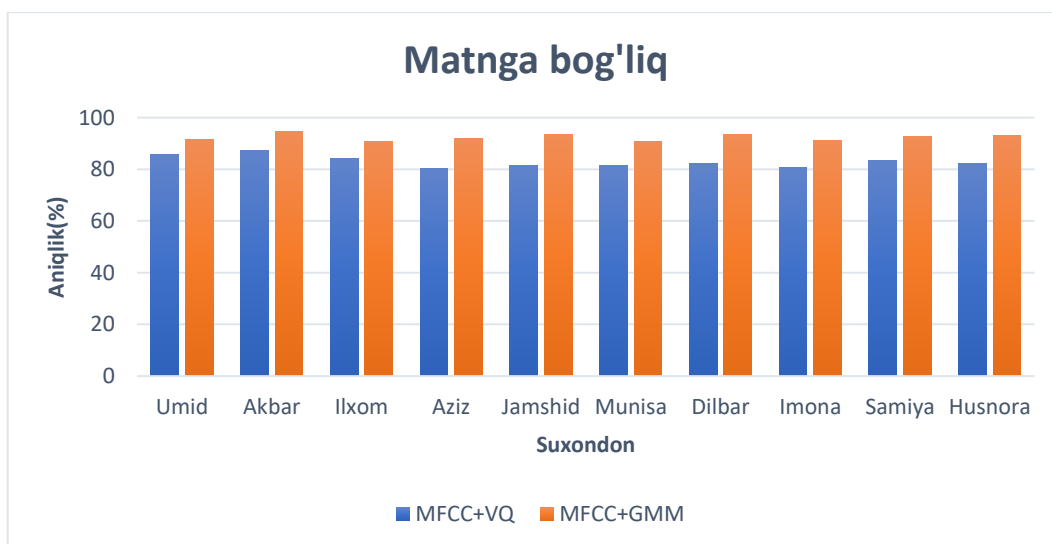
2-jadval.

Matnga bo'g'liq shaxsni ovozi asosida tanib olish.

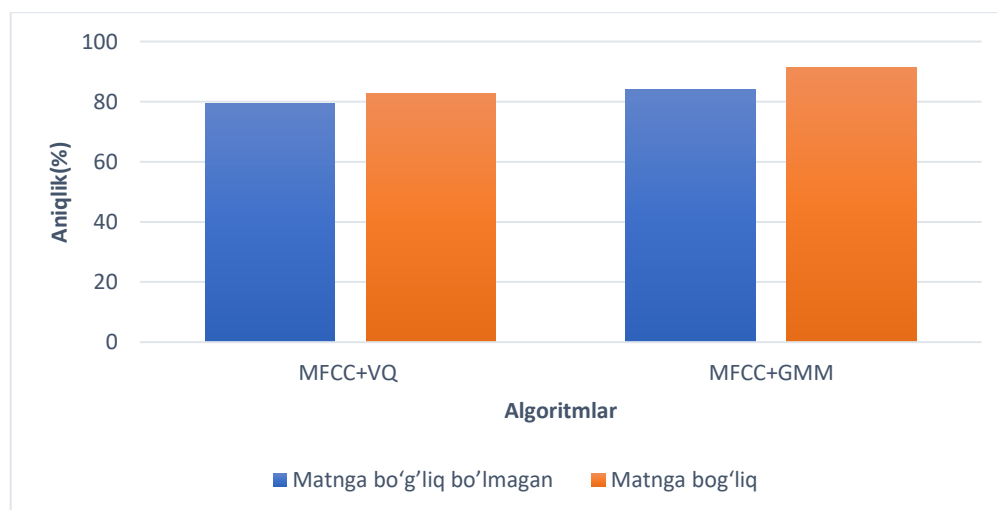
Suxan-don	Algoritm	Umid	Akbar	Ilxom	Aziz	Jamshid	Munisa	Dilbar	Imona	Samiya	Husnora	Natija
Matnga bog'liq	MFCC+VQ	85,7	87,4	84,2	80,5	81,6	81,4	82,2	80,7	83,5	82,1	82,8
	MFCC+GMM	91,4	94,5	90,9	91,9	93,6	90,6	93,4	91,3	92,8	92,9	91,3



3-rasm. Matnga bo'g'liq bo'lmagan shaxsni ovozi asosida tanib olish uchun har bir suxondonning aniqlik grafigi.



4-rasm. Matnga bo'g'liq shaxsni ovozi asosida tanib olish uchun har bir suxondonning aniqlik grafigi.



5-rasm. MFCC+VQ va MFCC+GMM algoritmlarining umumiy aniqlik grafigi.

IV.Xulosa

Natijalar shaxsni ovozi asosida identifikatsiyalash masalasini yechishda mel-chastotali kestral koeffitsientlar (MFCC) belgilar to'plamini ajratish algoritmi va vektorli kvantlash(VQ) modellashtirish algoritmi natijalariga qaraganda mel-chastotali kestral koeffitsientlar (MFCC) belgilar to'plamini ajratish algoritmi va Gauss aralashma modeli (GMM) modellashtirish algoritmidan foydalanish yuqori natija berishini ko'rsatdi.

V.Foydalanilgan adabiyotlar

1. Kenny, P., Ouellet, P., Dehak, N., Gupta, V., & Dumouchel, P. (2008). A study of interspeaker variability in speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(5), 980-988.
2. Bao, L., & Shen, X. (2016, April). Improved Gaussian mixture model and application in speaker recognition. In *Control, Automation and Robotics (ICCAR), 2016 2nd International Conference on* (pp. 387-390). IEEE.
3. Bimbot F., Bonastre, J. F., Fredouille C., Gravier G., Magrin-Chagnolleau, I., Meignier, S., Reynolds D. A. (2004). A tutorial on textindependent speaker verification. *EURASIP journal on applied signal processing*, 2004, 430-451.
4. Mamatov N., Niyozmatova N., Nurimov P., Samijonov A., Samijonov B. Development Software for Preprocessing Voice Signals In: Sandip A. Kale editor, *Advanced Research in Computer Engineering*, Pune: Grinrey Publications, 2021, pp. 53-66
5. Narzillo, M., Abdurashid, S., Parakhat, N., & Nilufar, N. (2019). Karakalpak speech recognition with CMU sphinx. *International Journal of Innovative Technology and Exploring Engineering*, 8(10), 2446–2448. <https://doi.org/10.35940/ijitee.J9524.0881019>
6. Wiedecke, B., Narzillo, M., Payazov, M., & Abdurashid, S. (2019). Acoustic signal analysis and identification. *International Journal of Innovative Technology and Exploring Engineering*, 8(10), 2440–2442. <https://doi.org/10.35940/ijitee.J9522.0881019>
7. Ниёзматова Н.А., Нуримов П.Б., Юлдошев Ю.Ш, Абдуллаев Ш.Ш. Автоматическая идентификация диктора по голосу с использованием векторного квантования. *Халқаро илмий-амалий анжуман, Андижон давлат университети, Андижон, 2019й. Б. 69-71*
8. Ниёзматова Н.А., Нуримов П.Б., Юлдошев Ю.Ш, Абдуллаев Ш.Ш. Шахсни овози асосида биометрик идентификациялаш. *Проблемы вычислительной и прикладной математики №2/1(40) 2022*