УДК 004.032.26

РАЗРАБОТКА И ПРИМЕНЕНИЕ МЕТОДОВ РАСПОЗНАВАНИЯ ЗАШУМЛЕННЫХ АУДИОФАЙЛОВ ПОСРЕДСТВОМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

Ю. Л. Леохин, д.т.н., профессор, проректор по научной работе МТУСИ, Москва, Россия;
orcid.org/0000-0003-3321-4497, e-mail: y.l.leokhin@mtuci.ru
Т. Д. Фатхулин, к.т.н., доцент кафедры МК и ИТ МТУСИ, Москва, Россия;
orcid.org/0000-0003-0998-1055, e-mail: t.d.fatkhulin@mtuci.ru
М. В. Ментус, студент МТУСИ, Москва, Россия;
orcid.org/0009-0005-8300-6954, e-mail: maksimmentus@mail.ru

Рассматривается задача распознавания речи в условиях присутствия посторонних шумов различного происхождения. ***Целью работы*** *являются разработка и оценка эффективности методов, дающих возможность распознать речь при наличии шумов с помощью нейросетевых технологий. Актуальность работы обусловлена тем, что благодаря развитию нейросетевых технологий значительно расширился круг отраслей, в которых распознавание речи стало значительно проще и эффективнее. Рассматриваются программные решения «Whisper» и «Vosk», позволяющие транскрибировать (распознавать) речь. Приводится классификация аудиошумов, описаны существующие методы борьбы с ними. Показано влияние шумов на обучение системы распознавания речи. Разработаны методы обучения системы распознавания речи с использованием синтетически сгенерированного датасета с зашумлением. Спроектирован и разработан модуль зашумления данных, собран тестовый стенд. Приведена апробация разработанных методов. В заключении представлены результаты анализа данных, полученных в ходе экспериментов, сделаны выводы.*

Ключевые слова: *системы распознавания речи, системы синтеза речи, обучение нейросетей, шум, эффективность, Word Error Rate, генерация датасета, зашумленные данные*

DOI: 10.21667/1995-4565-2024-88-65-73

Введение

Методы машинного обучения (МО) сейчас находят все большее применение в самых разнообразных отраслях деятельности человека. Распознавание речи является одним из наиболее перспективных направлений в области искусственного интеллекта (ИИ) [1-3, 6-8, 17, 18]. С развитием технологий и совершенствованием алгоритмов нейронные сети (НС), являющиеся одним из методов МО, стали основным инструментом для решения задач распознавания речи.

Распознавание речи в настоящее время используется в различных сферах жизни человека, таких как [1, 3-5, 9-11]:

– *голосовые помощники*: распознавание речи позволяет голосовым помощникам, таким как «Алиса», понимать и выполнять команды пользователя;

– *перевод речи в текст*: распознавание речи может быть использовано для преобразования устной речи в письменный текст, что полезно для создания субтитров, расшифровки интервью и т.д;

– *медицинское обслуживание*: в некоторых случаях, людям с ограниченными возможностями коммуникации может помочь программное обеспечение для распознавания речи;

– *образование*: технологии распознавания речи могут использоваться для создания обучающих программ, которые могут адаптировать уроки в соответствии с ответами ученика;

– *многоязычное распознавание*: некоторые системы распознавания речи созданы для работы с несколькими языками при переводе в режиме онлайн.

Однако существуют факторы, которые могут негативно сказаться на качестве распознавания речи нейросетью, в частности имеют место различные шумы, которые накладываются на нее [1, 2, 17, 18]. Под *шумом* в общем случае понимают отличающиеся по сложности временной и спектральной структуры беспорядочные колебания разнообразной физической природы [2]. Под *акустическими* или *аудиошумами* понимают беспорядочные звуковые колебания в атмосфере в диапазоне частот от 20 Гц до 20 кГц [1]. Наличие посторонних шумов в аудиозаписях всегда было проблемой записывающих устройств [1]. Зачастую это постоянные, стабильные шумы на определенных частотах на всей длине аудиодорожки. Однако существует другая, более реальная категория шумов – шум поездов, шум ветра, разговоры других людей. Такие шумы тяжело выделить детерминированными алгоритмами – они состоят из множества простых шумов, меняющих свои параметры во времени. Если правильно предсказывать и выделять полезный сигнал (речь), то можно получить шумоустойчивую систему распознавания речи, которая будет совершать минимум ошибок.

