

FTAMP 50.10.41

Д. Сұлтанбекұлы<sup>1</sup> – негізгі автор, | ©  
С.А. Кульмамиров<sup>2</sup><sup>1</sup>Студент, <sup>2</sup>Техн. ғылым. канд., доцент

ORCID

<sup>2</sup><https://orcid.org/0000-0003-0912-7836>

Әл-Фараби атындағы Қазақ ұлттық университеті,



Алматы қ., Қазақстан Республикасы

<sup>1</sup>[raevdauir00@gmail.com](mailto:raevdauir00@gmail.com)<https://doi.org/10.55956/EXAO9516>

## СӨЙЛЕУДІ ТАҢУ ЖҮЙЕСІНІҢ АЛГОРИТМДЕРІН РӘСІМДЕУ

**Аңдатпа.** Мақалада сөйлеуді тану алгоритмдерінің формасы мен мазмұны сипатталған. Сөйлеуді тану алгоритмін жүзеге асыру дыбыстық сигнал үлгілерінің мелжиілік коэффициенттерін салыстыруға негізделген. Сөйлеу белсенділігінің бағдарламалық детекторы салынды және зерттелді, бұл есептеу ресурстарының көлемін едәуір азайтуға мүмкіндік береді.

**Тірек сөздер:** сөйлеуді тану, мелжиілікті коэффициенттер, кепстр.



Сұлтанбекұлы, Д. Сөйлеуді тану жүйесінің алгоритмдерін рәсімдеу [Мәтін] / Д. Сұлтанбекұлы, С.А. Кульмамиров // Механика және технологиялар / Ғылыми журнал. – 2022. – №1(75). – Б.93-102. <https://doi.org/10.55956/EXAO9516>

**Кіріспе.** Соңғы уақытта цифрлық технологиялар мен ғылыми-техникалық прогрестің негізгі бағыттарының бірі болып табылатын автоматтандыру саласы тез дамып келеді. Автоматтандыру сипатталған процестерді мүмкіндігінше жеңілдетуге және оларды виртуалды құралдармен ұйымдастырудың тиімділігін едәуір арттыруға арналған. Бүгінгі таңда цифрландыру - сандық технологияларды физикалық процестерге интеграциялау құралы болып табылады. Адам мен машинаның өзара әрекеттесу тиімділігінің жоғары жылдамдықты жарысына қатыса отырып, бізге таныс диалог және графикалық интерфейстер біртіндеп шегініп, сөйлеу интерфейсіне кейбір позицияларды береді. Сөйлеу интерфейсі бар жүйелер адамның әртүрлі құрылғылармен өзара әрекеттесуін жеңілдетіп қана қоймайды, сонымен қатар бірқатар түрлі мүмкіндіктерді ашады: дауыстық биометрия, сөйлеуді синтездеу және талдау, оқыту және білім беру, телефония.

Енді сөйлеу технологиялары үйді автоматтандыруға жетті, бұл сізге үнемі күнделікті міндеттерден, күтпеген форс-мажорлық жағдайлардан арылуға, күнделікті өмірге ыңғайлылық пен жайлылық әкелуге мүмкіндік береді. Үйді автоматтандыруға арналған құрылғылар нарығын қарастыра отырып, жүйелердің екі класын ажыратуға болады:

- интернет желісіне үздіксіз қол жеткізуді талап ететін онлайн жүйелер және интернет-қосылусыз автономды режимде жұмыс істей алатын онлайн жүйелер. Қазіргі уақытта коронавирустық пандемия кезінде сөйлеуді танудың онлайн жүйелері көп қолданылады, өйткені интернет-технологиялар кең

таралған. Мысал ретінде PocketSphinx ашық бастапқы бағдарламалық жасақтамасының функцияларын қарастырып көрейік.

PocketSphinx-тің кемшілігі – функциялардың бастапқы жиынтығында орыс тіліндегі сөйлеуді тану мүмкіндігінің болмауы. Мұндай функционалдылықты қарапайым қолданушы жасай алмайтын бағдарламалық кодты өзгерту арқылы ғана жүзеге асыруға болады. Мұндай шешім жаппай нарыққа бағытталуы мүмкін емес.

