АННОТАЦИЯ
Дархан Куанышбай

Диссертационной работы по специальности
6D070400 - “Вычислительная техника и программное обеспечение” на тему
“Разработка методов, алгоритмов машинного обучения и мобильных приложений для распознавания казахской речи”

Актуальность

Автоматическое распознавание речи (ASR) - динамичное направление в области искусственного интеллекта. За последние полвека в этой области был достигнут значительный прогресс - существует множество коммерческих приложений, которые делают инвестиции в эту область оправданными и прибыльными. Среди таких приложений, в первую очередь, можно отметить внедрение колл - центров или систем IVR (Interactive Voice Response) - систем автоматического доступа к информации в обход оператора. В современных колл - центрах вопросы формулируются пользователем на естественном языке, а ответ синтезируется компьютером также на языке пользователя. Внедрение колл - центров освободило огромное количество операторов и улучшило качество обслуживания во многих аэропортах и на железнодорожных вокзалах. Системы автоматического распознавания речи широко используются в медицинских исследованиях, требующих ввода, когда руки оператора заняты (рентген), или когда требуется управлять автономным аппаратом для исследования внутренних органов. Даже заполнение медицинских карт персоналом среднего звена в передовых медицинских учреждениях осуществляется с помощью голоса.

Важной областью применения систем автоматического распознавания и синтеза речи является помощь людям с ограниченными возможностями, такими как проблемы с костно-мышечной системой и ухудшение зрения (вспомогательные технологии).

Следует отметить, что системы автоматического распознавания речи в Казахстане практически никогда не применялись, что приводит к активным исследованиям и изысканиям в этой области для разработчиков. Основной причиной слабых исследований и исследований в области распознавания речи на казахском языке является отсутствие речевых данных. Как было замечено на популярных языках, таких как английский, испанский и китайский, системы ASR хорошего качества требуют огромного объема данных. Популярные речевые корпуса, такие как TIMIT или коммутатор, содержат огромное

количество расшифрованных аудиозаписей с различными типами речей, такими как телефонные речи, разговорная речь или речи с чистым микрофоном. На казахском языке в веб - источниках практически нет приличных речевых корпусов. Доступные из них, как правило, не бесплатны для использования и, безусловно, недостаточны для получения мощных и эффективных моделей ASR. Для того, чтобы создать достойный корпус речи для системы ASR, требуется много времени, хорошо структурированная среда и надежная система мониторинга. Однако, чтобы полностью избавиться от проблемы, связанной с дефицитом данных, следует также рассмотреть структуру и подход нейронной сети.

Речевых данных на большинстве языков с низким ресурсом даже не существует. Поэтому создание речевых корпусов является очень сложной задачей и требует огромного количества времени. Казахский язык из-за своей мало популярности считается малообеспеченным языком.

Общение между людьми может осуществляться в различных формах, таких как речь, изобразительный язык, жесты, язык жестов и т.д. Среди этих форм общение с использованием речи считается более эффективным и популярным. Это приводит к тому, что взаимодействие человека с компьютером должно осуществляться с использованием речевого общения. Поэтому этот факт подчеркивает важность разработки системы автоматического распознавания речи.

Производительность системы ASR напрямую зависит от качества речевых данных. Однако речевые корпуса могут основываться на общем языке или языке конкретной предметной области. Использование набора данных, специфичного для конкретной области, имеет преимущество в расширении возможностей процесса распознавания. Более того, использование реальной речи пользователей в качестве набора данных повышает общую производительность ASR.

Это очень важный и решающий фактор для поиска адекватного набора данных речи при построении системы ASR для языков с низким уровнем ресурсов, таких как казахский. Однако отсутствие необходимого объема речевых данных для языков с низким уровнем ресурсов очевидно. Таким образом, инструменты для сбора речевых данных могут сыграть значительную роль в создании систем распознавания речи.

Цели и задачи исследования.

Для того чтобы построить надлежащую систему ASR, избегая проблемы дефицита данных, наши основные цели заключаются в следующем:

- Создать хорошо продуманную среду для сбора речевых данных с использованием веб - платформы
- Разработать модель синтеза речи для создания автоматической системы сбора речи для небольших предложений
- Собрать значительный объем речевых данных с транскрипциями для казахского языка
- Для последующей обработки собранных данных и структурирования файлов для работы нейронной сети
- Построить нейронную сеть с использованием рекуррентных нейронных сетей на основе функции потери CTC
- Построить многоязычную методику с русским языком с использованием подхода, основанного на передаче знании

Объект исследования

Исследование сосредоточено на методах и приемах автоматического сбора речевых данных для систем распознавания речи.

