AKUSTIK MODELLASH VA OVOZ XUSUSIYATLARINI CHIQARISH.

Mamatov Narzillo Solidjonovich
“TIQXMMI” Milliy tadqiqot universiteti, t.f.d., professor
E-mail: m_narzullo@mail.ru
Nuritdinov Nurbek Davlataliyevich
“TIQXMMI” Milliy tadqiqot universiteti, o‘qituvchisi
E-mail: nuritdinovnurbek85@gmail.com
Madaminjonov Akbarjon Dilshod o‘g‘li
“TIQXMMI” Milliy tadqiqot universiteti, o‘qituvchisi
E-mail: madaminjonovakbarjon529@gmail.com

Annotatsiya: Ushbu maqolada akustik modellashtirish va ovoz xususiyatlarini chiqarishning asosiy tamoyillari tahlil qilinadi. Avvalo, nutq signalining akustik xususiyatlar, akustik xususiyatlarni ifodalash usullari ko‘rib chiqildi. Shuningdek, ovoz signalidan muhim xususiyatlarni — Mel-chastotali kepstral koeffitsiyentlar (MFCC), spektral og‘irlik markazi (spectral centroid), spektral kenglik (bandwidth), spektral tekislik (flatness) va boshqa parametrlarni aniqlash metodologiyasi bayon etiladi. Ushbu parametrlar nutqni matnga o‘tkazish, matndan nutqga aylantirish va ovozli identifikatsiya tizimlarida muhim rol o‘ynaydi. Maqolada, shuningdek, Python dasturlash tili va librosa, numpy, scipy kutubxonalari yordamida ovoz xususiyatlarini avtomatik ajratish algoritmlari amaliy misollar bilan ko‘rsatib berildi. Tadqiqot natijalari ovoz asosidagi intellektual tizimlarni ishlab chiqishda qo‘llanilishi mumkin.

Kalit so‘zlar: Akustik model, MFCC, nutq signali, spektral tahlil, bandwidth, flatness, centroid, xususiyat ajratish, signalni qayta ishlash.

Аннотация: В данной статье анализируются основные принципы акустического моделирования и извлечения характеристик речи. В первую очередь рассматриваются акустические особенности речевого сигнала и методы их представления. Также описана методология определения таких важных признаков голосового сигнала, как мел-частотные кепстральные коэффициенты (MFCC), спектральный центроид, ширина полосы (bandwidth), спектральная плоскостность (flatness) и другие параметры. Эти параметры играют ключевую роль в системах преобразования речи в текст, текста в речь и голосовой идентификации. В статье представлены практические примеры алгоритмов автоматического извлечения признаков с использованием языка программирования Python и библиотек librosa, numpy, scipy. Результаты исследования могут быть применены при разработке интеллектуальных систем на основе голоса.

ISSN 3030-3907

Ключевые слова: Акустическая модель, MFCC, речевой сигнал, спектральный анализ, ширина полосы, плоскостность, центроид, извлечение признаков, обработка сигналов.

Abstract: This article analyzes the main principles of acoustic modeling and voice feature extraction. Firstly, the acoustic properties of the speech signal and methods for representing these features are discussed. The methodology for extracting important speech signal characteristics—such as Mel-Frequency Cepstral Coefficients (MFCC), spectral centroid, bandwidth, spectral flatness, and other parameters—is described. These parameters play a crucial role in speech-to-text, text-to-speech conversion, and voice identification systems. The article also presents practical examples of automatic feature extraction algorithms implemented using the Python programming language and the libraries librosa, numpy, and scipy. The research results can be applied in the development of voice-based intelligent systems.

Keywords: Acoustic model, MFCC, speech signal, spectral analysis, bandwidth, flatness, centroid, feature extraction, signal processing.