**Зерттеу шарттары мен әдістері.** Мақалада үйде автоматтандыру мақсатында кеңінен қолданылатын компьютерде жұмыс істеуге жарамды, оңай конфигурацияланатын онлайн сөйлеуді тану жүйесін жасауға әрекет жасалды. Адамның есту аппаратының АЖС сызықты емес сипаты бар екені белгілі [1]. Осыған байланысты бізге таныс физикалық шамаларды, мысалы, амплитуда мен дыбыс биіктігін пайдалану нәтижесіз өлшем болып табылады.

Бұл мәселені шешу үшін балама өлшем бірліктері енгізілді: фон және мел.

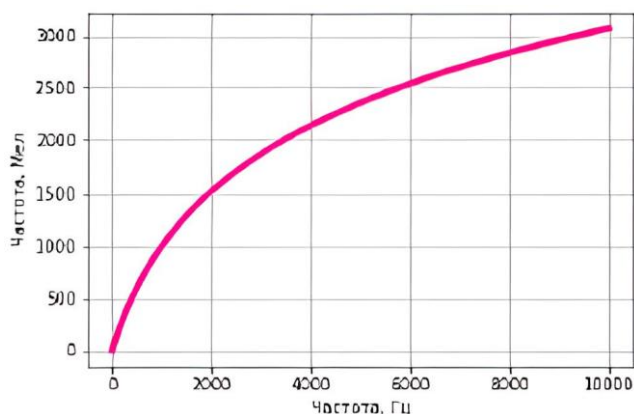
Мел – бұл қабылдаудың психофизикалық параметрлеріне негізделген эмпирикалық жолмен алынған дыбыс өлшеу бірлігі. Фон – әртүрлі жиіліктегі адамның есту сезімталдығын ескеретін дыбыс деңгейін бағалауға арналған логарифмдік бірлік.

Мел сөйлеуді талдау жүйелерінде қолдануға ыңғайлы, өйткені оны қолдану адамның есту анализаторының бірқатар ерекшеліктерін ескереді, алгоритмдердің сезімталдығын адамның қабылдау параметрлеріне жақын етеді [2]. Жиілікті Гц-тен мелге ауыстыру (1) өрнегі арқылы, ал кері түрлендіру (2) өрнегі арқылы жүзеге асырылады.

$$m = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right) = 1127 \cdot \ln\left(1 + \frac{f}{700}\right). \quad (1)$$

$$f = 700 \cdot (10^{m/2595} - 1). \quad (2)$$

мұндағы  $f$  – герцпен өлшенетін жиілік,  $m$  – мелдегі жиілік. Мел жиілігі шкаласын көрсететін график 1-суретте көрсетілген.



1-сурет. Мел жиілігі шкаласы

Енді сөйлеуді тану жүйесінің тағы бір параметрін қарастырайық - бұл сөйлеу сигналының спектрі. Сигналдарды өңдеу теориясында спектр сигнал энергиясын жиіліктерге бөлу ретінде түсіндіріледі. Спектрді зерттеу сигналдың жиілік құрамын сапалы бағалауға мүмкіндік береді [3].

(3) өрнекке сәйкес дыбыс сигналының спектрі стандартты дискретті Фурье түрлендіру (DFT) әдісімен жазылады, құрастырылады және бағаланады, оның нәтижесі жиілік аймағындағы дискретті  $X(m)$  реттілігі ретінде анықталады.  $N$  сигнал үлгілері:

$$X(m) = \sum_{n=0}^{N-1} x(n) \exp(-i \frac{2\pi mn}{N}) \quad (3)$$

Мұндай жағдайда DFT цифрлық жүйенің жадты пайдалануы және есептеу уақыты тұрғысынан тиімсіз құрал болып табылады. DFT нүктелерінің саны мыңдаған мәндерге дейін өскенде, есептеу операцияларының саны өте үлкен болады.

