

Д.Р. Рахимова^{1,2} , Ж. Дүйсенбекқызы^{1*} , О.А. Сағынтай³ ,
Ә. Турарбек^{1,2} , Ш. Төлеуғали² 

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

²Ақпараттық және есептеу технологиялары институты, Алматы қ., Қазақстан

³NARXOZ University, Алматы қ., Қазақстан

*e-mail: jansaya_sugirbaeva@mail.ru

БАЛАЛАРДЫҢ СӨЙЛЕУІН ТАҢУ ЖҮЙЕЛЕРІНДЕГІ ШУДЫ АЗАЙТУ АЛГОРИТМДЕРІН ТАЛДАУ ЖӘНЕ ЖАҚСARTУ

Аңдатпа

Бұл жұмыста нақты акустикалық ортада жазылған балалардың сөйлеу сапасын жақсартуға арналған гибриді шуды азайту алгоритмі ұсынылған. Алгоритм дәстүрлі спектрлік талдау әдістерін – жылдам Фурье түрлендіруі (FFT) және Винер сүзгісін – терең оқытудың тәсілдерімен, атап айтқанда қайталанатын нейрондық желілер (RNN) мен ұзақ қысқа мерзімді жады (LSTM) модельдерімен біріктіреді. Ұсынылған әдістің тиімділігін бағалау үшін қазақ тіліндегі балалар сөйлеуінің аудио деректерінен құралған арнайы корпус Telegram-бот арқылы жиналды. Эксперимент нәтижелері алгоритмді қолданғаннан кейін сөйлеуді тану көрсеткіштерінің айтарлықтай жақсарғанын көрсетті: сөзді тану қателігі (WER) 32.4%-дан 18.7%-ға дейін төмендеді, ал F1-score 0.72-ден 0.88-ге дейін өсті. Спектрограммалық талдау фондық шудың азайғанын және сөйлеу сигналдарының анықтығының артқанын көрсетті. Зерттеу нәтижелері балалардың сөйлеуін тану жүйелерінде шуды азайту үшін гибриді әдістерді қолданудың тиімді екенін дәлелдеді.

Түйін сөздер: сөйлеуді жақсарту, терең оқыту, нейрондық желілер, шуды болдырмау, сөйлеу сигналдарын өңдеу, стационарлық емес шу, баланың сөйлеуі, гибриді алгоритмдер, деректерді өңдеу, қазақ тілі.

Д.Р. Рахимова^{1,2}, Ж. Дүйсенбекқызы¹, О.А. Сағынтай³, Ә. Турарбек^{1,2}, Ш. Төлеуғали²

¹Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

²Институт информационных и вычислительных технологий, Алматы, Казахстан

³NARXOZ University, Алматы, Казахстан

АНАЛИЗ И УЛУЧШЕНИЕ АЛГОРИТМОВ ШУМОПОДАВЛЕНИЯ В СИСТЕМАХ РАСПОЗНАВАНИЯ РЕЧИ ДЕТЕЙ

Аннотация

В данной работе представлен гибридный алгоритм шумоподавления, предназначенный для улучшения качества речи детей, записанных в конкретной акустической среде. Алгоритм сочетает в себе традиционные методы спектрального анализа, такие как быстрое преобразование Фурье (FFT) и фильтр Винера, с подходами глубокого обучения, включая повторяющиеся нейронные сети (RNN) и сети с долговременной кратковременной памятью (LSTM). Для оценки эффективности предложенного метода был собран аудиофайл детской речи на казахском языке с использованием специально разработанного Telegram-бота "Dataset Loader". Экспериментальная часть исследования показала значительное улучшение показателей распознавания речи после использования алгоритма шумоподавления. В частности, частота ошибок распознавания слов (WER) снизилась с 32.4% до 18.7%, а F1-score выросла с 0.72 до 0.88. Анализ спектрограмм показал явное снижение уровня фонового шума и улучшение читаемости речи. Результаты исследования подтверждают эффективность гибридного подхода для улучшения качества речевых данных и повышения точности систем автоматического распознавания речи в сложных акустических условиях.

Ключевые слова: улучшение речи, глубокое обучение, нейронные сети, шумоподавление, обработка речевых сигналов, нестационарный шум, речь ребенка, гибридные алгоритмы, обработка данных, казахский язык.

D. R. Rakhimova^{1,2}, Zh. Duisenbekkyzy¹, O. A. Sagintayev³, A. Turarbek^{1,2}, Sh. Toleugali²

¹ Al-Farabi Kazakh National University, Almaty, Kazakhstan

² Institute of information and computing technologies, Almaty, Kazakhstan,

³NARXOZ University, Almaty, Kazakhstan

ANALYSIS AND IMPROVEMENT OF NOISE REDUCTION ALGORITHMS IN CHILDREN'S SPEECH RECOGNITION SYSTEMS

Abstract

This article presents a hybrid noise reduction algorithm designed to improve the quality of children's speech recorded in a specific acoustic environment. The algorithm combines traditional spectral analysis techniques such as fast Fourier transform (FFT) and Wiener filter with deep learning approaches including repetitive neural networks (RNNs) and long-term short-term memory (LSTM) networks. To evaluate the effectiveness of the proposed method, an audio file of children's speech in Kazakh was compiled using a specially developed Telegram bot "Dataset Loader". The experimental part of the study showed a significant improvement in speech recognition after using the noise reduction algorithm. In particular, the word recognition error rate (WER) decreased from 32.4% to 18.7%, and the F1-score increased from 0.72 to 0.88. Spectrogram analysis showed a clear reduction in background noise and improved speech readability. The results of the study confirm the effectiveness of the hybrid approach to improve the quality of speech data and improve the accuracy of automatic speech recognition systems in difficult acoustic conditions.

Keywords: speech enhancement, deep learning, neural networks, noise reduction, speech signal processing, non-stationary noise, child's speech, hybrid algorithms, data processing, Kazakh language.

Кіріспе

Бұл зерттеу қазақ тіліндегі балалар сөйлеуіндегі шуды азайту мәселесін шешуге бағытталған. Дәстүрлі спектрлік әдістер мен терең оқыту модельдерін (RNN, LSTM) біріктіретін гибриді алгоритм ұсынылды. Эксперименттер нәтижесінде сөйлеуді тану сапасы айтарлықтай жақсарғаны анықталды (WER 32.4%-дан 18.7%-ға дейін төмендеді). Нәтижелер күрделі акустикалық ортада ASR жүйелерінің сапасын арттыруға жол ашады.

Сөйлеу – адамдардың қарым-қатынасының негізгі тәсілі. Қарым-қатынас процесін екі кезеңге бөлуге болады: сөйлеуді құру және сөйлеуді қабылдау [1]. Сөйлеуді құру адам өз ойын дыбыстарда білдірген кезде пайда болады. Сөйлеуді қабылдау-бұл басқа адам бұл дыбыстарды естіп, сөйлеушінің не айтқысы келетінін түсінеді.