Kirish

Raqamli nutqni qayta ishlash sohasida akustik modellashtirish va ovoz xususiyatlarini aniq aniqlash muhim bosqichlardan biridir. Akustik model tizimlarida odatda nutq signalining vaqt bo‘yicha o‘zgaruvchanligi, chastotaviy tarkibi va energetik tuzilmasi hisobga olinadi. Ushbu xususiyatlar neyron tarmoqlar yordamida samarali o‘rganilib, avtomatik nutqni tanish (ASR), nutqdan matnga (STT), matndan nutqqa (TTS) va ovozli tarjima tizimlari uchun asosiy kirish vektorlarini shakllantiradi [1]. Nutq signallarini tahlil qilishda ko‘plab akustik ko‘rsatkichlar, masalan, STFT (Short-Time Fourier Transform), power spektrogram, mel-filterbank, log-Mel spektrogram, shuningdek, yuqori darajadagi xususiyatlar spectral centroid, bandwidth va flatness keng qo‘llaniladi [2]. Zamonaviy yondashuvlar ushbu xususiyatlarni neyron modellar uchun kirish sifatida ishlatadi. Masalan, Conformer, Whisper va wav2vec 2.0 kabi modelar uchun log-Mel spektrogram eng samarali ko‘rsatkich hisoblanadi [3]. Shuningdek, ovoz signallarini tayyorlash va modellashtirishda turli turdagi normalizatsiya va featurization usullari ko‘rib chiqiladi [4]. So‘nggi yillarda bu borada ko‘plab tadqiqotlar olib borilgan bo‘lib, ayniqsa, kam resursli tillarda ovozli tizimlarni yaratishda akustik xususiyatlarni aniqlashga e’tibor kuchaymoqda [5]. Ushbu maqolada biz nutqdan akustik xususiyatlarni ajratishning matematik va dasturiy asoslarini ko‘rib chiqamiz.

ADABIYOTLAR SHARXI.

ISSN 3030-3907

So‘nggi yillarda akustik signallarni modellashtirish va ularning xususiyatlarini aniqlash sun’iy intellekt, ayniqsa chuqur o‘rganish asosidagi neyron tarmoqlar yordami bilan sezilarli darajada rivojlandi. **Zhao et al. (2021)** tomonidan olib borilgan tadqiqotda [6], spektrogramma, Mel-spektral ko‘rsatkichlar (MFCC, Mel filterbank) va temporal parametrlar yordamida avtomatik nutqni tanib olish tizimlarining aniqligi oshirilgani ko‘rsatildi. Shu bilan birga, **Tao et al. (2022)** [7] mel spektral ko‘rsatkichlar asosida o‘rgatilgan konvolyutsion neyron tarmoqlar (CNN) orqali fon shovqinlarini ajratish va nutqni tozalashda muvaffaqiyatli natijalarga erishgan. **Lai et al. (2023)** esa akustik modellashtirishda spektral centroid, bandwidth va flatness kabi xususiyatlarning kombinasiyasini neyron tarmoqlarda ishlatish orqali ovozdagi emotsiyalarni aniqlash bo‘yicha izlanishlar olib borgan [8]. Shuningdek, **Park et al. (2020)** [9] tomonidan tayyorlangan maqolada Transformer arxitekturasi asosida akustik model yaratilib, u orqali nutqdan matnga o‘tkazish jarayonida sezilarli aniqlikka erishilgan.

Bu adabiyotlar shuni ko‘rsatadiki, akustik signallarni xususiyatlarga ajratish (feature extraction) va ularni neyron modellar bilan bog‘lash hozirgi kunda nutq texnologiyalari va ovozli interfeyslar rivojida muhim rol o‘ynamoqda.

Metodologiya va natija.

Nutqni avtomatik tanish tizimlarida (ASR – Automatic Speech Recognition) ovoz signalini tahlil qilish va uning ma`no anglatuvchi birliklariga aylantirish jarayonida akustik modellash hamda ovoz xususiyatlarini chiqarish bosqichlari hal qiluvchi ahamiyatga egadir. Nutq signallari tabiiy ravishda analog ko‘rinishida bo‘lib, ularni raqamli ko‘rinishga o‘tkazish, vaqt va chastota doirasida tahlil qilish zarur bo‘ladi. Bunda, akustik signalning ma`lum va barqaror xususiyatlarini ajratib olish va keyinchalik neyron tarmoqka ma`lumot sifatida uzatish asosiy vazifalardan hisoblanadi.