Бұл мәселені жоюды Кули және Тьюки ұсынды, ол қазір стандартты DFT-ке қарағанда бірнеше есе аз есептеу операцияларын қажет ететін жылдам Фурье түрлендіруі (бұдан әрі - FFT) деп аталады. Бұл жағдайда FFT нәтижесі DFT-ге жақын емес, бірақ оны дәл қайталайды. Есептеу жүйесінің ресурстарын үнемдеудің арқасында FFT контроллерлерде жүзеге асырылуы мүмкін.

Әдетте, FFT есептеу үшін  $N$  екінің дәрежесі ретінде таңдалады. Бұл  $N/2$  нүкте DFT-ті әрқайсысы 2 үлгіден тұратын қарапайым блоктарға бөлу, олар үшін түрлендіруді есептеу, содан кейін айналу факторларының таңбасын жай ғана қайта реттеу арқылы, қалған шарттарды есептеу және барлық синтезделген деректерді жинау үшін қажет.

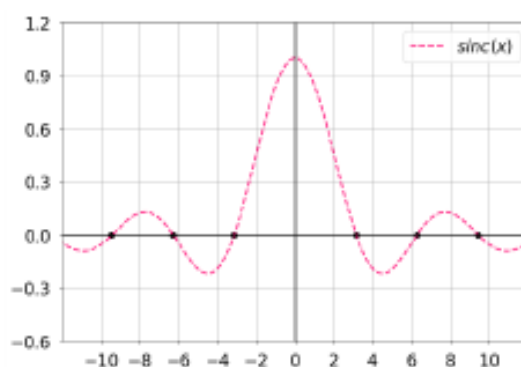
Тағы бір қолданылатын параметр (спектрдің ағуы) - егер кіріс сигналының жиілігі дәл биндердің біреуінің орталық жиілігіне тең болмаса, түрлендіру бин (жиілік шкаласындағы бин - интервал) бойынша спектрдің таралуынан тұратын құбылыс.

Мысалы, егер жиілік шкаласында  $N$  кГц қадам болса, ал кіріс сигналы 1,5 кГц жиілікте болса, онда ағып кету пайда болады. Бұл DFT-ді нақты ұзындықтың нақты тізбегін орындау кезінде сөзсіз әсер етеді, өйткені уақыт бойынша сигналды шектеу – бұл тікбұрышты терезе спектрімен керемет шексіз сигнал спектріні жинақтау. Терезе – бұл сигналдың белгілі бір бөлігін бөлуге мүмкіндік беретін арнайы салмақ функциясы [4]. Атап айтқанда, тікбұрышты терезе шексіз тізбекті шектеуге мүмкіндік береді.

Жалпы жағдайда, тікбұрышты тереземен өлшенген косинусоидтардың уақытпен шектелген АЖС  $\text{sinc}(x)$  функциясымен жақындастырылады:

$$X(m) = \frac{N}{2} \cdot \left[ \frac{\sin(\pi k - \pi m)}{\pi k - \pi m} \right] = \frac{N}{2} \cdot \text{sinc}(\pi k - \pi m) \quad (4)$$

$$\text{sinc}(x) = \begin{cases} \frac{\sin(x)}{x}, & x \neq 0; \\ 1, & x = 0 \end{cases} \quad (5)$$



2-сурет.  $\text{sinc}(x)$  функциясының графигі

Мұнда  $\text{sinc}(x)$  функциясы тікбұрышты терезенің Фурье түрлендіруінің нәтижесі болып табылады (2-сурет). Бұл  $\text{sinc}(x)$  функциясының бүйірлік жапырақшаларының пайда болуына себеп болатын 0-ден 1-ге дейінгі өткір ауысулар. Бүйірлік жапырақшаларға байланысты ағып кетумен күресу үшін тіктөртбұрыштан басқа терезелерді пайдалану керек. Мұндай терезелер көп, олардың әрқайсысының өз артықшылықтары мен кемшіліктері бар. Терезені таңдағанда, негізгі жапырақшаның ені, бірінші жапырақшаның деңгейі және жиіліктің жоғарылауымен бүйірлік жапырақшалардың биіктігінің төмендеу жылдамдығы арасындағы ымыраға келу керек.

