

НУТҚ ТОВУШЛАРИНИ ТАНИШ АЛГОРИТМИНИ ИШЛАБ ЧИҚИШ

Хужаяров И.Ш, Очилов М.М.

Ушбу мақолада ўзбек тили нутқ товушларини таниш усулларидан бири келтирилган. Биринчи босқичда нутқ сигналларини частота соҳасида қайта ишлаш асосида спектрограмма тасвирлари ҳосил қилинган. Иккинчи босқичда ҳосил қилинган спектрограмма тасвирларини дастлабки қайта ишлаш асосида параметрлаш ва улар асосида нутқ товушларини бир бири билан солиштиришнинг корреляцион таҳлил амалга оширилган.

Таянч иборалар: Нутқни таниш, фрейм, дискрет косинус алмаштириш, спектрограмма, тасвир, корреляцион солиштириш.

В данной статье описан один из способов распознавания речи узбекского языка. На первом этапе была создана спектрограмма речевого сигнала на основе обработки речевых сигналов в частотной области. На втором этапе была выполнена параметризация на спектрограмме и проведен корреляционный анализ с использованием сравнения речевых звуков.

Ключевые слово: распознавание речи, фрейм, дискретное косинусное преобразование, спектрограмма, изображение, сравнение корреляций.

This article outlines the algorithm for speech recognition. The article outlines the stages of the speech spectrogram image. The speech spectrum image is the richest parameter that characterizes speech. High accuracy can be achieved by recognizing the speech in other ways by processing the resulting spectrogram image.

Today, many speech recognition algorithms use the speech spectrogram. An important aspect of familiarity with the speech spectrogram is the transition from one-dimensional signal to two-dimensional. The main parameters characterizing the speech during the transition are separated using spectral transformation methods. As the basic parameters, the main tone frequency serves as a form.

There are many effective ways to get to knowing speech, and most commonly, this speech is familiar to them through the processing of the spectrogram image. We have already mentioned the familiar ways of using point-to-speech sound, speeches of the speech spectrogram, and the steps to reproduce the image.

The article discusses the topics of modern speech recognition algorithms, initial processing methods for speech signals, algorithm of speech spectrogram image generation, stages of spectrum image processing (filtration, site allocation, correlation comparison). Any speech (word or phrase) consists of small phonemes (letters or combination of letters). Given this fact, this article considers the most

commonly used phonemes in the Uzbek language, namely, the development of six vowel letters.

Key words: speech recognition, frame, discrete cosine transform, spectrogram, image, correlation comparison.

I. КИРИШ

Ҳозирги кунда нутқ сигналларини автоматик таниб олиш масалалари жуда муҳим муаммолардан биридир. Нутқ сигналларини таниш масалалари замонавий ахборот технологиялари, телекоммуникация ва рақамли алоқа тизимлари, IP-телефония ва видеоконференция, ақилли уй лойиҳалари, бир суткада минглаб чақирувларга автоматик режимда жавоб берувчи автоответчиклар(мисол учун call центрлар, авиабилетларни сотиш тизимлари), саноат ва ишлаб чиқариш ҳамда имконияти чекланган инсонлар учун овозли бўйруқ бериш орқали техникани бошқариш соҳаларидаги масалаларни ечишда қўлланилмоқда.

Нутқ товушини танишни жуда кўп самарали усуллари мавжуд бўлиб улардан энг оммалашгани бу нутқ спектограмма тасвирини қайта ишлаш орқали уни таниш ҳисобланади[3-17]. Мақолада нутқ товушини таниш усуллари, нутқ спектограмма тасвирини ҳосил қилиш босқичлари, ҳосил килинган тасвирини қайта ишлаш қадамлари ҳақида тўхталиб ўтганмиз.

Ҳар қандай нутқ товуши (сўз ёки гап) албатта кичик фонема(ҳарф ёки ва ҳарф бирикмаси) дан ташкил топган. Шунини инобатга олиб, ушбу мақолада ўзбек тилида мавжуд сўзларда энг кўп учрайдиган фонемалар яъни, олти унли ҳарфларни танишга мўлжалланган алгоритм ишлаб чиқишни кўриб ўтганмиз.