Методы исследования

Исследование проводилось путем анализа и интерпретации существующих результатов современных работ в области распознавания речи, синтеза речи, обработки естественного языка, определяющих преимущества и недостатки. Поставленные цели были достигнуты за счет применения алгоритмов машинного обучения, последних достижений рекуррентных нейронных сетей и методов моделирования последовательности.

Научная новизна

Новинками, которые были получены в ходе выполнения поставленных задач, являются следующие:

- Применение нового подхода к сбору речевых данных для любого языка путем создания надежного веб - сайта с хорошо продуманной системой мониторинга и контроля
- Построение автоматического этапа постобработки речевых данных после процесса сбора, который структурирует речь и транскрипцию в отношении обучения нейронной сети на основе CTC

- Обучение нейронной сети с использованием многоязычного подхода путем передачи знаний, полученных из предварительно обученной модели русского языка

- Автоматическое выравнивание последовательности ввода речи с последовательностью вывода транскрипции

Научные утверждения подлежат защите

- Платформа методов автоматического сбора речевых данных, основанных на моделировании синтеза речи
- Автоматическая предварительная обработка и структурирование речевых данных с соответствующей текстовой транскрипцией
- Метод коннекционистских алгоритмов временной классификации для обучения рекуррентной нейронной сети
- Методы и алгоритмы построения подхода к обучению передаче с использованием модели распознавания русской речи
- Сравнительный анализ моделей долговременной кратковременной памяти (LSTM) и двунаправленных рекуррентных нейронных сетей на основе LSTM

Апробация работы

Результаты исследования были доложены и представлены на международных конференциях: “III Международная научная конференция “Информатика и прикладная математика”, посвященная 80-летию профессора Р. Г. Бияшева и 70-летию профессора М. Б. Айдарханова”(Казахстан, Алматы, 2018); “IV Международная научно-практическая конференция “Информатика и прикладная математика”, посвященная 70-летию процессора Биярова Т. Н., Вальдемара В. И 60-летию профессора Амиргалиева Е. Н.” (Казахстан, Алматы, 2019); “Тенденции современной науки” (Великобритания, Шеффилд, 2019)

Публикации

По теме диссертации представлено 14 опубликованных работ, в том числе:

- 1 монография
- 2 статьи в международном журнале, индексируемом Scopus
- 4 статьи, соответствующие требованиям высшей аттестационной комиссии Министерства образования и науки Республики Казахстан
- 4 статьи, опубликованные на Международных конференциях
- 2 свидетельства об авторстве/патентах
- 1 публикация в иностранном журнале

1. Амиргалиев Е.Н., Куанышбай Д.Н., “Вербальный робот”, Алматы: ИИВТ, 2020. -143 с, ISBN 978-601-08-0090-8
2. Kuanyshbay D.N., Amirgaliyev Y, Shoiynbek A., Yedilkhan D., “Automatic speech recognition system for kazakh language using connectionist temporal classifier”, Journal of Theoretical and Applied Information Technology, Vol. 98 No.04, PP 703-713, ISSN: 1992-8645(print), ISSN: 1817-3195(online) (SCOPUS: CiteScore:1.2, Quartile: Q3, percentile: 37)
3. Kuanyshbay D.N., Amirgaliyev B., Kutubayeva M., Baimuratov O., “Development of Automatic Speech Recognition for Kazakh Language using Transfer learning”, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9 No.04, ISSN: 2278-3091(online) (SCOPUS: CiteScore:1.2, percentile: 36, Quartile: Q3)
4. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Comparison of two speech signal features for deep neural networks to classify emotions”, Вестник КазНИТУ. Том 132 №2 2019, с.343-350 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
5. Kuanyshbay D.N., Amirgaliyev Y.N., Musabaev T.R., Kenshimov Sh., “Разработка голосовой идентификации диктора с использованием статистик контура основного тона для применения в робото-вербальных системах”, Вестник КазНИТУ. Том 132 №2 2019, с.424-433 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
6. Куанышбай Д.Н., Амиргалиев Е, Едилхан Д., Баймуратов О., “Об одном улучшенном подходе автоматического распознавания казахской речи с использованием трансферного обучения”, Вестник КазНИТУ. Том 141 №5 2019, с.183-189 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
7. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Various Languages impact on the problem of emotion recognition in speech”, Вестник КазНИТУ. Том 141 №5 2019, с.182-185 ISSN 1680-9211 (Импакт-фактор по Казахстанской базе цитирования за 2017 год = 0,045)
8. Kuanyshbay D.N., Amirgaliyev Y.N., Kozhakhmet K, Shoiynbek A., “Comparison of optimization algorithms of connectionist temporal classifier for speech recognition system”, "Informatyka, Automatyka, Pomiary w Gospodarcei-OchronieŚrodowiska" - IAPGOS, Vol. 9 No.3, PP 54-57, ISSN: 2083-0157(print), ISSN: 2391-6761 (online)
9. Куанышбай Д.Н., Кожахмет К.Т., Шойынбек А., “Создание корпуса эмоций на казахском и русском языке для голосового распознавания эмоций”, MATERIALS OF XV INTERNATIONAL RESEARCH AND PRACTICE CON-