Сөйлеуді автоматты түрде тану (ASR), дауыстық интерфейстер (мысалы, теру немесе деректерді енгізу үшін) және есту аппараттары сияқты сөйлеуге негізделген көптеген қосымшалар бар. Алайда, нақты жағдайда сөйлеу сапасы көбінесе фондық шу, кедергі сияқты сыртқы факторларға байланысты нашарлайды. Мұндай жағдайларда тыңдаушы сөйлеуді қабылдауда қиындықтарға тап болады, әсіресе шуға қатысты сигнал деңгейі (SNR) төмендеген кезде. Осыған байланысты сөйлеу сапасын жақсарту (SE) міндеті аудио сигналдарды өңдеу саласында өте маңызды болып табылады. SE бұрмаланған сигналдан шуды жою арқылы түсінікті және сөйлеу сапасын арттыруға бағытталған [2]. Дегенмен, шуды азайту және бастапқы сигналды қалпына келтіру процесі әлі де әдістер мен технологияларды жетілдіруді қажет ететін күрделі тапсырма болып табылады.

Сөйлеу сапасын жақсарту және шуды азайту міндеті аудио сигналдарды өңдеу саласындағы ең күрделі және өзекті мәселелердің бірі болып қала береді [3]. Соңғы ондаған жылдарда сөйлеу мен шудың өзара әрекетін талдау үшін статистикалық модельдерден бастап терең оқытуға негізделген заманауи тәсілдерге дейін әртүрлі әдістер қолданылды [4]. Дәстүрлі әдістер жоғары сигнал/шу қатынасы (SNR) мен тұрақты шу жағдайында тиімді болғанымен, төмен SNR және өзгермелі шу кезінде пайдалы сигналды ажыратуда әлсіз екенін көрсетті [5,6]. Сөйлеудің оқылуы мен шуды азайту мәселесі әлі де өзекті. Бұл әсіресе автоматты тану жүйелері үшін маңызды, себебі мұнда тек шуды жою ғана емес, пайдалы сигналды сақтау және нақты уақыт режимінде жұмыс істеу де талап етіледі.

Шуды болдырмау – пайдалы сигналды көшедегі шу, фондық сөйлеу немесе қоршаған орта дыбыстары сияқты қажетсіз шулардан бөлу мақсатында аудио сигналдарды өңдеудегі

маңызды міндет. Шуды азайту міндеті сөйлеуді тану жүйелері, медиа дыбыс сапасын жақсарту, сондай-ақ таза аудио файлдарды жасау сияқты көптеген технологиялардың орталығында тұр. Соңғы онжылдықтарда дәстүрлі сүзгілермен салыстырғанда тиімділігі дәлелденген машиналық оқыту мен терең оқытуға негізделген күрделі әдістерге назар аударылды. Төтенше оқиғаларды анықтау апаттық сигналдарды жедел әрі жоғары дәлдікпен анықтауды талап ететін маңызда тапсырма болып табылады. Заманауи терең оқыту технологиялары аудио сигналдар талдауына арналған тиімді шешімдерді алуға мүмкіндік береді. Бұл өз кезегінде төтенше оқиғаларды анықтаудың жылдамдығы пен дәлдігін арттырады.

Шуды болдырмаудың заманауи әдістерін бірнеше санатқа бөлуге болады:

1. Дәстүрлі әдістер (мысалы, спектрлік тығыздықты сүзу немесе спектрлік азайту әдісі);
2. Статистикалық тәсілдерге негізделген әдістер (SNR ұпайы, Винер сүзгілері);
3. Терең нейрондық желілерге негізделген заманауи Алгоритмдер (RNN, CNN және басқалары).

Дыбыстық сигналдағы шу – бұл бастапқы пайдалы сигналды бұрмалайтын дыбыс толқындарының кездейсоқ тербелісі. Ғылыми әдебиеттерде пайдалы сигнал мен шудың өзара әрекеттесуі көбінесе келесі математикалық модельді қолдану арқылы сипатталады:

$$Y_n = S_n + N_n \quad (1)$$

мұндағы:

Y_n – n уақытындағы шулы сигнал,

S_n – пайдалы ақпарат болып табылатын таза сигнал,

N_n – пайдалы сигналға қосылатын Шу,

n – жазу процесінде уақыт моментін сипаттайтын уақыт индексі.

Шуды азайту немесе дыбыс сапасын жақсарту міндетінің мақсаты-таза S_n бастапқы сигналына мүмкіндігінше жақын шулы Y_n сигналын қалпына келтіруді табу, ал міндет шудың сигналды қабылдаудың соңғы сапасына әсерін азайту болып табылады [7,8].

Шуды болдырмау әдістері сапаны жақсарту және сигналды кейінгі ойнатуға дайындау мақсатында дыбыстық жазбаларды қажетсіз дыбыстық кедергілерден тазартуда маңызды рөл атқарады. Бұл тәсілдер дыбысты өңдеу процесінде фильмдер, музыка, подкасттар және басқа мазмұн түрлері сияқты әртүрлі медиа форматтарда кеңінен қолданылады. Мұндай тапсырмаларда шуды жоюдан басқа, жазудың жалпы тазалығын жақсарту қажет, бұл тек шуды сүзуді ғана емес, сонымен қатар сөйлеуді түсінікті ету үшін сигналды өзгертуді де қамтуы мүмкін.

Осы мақсаттар үшін музыкалық және дыбыстық тректерді жасау үшін қолданылатын аудио редакторлар мен араластырғыш бағдарламаларда қол жетімді арнайы құралдар бар. Танымал аудио редакторлардың бірі Audacity "Шу қақпасы" (noise gate) деп аталатын әдісті қолданады. Бұл бағдарлама жылдам Фурье түрлендіруін (FFT) орындағаннан кейін қолданылатын шу қақпасының спектрлік нұсқасын пайдаланады. Бұл әдіс пайдалы сигналға әсер етпестен төмен амплитудасы бар дыбыстық кедергілерді тиімді сүзуге мүмкіндік береді. Сонымен қатар, Audacity дыбыстық сигналдағы артефактілер мен ұсақ бұрмалануларды жоюға көмектесетін терезені тегістеу механизмдерін қамтиды. Audacity сияқты бағдарламалар, әсіресе ескірген немесе бүлінген микрокассеталық жазбалармен жұмыс істегенде, өздерін аудио жазбаларды қалпына келтірудің тиімді құралы ретінде көрсетті [9].

Нақты уақыттағы шуды азайту міндеті қазіргі заманғы аудио сигналдарды өңдеу технологияларында ең күрделі және сұранысқа ие мәселелердің бірі болып табылады. Мұндай шуды жоюдың негізгі мақсаты - адам айтқан маңызды ақпаратқа қатысы жоқ және сөйлеуді қабылдауға кедергі келтіретін дыбыстарды бөлу. Бұл дыбыстарға фондық шу кіруі мүмкін, мысалы, көлік қозғалысының дыбыстары, бейтаныс адамдардың сөйлесулері немесе жұмыс істейтін желдеткіштер немесе бұрғылар сияқты механикалық ішкі Шу. Көбінесе шуды

болдырмау технологиялары Skype, Zoom, Discord және басқалары сияқты аудио конференцияларға арналған жүйелерде қолданылады, мұнда сырттан келгендер болған кезде сөйлеудің жоғары оқылуы маңызды. дыбыстық, мұндай жүйелер белгілі бір дыбыс деңгейінен тыс дыбыстарды блоктайтын немесе сүзетін "Шу қақпасы" принципіне негізделген тәсілдерді қолданады. Алайда, жақында мұндай жүйелердің тиімділігін арттыру үшін нақты уақыт режимінде сигналды жоғары дәлдікпен тазартуға мүмкіндік беретін Машиналық оқыту әдістері көбірек қолданыла бастады. Машиналық оқытуды сәтті қолданудың жарқын мысалдарының бірі-шуды сүзу үшін ұзақ мерзімді қысқа мерзімді жад блоктары (LSTM) және конволюциялық нейрондық желілер (CNN) бар қайталанатын нейрондық желілерді (RNN) пайдалану [10,11].