**Зерттеу нәтижелері.** Әдетте сигналды спектрлік талдау арқылы алынған мәліметтер сигналды жақсы бағалау үшін жеткіліксіз. Талдаудың дәлдігін арттыруға көмектесетін құралдардың бірі – кепстр (спектр спектрі). Бұл жағдайда бастапқы спектр тәуелсіз сигнал ретінде ұсынылады. Кепстрларды қолдана отырып, ақпаратты ықшам түрде алуға болады, бұл деректерді талдауды жеңілдетеді.

Конволюция пішінінің өрнегі:

$$f_2(t) = \int_0^t f_1(\tau) \cdot w(t - \tau) d\tau \quad (6)$$

Конволюция  $f_2(t)$  жүйесінің ерікті әрекеті кезінде  $f_1(t)$  реакциясын табуға мүмкіндік береді, егер оның импульстік жауабы  $w(t)$  белгілі болса. Жұмысын сүзгі әрекеті ретінде түсіндіруге болатын дауыстық жолдың шығысындағы сигналдың конволюциясын келесі түрде ұсынуға болады:

$$f(t) = s(t) \otimes w(t) \quad (7)$$

мұндағы  $s(t)$  - сигналдың бастапқы түрі,  $w(t)$  - дауыс жолының параметрлерімен анықталатын сүзгінің сипаттамасы.

Жиілік аймағына өту арқылы мыналарды ала аламыз:

$$F(t) = S(\omega) \cdot W(\omega) \quad (8)$$

Осы өрнекті логарифмдеу арқылы қосындыға ауысамыз:

$$\ln[S^2(\omega) \cdot W^2(\omega)] = \ln S^2(\omega) + \ln W^2(\omega) \quad (9)$$

Енді Фурье түрлендіруін қолдана отырып, кепстр үшін соңғы өрнекті аламыз:

$$C(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln[F(\omega)]^2 \cdot e^{i\omega q} d\omega \quad (10)$$

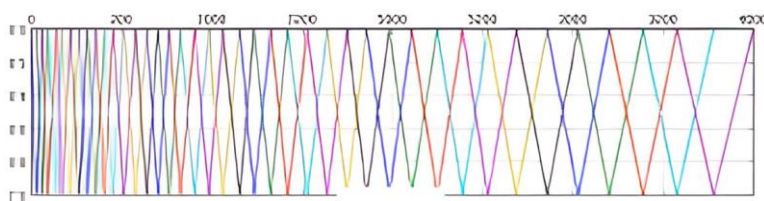
Мұнда  $q$  уақыт өлшеміне ие, бірақ жүргізілген түрлендірулерге байланысты бұл қарапайым уақыт емес, яғни кепстралды уақыт. Кейде  $q$  жиілік (немесе кьюфренси) деп аталады.

Жоғарыда келтірілген өрнектер ұсақ жиілікті коэффициенттердің мәндерін синтездеу алгоритмін жасауға мүмкіндік береді (MFCC - Mel-Frequency Cepstrum Coefficients). Бұл коэффициенттер акустикалық параметрлер мен сөйлеу белгілерін ажыратуға мүмкіндік береді. Мелде сигнал жиілігін дыбысқа аудару шкаласын қолдана отырып, көпөлшемді белгілер векторларын алуға болады, олар өз кезегінде салыстыру алгоритмдеріне беріледі [5].

MFCC есептеу кіріс сөйлеу сигналын жеке қабаттасатын аралықтарға - жақтауларға бөлуден басталады. Фреймдерге бөлу тапсырманың есептеу күрделілігін азайту үшін жасалады, ал қабаттасу оларды дәйекті түрде жоғалтқан кезде жоғалуы мүмкін фреймдердің шекараларында ақпарат жинау үшін қажет. Әдетте, рамалардың ұзындығы 20-дан 40 миллисекундқа дейін таңдалады, ал қабаттасу 50% құрайды.