FERENCE «TRENDS OF MODERN SCIENCE - 2019» Vol. 14, PP 9-11, Science and Educational LTD, Sheffield, UK ,2019 ISBN 978-966-8736-05-6

10. Kuanyshbay D.N., Kozhakhmet K, Shoiynbek A., “Comparison of classification algorithms svm vs logistic regression for detecting crime”, МАТЕРИАЛЫ III Международной научной конференции «Информатика и прикладная математика», посвященная 80-летнему юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. 2018 года, Алматы, Казахстан, Часть 2 , с 37-41, ISBN 978-601-332-165-3
11. Kuanyshbay D.N., Amirgaliyev Y, Shoiynbek A., “Speech recognition preprocessing, background removal”, МАТЕРИАЛЫ III Международной научной конференции «Информатика и прикладная математика», посвященная 80-летнему юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. 2018 года, Алматы, Казахстан, Часть 2 , с 7-18, ISBN 978-601-332-165-3
12. Куанышбай Д.Н., Амиргалиев Е.Н., Шойынбек А., “Построение языковой модели для казахского языка на основе рекуррентных нейронных сетей”, МАТЕРИАЛЫ IV международной научно -практической конференции "Информатика и прикладная математика", посвященной 70-летнему юбилею профессоров Биярова Т.Н., Вальдемара Вуйцика и 60-летию профессора Амиргалиева Е.Н. 2019, Алматы, Казахстан Часть 1 с. 401 – 406, ISBN 978-601-332-384-8
13. Куанышбай Д.Н., Баймуратов О., “Алгоритм синтеза речи казахского языка”, Свидетельство о внесении сведений в государственный реестр прав на объекты охраняемые авторским правом. Вид объекта авторского права: программа для ЭВМ. Запись в реестре №15029 от 10 февраля 2021 года
14. Куанышбай Д.Н., Амиргалиев Е.Н, Козбакова А.Х., “Автоматизированная система формирования корпуса речевых данных”, Свидетельство о внесении сведений в государственный реестр прав на объекты охраняемые авторским правом. Вид объекта авторского права: программа для ЭВМ. Запись в реестре №15236 от 17 февраля 2021 года

Структура диссертации

Диссертация состоит из введения, 5 глав, заключения и списка литературы. Объем диссертации состоит из 120 страниц, 30 рисунков, 124 ссылок и 10 приложений.

Первая глава состоит из обзора самой обработки речевого сигнала и методов обработки речевого сигнала. Он предоставляет общую информацию о типах

шумов, таких как широкополосный шум, мешающие речи и периодические шумы. Кроме того, в главе показана и проиллюстрирована система улучшения речи (рис. 1) и приведен анализ различных типов методов улучшения речи.