Постановка задачи

С учетом указанной ранее актуальности распознавания речи нейросетевыми технологиями в различных областях человеческой деятельности возникает необходимость в разработке методов, которые позволят повысить эффективность существующих технических решений. Исходя из этого, рассмотрим некоторые наиболее популярные решения по распознаванию речи, выявим их преимущества и недостатки, а также определим вектор дальнейших исследований. В качестве примера систем распознавания речи можно привести следующие программные продукты, получившие наибольшую популярность в настоящее время.

«*Whisper*» – модель для перевода устной речи в текст [12]. Она использует искусственный интеллект и машинное обучение для анализа звука и определения того, что было сказано. «*Whisper*» может использоваться для различных целей, таких как создание субтитров для видео, расшифровка аудиозаписей и т. д. В качестве преимущества рассматриваемого решения стоит отметить, что данная модель работает с различными языками, в том числе и с русским. Однако главный недостаток модели – низкая точность перевода аудиодорожки в текст на русском языке при использовании более легких, простых вариантов модели.

«*Vosk*» – открытый исходный код для распознавания речи, основанный на технологии глубокого обучения [13]. Он может использоваться для создания приложений, которые могут распознавать речь и переводить ее в текст. «*Vosk*» также может использоваться для создания систем голосового поиска и других приложений, связанных с обработкой речи. В качестве преимущества данного продукта можно отметить его свободное распространение. В качестве недостатков выступает низкая точность распознавания речи – сильная зависимость от акцента, небольшой словарь модели.

Таким образом, принимая во внимание преимущества и недостатки указанных программных решений, в настоящем исследовании требуется разработать систему, позволяющую дополнительно обучить данные модели для повышения точности преобразования аудиодорожек в текст с учетом отсутствия и наличия шумов различного происхождения.

Теоретические исследования

Шумы и их классификация

Аудиошумы можно классифицировать по различным параметрам, включая источник возникновения, частоту, амплитуду и т.д. В контексте распознавания речи нас интересуют следующие типы шумов [17, 18]:

– *воздушные* – возникают при передаче звука через воздух, например уличный шум, звук проезжающих автомобилей, голоса других людей и т.д.;

– *ударные* – механические колебания, передающиеся через твердые тела (например, стук в дверь, звук шагов, вибрация от строительных работ);

– *структурные* – распространяются через строительные конструкции (например, звук от работы лифта, вибрации от проезжающего поезда);

– *атмосферные* – связаны с природными или погодными условиями (например, дождь, ветер).

Данные категории шумов будут учтены и использованы при дальнейшем проектировании и разработке системы зашумления данных.

Методы борьбы с шумом

Существуют различные методы, позволяющие улучшить работу нейросети, распознающую речь при наличии шумов [1, 17, 18]. Ниже приведены некоторые из них:

– *подавление шума* – использование алгоритмов для удаления шума или снижения его уровня в звуковых данных;

– *добавление шума* – добавление определенного уровня искусственного шума в обучающие данные для улучшения устойчивости нейросети к реальным шумам;

– *использование нескольких микрофонов* – запись звука с разных источников для получения более полной информации о звуковых сигналах;

– *обучение на больших объемах данных* – использование больших наборов данных для обучения нейросети помогает сделать ее более устойчивой к разным видам шумов.

Влияние шумов на обучение системы распознавания речи

Шум может негативно влиять на качество работы нейросети по следующим причинам [1, 17, 18]:

– *затруднение обучения* – нейросеть может испытывать трудности в обучении, если шумы маскируют или искажают полезные сигналы;

– *увеличение ошибок распознавания* – в случае, если нейросеть не может адекватно отделить полезные сигналы от шумов, точность распознавания может снижаться;

– *ухудшение качества звука* – шумы могут ухудшить общее качество звука, что также может сказаться на работе нейросети.