Microsoft DNS-Challenge [12] нәтижелеріне сүйене отырып, осы озық тәжірибелерді Skype және Teams қосымшаларына біріктірді. Нақты уақыттағы шуды азайту алгоритмдері агрессивті шудың (мысалы, бұрғылау, желдеткіш, сыртқы жел) жағдайында да дауыстық сигнал сапасын жақсартып алады. Бұл нәтижеге қоршаған шудың өзгеруіне икемделетін күрделі нейрондық желілік архитектуралар арқылы қол жеткізіледі. Мұндай тәсіл нақты жағдайда сөйлеуді қабылдауды едәуір жақсартады.

Осылайша, жаңа оқыту, соның ішінде LSTM және CNN, сыртқы шудың жоғары деңгейінде дауыстық қосымшалардың сапасын едәуір жақсартып отырып, нақты уақыттағы шуды болдырмаудың жаңа көкжиегін ашады. Шуды болдырмаудың негізгі міндеттерінің бірі-кейінгі танудың дәлдігін жақсарту үшін дыбыстық сигналды сөйлеуді автоматты тану жүйесіне (ASR) жібермес бұрын оны өңдеу және тазарту. Бұл тапсырма бірқатар қиындықтармен бірге жүреді, өйткені шуды жою процесінде шуды жою ғана емес, сонымен қатар тану сапасына теріс әсер етуі мүмкін сигналға жасанды бұрмалануларды енгізбеу маңызды. Мысалы, Зерттеулер [13] маскаларды қолданатын шуды азайту жүйелері кейде сөйлеуді тану нәтижесін жақсартып қоймай, керісінше нашарлатуы мүмкін екенін көрсетеді. Себебі мұндай әдістер жиі спектрлік сипаттамаларды бұрмалап, декодтау сапасына кері әсер етеді. Ал заманауи терең нейрондық желілер алдын ала өңдеу кезеңінде қолданылғанда, сөйлеу сигналын тануға дейін оның сапасын тиімді түрде жақсартуға мүмкіндік берді.

Зерттеу әдіснамасы

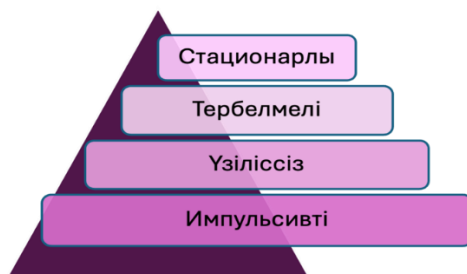
Дыбыстық сигналдағы шуды спектрлік сипаттамалар немесе жиілік диапазоны сияқты әртүрлі белгілер бойынша жіктеуге болады [14]. Алайда, сөйлеуді өңдеуде және қажетсіз шуды жоюға тырысқанда, шуды олардың уақыт сипаттамаларына қарай жіктеу өте маңызды [15]. 1-суретте шудың жіктелуі олардың динамикасы мен уақыттық өзгергіштігіне байланысты бірнеше санатқа бөлінеді.



Сурет 1. Шудың уақыт сипаттамалары бойынша жіктелуі

1-суретте көрсетілгендей, шудың уақыт сипаттамалары оның көздерімен және қалыптасу процестерімен тығыз байланысты. Стационарлық және тербелмелі шу әдетте табиғи және жасанды жүйелерде тұрақты және болжамды сипаттамалары бар процестерден туындайды.

Шудың бұл түрлері тұрақты немесе мезгіл-мезгіл қайталанатын құрылымға ие, бұл оларды уақыт бойынша салыстырмалы түрде тұрақты етеді. Керісінше, үзік-үзік және импульстік шу қысқа мерзімді сипаттағы және уақыттың белгілі бір нүктелерінде ғана болатын бір реттік және өткір оқиғалардан туындайды. Бұл шулар ұзақ уақыт сақталмайды және пайда болғаннан кейін бірден жоғалады. Үзіліссіз шуды белгілі бір уақыт аралығында қайталанатын импульстік шудың бір түрі ретінде қарастыруға болады, бұл ауыспалы тыныштық кезеңдерін және шудың қысқа жарылыстарын тудырады. Шудың жіктелуі оларды басудың күрделілігіне байланысты әр түрлі шу түрлерін бөліп көрсету үшін қажет. Шуды болдырмау мәселесін шешудің негізгі қиындығы - болжау мүмкін еместігі және дыбыстық сигналда болуы мүмкін Шу түрлерінің әртүрлілігі. Шу түрі алдын-ала белгілі болған кезде, шуды жою міндеті басқарылатын болады. Мысалы, стационарлық шуды салыстырмалы түрде оңай жоюға болады, өйткені оның сипаттамаларын болжауға және анықтауға болады. Ақ шу әдетте спектр бойынша біркелкі бөлінеді, бұл тыныштық аймақтарында шудың шекті амплитудасын орнатуға және оны пайдалы сигналдан дәл оқшаулауға мүмкіндік береді. Осылайша, 2-суретте көрсетілгендей шуды азайту мәселелерін шешуде күрделілік иерархиясын құруға болады, мұнда күрделі шу түрлері күрделі өңдеу әдістерін қолдануды қажет етеді, ал қарапайым және стационарлық шуды сүзу оңайырақ.



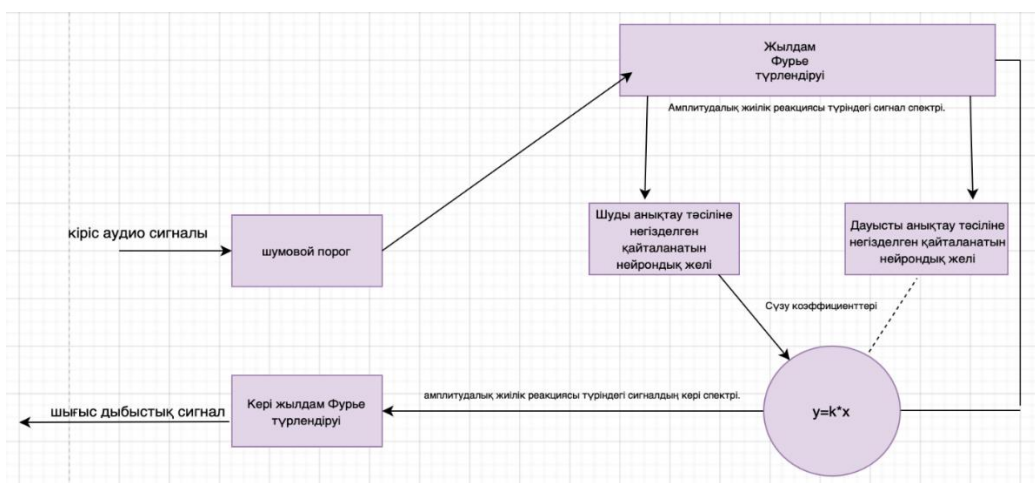
Сурет 2. Уақыт сипаттамаларына және басудың күрделілігіне байланысты шу түрлерінің иерархиясы

1-кестеде дыбыстық сигналдарда пайда болатын шудың негізгі түрлерінің жіктелуі, олардың көріну мысалдары және сәйкес басу әдістері келтірілген. Шудың әр түрі өзінің уақыттық және спектрлік сипаттамаларымен ерекшеленеді, бұл оны жою тәсілін таңдауды анықтайды [16].