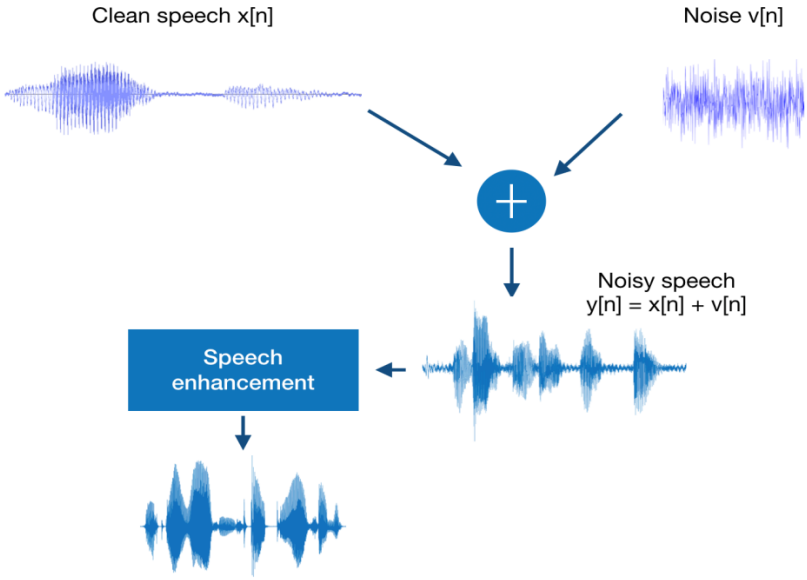


Рисунок 1. Система улучшения речи

Были представлены различные типы методов шумоподавления и улучшения речи, такие как линейное прогнозирующее кодирование, метод подпространства сигналов, методы на основе DFT и т.д.

Во второй главе представлен анализ существующих систем автоматического распознавания речи. Сначала она охватывает акустическую модель на основе скрытой модели Маркова, построенную с использованием вероятностной методологии. В процессе распознавания невидимые слова распознаются путем сравнения их с уже виденными шаблонами и нахождения ближайшего из них. Но эти шаблоны не могут соответствовать акустическим изменениям. Поэтому акустические модели строятся путем распределения вероятностей по акустике. Самый простой способ - использовать распределение Гаусса, которое находит параметры представления.

Задача распознавания речи первоначально рассматривалась как задача статистической классификации. Классы, определенные как последовательность слов W из большого набора словарного запаса, и входные данные, определенные как особенности речи X. Уравнение выглядит следующим образом:

$$P(W \mid X) = \frac{P(X \mid W)P(W)}{P(X)}$$

Во-вторых, в главе рассматривается акустические модели на основе ANN. Одним из главных преимуществ ANN перед другими методами моделирования в распознавании речи является возможность аппроксимации нелинейных динамических систем. Речь-это нелинейный сигнал, создаваемый нелинейной системой.

ANN в основном представляет собой взаимосвязь вычислительных элементов (нейронов), и эта нелинейная система распределена по сети. Базовая нелинейная модель нейрона показана на рис. 2.

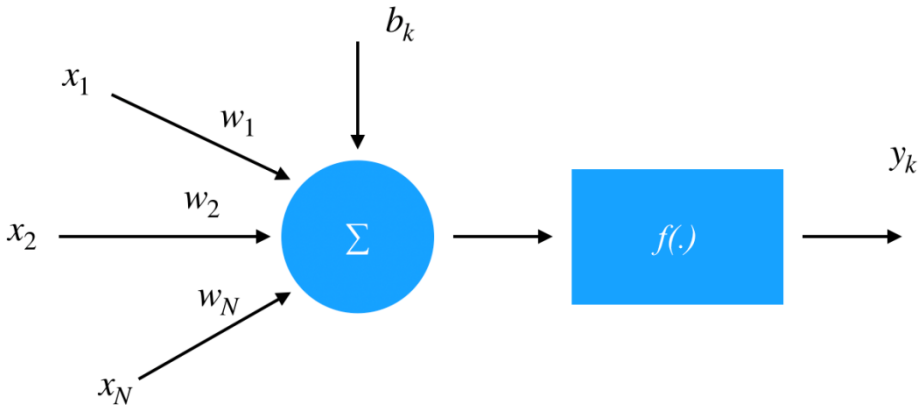


Рисунок 2. Нелинейная нейронная модель

На ранней стадии была успешно предпринята попытка использования ANN для распознавания речи при распознавании простых цифр, нескольких фонем и слов с использованием многослойного персептрона.