Шумы, встречающиеся в повседневной жизни, являются наложением множества простых шумов и частот, что делает их выделение проблематичным. Для задачи распознавания речи важно, чтобы искусственный сгенерированный шум был максимально приближен к реальному, так как только один вид шума, наложенный поверх аудиодорожки для обучения нейронной сети, не даст должного результата. То же можно сказать и про полное отсутствие шумов – нейронная сеть, обученная на датасете (от англ. *dataset* – набор данных), записанном в идеальных условиях, также не даст хороших результатов при попытке распознавания реальной аудиодорожки с шумами. Поэтому датасет для обучения таких систем должен быть максимально разнообразным – содержать как можно больше видов различных шумов, а также не содержать их. В настоящем исследовании датасет будет содержать 100000 экземпляров, что обусловливается началом переобучения НС при его большем размере.

Разработка методов обучения системы распознавания речи с использованием синтетически сгенерированного датасета с зашумлением

Для повышения эффективности распознавания речи в зашумленных аудиофайлах в настоящей работе предполагается разработка метода генерации аудиодорожки, метода зашумления аудиодорожки и метода «дообучения» нейронной сети «Whisper». На рисунке 1 представлена структурная схема работы тестового стенда. Предложенные в настоящем исследовании методы подразумевают выполнение следующих нескольких ***этапов.***

На 1 этапе генерируются аудиодорожки с использованием Variational Inference with adversarial learning for end-to-end Text-to-Speech (VITS) [14] нейронной сети (НС, использующая вариативный вывод с состязательным обучением для сквозного преобразования текста в речь), причем генерация выполняется с разными параметрами голоса диктора, чтобы Voice-To-Text нейронная сеть могла распознать как можно больше вариаций голосов.

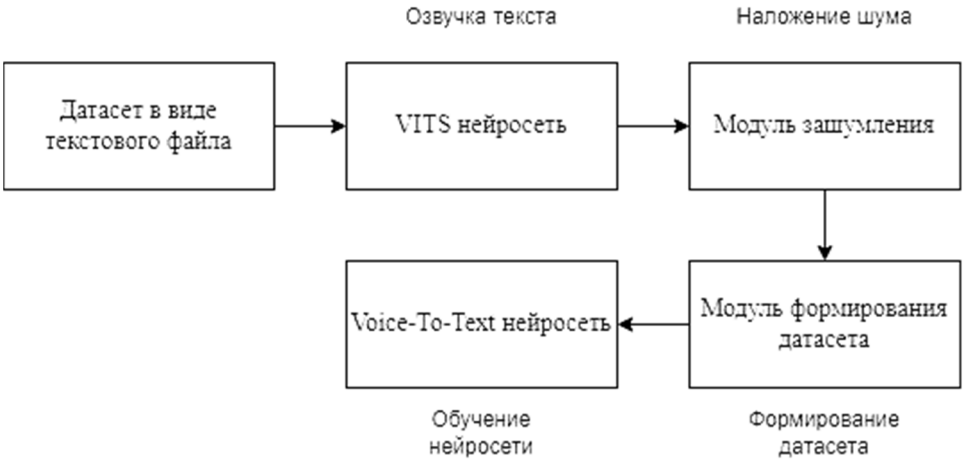

Рисунок 1 – Структурная схема работы тестового стенда
Figure 1 – Block diagram of the test bench operation

На 2 этапе сгенерированная аудиодорожка зашумляется с использованием реального шума. Все шумы представляют собой достаточно длинные аудиодорожки (около 10 минут каждая). При наложении шума выбирается случайный отрезок данной аудиодорожки, после чего происходит модификация данного отрезка – изменение высоты тона, изменение скорости, изменение тембра и т.п. Далее отрезок накладывается на сгенерированную на 1 этапе дорожку.

На 3 этапе зашумленная дорожка добавляется в датасет с расшифровкой того, что было сказано.

Такой подход к составлению датасета для обучения нейронной сети имеет следующие *достоинства*:

– быстрый сбор датасета;
– возможность сборки 1 специалистом;
– простота использования.

Среди *недостатков* рассматриваемого подхода можно отметить следующее:

– необходимость использования VITS нейросети;
– необходимость постоянно менять множество параметров для заполнения датасета разнообразными данными;
– возможность ошибок озвучивания текста.

Таким образом, использование разработанных методов позволит в теории достичь необходимых результатов. Далее приведены экспериментальные исследования для апробации разработанных методов.