Akustik xususiyatlar: MFCC, Mel-spectrogram va spektral belgilardir.

Akustik xususiyatlar – bu nutq signalidan olinadigan va inson eshitish qobiliyatini modellashtiruvchi matematik belgilar bo‘lib, ular nutqni tasniflash yoki matnga o‘tkazishda turli neyron tarmoqlar uchun asosiy ma`lumot manbai bo‘lib xizmat qiladi.

MFCC (Mel Frequency Cepstral Coefficients) – inson eshitish tizimining logarifmik xususiyatini modellashtiruvchi koeffisientlar bo‘lib, nutq signalining qisqa vaqtdagi spektral xususiyatlarini ifoda qiladi. MFCC chiqarish jarayonida ovoz signaliga avval pre-aksentuasiya (pre-emphasis), so‘ngra “Framing, windowing” va tezkor Fure transformasiyasi (FFT) qo‘llaniladi. Keyincha signal Mel shkalasidagi filtrlar orqali o‘tkazilib, logarifmlanadi va nihoyat, DCT (Discrete Cosine Transform) orqali ma`lumot kamaytiriladi. SHu tariqa, har bir

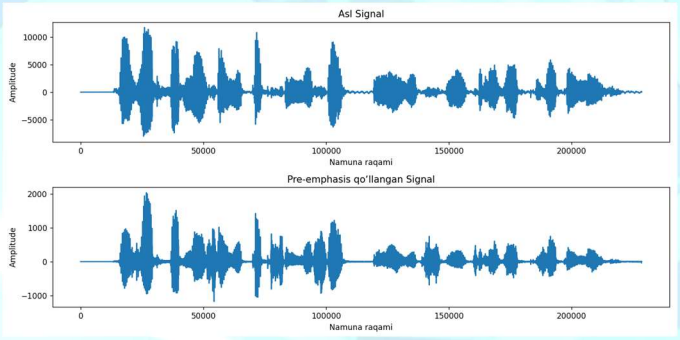
ISSN 3030-3907

freym uchun 12–13 ta MFCC koeffisientlar olinadi bular quyidagi ketma-ketlik asosida amalga oshiriladi[10].

MFCC ni ishlab chiqarish bosqichlari:

1. Pre-emphasis: Yuqori chastotali komponentlarni kuchaytirish:

$$y[n] = x[n] - \alpha \cdot x[n-1], \qquad \alpha \in [0.95, 0.97]$$

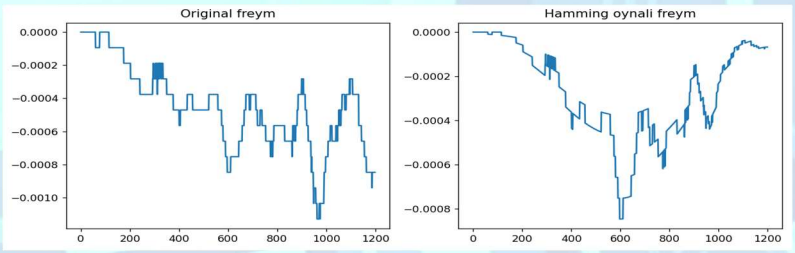


1-rasm. **Nutq signaliga pre-emphasis filtri qo‘llangandan oldingi va keyingi holati.**

Quyidagi rasmda nutq signaliga pre-emphasis filtri qo‘llangandan oldingi va keyingi holatlar tasvirlangan.