II. АСОСИЙ ҚИСМ

Нутқ сигналларини автоматик таниш бўйича изланишлар ўтган асрнинг 60-йилларидан бошланиб то ҳозиргача интенсив равишда давом этиб келмоқда. Ҳозиргача олинган натижалар яъни яратилган усул ва алгоритмлар реал шароитларда ишловчи нутқни таниш тизимларида(НТТ) етарлича самарадорлик кўрстакларига эга эмас. Шундай экан НТТнинг самарадорлик кўрсаткичларини яхшилаш учун турли нутқ сигналларини таниб олиш усуллари биргаликда қўллаш мақсадга мувофиқ саналади[1,5].

Бугунги кунда бир қанча нутқ сигналларини таниш усуллари мавжуд бўлиб, қўйида уларнинг бир нечта асосий синфларига статистик таҳлил асосида ётувчи усуллар(яширин Марков моделлари)[5,7,13], спектрал соҳада сигнал элементларини қайта ишлаш усуллари(Фурье таҳлил, Вейвлет таҳлил) [6], нейрон тармоқларига асосланган усуллар(ўз-ўзини ўқитиш ва такомиллаштирувчи алгоритмлар ва усуллар)[11] киради .

Бундан кўриниб турибдики, юқорида келтирилган битта синф усуллари

асосида яхши самарадорликни берувчи НТТ ни яратиб бўлмайди. Шунинг учун бугунги кунда нутқни сигналларини таниш масалаларида юқорида келтирилган усуллар синфини биргаликда қўллаш асосида олинган натижаларига интеллектуал ишлов бериш НТТ нинг самарадорлик кўрсаткичларини оширади[6].

Нутқ сигналларини қайта ишлаш – бу нутқ сигналлари устида филтрлаш, кодлаш, фойдали ахборотларни ажратиб олиш, сиқиш ва тиклаш алгоритмларини амалга оширувчи фан соҳасидир. Нутқни қайта ишлаш ва таниб олиш тизимлари қўйидаги масалаларни ўз ичига олади:

- филтрлаш ва шовқинларни камайтириш;
- сигналдаги ахборотли қисмларни сегментациялаш;
- ахборотли параметрларни аниқлаш;
- Таниб олиш.

Ҳар бир нутқ сигналларини қайта ишлаш масалаларини фақат муайян усуллар ёрдамида амалга ошириш мумкин. Қўлланилиш соҳасига қараб ишлов бериш усулларини 3 та соҳага ажратилади: частотали, вақт ва частота-вақт соҳаси[1].

Вақт соҳасида ишлов бериш усуллари нутқ сигналидан характерли нуқталарни аниқлайди. Аниқланган характерли нуқталар нутқ сигналларини таҳлил қилиш учун ишлатилади. Характерли нуқталар сифатида сигналнинг максимум(минимум) қийматлари ва оси ўқини кесиб ўтиш моментлари киради. Вақт соҳасида ишлов бериш усулларининг асосий камчилиги сифатида қўйидагиларни келтиришимиз мумкин:

- кўп сондаги параметрлар билан ишлаш қийинчилиги;
- олинган параметрлардан керакли ва ишончли бўлганларини ажратишнинг ноқулайлиги;
- танишда ва синфлаштиришда алгоритмларнинг аниқлигини пастлиги ва ҳк.

Частота-вақт (спектрал) соҳасида сигналларни қайта ишлаш вақт соҳасида учрайдиган муаммоларни ҳал қилиш билан бирга бизга қўйидаги масалаларни ечишда ёрдам беради[6]:

- сигнални кичик бўлак(фрейм)ларга ажратиш ва ҳар бир бўлакни алоҳида қайта ишлаш;
- спектрал ишлов бериш соҳаси сигналлар устида бажариладиган мураккаб алгоритмлар синфини кенгайтиради ва уларда қўлланиладиган сонли усуллар, бир вақтнинг ўзида бир нечта амалларни(филтрлаш, сиқиш ва ҳк) бажариш имкониятини тақдим этади.;
- сегментлаш натижасида ҳар бир бўлак учун маълумотлар устида параллелизмни қўллаш имкониятини яратилиши.