Деректерді одан әрі талдау үшін Фурье түрлендіруі қажет. Мұны жасамас бұрын, спектрдің ағып кетуін азайту және сигналдың үзілуін кеміту үшін әр жақтау үшін терезе функциясын қолдану керек. Терезе функциясын қолданғаннан кейін қуат спектрін есептеу кезінде DFT-ге өтуге болады.

Алдыңғы кезеңде алынған спектрлік коэффициенттер мел-жиілік терезелеріне (3-сурет) қолданылады, олар төмен жиіліктер диапазонында тығыз орналасады және жоғары жиіліктерге көшу кезінде олардың айырмашылығын арттырады. Терезелердің бұл орналасуы дыбысты қабылдауға сәйкес келеді: жиілік неғұрлым төмен болса, көрші жиіліктер арасындағы айырмашылық солғұрлым аз болады.



3-сурет. Мел-шкаладағы терезелер

Алынған мел жиілік спектрлік коэффициенттері әр үшбұрышты терезеге қанша сигнал энергиясы түскенін көрсетеді. Енді "спектр спектрін" алу арқылы кепстралды коэффициенттерді алу қажет. Бұл үшін әдетте дискретті косинус түрлендіруі қолданылады:

$$a_k = \sum_{n=0}^{N-1} x_n \cos \left[ \frac{\pi}{N} \left( 1 + \frac{1}{2} \right) k \right], \quad k \in [0, N-1] \quad (11)$$

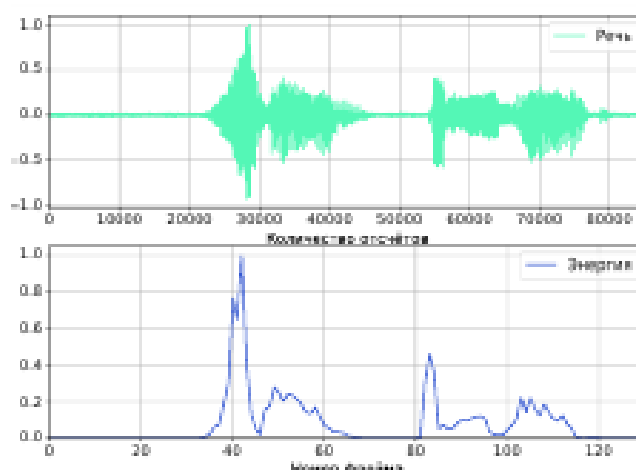
Сөйлеудің барлық осы спектрлік коэффициенттері сөйлеу белсенділігінің детекторын құруға мүмкіндік береді (ағылш. Voice Activity Detector, VAD) - бұл белсенді сөйлеуді фондық шу мен үнсіздіктен бөліп, кіріс акустикалық сигналдағы дауыстық белсенділікті анықтауға мүмкіндік беретін бағдарламалық алгоритм [6]. Мұндай жүйені пайдалану есептеу операцияларының санын едәуір азайтады, сонымен қатар сөздікте тыныштықты жазу мүмкіндігін жояды.

**Ғылыми нәтижелерді талқылау.** VAD-ын көптеген жолдармен жүзеге асыруға болады. Осы жобаны іске асыру барысында жалпы аудио ағыннан сөйлеу деректерін бөлудің жеке алгоритмі жасалды. Ол үшін сөйлеу әрекетінің кезеңдерін нақты анықтауға мүмкіндік беретін үш түрлі сипаттама қолданылды:

1) қысқа мерзімді энергия (ағылш. Short Term Energy, STE) - бұл әр кадр үшін бөлек есептелетін қарапайым энергия. Жоғарыда айтылғандай, рамалардың ұзындығы шамамен 10-40 мс таңдалады, сондықтан мұндай атау толығымен негізделген. Қысқа мерзімді энергияны келесідей есептеуге болады [6]:

$$E = \sum_{m=0}^{N-1} s^2(m) \quad (12)$$

мұндағы  $s(m)$  -  $m$  жақтауының қуаты,  $N$  - кадрлар саны. Қысқа мерзімді энергияның кадр нөміріне тәуелділік графигі 4-суретте көрсетілген.