Кесте 1. Шудың жіктелуі және оларды басу әдістері

Шу түрі	Мысалдар	Басу әдістері	Өңдеу ерекшеліктері
Стационарлы	Ақ шу, әселдеткіштің дыбысы	Спектрлік тығыздық бойынша сүзу (Spectral Subtraction), Винерді сүзу (Wiener Filter)	Болжамды спектрі бар шуды тиімді басу, жоғары SNR кезінде минималды бұрмаланулар
Тербелмелі	Машиналардан ауыспалы шу, дыбыс деңгейінің ауытқуы	Адаптивті сүзгілер (LMS алгоритмі), SNR бағалау әдістері	Шу деңгейінің өзгеруіне байланысты сүзгілерді динамикалық реттеу қажет
Үзіліссіз	Мерзімді шертулер, баспа машинкасының дыбысы	Уақыт терезелерімен спектрлік азайту, уақыт масқаларымен сегменттеу және терезені тегістеу	Артефактілерді жою және жою үшін уақытша Шу құрылымын қолдану
Импульсивті	Шапалақтау, заттардың құлауы, қатты дыбыстар	Конволюциялық нейрондық желілер (CNN), LCM блоктары бар қайталанатын нейрондық желілер (RN)	Қысқа мерзімді Шу жағдайында күрделі заңдылықтарды анықтау үшін терең оқыту әдістерін қолдану

Кестеден шуды азайту әдістерінің тиімділігі шудың түрі мен сипаттамаларына тікелей байланысты деген қорытынды жасауға болады. Шудың әртүрлі түрлерінің пайда болуының уақыт сипаттамалары мен табиғатын талдау негізінде шуды болдырмаудың гибриді алгоритмі жасалды, ол нақты уақыттағы шу сипаттамаларына бейімделеді. Бұл тәсіл шудың күрделілігі мен түріне байланысты дәстүрлі есептеу әдістері мен заманауи нейрондық желілік архитектуралардың тіркесімін қолдануды қарастырады. 3-суретте ұсынылып отырған гибриді шуды азайту әдістемесінің жалпы схемасы берілген. Схема кіріс аудио сигналдан бастап, шуды анықтау, дауысты анықтау және кері түрлендіру арқылы тазартылған дыбыс сигналын алу үдерісін көрсетеді.



Сурет 3. Гибриді шуды азайту әдістемесінің схемасы

Шуды азайту процесін сипаттайтын негізгі формула келесідей:

$$Y(b, t) = k(b, t) * x(b, t) \quad (2)$$

мұндағы: $Y(b, t)$ – t уақытындағы b жиілік диапазонындағы сигналдың шығыс амплитудасы, $x(b, t)$ – t уақытындағы b жиілік диапазонындағы сигналдың кіріс амплитудасы, $k(b, t)$ – b жолағындағы сигналдың басылу немесе сақталу дәрежесін анықтайтын $[0, 1]$ диапазонындағы сүзу коэффициенті.

Блок-схемада көрсетілген алгоритм қадамдары:

1. Шекті пайдаланып әлсіз шу шығару.
2. FFT көмегімен сигналды жиілік аймағына түрлендіру.
3. Шу мен сөйлеуді анықтау үшін нейрондық желілер арқылы спектрді талдау.
4. Шуды басу және сөйлеуді сақтау үшін сүзуді қолдану.
5. Тазартылған сигналды уақыт аймағына кері түрлендіру.
6. Тиімділікті арттыру үшін барлық кезеңдерді бір өңдеуге біріктіру.

Ұсынылған гибридік тәсілді егжей-тегжейлі сипаттау үшін алгоритмді бірнеше дәйекті кезеңдерге бөлуге болады, олардың әрқайсысы шуды тиімді басуға және сөйлеу сапасын жақсартуға ықпал етеді:

1. *Деректерді өңдеу:* бірінші қадамда Шу шегі арқылы сүзілетін әлсіз шу шығады. Бұл маңызды сөйлеуге әсер етпестен төмен амплитудалық шуды жоюға мүмкіндік береді. Бұған келесі формуланы қолдану арқылы қол жеткізіледі:

$$x_{threshold} = \begin{cases} x(b, t), & \text{если } x(b, t) > \theta \\ 0, & \text{если } x(b, t) \leq \theta \end{cases} \quad (3)$$

мұндағы: θ — фон деңгейіне байланысты таңдалған шу шегі.

2. *Жиілікті талдау*: жиілік аймағына түрлендіру $x_{threshold}(t)$ сигналын спектрлік компоненттерге бөлуге мүмкіндік беретін жылдам Фурье түрлендіруі (FFT) арқылы жүзеге асырылады:

$$X(b, t) = F \{ x_{threshold}(t) \} \quad (4)$$

мұндағы: F – Жылдам Фурье түрлендіру операциясы (FFT), $x(b, t)$ – Барк шкаласы бойынша 24 жолаққа бөлінген сигнал спектрі [17].

3. *RNN көмегімен шуды талдау және анықтау*: сигнал спектрін өңдеу және шуды анықтау үшін қайталанатын нейрондық желі (RNN) қолданылады. Модель шудың динамикалық сипаттамаларын анықтау және әр жиілік диапазоны үшін сүзгі коэффициенттерін орнату үшін мәліметтерден үйренеді:

$$k_{noise}(b, t) = RNN_{noise}(X(b, t)) \quad (5)$$

мұндағы:

RNN_{noise} – шуды анықтауға арналған қайталанатын нейрондық желі моделі

$x(b, t)$ – шуды басуға арналған сүзу коэффициенттері.

Егер $k_{noise}(b, t) \approx 1$, сигнал өзгермейді және келесі нейрондық желі қолданылады. Егер шу анықталмаса, сөйлеу компоненттерін оқшаулау үшін екінші РНС қолданылады:

$$k_{voice}(b, t) = RNN_{voice}(X(b, t)). \quad (6)$$

Қорытынды сүзу коэффициенті келесідей анықталады:

$$x(b, t) = \min(k_{noise}(b, t), k_{voice}(b, t)) \quad (7)$$

Бұл сигнал шу ретінде қате танылмаған болса, ең қатаң сүзгінің қолданылуын қамтамасыз етеді.

4. *Сигналды сүзу*: Келесі қадам шуды басу немесе сөйлеу сигналын сақтау үшін бұрын есептелген коэффициенттерді қолдана отырып сүзуді қолданады:

$$Y(b, t) = k(b, t) * x(b, t) \quad (8)$$

мұндағы:

$Y(b, t)$ – өңделген сигнал спектрі.