Қўйида нутқ сигналларни спектрал соҳада қайта ишлаш натижасида ҳосил қилинган спектограмма тасвирини нуқтни таниш масаласида

қўлланишини кўриб ўтамиз.

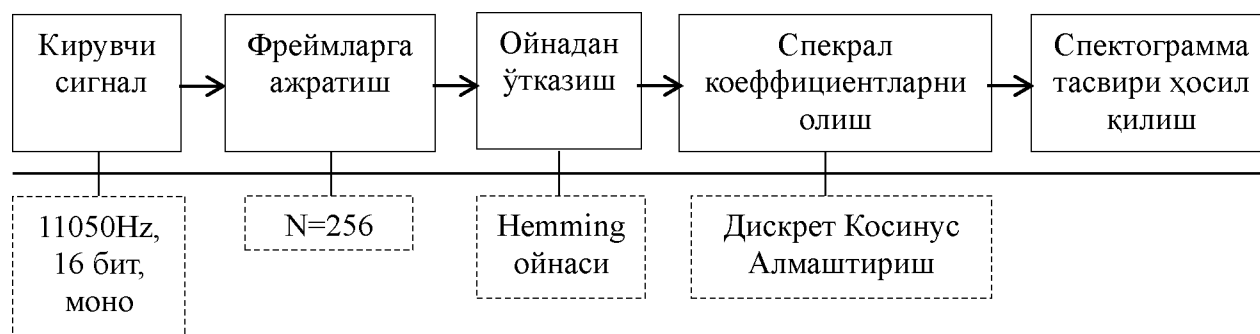
Юқорида айтиб ўтганимиздек нутқни танишда бир нечта таниш алгоритмлари ишлатилади. Шулар орасида энг самарали усулларида бири бу нутқ спектрограмма тасвирини ҳосил қилиш ва таҳлил қилиш асосида уни таниш ҳисобланади. Ҳозирги кунда яратилаётган нутқни таниш алгоритмларнинг аксарияти нутқ спектрограмма тасвирни қайта ишлашга асосланган. Негаки нутқ спектрограмма тасвири уни энг яхши тавсифлайдиган восита ҳисобланади[8-18].

Шу ўринда нутқ спектрограмма тасвирини танлашдан асосий мақсадларни келтириб ўтадиган бўлсак, улар қуйидагилар:

- нутқ спектрограмма тасвирни сигналдан фарқли ўлароқ икки ўлчамли кўринишда. Бунинг натижасида икки ўлчамли сигналларни қайта ишлаш усулларида ҳам бемалол фойдалансак бўлади;
- нутқ спектрограмма тасвирни ҳосил қилиш орқали масалани тасвирларни таниш масаласига айлантиришга эришилади;
- спектрограмма тасвирини тескари алгоритм ёрдамида яна товушга ўзгартира олиш имконияти;
- спектрограмма тасвири ёрдамида асосий параметрик кўрсаткичлар (асосий, биринчи ва иккинчи форманталар жойлашуви) ни аниқлаш имкониятининг мавжудлиги ва бошқалар.

Нутқ товушини спектал анализ қилишда спектрограмма тасвирни таҳлил қилиш самарали усуллардан бири ҳисобланади. Тасвир ёрдамида нафақат нутқни таниш ва шу билан бир қаторда сўзловчи ҳақида кўплаб маълумотларни олишимиз мумкин бўлади. Масалан сўзловчининг жинсини, ёшини ва ҳаттоки эҳмоционал ҳолатини ҳам аниқлашимиз мумкин[7].