Экспериментальные исследования

Описание тестового стенда

В качестве VITS нейронной сети в данном исследовании используется "silero-v4_ru" [15]. Данная нейросеть позволяет генерировать аудиодорожки с использованием русского языка в хорошем качестве, также позволяет выбрать разных дикторов, что лучшим образом повлияет на итоговый результат обучения.

В качестве обучаемой Voice-To-Text нейросети в данном исследовании выступает "Whisper" [16]. Данная нейросеть позволяет распознать текст на аудиодорожке с высокой точностью, также она поставляется с различным количеством учитываемых параметров, что позволяет подобрать вариант нейросети как по размеру занимаемого места и скорости работы, так и по точности распознавания. Для данного исследования был выбран вариант нейросети «medium».

В разрабатываемом модуле зашумления присутствует ограниченный набор шумов. Обучение нейросети осуществляется на сгенерированных данных. В модуле зашумления были использованы следующие шумы:

– *транспортный шум* (автобусный шум, автомобильный шум, шум поездов);
– *шум*, вызванный присутствием *другого диктора* в аудиодорожке;
– *природный шум* (шум леса, шум воды, шум ветра);
– *технический шум* (шум звукозаписывающего устройства);
– *индустриальный шум* (шум стройки).

В структуру тестового стенда входят персональный компьютер (ПК) с установленной операционной системой Windows 10 22H2 (19045.4291), средства разработки программного обеспечения, набор библиотек, а также нейросетевая модель «Whisper» в конфигурации «medium». Конфигурация оборудования (персональный ПК) для обучения нейросети включает:

– процессор Intel Core i5-11400F;
– графический ускоритель Gigabyte Eagle GeForce RTX 3050;
– оперативную память 16GB DDR4 3200MHz.

Обучение модели производилось с использованием инструмента "CUDA" (версии 11.6) на графическом ускорителе. Используемые программные инструменты (средства) разработки включают:

– базу данных SQLite3;
– систему управления базами данных (СУБД) DBeaver;
– язык программирования Python.

Использовались следующие программные библиотеки:

– *"pandas";*
– *"torch";*
– *"whisper";*
– *"sqlite3".*

Представленная структура тестового стенда позволяет оценить эффективность разработанных методов.

Оценка результатов обучения НС Voice-To-Text

В качестве оценки результатов обучения Voice-To-Text нейронной сети и эффективности разработанных методов выступает метрика Word Error Rate (WER) (отношение ошибочных слов) [19]. Метрика WER применяется только к самим словам, орфография игнорируется. Пример:

Исходный текст: “*Карл у Клары украл кораллы*”.
Распознанный текст: “*Клал у Клары украл кораллы*”.
WER для данного примера – 20 %, так как одно из пяти слов распознано неправильно.

Результаты обучения НС Voice-To-Text

Ниже приведены результаты распознавания обученной Voice-To-Text модели в шести различных ситуациях:

– в исходной аудиодорожке присутствует *транспортный шум* (шум метро, шум автобусов);

– в исходной аудиодорожке присутствует *индустриальный шум* (шум со стройки, шум тяжелой техники);

– в исходной аудиодорожке присутствует *природный шум* (лесной шум, шум от животных);

– в исходной аудиодорожке присутствует *технический шум* (шум записывающей техники, помехи);

– в исходной аудиодорожке присутствует голос *другого диктора* (спикера);

– в исходной аудиодорожке *отсутствуют посторонние шумы*.

Оценка результатов распознавания аудиодорожек в ходе обучения во всех перечисленных выше ситуациях представлена в таблице 1. Визуализация полученных результатов представлена на рисунке 2.

Таблица 1 – Результаты изменения WER в ходе обучения модели
Table 1 – Results of WER changes during model training

Размер датасета	WER, %						
	Средний для системы	*Для транспортного шума*	*Для индустриального шума*	*Для природного шума*	*Для технического шума*	*Шум от нескольких дикторов*	*Без шумов*
0	52,7	69,9	49,4	43,4	29,7	81,2	9,4
10000	52,1	67,3	48,9	42,6	28,4	80,2	9,3
20000	51	67	48,1	42,1	27,9	79,6	9,3
30000	50,4	66	47,3	41,4	26,9	78,4	9,1
40000	49,6	65,3	45,9	40,9	25,4	77,7	9
50000	48,3	64,1	44,8	40,1	24,3	77,1	8,7
60000	45,6	63,3	42,1	39,6	23,9	76,9	8,6
70000	44	62,8	41,3	39,1	23,6	76,2	8,4
80000	43,6	62,2	40,7	38,6	22,9	75,1	8,3
90000	41,1	61,9	39,9	38,4	22,7	74,3	8,1
100000	40,1	61,7	39	38,1	22,5	73,9	8