Бұл сөйлеуді сақтау және шуды жою үшін ең жақсы сүзгіні қолдануды қамтамасыз етеді.

5. *Уақыт аймағына кері түрлендіру*: Спектрді сүзгеннен кейін кері жылдам Фурье түрлендіруі (IFFT) арқылы жиілік аймағынан уақытша аймаққа кері түрлендіру орын алады:

$$y(b, t) = F^{-1}(Y(b, t)) \quad (9)$$

мұндағы:

F^{-1} – кері Жылдам Фурье түрлендіру операциясы (IFFT)

$y(t)$ – соңғы тазартылған аудио сигнал.

Барлық кезеңдерді бір формулаға біріктірейік:

$$y(t) = F^{-1} \{ \min(RNN_{noise}(F \{ x_{threshold}(t) \}), RNN_{voice}(F \{ x_{threshold}(t) \})) * F \{ x_{threshold}(t) \} \} \quad (10)$$

мұндағы:

Ішкі бөлігі нейрондық желілер арқылы спектрді талдауға жауап береді, сыртқы Int операциясы сигналды бастапқы уақыт формасына қайтарады. Бұл әдістеме шуды тиімді

жоюды және сөйлеу сапасын жақсартуды қамтамасыз етеді, бұл сөйлеуді автоматты түрде тану жүйелерінде, әсіресе балалар сөйлеуі сияқты күрделі акустикалық ортада қолдану үшін маңызды.

Зерттеу нәтижелері

Бұл бөлімде зерттеу жағдайының сипаттамасы, экспериментті бастамас бұрын модель параметрлері, эксперименттік деректерді ұсыну, алынған нәтижелер және алынған деректерді салыстырмалы талдау берілген. Ұсынылған шуды болдырмаудың гибриді моделінің тиімділігін бағалау үшін қазақ тіліндегі балалар сөйлеуінің аудио корпусы құрылды. Деректерді жинау дауыстық хабарламаларды жазу және жүктеу үшін ыңғайлы интерфейсті қамтамасыз ететін арнайы әзірленген "Dataset Loader" Telegram-ботының көмегімен жүзеге асырылды. Аудио жазбалар сыныптар, үй жағдайлары және ойын алаңдары сияқты нақты акустикалық ортада жасалды, бұл көшедегі Шу мен қоршаған ортадағы әңгімелерден бастап механикалық дыбыстарға дейін (мысалы, тұрмыстық техниканың жұмысы) фондық шудың кең ауқымын қамтамасыз етті.

Нәтижесінде 2 жастан 8 жасқа дейінгі балалардың сөйлеу деректері бар 800-ге жуық аудиожазба жиналды. Корпустың жалпы көлемі шамамен 1000 мәтінді және шамамен 12 минуттық аудио материалды құрады. Деректер базасында 30-ға жуық баланың (17 қыз және 9 ұл) балалар аудиториясының тақырыбы мен қиындық деңгейіне сәйкес таңдалған 75 ерекше сөздер мен сөз тіркестері бар жазбалары бар. Әр жазбаның орташа ұзақтығы 3-4 секундты құрады, бұл 2 сағаттан астам аудио деректерді біріктірді. Модельді кейінгі талдау мен оқытудың жоғары сапасын қамтамасыз ету үшін барлық жазбалар қолмен транскрипцияланды. Аудио жазбалар бастапқыда OGG форматында сақталады, ол деректер жүктелгеннен кейін WAV форматына түрлендіріледі, әрі қарай оқыту мен талдау үшін қолайлы. 2-кестеде жиналған аудиожазбалар туралы ақпарат, соның ішінде балалардың жасына, жынысына және әр жас тобына арналған жазбалар санына байланысты мәліметтер таратылады.

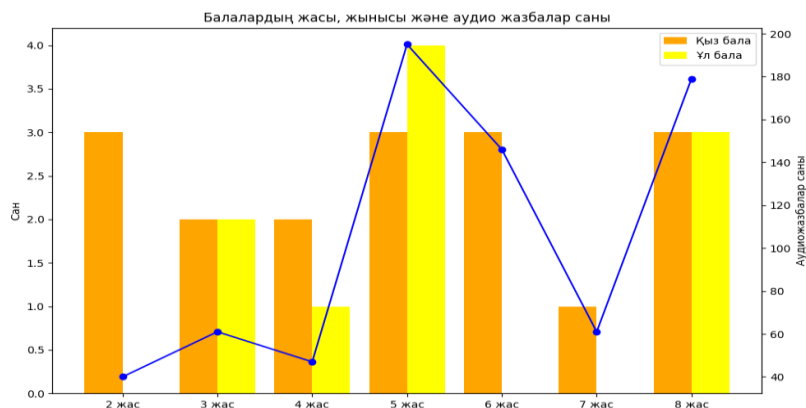
Кесте 2. Шудың жіктелуі және оларды басу әдістері

Баланың жасы	Қыз бала	Ұл бала	Аудиожазбалар саны
2 жас	3	0	40
3 жас	2	2	61
4 жас	2	1	47
5 жас	3	4	195
6 жас	3	0	146
7 жас	1	0	61
8 жас	3	3	179

Бұл деректер корпусының құрамын және оның білім беру және зерттеу мақсатында пайдалануға жарамдылығын түсінуге мүмкіндік береді. 4-суретте көрсетілген диаграмма қыздар мен ұлдар арасындағы жазбалардың санын көрсететін әр жас тобына арналған аудио жазбалардың санын көрсетеді. Деректердің бұл таралуы жасына және жынысына байланысты корпусының өкілдігін оңай бағалауға, сондай-ақ деректерді одан әрі толықтыру үшін мүмкін болатын олқылықтарды анықтауға мүмкіндік береді.

Осылайша, корпусының жас құрылымы мен жыныстық құрамы әр түрлі жастағы сөйлеу деректерін кең қамтуды қамтамасыз етеді, бұл оны балалардың сөйлеуін оқытуға және талдауға байланысты тапсырмалар үшін пайдалы етеді.

Гибриді шуды азайту алгоритмінің тиімділігін бағалау үшін жиналған аудио деректер екі кіші топқа бөлінді: шикі деректер және тазартылған деректер.



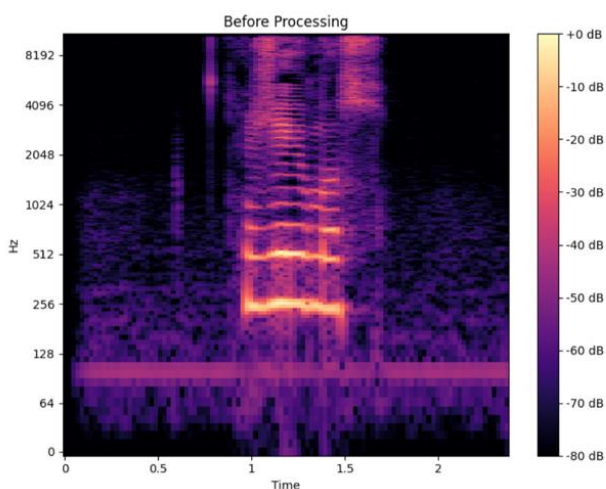
Сурет 4. Балалардың жасы, жынысы және аудио жазбалар саны

Бірінші топқа нақты акустикалық ортаны (көше шуы, әңгімелер, механикалық дыбыстар) көрсететін сақталған фондық шулары бар түпнұсқа жазбалар кірді. Екінші топқа спектрлік талдау әдістерін (Жылдам Фурье түрлендіруі (FFT), винерді сүзу) және қайталанатын нейрондық желілерге (RNN) және ұзақ мерзімді қысқа мерзімді жады желілеріне (LSTM) негізделген терең оқыту тәсілдерін біріктіретін гибриді шуды азайту алгоритмімен өңделген жазбалар кірді.