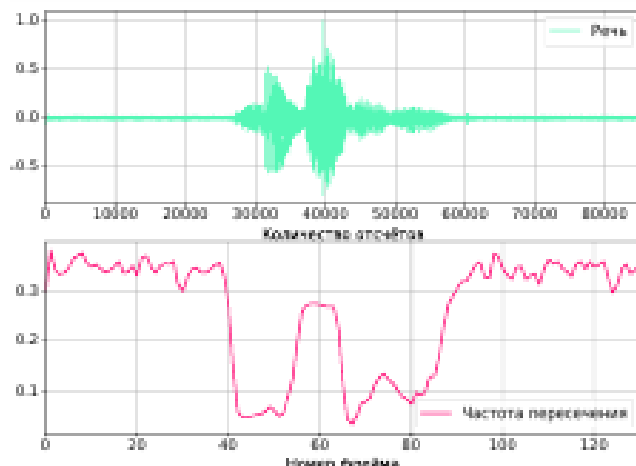
4-сурет. Сөйлеу сигналының қысқа мерзімді энергиясы

2) нөлдік қиылысу жиілігі (ағылш. Zero Crossing Rate, ZCR) - бұл мән кадр ішінде сигнал өзінің белгісін қанша рет өзгерткенін көрсетеді. Сөйлеу сигналының функциясы үшін қиылыстар саны шуды сипаттайтын функцияға қарағанда едәуір аз болады, өйткені сөйлеу диапазоны шуымен салыстырғанда негізінен төмен жиіліктермен сипатталады [7].

ZCR-ді келесі формула бойынша есептеуге болады [8]:

$$Z(n) = \sum_{m=1}^N |sign[s(m)] - |sign[s(m-1)]| \quad (13)$$

5-суретте нөлдік қиылысу жиілігінің графигі көрсетілген. Суреттен шамамен 40 кадрдан бастап, нөлден өту жиілігінің мәні төмендейтінін көруге болады, бұл сөйлеу белсенділігінің болуын көрсетеді.



5-сурет. Нөлдік қиылысу жиілігі

3) спектрлік жазықтық метрі (ағылш. Spectral Flatness Measure, SFM) - сигналдарды сандық өңдеуде қолданылатын спектрлік сипаттама, ол дыбыстық деректер ағынында сөйлеу фрагменттерінің болуын жақсы сандық бағалауға мүмкіндік береді [6].

Бұл сипаттаманы төмендегі формула бойынша есептеуге болады:

$$F = 10 \lg \frac{G}{A} = 10 \lg \frac{(\prod_{m=0}^{N-1} s(m))^{\frac{1}{N}}}{\frac{1}{N} \sum_{m=0}^{N-1} s(m)}, \quad (14)$$

мұндағы  $G$  және  $A$  сәйкесінше геометриялық және арифметикалық орта болып табылады.

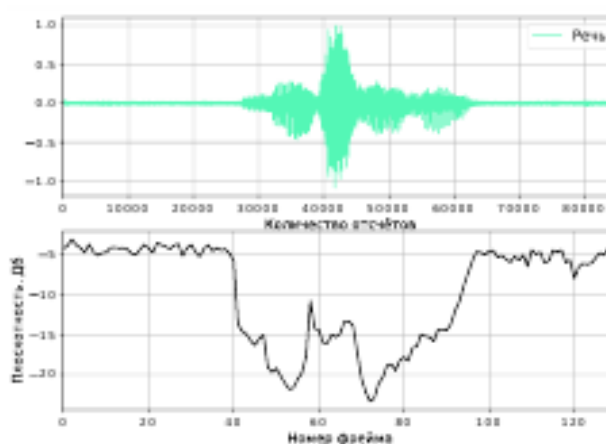
Спектрлік жазықтықтың жоғарғы шегі (Дб шкаласын ескерместен 1-ге жуықтайды) спектрдің ақ шуылға ұқсас барлық жиілік диапазонында бірдей қуат мөлшері бар екенін көрсетеді. Бұл жағдайда спектрдің графигі салыстырмалы түрде тегіс көрінеді.