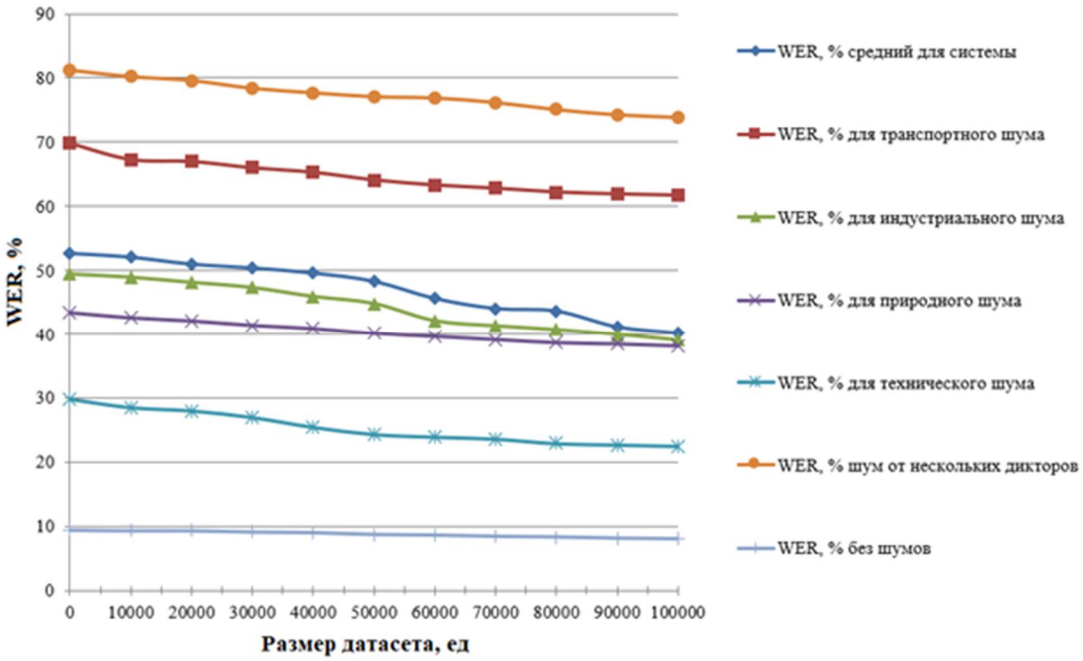

Рисунок 2 – Визуализированные результаты изменения WER в ходе обучения модели
Figure 2 – Visualized results of WER changes during model training

Из результатов проведенного эксперимента можно отметить следующее. Присутствует улучшение в распознавании по всем пунктам тестов. Особенно заметное улучшение присутствует при наличии индустриального шума в исходной аудиодорожке. Однако полученный результат не является достаточным, так как модель все еще плохо распознает слова в аудиодорожках (средний WER после обучения ~40 %).

Заключение

В ходе проведенного исследования были рассмотрены теоретические аспекты задачи распознавания речи нейросетевыми технологиями в условиях присутствия посторонних шумов различного происхождения. Определены сферы, в которых применяется распознавание речи, представлены наиболее популярные используемые программные решения, выявлены

их преимущества и недостатки. Поставлена задача исследования. Осуществлена классификация шумов, приведены существующие методы борьбы с ними. Разработаны и исследованы методы обучения нейронных сетей для транскрибировании речи с использованием искусственно сгенерированных и зашумленных данных. Спроектирован и разработан модуль зашумления данных, собран тестовый стенд. В ходе экспериментальных исследований было выявлено незначительное улучшение в распознавании текста нейронной сетью для транскрибировании речи, были предложены варианты улучшения качества распознавания.

Предполагается в дальнейших исследованиях добавить дополнительные шумы в модуль зашумления аудиодорожки, что позволит научить модель распознавать большее количество шумов или большую вариацию существующих в модуле шумов.