В-третьих, далее следует глава с глубокими убеждениями, основанными на акустических моделях нейронных сетей.

Нейронная сеть глубокого убеждения (DBNN) была обнаружена как мощная и эффективная для многих задач машинного обучения, а также для акустического моделирования в системе распознавания речи на основе HMM/ ANN. Сначала DBNN был представлен для задач акустического моделирования, поскольку он обладал гораздо большей способностью к моделированию, чем обычный GMM. Кроме того, DBNN также продемонстрировал очень эффективную способность к обучению, которая объединяет неконтролируемое

обучение для обнаружения функций и контролируемое обучение для оптимизации и точной настройки функций. Еще одна причина эффективности DBNN заключается в том, что характеристики объектов низкого уровня вычисляются в нижних слоях, а сильно нелинейная структура входных данных учитывается в более высоких слоях. Это может быть похоже на распознавание человеческой речи, которое использует множество слоев для извлечения функций.

В третьей главе рассматриваются новые подходы к обучению нейронной сети, такие как CTC(Connectionist Temporal Classifier) и сквозные модели на основе RNN-преобразователей. Были проанализированы и объяснены тренинги по моделям распознавания речи на основе CTC. CTC вводит пустой символ, чтобы сопоставить две последовательности вместе, вычисляя вероятность пути. После этого выполняется алгоритм агрегирования путей (рис. 3).

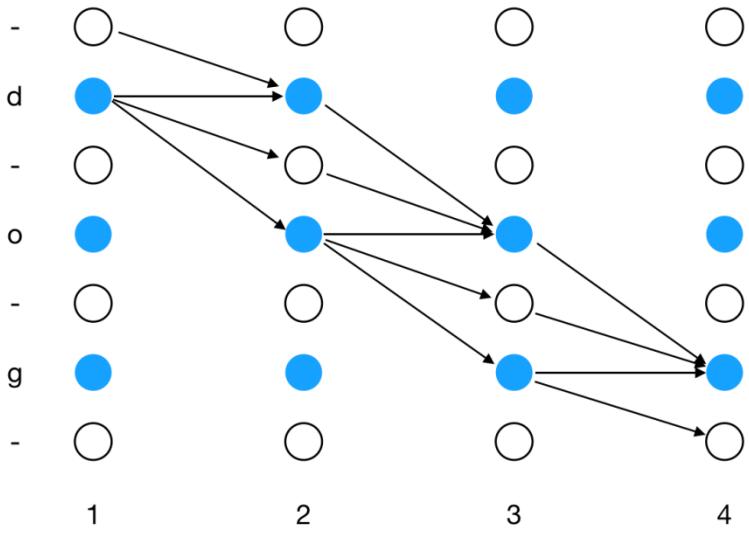


Рисунок 3. Пример пути для слово “dog”

В следующем разделе представлена сквозная модель на основе преобразователя RNN. Модель на основе RNN-преобразователя с точки зрения структуры имеет много общего с моделью на основе CTC. Например, они оба используют одну и ту же функцию потерь; они оба решают проблему ручной сегментации между входной последовательностью и выходной последовательностью; они используют элемент “черный символ”; они оба оценивают вероятности всех путей и агрегируют пути для получения последовательности меток. Но процессы генерации пути и оценки вероятности пути очень разные. В главе также определяются преимущества и недостатки этого алгоритма по сравнению с моделью, основанной на CTC.

Модель на основе RNN-преобразователя содержит три важных компонента: Сеть для транскрипции *(F(x))*; сеть прогнозирования *((P(y, g))*; совместная сеть *(J(f,g))*. Построение модели RNN-преобразователя проиллюстрировано на рисунке 4.