Қўйида нутқ товуши спектрограммасини ҳосил қилиш алгоритми келтирилган. Унга кўра нутқ спектрограммасини ҳосил қилиш қўйидаги босқичларда амалга оширилади (1-расм).



1-расм. Нутқ спектрограммасини ҳосил қилиш босқичлари.

Юқоридаги 1-расмни кўриб ўтадиган бўлсак, кирувчи сигнал бир нечта алмаштиришлардан ўтиб спектрограмма тасвири ҳосил қилинади.

Илк босқичлардан бири сифатида кирувчи сигнални бўлақларга бўлиш

келтирилган. Бунда сигналнинг хусусиятидан келиб чиқиши керак, одатда фрейм узунлиги этиб $N=256$ (нутқ товишининг 25 миллисекунддаги қийматлар сони. Мисол учун $N=256$, агар частота 11050 Hz бўлса) танланади. Ҳар бир бўлак алоҳида-алоҳида ойнадан ўтказилади. Ойна сифатида Hemming ойнаси танланган бўлиб, у сигнални филтрлайди ва ундаги узилишларни йўқотишга олиб келади. Hemming ойнасини қуйидаги (1) формула орқали ифодалашимиз мумкин.

$$H(n) = 0.56 + 0.46 * \cos\left(\frac{2*\pi*n}{N}\right), n=1..N \quad (1)$$

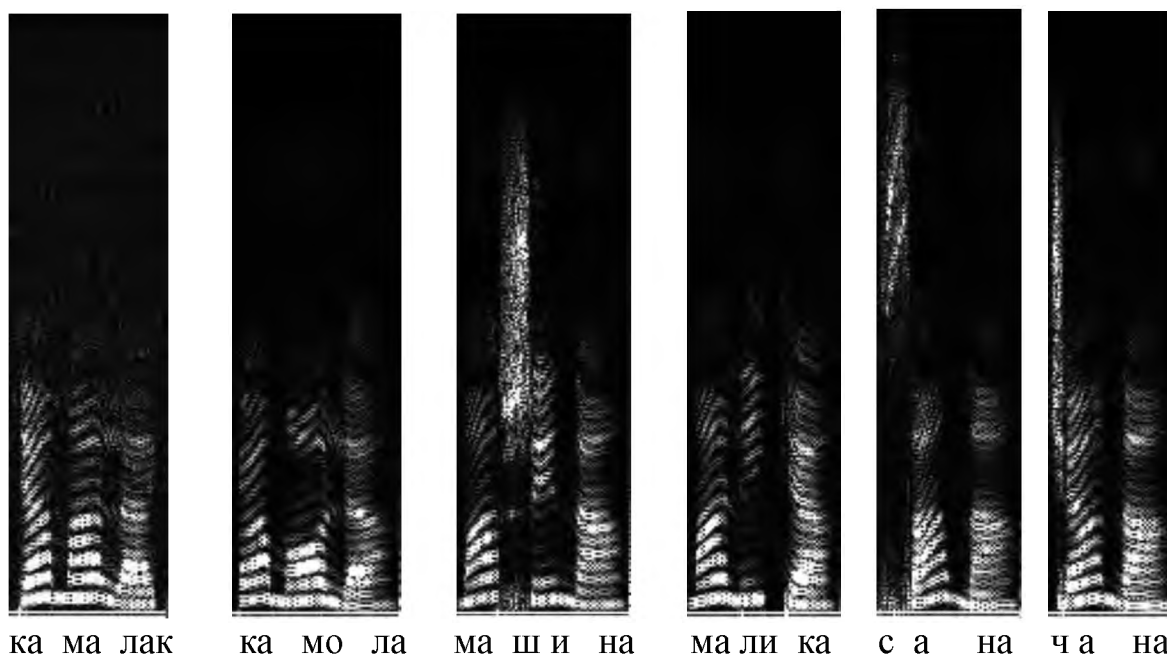
Hemming ойнасидан ўтказилганидан сўнг ҳар бир сигналнинг бўлаги учун спектрал коэффицентлар аниқланади. Одатда Дискрет Фуре Алмаштириш (ДФА), Тезкор Фуре Алмаштириш (ТФА) ёки Дискрет Косинус Алмаштириш (ДКА) лар орқали спектрал коэффицентлар аниқланади[2,3]. Биз ДКА дан фойдаланганмиз ва у (2) формулада келтирилган.