Библиографический список

1. **Gorodnichev M., Mkrtchian G., Furletov Yu.** On the Task of Classifying Sound Patterns in Transport // Systems of Signals Generating and Processing in the Field of on Board Communications. 2023. vol. 6, no. 1. pp. 159-167. DOI 10.1109/IEEECONF56737.2023.10092004. EDN SJVVRH.

2. **Городничев М.Г., Токарев Д.М.** Обнаружение аномалий на основе машинного обучения с использованием сочетания алгоритмов K-MEAN и SMO // Телекоммуникации и информационные технологии. 2023. Т. 10, № 1. С. 5-13. EDN ILCJZP.

3. **Леохин Ю.Л., Фатхулин Т.Д.** Разработка методов и алгоритма формализации текстового запроса к онлайн-сервисам, генерирующим изображения посредством нейросетевых технологий // Вестник Рязанского государственного радиотехнического университета. 2023. № 85. С. 82-95. DOI 10.21667/1995-4565-2023-85-82-95. EDN PZWYZV.

4. **Фатхулин Т.Д., Климов Н.Ю., Гежин С.А.** Анализ нейросетевых технологий, позволяющих генерировать текст // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 123-127. EDN MXCTVO.

5. **Фатхулин Т.Д., Лушин Е.А.** Анализ развития автоматической генерации кода для web-сервисов // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 128-132. EDN JUEGXP.

6. **Фатхулин Т.Д., Фатхулина Г.Г., Ментус М.В.** Разработка методики формирования запроса к нейросети с целью генерации изображений с учетом рекомендаций компьютерной лингвистики // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 133-139. EDN PPRTOM.

7. **Leokhin Y., Fatkhulin T. and Kulikova A.** Research of the Influence of the Semantics and Syntax of a Text Request on the Image Quality Generated by Neural Network Techniques // 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Moscow, Russian Federation, 2023, p. 1-7, doi: 10.1109/TIRVED58506.2023.10332723.

8. **Fatkhulin T., Leokhin Y., Mentus M., Kulikova A. and Alshawi R.** Analysis of the Basic Image Generation Methods by Neural Networks // 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Moscow, Russian Federation, 2023. pp. 1-7. DOI: 10.1109/TIRVED58506.2023.10332668.

9. **Перепелкин Д.А., Ткачев Д.Д.** Разработка шлюза и облачной платформы программно-конфигурируемой сети устройств Интернета вещей // Вестник Рязанского государственного радиотехнического университета. 2023. № 84. С. 88-98. DOI 10.21667/1995-4565-2023-84-88-98. – EDN HHGXCU.

10. **Перепелкин Д.А., Завалишин Г.А.** Программно-конфигурируемый Интернет вещей // Информационные технологии в прикладных исследованиях: межвуз. сб. науч. тр. – Рязань: ИП Коняхин А.В. (Book Jet), 2023. – С. 169-173. – EDN KVQOWS.

11. **Перепелкин Д.А., Ткачев Д.Д.** Разработка устройства сбора параметров для системы Интернета вещей // Информационные технологии в прикладных исследованиях: межвуз. сб. науч. тр. – Рязань: ИП Коняхин А.В. (Book Jet), 2023. – С. 176-180. – EDN UBRLBW.

12. Whisper [Электронный ресурс]: // Сайт с исходными кодами нейронной сети «Whisper». – Режим доступа: https://github.com/openai/whisper, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

13. Vosk [Электронный ресурс]: // Сайт с документацией и исходными кодами нейронной сети «Vosk». – Режим доступа: https://alphacephei.com/vosk/, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

14. VITS [Электронный ресурс]: // Сайт с исходными кодами нейронной сети «VITS». – Режим доступа: https://huggingface.co/docs/transformers/model_doc/vits, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

15. Silero models [Электронный ресурс]: // Сайт с исходными кодами нейронной сети «silero-models». – Режим доступа: https://github.com/snakers4/silero-models#v4, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

16. **Radford, A.** Robust Speech Recognition via Large-Scale Weak Supervision [Электронный ресурс] / A. Radford, J. Wook Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever // cdn.openai. – Режим доступа: https://cdn.openai.com/papers/whisper.pdf, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

17. **Вахитов Ш.Я., Ковалгин Ю.А., Фадеев А.А., Щевьев Ю.П.** Акустика: учебник для вузов. М.: Горячая линия-Телеком, 2009. 660 с.