Спектрлік жазықтықтың төменгі шегі (Дб шкаласын ескерместен 0-ге жақын) сигналдың қуаты салыстырмалы түрде аз жиілік диапазонында шоғырланғанын көрсетеді. Бұл жағдайда спектр графигінде нақты белгіленген белсенділік шындары болады.

Осылайша, әр кадр үшін спектрлік жазықтықты есептеу арқылы сөйлеу белсенділігінің айқын көрінетін кезеңдерін байқауға болады. Тесттік сөйлеу сигналының спектрлік жазықтығының графигі 6-суретте көрсетілген.

Графиктерден көрініп тұрғандай, әр мән аудио деректер ағынында сөйлеудің болуын бағалауға көмектеседі.

Енді біз сөйлеуді танудың дамыған алгоритмінің мүмкіндіктері мен артықшылықтарын сипаттайық. Ең алдымен, есептеу тапсырмасын жеңілдету және сигналдың өзгеруін дәлірек бақылау үшін дыбыстық сигнал жақтауларға бөлінеді. Ұзақтығы 30 миллисекунд, қабаттасуы 50%.



Сурет 6. Сөйлеу сигналының спектрлік жазықтығы

Өнімділікті арттыру үшін тапсырманы әртүрлі процестер арасында бөлуге болады, олардың әрқайсысы өз шеңберін өңдейді.

Келесі қадам – әр кадрдың негізгі сипаттамаларын есептеу: қысқа мерзімді энергия, нөлдік қиылысу жиілігі және спектрлік жазықтық. Жазықтықты есептеу үшін әр кадр үшін спектрді есептеу керек. Әрі қарай, алғашқы 30 кадр таңдалады, олар әдетте үнсіздік болып табылады.

Егер кез келген кадр үшін барлық үш сипаттаманың мәні шекті мәннен асып кетсе, онда кадр сөйлеу деп саналады және массивке жазылады. Олай болмаған жағдайда, жақтау ретінде атап өтіледі. Алынған сөйлеу рамалары WAV форматындағы файлға жазылатын жалпы сигналға біріктіріледі.

**Қорытынды.** Сигналмен осындай манипуляциялардан кейін кіріс сигналын салыстыратын сөйлеу үлгілерінің сөздігін жасау керек. Жүйені тексеру үшін әртүрлі сөйлеу фрагменттері жазылды. Жазу сөздікке энтропияны қосу үшін әр сөз үшін 2 рет жасалды, өйткені бір сөзді бірдей интонациямен, жылдамдықпен қайталау мүмкін емес. Сайып келгенде, сөздіктің ұзындығы 20 сөз болды. Кіріс сигналы ретінде дербес компьютердің дыбыстық картасына қосылған микрофоннан тікелей сигнал қолданылды.

Екі уақыттық қатарды салыстыру үшін уақыт шкаласының динамикалық өзгеру алгоритмі қолданылады. Шығу кезінде бұл алгоритм екі тізбектің ұқсастық дәрежесін арнайы айнымалы түрінде береді, оның мәні массивке жазылады. Айнымалының мәні неғұрлым аз болса, соғұрлым тізбектер бір-біріне ұқсас болады. Екі бірдей тізбекті салыстырған кезде көрсетілген айнымалының мәні нөлге тең болатыны анық. Салыстыру сөздіктің барлық қол жетімді элементтерімен жасалады. Келесі қадам – массивтің минималды элементін және оның индексін табу.

Авторлық зерттеулер барысында келесі нәтижелер алынғанын атап өтеміз:

- сөйлеуді тану дәлдігі 86% ретінде анықталды;
- сөйлеу әрекеті детекторының іске қосылу дәлдігі 97% ретінде есептелді;
- тану уақыты 2-4 секундты құрады.

Алгоритмдер Python-да жүзеге асырылды. Құрастырылған бағдарламалық кешен үйді автоматтандырудың ақылды жүйелерін құру үшін кеңінен қолданылатын Raspberry Pi IoT класты компьютерлерде жұмыс істеуге жарамды болып табылады.