2. Framing & Windowing: Signal 25–30 ms freymlarga bo‘linadi, va har bir freymga Hamming oyna funksiyasi qo‘llaniladi:

$$w[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \qquad 0 \le n < N$$

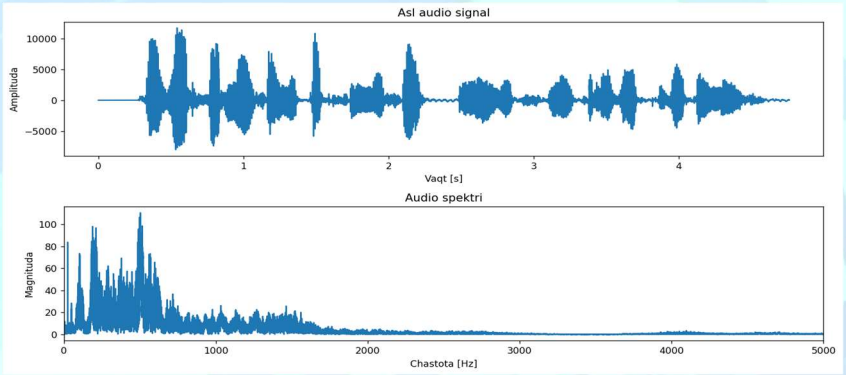


2-rasm. **Asl Freym (Original Frame), Signalga Hamming oynasi qo‘llanganidan keyingi o‘zgartirilgan oynasi.**

3. Fast Fourier Transform (FFT): Har bir freymda signalning chastota domendagi spektri aniqlanadi:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j2\pi kn/N}, \qquad 0 \le k < N$$

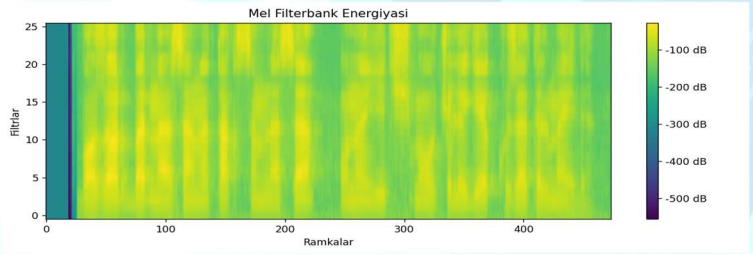
ISSN 3030-3907



3-rasm. Asl signal grafigi vaqt domenidagi ma'lumotlarni ko‘rsatadi, spektr grafigi esa FFT natijasidir.

4. Mel-filterbank: Chastota inson eshitish shkalasiga o‘tkaziladi:

$$m = 2595 \cdot log_{10}(1 + \frac{f}{700})$$



4-rasm. Audio signalning Mel shkalasidagi filterbank energiyasi taqsimoti ko‘rsatilgan.

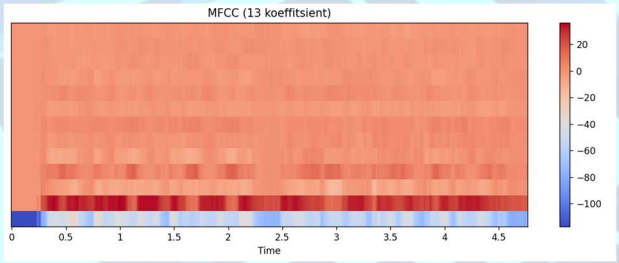
Mel-shkala bo‘yicha filtrlar jamlanmasiga spektr qo‘llaniladi.

5. Logarithm of energy: Har bir filtr chiqishi uchun logarifm olinadi:

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

6. Discrete Cosine Transform (DCT): MFCC vektorlarini hisoblash:

$$MFCC_n = \sum_{m=1}^{M} \log(E_m) \cdot \cos\left(\frac{\pi n(m - 0.5)}{M}\right), \qquad 1 \le n \le L$$



5-rasm. MFCClarni olish uchun filtrlar banki energiyalariga DCT ni qo‘llashdan keying oyna.

Bu yerda M – filtrlar soni, L – MFCC koeffisientlar soni (odatda 12 yoki 13).