18. **Щевьев Ю.П.** Основы физической акустики: учеб. пособие. СПб.: Лань, 2017. 367 с.

19. Fine-Tune Whisper For Multilingual ASR with HuggingFace Transformers [Электронный ресурс]: // Руководство по работе с ASR. – Режим доступа: https://huggingface.co/blog/fine-tune-whisper, свободный. – Загл. с экрана. (дата обращения 28.01.2024).

UDC 004.032.26

DEVELOPMENT AND APPLICATION OF METHODS OR RECOGNIZING NOISY AUDIO FILES USING NEURAL NETWORK TECHNIQUES

Yu. L. Leokhin, Dr. in technical sciences, full professor, deputy rector for scientific work, MTUCI, Moscow, Russia; orcid.org/0000-0003-3321-4497, e-mail: y.l.leokhin@mtuci.ru
T. D. Fatkhulin, PhD (in technical sciences), associate professor, department of MC and IT, Moscow, Russia; orcid.org/0000-0003-0998-1055, e-mail: t.d.fatkhulin@mtuci.ru
M. V. Mentus, student, MTUCI, Moscow, Russia;
orcid.org/0009-0005-8300-6954, e-mail: maksimmentus@mail.ru

The problem of speech recognition in the presence of extraneous noises of various origins is considered. The aim of the work is to develop and evaluate the effectiveness of methods that make it possible to recognize speech in the presence of noise using neural network techniques. The relevance of the ***work is*** *a significant expansion of the range of industries in which speech recognition has become much simpler and more efficient due to the development of neural network techniques. Software solutions "Whisper" and "Vosk" which allow transcribing (recognizing) speech are considered. A classification of audio noise is given, and existing methods of dealing with them are described. The influence of noise on training a speech recognition system is shown. Methods for training speech recognition system using synthetically generated dataset with noise have been developed. A data noise module was designed and developed, and a test bench was assembled. Approbation of the developed methods is given. Finally, the results of analyzing the data obtained during the experiments are presented and conclusions are drawn.*

Keywords: *speech recognition systems, speech synthesis systems, neural network training, noise, efficiency, Word Error Rate, dataset generation, noisy data.*

DOI: 10.21667/1995-4565-2024-88-65-73

References

1. **Gorodnichev M., Mkrtchian G., Furletov Yu.** On the Task of Classifying Sound Patterns in Transport. *Systems of Signals Generating and Processing in the Field of on Board Communications*. 2023, vol. 6, no. 1, pp. 159-167. DOI 10.1109/IEEECONF56737.2023.10092004. EDN SJVVRH.

2. **Gorodnichev M.G., Tokarev D.M.** Obnaruzhenie anomalij na osnove mashinnogo obuchenija s ispol'zovaniem sochetanija algoritmov K-MEAN i SMO. *Telekommunikacii i informacionnye tehnologii*. 2023, vol. 10, no. 1, pp. 5-13. EDN ILCJZP (in Russian).

3. **Leokhin Yu.L., Fatkhulin T.D.** Razrabotka metodov i algoritma formalizacii tekstovogo zaprosa k onlajn-servisam, generirujushhim izobrazhenija posredstvom nejrosetevyh tehnologij. *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta*. 2023, no. 85, pp. 82-95. DOI 10.21667/1995-4565-2023-85-82-95. EDN PZWYZV (in Russian).

4. **Fatkhulin T.D., Klimov N.Ju., Gezhin S.A.** Analiz nejrosetevyh tehnologij, pozvoljajushhih generirovat' tekst. *Trudy Severo-Kavkazskogo filiala Moskovskogo tehnicheskogo universiteta svjazi i informatiki*. 2023, no. 1, pp. 123-127. EDN MXCTVO (in Russian).

5. **Fatkhulin T.D., Lushin E.A.** Analiz razvitija avtomaticheskoj generacii koda dlja web-servisov. *Trudy Severo-Kavkazskogo filiala Moskovskogo tehnicheskogo universiteta svjazi i informatiki*. 2023, no. 1, pp. 128-132. EDN JUEGXP (in Russian).