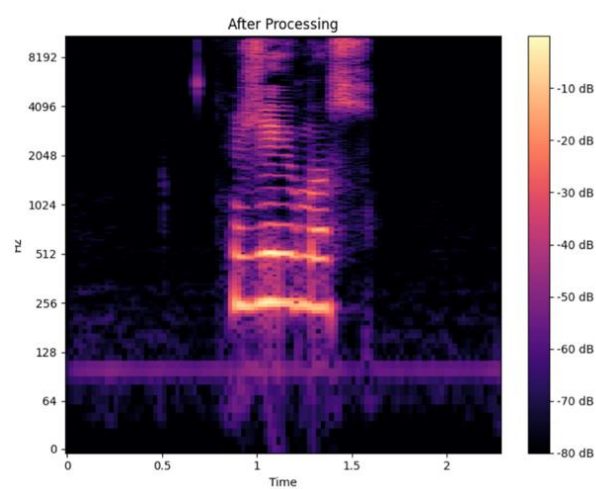
Шуды болдырмаудың сөйлеуді тану дәлдігіне әсерін анықтау үшін қайталанатын нейрондық желіге (LSTM) негізделген модель қолданылды. Модель архитектурасына мыналар кірді:

- Дыбыс сигналдарының спектрограммаларын өңдейтін кіріс қабаты;
- Сөйлеуді тану тапсырмалары үшін оңтайландырылған нейрондар саны бар екі жасырын LSTM қабаты;
- Сөйлеудің кейінгі транскрипциясы үшін болжамды таңбалар тізбегін құрайтын Шығыс қабаты.

Ұсынылған гибриді шуды азайту алгоритмінің тиімділігін көрнекі түрде көрсету үшін өңдеуден бұрын және кейін бірдей аудиожазбаның спектрограммалары жасалды. Жазба ретінде 5 жасар баланың 1 секундқа созылатын "жеміс" сөзі қолданылды. 5-суретте осы сөздің өңдеуге дейінгі спектрограммасы көрсетілген. 6-суретте осы "жеміс" сөзінің өңдеуге кейінгі спектрограммасы көрсетілген.



Сурет 5. Өңдеуден бұрын "жеміс" сөзінің спектрограммасы



Сурет 6. Өңдеуден кейін "жеміс" сөзінің спектрограммасы

Өндеуге дейінгі және кейінгі спектрограммаларды салыстырмалы талдау гибриді алгоритмінің аудио сигналдардың сапасын жақсартудағы тиімділігін растайды. Фондық шуды жою және сөйлеу компоненттерін бөлектеу сөйлеуді тану дәлдігін жақсартуға және аудио материалдардың субъективті қабылдауын жақсартуға көмектеседі. Бұл нәтижелер сөйлеуді автоматты түрде тану мәселелерінде, әсіресе күрделі акустикалық ортада ұсынылған тәсілдің қолданылуын растайды [18].

3-кестеде екі деректер жиынтығы үшін сөйлеуді тану тиімділігін салыстырмалы талдау нәтижелері келтірілген: шикі аудио жазбалар және гибриді шуды болдырмау алгоритмін қолдана отырып алдын ала өңделген жазбалар.

Кесте 3. Шуды болдырмас бұрын және одан кейінгі сөйлеуді тану тиімділігінің салыстырмалы көрсеткіштері

Метрика	Шикі деректер	Тазартылған деректер (гибриді алгоритм)
Word Error Rate (WER)	32.4%	18.7%
F1-score	0.72	0.88
PESQ	2.1	3.4
SNRi (дБ)	+5.2	+13.1

Бағалау төрт негізгі метрика бойынша жүргізілді: Word Error Rate (WER) [19], F1-score [20], Speech Quality Perceptual Evaluation (pesq)[21] және signal-to-Noise Ratio Improvement (SNRi) [22]. Нәтижелерден көріп отырғанымыздай, деректерді алдын-ала тазарту модель көрсеткіштерін едәуір жақсартты. Сонымен, сөзді тану қателіктерінің жиілігі (WER) 32.4% - дан 18.7% - ға дейін төмендеді, бұл тану дәлдігінің жоғарылауын көрсетеді. Дәлдік пен толықтық арасындағы гармоникалық орташа мәнді көрсететін F1-score 0.72-ден 0.88-ге дейін өсті, бұл модельдің сөйлеу сегменттерін дұрыс бөлектеу қабілетінің жақсарғанын көрсетеді. Pesq сөйлеу сапасын объективті бағалау 2.1-ден 3.4-ке дейін өсті, бұл тазартылған жазбаларды қабылдаудың жақсарғанын көрсетеді. Ақырында, сигнал/шу қатынасының жақсаруы (SNRi) +5.2 дБ-ден +13.1 дБ-ге дейін өсті, бұл сөйлеу сигналының сапасына нұқсан келтірместен фондық шуды басуда ұсынылған алгоритмнің тиімділігін растайды. Ұсынылған гибриді шуды азайту алгоритмі әр түрлі Шу жағдайында сөйлеу сапасын жақсартуға арналған. Алайда, алгоритмнің тиімділігі аудио сигналдағы шудың түрі мен сипаттамаларына байланысты екенін ескеру маңызды. Бұл тәсіл дәстүрлі спектрлік талдау әдістері мен терең оқытудың заманауи әдістерінің жиынтығын қолданады, бұл шудың белгілі бір түрлерімен күресуге мүмкіндік береді, бірақ шудың барлық түрлері үшін бірдей тиімділікке кепілдік бермейді.

1. *Стационарлық шу* (мысалы, ақ шу, желдеткіштің дыбысы). Шудың бұл түрлері болжамды құрылымға және салыстырмалы түрде тұрақты спектрлік сипаттамаларға ие. Оларды басу үшін спектрлік азайту әдістері мен винерді сүзу тиімді қолданылады, олар жоғары сигнал/шу қатынасында (SNR) шуды жақсы басқарады. Алгоритм сөйлеудің бұрмалануын азайту арқылы мұндай шуды тиімді түрде жоя алады.

2. *Тербелмелі шу* (мысалы, машиналардан шығатын айнымалы шу, қоршаған дыбыстың көлемінің өзгеруі). Бұл шулар динамикалық тәсіл мен бейімделу әдістерін қажет етеді. Мұндай жағдайларда SNR бағалау әдістері мен адаптивті сүзгілер қолданылады (мысалы, LMS алгоритмі). Жақсы нәтижеге қол жеткізу үшін сүзгілер шу деңгейінің өзгеруіне байланысты динамикалық түрде реттелуі керек. Алгоритм мұндай шулармен тиімді жұмыс істей алады, егер олардың сипаттамалары белгілі бір уақыт шегінде тұрақты болса.