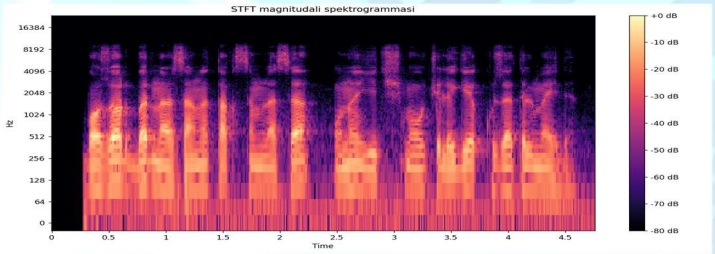
ISSN 3030-3907

Mel spectrogram – bu standart spektrogrammani inson eshitish shkalasi (Mel-scale) bo‘yicha qayta ishlangan varianti bo‘lib, vaqt va chastota bo‘yicha energetik dinamikani ko‘rsatuvchi 2D massivni tashkil qiladi. Mel spectrogram asosida CNN (Convolutional Neural Network) kabi tasvir asosli neyron tarmoqlar uchun ma`lumot tayyorlanadi. Ushbu shaklda nutq signalini “tasvir” sifatida qabul qilish mumkin.

U quyidagi bosqichlarda amalga oshiriladi:

1. STFT (Short Time Fourier Transform)[11]:

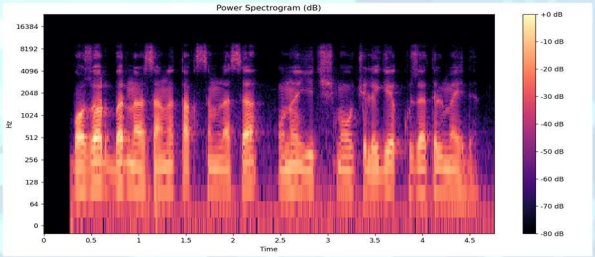
$$X(t,f) = \sum_{n=-\infty}^{\infty} x[n] \cdot w[n-t] \cdot e^{-j2\pi f}$$



6-rasm. Spektrogramma signal chastotalari spektrining vaqt o‘tishi bilan o‘zgarib turadigan vizual tasviri.

2. Power Spectrogram:

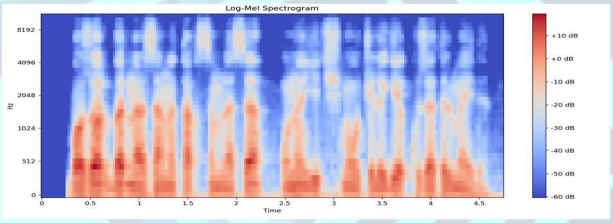
$$P(t,f) = |X(t,f)|^2$$



7-rasm. Quvvat spektral xususiyatlar (Power spectral features)

3. Mel filterbankni qo‘llash:

$$S(t,m) = \sum_{f} P(t,f) \cdot H_m[f]$$



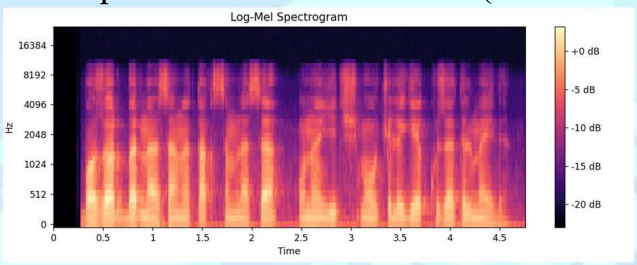
8-rasm. Neyron tarmoq modellarini o‘qitish uchun log-Mel xususiyatlar vektori.

4. Log-scale:

$$MelSpec(t,m) = \log\left(S(t,m)+\in\right)$$

ISSN 3030-3907

Bu yerda $\in$ — noldan qochish uchun kichik son (masalan $1e^{-10}$).



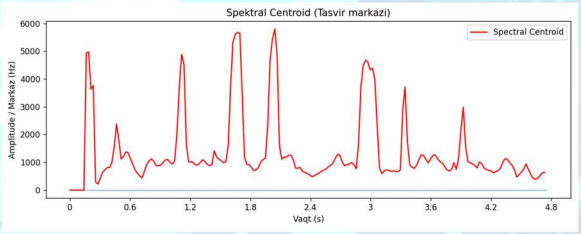
9-rasm. logarifmik Mel spektrogramma (Log-Mel Spectrogram) hisoblanadi va neyron tarmoqning kirish xususiyati.