**Әдебиеттер тізімі**

1. Линдсей П., Норман Д. Переработка информации у человека. М.: Мир, 1974. 546 с.
2. Huang X., Acero A. Spoken Language Processing: A Guide to Theory Algorithm, and System Development. Prentice Hall, 2001. 965 p.
3. Lyons R. G. Understanding Digital Signal Processing. Addison Wesley Pub. Co, 2006. 656 p.
4. Bracewell R. N. The Fourier Transform and its Applications. McGraw Hill, 2000. 620 p.
5. Ganchev T., Fakotakis N. Comparative evaluation of various MFCC implementations on the speaker verification task // 10<sup>th</sup> International Conference on Speech and Computer. Patras, Greece, 2005.
6. Moattar M. H., Homayounpour M. M. A e-cient real-time voice activity detection algorithm // Laboratory for Intelligent Sound and Speech Processing (LISSP), Computer Engineering and Information Technology Dept. , Amirkabir University of Technology, Tehran, Iran. 24 августа 2009.
7. Nandhini S., Shenbagavalli A. Voiced/Unvoiced Detection using Short Term Processing // International Journal of Computer Applications (0975-8887). 2014.
8. Bachu R., Kopparthi S., Adapa B., Barkana B. Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy // Advanced Techniques in Computing Sciences and Software Engineering, 2010. Pp. 279-282.

*Материал редакцияға 11.03.22 түсті.*

**Д. Султанбекулы, С.А. Кульмамиров**

*Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан*

**РАЗРАБОТКА АЛГОРИТМОВ РАСПОЗНАВАНИЯ УСТНОЙ РЕЧИ**

**Аннотация.** В статье описаны форма и содержание алгоритмов систем распознавания речи. Реализация алгоритма распознавания речи основан на сравнении мелкочастотных кепстральных коэффициентов выборок звуковых сигналов. Построен и исследован программный детектор речевой активности, позволяющий существенно снизить объем вычислительных ресурсов для рассматриваемой задачи.

**Ключевые слова:** распознавание речи, мелкочастотные коэффициенты, кепстр.

**D. Sultanbekuly, S.A. Kulmamirov**

*Al-Farabi Kazakh National University, Almaty, Kazakhstan*

**DEVELOPMENT OF ALGORITHMS FOR ORAL SPEECH RECOGNITION**

**Abstract.** The article describes the form and content of algorithms for speech recognition systems. The implementation of the speech recognition algorithm is based on the comparison of small-frequency cepstral coefficients of audio signal samples. A software speech activity detector has been constructed and studied, which allows to significantly reduce the amount of computing resources for the problem under consideration.

**Keywords:** speech recognition, small frequency coefficients, cepstrum.

**References**

1. Lindsay P., Norman D. Pererabotka informatsii u cheloveka [Human information processing]. - Moscow: Mir, 1974. 546 p. [in Russian].

2. Huang X., Acero A. Spoken Language Processing: A Guide to Theory Algorithm, and System Development. Prentice Hall, 2001. 965 p.
3. Lyons R. G. Understanding Digital Signal Processing. Addison Wesley Pub. Co, 2006. 656 p.
4. R. N. Bracewell, The Fourier Transform and its Applications. McGraw Hill, 2000. 620 p.
5. Ganchev T., Fakotakis N. Comparative evaluation of various MFCC implementations on the speaker verification task // 10th International Conference on Speech and Computer. Patras, Greece, 2005.
6. Moattar M. H., Homayounpour M. M. A efficient real-time voice activity detection algorithm // Laboratory for Intelligent Sound and Speech Processing (LISSP), Computer Engineering and Information Technology Dept. , Amirkabir University of Technology, Tehran, Iran. August 24, 2009.
7. Nandhini S., Shenbagavalli A. Voiced/Unvoiced Detection using Short Term Processing // International Journal of Computer Applications (0975-8887). 2014.
8. Bachu R., Kopparthi S., Adapa B., Barkana B. Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy // Advanced Techniques in Computing Sciences and Software Engineering, 2010. P. 279-282.