$$Y(x) = w(n) * \sum_{n=1}^N x(n) * \frac{\cos \pi(2n - 1)(k - 1)}{2N}, \quad k = 1..N \quad (2)$$

Бу ерда: $w(n) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leq k \leq N \end{cases}$

Ҳар бир Hemming ойнасидан ўтган фрейм, ДКА орқали частота соҳасида ёйилади, яъни спектрал коэффицентлар аниқланади. Олинган коэффицентлар орқали спектрограмма тасвири ҳосил қилинади. Тасвирни ҳосил қилиш бир мунча қийинчилик талаб қилади. Яъни биз топган ҳар бир бўлакдаги коэффециентлар 0 дан 255 ораликқа келтирилади. Бу эса кулранг тасвирни қиймати сифатида қаралади.

Мисол сифатида биз бир нечта сўзларни спектрограммасини ҳосил қилиб кўрганмиз. Улар қуйидаги кўринишда ҳосил бўлди:



2-расм. Нутқ спектрограмма тасвири

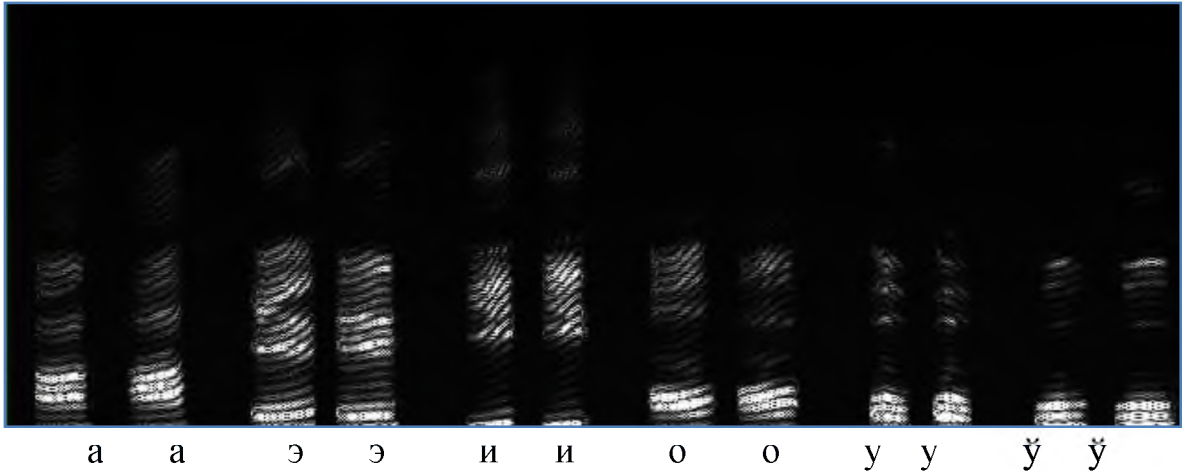
Спектрограмма тасвир 256xM ўлчамда бўлади. Бу ерда M бўлақлар сони. Спектрограмма тасвирининг i - устунига i - бўлақ қийматлари тўғри келади.