3. *Үзіліссіз және импульсивті шу* (мысалы, шертулер, заттардың түсуі, механикалық дыбыстар). Қысқа мерзімді және өткір шуды өңдеу үшін конволюциялық нейрондық желілер (CNN) және ұзақ мерзімді қысқа мерзімді жады (LSTM) қайталанатын нейрондық желілер (RNN) сияқты нейрондық желі әдістерін қолдану арқылы күрделі өңдеу қолданылады. Бұл

әдістер кездейсоқ уақыт аралықтары болса да, мұндай шуды тануға және тиімді түрде басуға қабілетті. Алайда, мұндай әдістердің тиімділігі оқытудың дәлдігіне және нақты шуды тану үшін деректердің сапасына байланысты.

Осылайша, алгоритм шудың әртүрлі түрлерімен тиімді жұмыс істей алады, бірақ оның өнімділігі аудио сигналда болатын шудың нақты түріне байланысты болады. Нақты жағдайларда, сигналда шудың аралас түрлері болған кезде (мысалы, стационарлық және импульсивті комбинация), алгоритмнің әмбебаптығы мен бейімделуін арттыру үшін қосымша түзетулер қажет болуы мүмкін.

Дискуссия

Нәтижелер көрсеткендей, ұсынылған гибриді алгоритм сөйлеуді тану сапасын едәуір жақсартты. WER төмендеуі және F1-score өсуі нақты акустикалық ортада алгоритмнің тиімділігін дәлелдейді. PESQ және SNRi көрсеткіштері де субъективті және объективті сапаның артқанын көрсетеді. Бұл көрсеткіштер [2,4] еңбектеріндегі нәтижелермен сәйкес келеді, онда терең оқытуға негізделген тәсілдер стационарлық емес шуды тиімді басатыны дәлелденген. Сонымен қатар, 5–6 суреттегі спектрограмма анализі визуалды түрде де сөйлеу сигналдарының айқындылығын растайды. Алгоритмнің артықшылығы – нейрондық желілердің нақты уақыттағы бейімделгіштігі және әртүрлі шу түрлеріне икемделуі. Алайда, шектеуі – нейрондық модельдердің өнімділігі деректердің сапасына және оқыту көлеміне тәуелді. Шудың аралас түрлері кезінде қосымша бейімдеу қажет болуы мүмкін. Болашақта корпус көлемін арттыру, нақты уақыттағы өңдеу жылдамдығын жоғарылату және жүйені толыққанды ASR жүйесіне енгізу жоспарлануда.

Қорытынды

Бұл зерттеу дәстүрлі спектрлік талдау әдістерін (мысалы, жылдам Фурье түрлендіруі және Винер сүзгісі) қайталанатын нейрондық желілер (RNN) және ұзақ мерзімді қысқа мерзімді жады (LSTM) желілері сияқты заманауи терең оқыту тәсілдерімен біріктіретін гибриді шуды азайту алгоритмін әзірледі және сынады. Негізгі мақсат әр түрлі фондық шулармен жазылған балалардың сөйлеуін қоса, күрделі акустикалық ортада сөйлеуді автоматты түрде тану (ASR) дәлдігін арттыру болды. Эксперимент жүргізу үшін "Dataset Loader" Telegram-ботының көмегімен жиналған қазақ тіліндегі балалар сөйлеуінің аудио корпусы қалыптастырылды, бұл фондық шудың кең спектрін – көшеден механикалық Шуға дейін ескеруге мүмкіндік берді. Нәтижелер барлық негізгі көрсеткіштердің айтарлықтай жақсарғанын көрсетті: сөзді тану қателерінің (WER) жиілігі 32.4%-дан 18.7% - ға дейін төмендеді, F1-score 0.72-ден 0.88-ге дейін өсті, сөйлеу сапасын объективті бағалау (pesq) 2.1-ден 3.4-ке дейін өсті және сигнал/шу қатынасының жақсаруы (SNRi) +13.1 дБ-ге жетті. Өңдеуге дейінгі және кейінгі спектрограммаларды талдау алгоритмнің тиімділігін растады, бұл фондық шудың айқын төмендеуін және сөйлеудің оқылуын жақсартты. Бұл нәтижелер ұсынылған гибриді тәсілді нақты әлемдегі сөйлеуді тану сапасын жақсарту үшін, әсіресе балалардың сөйлеуі үшін сәтті қолдануға болатындығын көрсетеді, бұл білім беру және зерттеу мақсаттары үшін, сондай-ақ дауыстық интерфейстер мен есту аппараттарын дамыту үшін өте маңызды. Алдағы уақытта аудио корпусы кеңейту, нақты уақыттағы өңдеу жылдамдығын арттыру үшін нейрондық желілердің архитектурасын оңтайландыру және алгоритмді қазақ тіліндегі балаларға арналған ASR жүйелеріне интеграциялау сипаттағы төтенше оқиғаларды сегіз класс бойынша анықтап, хабар беру үшін қолданылады.

Алғыс

Жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитетінің қолдауымен (AP 19577833 гранттық жобасы) орындалды.