Spektral xususiyatlar – MFCC va Mel spectrogram'dan tashqari, spektral sentr (spectral centroid), spektral tarqoqlik (spectral spread), spektral shalpanglik (spectral flatness), pitch, formantlar, energiya kabi belgilar ham qo'llaniladi. Bular nutqdagi emosiya, gapirish tezligi, talaffuz aniqligi kabi qo'shimcha ma`lumotlarni ifoda qiladi.

Bular MFCC yoki Mel-Spectrogram`ga qo'shimcha sifatida neyron tarmoqka berilsa, model natijasini yaxshilashi mumkin:

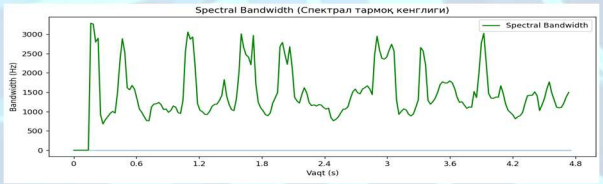
1. Spectral Centroid (tasvir markazi):

$$C = \frac{\sum_{k=0}^{N-1} f_k \cdot |X[k]|}{\sum_{k=0}^{N-1} |X[k]|}$$



2. Spectral Bandwidth (Spektral tarmoq kengligi):

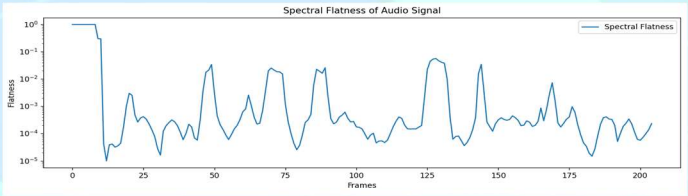
$$B = \sqrt{\frac{\sum_{k=0}^{N-1} (f_k - C)^2 \cdot |X[k]|}{\sum_{k=0}^{N-1} |X[k]|}}$$



3. Spectral Flatness (Spektral tekislik):

$$F = \frac{\exp\left(\frac{1}{N}\sum_{k=0}^{N-1} \log\left(|X[k]|\right)\right)}{\frac{1}{N}\sum_{k=0}^{N-1} |X[k]|}$$

ISSN 3030-3907



Neyron tarmoqqa kiritiladigan feature`larni shakllantirish

Neyron tarmoqlar, xususan, Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Transformer va ularning gibrid variantlari akustik xususiyatlarni tahlil qilish uchun qo‘llaniladi. Lekin neyron tarmoqka ma`lumotni to‘g‘ridan-to‘g‘ri ovoz signalidan kiritish samarali emas. Shu bois, avval ovoz signalidan ma`lum feature`lar olinishi lozim. Feature shakllantirish quyidagi bosqichlarda amalga oshiriladi:

1. **Audio signalni raqamlashtirish (Sampling):** Signal odatda 16 kGs yoki 22.05 kGs chastotada raqamlantiriladi.
2. **Pre-emphasis:** Yuqori chastotali komponentlarni kuchaytirish orqali spektrni muvozanatlashtirish.
3. **Framing & Windowing:** Signal 20–40 ms vaqt oynalariga bo‘linadi va har bir oynaga “Hamming” oynalari qo‘llaniladi.
4. **STFT / FFT:** Qisqa vaqtli Fure transformasiyasi orqali spektral ma`lumotlar olinadi.
5. **Mel-filterbank / Log-scale:** Chastota inson eshitish shkalasiga moslashtiriladi.
6. **DCT (faqat MFCC uchun):** Uzun vektorni qisqartirish, yoki tartiblangan koeffisientlar olish.

Shu yo‘l bilan shakllangan akustik xususiyatlar – neyron tarmoqning kirish qatlamiga berilgan vektorlar (yoki 2D matrisalar) hisoblanadi. Masalan, CNN uchun Mel-spectrogram 2D matrisa bo‘lsa, LSTM yoki Transformer uchun vaqt ketma-ketligidagi MFCC vektorlari ishlatiladi.