Бу тасвирлардан кўришимиз мумкинки, турли хил сўзнинг спектрограммаси ҳам турлича бўлади. Мисолларимизда талаффузи бир-бирига яқин бўлган сўзлар танланган, Лекин шунга қарамасдан уларнинг кўриниши бир-биридан фарқланади. “Камалак” ва “камола” сўзларининг спектрограмма тасвирига эътибор берадиган бўлсак, “ка” бўғини бир-бирига ўхшайди, “машина” ва “малика” сўзларида эса “ма” бўғини, “сана” ва “чана” сўзларида эса “на” бўғини бир-бирига мос келмоқда. Лекин бошқа қисмлари бир-биридан фарқ қилмоқда. Нутқ товуши спектрограммаси нутқ эгасига боғлиқ равишда фарқ қилиб боради. Мисол учун ёш болаларда бошқача бўлса, аёлларда бошқача бўлади. Лекин тасвирда кўриниб турган ρ ik (чўкки) лар унчалик фарқ қилмайди. Аммо унинг ёркинлиги(сон қиймати) паст ёки баланд бўлиши мумкин.

Энди ҳосил бўлган спектрограмма тасвири ёрдамида нутқ товушини таниш масаласини кўрадиган бўлсак, спектрограмма тасвирини таҳлил қилишни турли хил алгоритмлари мавжуд. Масалан олинган спектрограмма тасвирида форманталар жойлашуви, форманаталар орадидаги масофа, асосий форманталар учрайдиган частота ва бошқа кўрсаткичлар аниқланиб шулар ёрдамида таниш амалга оширилган. Биз тақдим қилган алгоритм фақат ўзбек тилидаги унли ҳарфлар учун мўлжалланган. Қўйида ўзбек тилидаги олти унли ҳарфларни спектрограмма тасвирини таҳлил қилиш орқали улар бир-бирларига қанчалик ўхшашлигини кўриб ўтамыз.

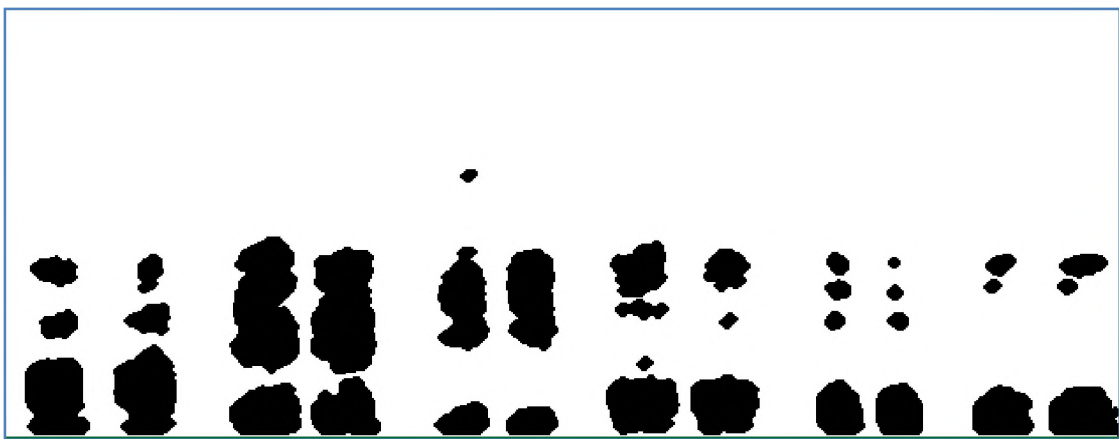
Нутқ сигналларини спектрограмма тасвири ёрдамида таниш бир неча кадамда амалга оширилади. Уларни кетма-кет кўриб ўтамиз.

1-кадам. Юқорида таърифланган дастур ёрдамида олти унли ҳарфлар ‘а’, ‘е’, ‘и’, ‘о’, ‘у’, ‘ў’ ларнинг спектрограмма тасвири хосил қилиб олдик. Қўйидаги расмда олти унли ҳарфлар кетма-кет икки мартадан такрорлаб талаффуз қилиниб уларни спектрограмма тасвири келтирилган:



3-расм. Ҳарфлар спектрограмма тасвири

2-кадам. Бу кадамда тасвири кулранг кўринишдан иккилик бинар яъни оқ қора кўринишга келтирилади ва тасвири бир неча усулларда филтрланади. Сўнгра қуйидаги тасвир хосил бўлади:



4-расм. Филтрлашлардан сўнг ҳосил бўлган спектрограмма тасвири

3-кадам. Сўнгра ҳосил бўлган тасвирда биз олти унли ҳарфни бира тўла спектрограмма тасвири хосил қилишни кўриб ўтдик. Аслида ҳар бир ҳарф алоҳида ҳолда ушбу кадамларган ўтказилиши керак. Мисол сифатида иккита унли ҳарфлар «а» ва «э» ҳарфларни спектрограмма тасвинини хосил қилиб кўрганмиз. Улар қуйидагича:



5-расм. «а» ва «э» ҳарфларининг спектрограмма тасвири

Энди биз ушбу иккита унли ҳарфларни бир-бирдан фарқлашни кўриб ўтамиз. Бу учун биз корреляция формуласидан фойдаланамиз. У қуйидаги формулада келтирилган.

$$f_{xy} = \left| \frac{\sum_i (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_i (x_i - M_x)^2} * \sqrt{\sum_i (y_i - M_y)^2}} \right| \quad (3)$$

Бу ерда X ва Y иккита кирувчи вектор бўлиб улар тасвирдан қандай олинади деган савол туғилади. Бунинг учун биз тасвирни қўйидаги кўринишда бўлақларга бўлиб чиқамиз. Бизга маълумки спектрограмма тасвири бўйи 256 пиксел бўлади. Лекин тасвирдан кўришимиз мумкин унинг вертикал бўйича 1-128 пикселларида қийматлар йўқ шунинг учун уларни ташлаб юборамиз ва қолган 129-256 пикселлар 8 пиксел интервал билан бўлақларга бўлиниб чиқилади. Ҳар бўлақдаги қора нуқталар сони аниқланади бу эса кирувчи X ва Y векторларни қийматлари бўлиб хизмат қилади.

Мисол учун:

$$X=[0,0,43,108,18,0,37,98,35,10,215,264,264,241,231,150]$$

$$Y=[0,105,256,254,274,261,242,278,274,295,169,12,198,312,321,152]$$

Бу ерда X векторда мос равишда 0 129-136, 0 137-144, 43 145-151, 108 152-159 пикселлар оралиғидаги қора нуқталар сони.

Ушбу алгоритм асосида олти унли ҳарфларни бир-биридан ва ўзинг иккинчи мартаба талафуз қилингани билан солиштириш орқали қуйидаги жадвални қўлга киритдик.

1-жадвал

Унли ҳарфларни ўхшашлик кўрсаткичлари

Унли ҳарфлар	A ₂	E ₂	I ₂	O ₂	Y ₂	Ў ₂
A ₁	93%	61%	25%	64%	56%	72%

E₁	42%	92%	68%	31%	34%	52%
И₁	13%	71%	94%	16%	7%	32%
O₁	72%	48%	10%	93%	63%	63%
У₁	62%	45%	12%	77%	93%	81%
Ў₁	81%	48%	19%	80%	84%	91%

Жадвалдан шуни кўришимиз мумкинки, инсон талаффуз қилганда 100% олдин талаффуз қилганидек бўлмайди. Мисол учун «а» ҳарфини оладиган бўлсак ўхшашлик 93% ни ташкил қилади. Бу солиштириш натижаси жиҳатидан «а» ҳарфи эканлигини кўрсатиши билан бир қаторда бошқа ҳарфлардан атрофлича фарқ қилишини ифодалайди. Масалан у талаффузи энг ўхшаш «о» ҳарифига ўртача 68% ўхшаш эканлигини кўришимиз мумкин. Корреляцион солиштириш хоссасидан келиб чиққан ҳолда солиштириладиган параметрлар сони қанча кам ва ишончли бўлса, аниқлик шунча юқори бўлади. Бу эса ҳарфларни солиштирганимизда кескин фарқларни юзга келтиришига олиб келади

III. ХУЛОСА

Хулоса қилиб шуни айтишимиз мумкинки, нутқ товушини танишда унинг спектрограмма тасвиридан фойдаланган ҳолда таниш, инсон нутқини танишда энг қулай воситалардан бири деб қарашимиз мумкин. Лекин шуни эсдан чиқармаслик керакки, бу усул ҳам ҳамма вақт масалани ягона ечими бўла олмайди. Бунда аниқланган натижалар қолган усуллар билан аниқланган натижаларга қўшимча сифатида қаралади ва охирида олинган натижалар умумлаштириш орқали нутқни таниш имконияти оширишга олиб келади.