Пайдаланылған дереккөздердің тізімі

- [1] Deller, J. R., Hansen, J. H. L., & Proakis, J. G. (2000). *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press.
- [2] Loizou, P. C. (2013). *Speech Enhancement: Theory and Practice*. CRC Press.
- [3] Hu, Y., & Loizou, P. C. (2007). Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, 49, 588–601.
- [4] Wang, D., & Chen, J. (2018). Supervised Speech Separation Based on Deep Learning: An Overview. *IEEE Transactions on Audio, Speech, and Language Processing*, 26, 1702–1726.
- [5] Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2), 443–445.
- [6] Kwan, C., Chu, S., Yin, J., Liu, X., Kruger, M., & Sityar, I. (2008). Enhanced speech in noisy multiple speaker environment. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, Hong Kong, 1–8 June 2008, 1640–1643.
- [7] Bouchard, M., et al. (2015). Speech Enhancement using Neural Networks. In *Proceedings of ICASSP*. IEEE.
- [8] Audacity Team. (n.d.). Missing Features - Audacity Wiki. Retrieved from <https://support.audacityteam.org/troubleshooting/missing-features#audacity-wiki>
- [9] Lim, Y., et al. (2020). Speech Enhancement Using LSTM-Based Neural Networks for Real-Time Applications. *arXiv preprint*. <https://arxiv.org/pdf/2010.15508>
- [10] Rao, Y., et al. (2020). Real-Time Noise Suppression Using Deep Convolutional Networks. *arXiv preprint*. <https://arxiv.org/pdf/2008.00264v4>
- [11] Zhang, Y., et al. (2020). Deep Noise Suppression: A Study of Denoising Techniques for Speech Enhancement. *arXiv preprint*. <https://arxiv.org/pdf/2005.13981>
- [12] Prodeus, A., & Didkovskiy, V. S. (2015). Оценивание априорного отношения сигнал-шум в алгоритмах шумоподавления. Системы обработки информации: сб. науч. пр. / Харк. ун-т Повітр. Сил ім. Івана Кожедуба, 29–34.
- [13] Zhang, Y., Hu, X., Zhao, H., Liu, H., & Liu, X. (2022). A Comprehensive Review of Speech Enhancement Using Deep Learning. *arXiv preprint*. <https://arxiv.org/abs/2208.09201>
- [14] Noise reduction. (n.d.). Retrieved from https://wiki.audacityteam.org/wiki/How_Audacity_Noise_Reduction_Works#algorithm
- [15] Reddy, C. K. A., Gopal, V., Cutler, R., Beyrami, E., Cheng, R., Dubey, H., Matusевич, S., Aichner, R., Aazami, A., Braun, S., Rana, P., Srinivasan, S., & Gehrke, J. (2020). The INTERSPEECH 2020 Deep Noise Suppression Challenge: Datasets, Subjective Testing Framework, and Challenge Results. *arXiv preprint*. <https://arxiv.org/pdf/2005.13981.pdf>
- [16] Hu, Y., Liu, Y., Lv, S., Xing, M., Zhang, S., Fu, Y., Wu, J., Zhang, B., & Xie, L. (2020). DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement. *arXiv preprint*. <https://arxiv.org/pdf/2008.00264v4.pdf>
- [17] Айфичер, Э. С. Цифровая обработка сигналов: практический подход: пер. с англ. / Э. С. Айфичер, Б.У Джервис. – изд. 2-е. – М.:Изд. дом Вильямс, 2004. – 992 с.
- [18] Morris, A. C., Maier, V., & Green, P. (2004). From WER and RIL to MER and WIL: Improved evaluation measures for connected speech recognition. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP)*, 2761–2764.
- [19] van Rijsbergen, C. J. (1979). *Information Retrieval (2nd ed.)*. London: Butterworths.
- [20] Топников, А. И., Нестеров, М. С., Новоселов, С. А., Приоров, А. Л. (2015). Шулы сөйлеу сигналдарының анықтығын эталондық емес бағалау. Сандық сигналдарды өңдеу, 1, 39–44.
- [21] ITU-T Recommendation P.862 (2001). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Retrieved from <https://www.itu.int/rec/T-REC-P.862-200102-I>
- [22] Берко, Г. А., Галич, С. А., Расюк, А. О., Семенов, Е. С. (2015). IP желілері арқылы сөйлеу сапасын бағалау үшін PESQ алгоритмін қолдану. Огарев-онлайн, 11(52), 3.

References

- [1] Deller, J. R., Hansen, J. H. L., & Proakis, J. G. (2000). *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press.

[2] Loizou, P. C. (2013). *Speech Enhancement: Theory and Practice*. CRC Press.

[3] Hu, Y., & Loizou, P. C. (2007). *Subjective comparison and evaluation of speech enhancement algorithms*. *Speech Communication*, 49, 588–601.

[4] Wang, D., & Chen, J. (2018). *Supervised Speech Separation Based on Deep Learning: An Overview*. *IEEE Transactions on Audio, Speech, and Language Processing*, 26, 1702–1726.

[5] Ephraim, Y., & Malah, D. (1985). *Speech enhancement using a minimum mean-square error log-spectral amplitude estimator*. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2), 443–445.

[6] Kwan, C., Chu, S., Yin, J., Liu, X., Kruger, M., & Sityar, I. (2008). *Enhanced speech in noisy multiple speaker environment*. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, Hong Kong, 1–8 June 2008, 1640–1643.

[7] Bouchard, M., et al. (2015). *Speech Enhancement using Neural Networks*. In *Proceedings of ICASSP*. IEEE.

[8] Audacity Team. (n.d.). *Missing Features Audacity Wiki*. Retrieved from <https://support.audacityteam.org/troubleshooting/missing-features#audacity-wiki>

[9] Lim, Y., et al. (2020). *Speech Enhancement Using LSTM-Based Neural Networks for Real-Time Applications*. *arXiv preprint*. <https://arxiv.org/pdf/2010.15508>

[10] Rao, Y., et al. (2020). *Real-Time Noise Suppression Using Deep Convolutional Networks*. *arXiv preprint*. <https://arxiv.org/pdf/2008.00264v4>

[11] Zhang, Y., et al. (2020). *Deep Noise Suppression: A Study of Denoising Techniques for Speech Enhancement*. *arXiv preprint*. <https://arxiv.org/pdf/2005.13981>

[12] Prodeus, A., & Didkovskiy, V. S. (2015). *Estimating the a priori signal-to-noise ratio in noise suppression algorithms*. *Information Processing Systems: Scientific Journal of the Ivan Kozhedub Air Force University*, 29–34.

[13] Zhang, Y., Hu, X., Zhao, H., Liu, H., & Liu, X. (2022). *A Comprehensive Review of Speech Enhancement Using Deep Learning*. *arXiv preprint*. <https://arxiv.org/abs/2208.09201>

[14] Noise reduction. (n.d.). Retrieved from https://wiki.audacityteam.org/wiki/How_Audacity_Noise_Reduction_Works#algorithm

[15] Reddy, C. K. A., Gopal, V., Cutler, R., Beyrami, E., Cheng, R., Dubey, H., Matuskevych, S., Aichner, R., Aazami, A., Braun, S., Rana, P., Srinivasan, S., & Gehrke, J. (2020). *The INTERSPEECH 2020 Deep Noise Suppression Challenge: Datasets, Subjective Testing Framework, and Challenge Results*. *arXiv preprint*. <https://arxiv.org/pdf/2005.13981.pdf>

[16] Hu, Y., Liu, Y., Lv, S., Xing, M., Zhang, S., Fu, Y., Wu, J., Zhang, B., & Xie, L. (2020). *DCCRN: Deep Complex Convolution Recurrent Network for Phase-Aware Speech Enhancement*. *arXiv preprint*. <https://arxiv.org/pdf/2008.00264v4.pdf>

[17] Ajficher, Je. C. (2004) *Cifrovaja obrabotka signalov: prakticheskij podhod [Digital Signal Processing: A Practical Approach]: per. s angl. / Je. C. Ajficher, B.U Dzhervis. izd. 2–e. M.: Izd. dom Vil'jams, 992.*

Ifeachor, E., & Jervis, B. (2017). *Digital Signal Processing: A Practical Approach (Translated Edition)*. Moscow: Williams Publishing. 992 pages.

[18] Morris, A. C., Maier, V., & Green, P. (2004). *From WER and RIL to MER and WIL: Improved evaluation measures for connected speech recognition*. In *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP)*, 2761–2764.

[19] van Rijsbergen, C. J. (1979). *Information Retrieval (2nd ed.)*. London: Butterworths.

[20] Topnikov, A. I., Nesterov, M. S., Novoselov, S. A., & Priorov, A. L. (2015). *Non-reference evaluation of intelligibility in noisy speech signals*. *Digital Signal Processing*, 1, 39–44.

[21] ITU-T Recommendation P.862 (2001). *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Retrieved from <https://www.itu.int/rec/T-REC-P.862-200102-I>

[22] Berko, G. A., Galich, S. A., Rasyuk, A. O., & Semenov, E. S. (2015). *Application of the PESQ algorithm to assess speech quality over IP networks*. *Ogarev Online*, 11(52), 3.