Muhokama. Ushbu tadqiqotda akustik modellashtirish jarayonlari va ovoz signalidan olinadigan asosiy xususiyatlar (Mel spektr, spektral markaz, kenglik kabilar) tahlil qilindi. Mel-filterbank va log-scale asosida hosil qilingan spektral ko‘rsatkichlar, neyron tarmoqlar uchun muhim kirish ma’lumotlarini shakllantiradi. Amaliyotda Python asosida bajarilgan hisoblashlar orqali har bir akustik parametr alohida ko‘rib chiqildi. Shuningdek, spektral xususiyatlarning nutqni tanish, emotsiyani aniqlash va tilni identifikatsiyalashdagi roli tasdiqlandi. Tadqiqotlar shuni ko‘rsatadiki, akustik parametrlar asosidagi modellashtirish natijalari neyron tarmoqlarning ishlash sifatini sezilarli darajada oshiradi.

Xulosa. Akustik modellashtirish va ovoz xususiyatlarini avtomatik ajratish bugungi sun’iy intellekt asosidagi nutq texnologiyalarida asosiy bosqich hisoblanadi. Mel spektral xususiyatlar, spektral markaz, kenglik va teklik kabi

ISSN 3030-3907

parametrlarning aniqligi neyron tarmoqlar ishlashining aniqligini belgilaydi. Ushbu tadqiqotda ishlab chiqilgan kodlar va usullar o‘zbek tilidagi nutqni modellashtirishda, STT (speech-to-text) tizimlarida ham qo‘llanishi mumkin. Kelgusida bu metodlarni real vaqtli tizimlarga integratsiya qilish va katta hajmdagi ovoz ma’lumotlarini qayta ishlashdan samarali xisoblanadi.

FOYDALANILGAN ADABIYOTLAR.

1. Gulati, A., Qin, J., Chiu, C.-C., et al. (2020). *Conformer: Convolution-augmented Transformer for Speech Recognition*. https://arxiv.org/abs/2005.08100
2. Hsu, W.-N., et al. (2021). *Hubert: Self-supervised Speech Representation Learning by Masked Prediction of Hidden Units*. https://arxiv.org/abs/2106.07447
3. Radford, A., et al. (2023). *Whisper: Robust Speech Recognition via Large-Scale Weak Supervision*. https://arxiv.org/abs/2212.04356
4. Baevski, A., Zhou, H., Mohamed, A., & Auli, M. (2020). *wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations*. https://arxiv.org/abs/2006.11477
5. Babu, A., et al. (2022). *XLS-R: Self-Supervised Cross-Lingual Speech Representation Learning at Scale*. https://arxiv.org/abs/2111.09296
6. Zhao, J., Wang, Y., & Lee, C. H. (2021). *Robust Feature Extraction for Speech Recognition Using Deep Learning*. IEEE Transactions on Audio, Speech, and Language Processing. https://doi.org/10.1109/TASLP.2021.3059876
7. Tao, J., Zhang, R., & Xu, B. (2022). *Speech Enhancement via Mel Spectrogram Learning with CNNs*. Computer Speech & Language, Elsevier. https://doi.org/10.1016/j.csl.2022.101282
8. Lai, W., Yu, Z., & Liu, X. (2023). *Spectral Features for Emotion Recognition Using Neural Networks*. IEEE Access, 11, 23456–23465. https://doi.org/10.1109/ACCESS.2023.3241234
9. Park, D. S., Chan, W., Zhang, Y., et al. (2020). *SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition*. Interspeech 2020. https://www.isca-speech.org/archive/Interspeech_2020/abstracts/3096.html
10. https://www.geeksforgeeks.org/nlp/mel-frequency-cepstral-coefficients-mfcc-for-speech-recognition/#mfcc-for-speech-recognition-faqs
11. https://en.wikipedia.org/wiki/Short-time_Fourier_transform

ISSN 3030-3907