АДАБИЁТЛАР

- [1] А.И.Солонина, Д.А.Улохович и др. Основы цифровой обработки сигналов.: Курс лекций. Изд. 2-с исправ. и перераб. СПб.: БХВ – Петербург, 2005. - 768с.
- [2] Nicholas W.D. Evans, John S.Mason and Matt J.Roach „Noise Compensation using Spectrogram Morphological Filtering”, Speech and Image Research Group, Department of Electrical and Electronic Engineering University of Wales Swansea, UK.
- [3] Rohini R. Mergu Dr.Shantanu K. Dixit. Multi-Resolution Speech Spectrogram. International Journal of Computer Applications (0975 – 8887) Volume 15– No.4, February 2011.
- [4] Хужаяров И.Ш., Очиллов М.М. “Нутқни қайта ишлаш масалаларини график процессорда амалга ошириш.” International conference on importance of information technologies in innovative development of real sectors of economy. April 5-6, 2018, Tashkent, Uzbekistan

- [5] Musaev M.M, Raximov M.F, Berdanov U.A. Parallel algorithms for acoustic processing of speech signals. : 2016 IEEE International Conference on Signal and Image Processing (ICSIP). 2017
- [6] Мусаев М.М. “Современные методы цифровой обработки речевых сигналов”. Научно-технический и информационно-аналитическая журнал ТУИТ, №.2(42)/2017, 2-13. Тошкент-2017
- [7] Berdanov U.A. “O’zbek tili nutqining qayta ishlashning korrelyatsion modeli tahlili”. TATUning ilmiy-texnika va axborot tahliliy jurnali, №.3(43)/2017, 10-18. Toshkent-2017.
- [8] Al-Darkazali, Mohammed. Image processing methods to segment speech spectrograms for word level recognition. Doctoral thesis (PhD), University of Sussex. 2017.
- [9] D. Polap, M. Woźniak. Image approach to voice recognition. 2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings. 2018. pp.1-7.
- [10] J. M. Borst, "The Use of Spectrograms for Speech Analysis and Synthesis," vol. 4, 1956.
- [11] B. Pinkowski, "Multiscale fourier descriptors for classifying semivowels in spectrograms," Pattern Recognition, vol. 30, p. 9, 1993.
- [12] B. Pinkowski, "Principal component analysis of speech spectrogram images," Pattern Recognition, vol. 30, pp. 777-787, 1997.
- [13] X. Xiong, L. Jinyu, C. Eng Siong, L. Haizhou, and L. Chin-Hui, "A study on hidden Markov model's generalization capability for speech recognition," in Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on, 2009, pp. 255-260.
- [14] W. B. Hussein, "Spectrogram Enhancement By Edge Detection Approach Applied To Bioacoustics Calls Classification," Signal & Image Processing : An International Journal, vol. 3, pp. 1-20, 2012.
- [15] L. D. Alsteris and K. K. Paliwal, Short-time phase spectrum in speech processing: A review and some experimental results, Digital Signal Process. 17(3) (2007) 578–616.
- [16] J. Allen, “Short term spectral analysis, synthesis, and modification by discrete fourier transform,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 25, no. 3, pp. 235–238, 1977.
- [17] Q. T. Nguyen et al., “Speech classification using sift features on spectrogram images,” Vietnam Journal of Computer Science, vol. 3, no. 4, pp. 247–257, 2016.
- [18] John G. Proakis, Dimitris Manolakis: Digital Signal Processing: Principles, Algorithms and Applications, 4th edition, Pearson, USA, 2006.