6. **Fatkhulin T.D., G.G., Mentus M.V.** Razrabotka metodiki formirovanija zaprosa k nejroseti s cel'ju generacii izobrazhenij s uchetom rekomendacij komp'juternoj lingvistiki. *Trudy Severo-Kavkazskogo filiala Moskovskogo tehnicheskogo universiteta svjazi i informatiki*. 2023, no. 1, pp. 133-139. EDN PPRTOM (in Russian).

7. **Leokhin Y., Fatkhulin T. and Kulikova A.** Research of the Influence of the Semantics and Syntax of a Text Request on the Image Quality Generated by Neural Network Techniques. *2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*, Moscow, Russian Federation, 2023, p. 1-7. DOI: 10.1109/TIRVED58506.2023.10332723.

8. **Fatkhulin T., Leokhin Y., Mentus M., Kulikova A. and Alshawi R.** Analysis of the Basic Image Generation Methods by Neural Networks. 2023 *Intelligent Technologies and Electronic Devices in Vehicle and Road Transport* Complex (TIRVED), Moscow, Russian Federation. 2023, pp. 1-7. DOI: 10.1109/TIRVED58506.2023.10332668.

9. **Perepelkin D.A., Tkachev D.D.** Razrabotka shljuza i oblachnoj platformy programmno-konfiguriruemoj seti ustrojstv Interneta veshhej. *Vestnik Rjazanskogo gosudarstvennogo radiotehnicheskogo universiteta*. 2023, no. 84, pp. 88-98. DOI 10.21667/1995-4565-2023-84-88-98. EDN HHGXCU (in Russian).

10. **Perepelkin D.A., Zavalishin G.A.** Programmno-konfiguriruemyj Internet veshhej. *Informacionnye tehnologii v prikladnyh issledovanijah: mezhvuz. sb. nauch. tr*. – Rjazan': IP Konjahin A.V. (Book Jet), 2023, pp. 169-173. EDN KVQOWS (in Russian).

11**. Perepelkin D.A., Tkachev D.D.** Razrabotka ustrojstva sbora parametrov dlja sistemy Interneta veshhej. *Informacionnye tehnologii v prikladnyh issledovanijah: mezhvuz. sb. nauch. tr*. – Rjazan' : IP Konjahin A.V. (Book Jet).2023. pp. 176-180. EDN UBRLBW (in Russian).

12. Whisper [Jelektronnyj resurs]: *Sajt s ishodnymi kodami nejronnoj seti «Whisper»*. – Rezhim dostupa: https://github.com/openai/whisper, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).

13. Vosk [Jelektronnyj resurs]: *Sajt s dokumentaciej i ishodnymi kodami nejronnoj seti «Vosk»*. – Rezhim dostupa: https://alphacephei.com/vosk/, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).

14. VITS [Jelektronnyj resurs]: *Sajt s ishodnymi kodami nejronnoj seti «VITS»*. – Rezhim dostupa: https://huggingface.co/docs/transformers/model_doc/vits, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).

15. Silero models [Jelektronnyj resurs]: *Sajt s ishodnymi kodami nejronnoj seti «silero-models»*. – Rezhim dostupa: https://github.com/snakers4/silero-models#v4, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).

16. **Radford, A.** Robust Speech Recognition via Large-Scale Weak Supervision [Jelektronnyj resurs] / A. Radford, J. Wook Kim, T. Xu, G. Brockman, C. McLeavey, I. Sutskever // cdn.openai. – Rezhim dostupa: https://cdn.openai.com/papers/whisper.pdf, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).

17. **Vahitov Sh.Ja., Kovalgin Ju.A., Fadeev A.A., Shhev'ev Ju.P.** *Akustika* (uchebnik dlja vuzov), M.: Gorjachaja linija-Telekom. 2009. 660 p. (in Russian).

18. **Shhev'ev Ju.P.** *Osnovy fizicheskoj akustiki* (ucheb. posobie), SPb.: Lan', 2017, 367 p. (in Russian).

19. Fine-Tune Whisper For Multilingual ASR with HuggingFace Transformers [Jelektronnyj resurs]: *Rukovodstvo po rabote s ASR*. – Rezhim dostupa: https://huggingface.co/blog/fine-tune-whisper, svobodnyj. – Zagl. s jekrana. (data obrashhenija 28.01.2024).