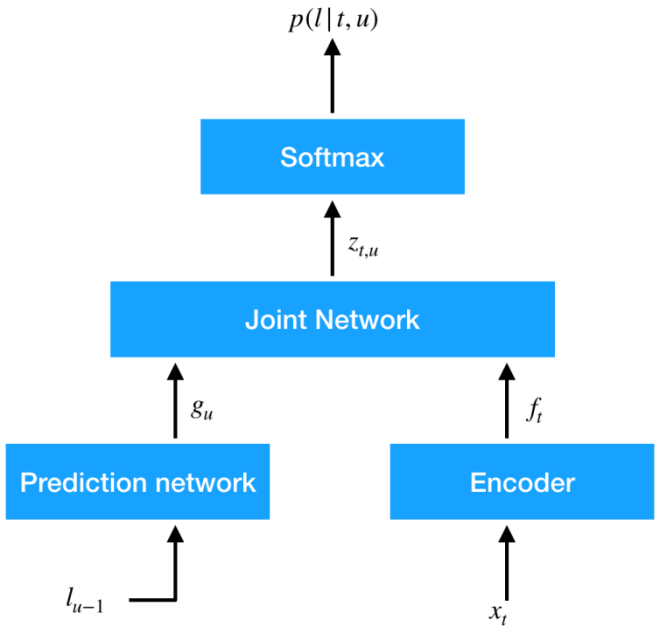


Рисунок 4. Структура RNN-преобразователя

В четвертой главе представлена информация о создании автоматической системы сбора речи, построенной на основе веб-приложения (рис. 5).

Это приложение было разработано с использованием новейших доступных технологий, которые подходят для любого устройства или компьютера, представленного сегодня на рынке. Бэкэнд приложения был выполнен с использованием yii2-фреймворка на языке php7.2. Разработка архитектуры для этого проекта основана на шаблоне Model-View-Controller (MVC).

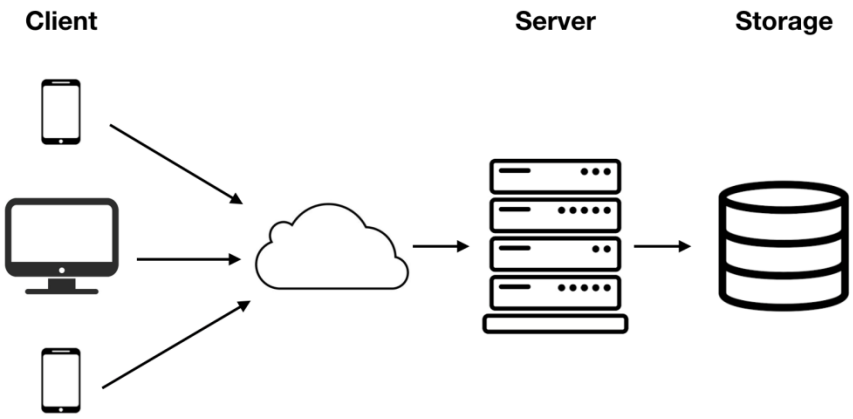


Рисунок 5. Представление архитектуры системы

В качестве HTTP-сервера мы использовали движок NGINX. Для управления базами данных в приложении мы использовали язык, основанный на MySQL. Панель управления была обработана phpMyAdmin и HeidiSQL. Для редактирования кода мы использовали технологию phpstorm. Каждая веб -страница в приложениях имеет возможность изменять разрешение в зависимости от текущего устройства.

Веб - приложение состоит из модели синтеза речи, которая автоматически произносит небольшие предложения, извлеченные из загруженной книги пользователем-администратором. Эта модель синтеза речи была обучена на книгах казахской литературы с использованием сверточной нейронной сети (CNN) (рис. 6).

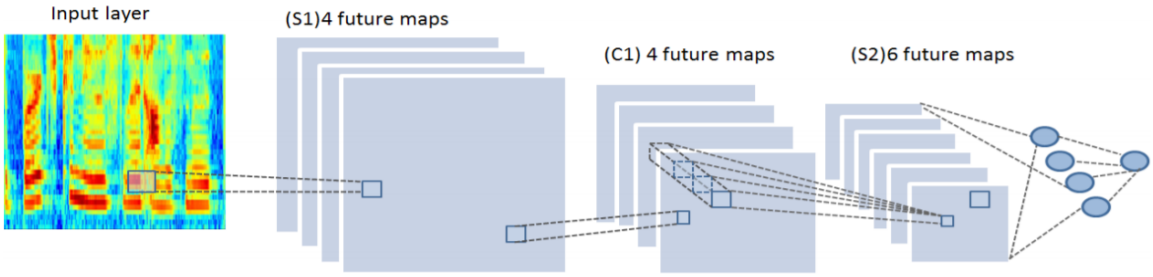
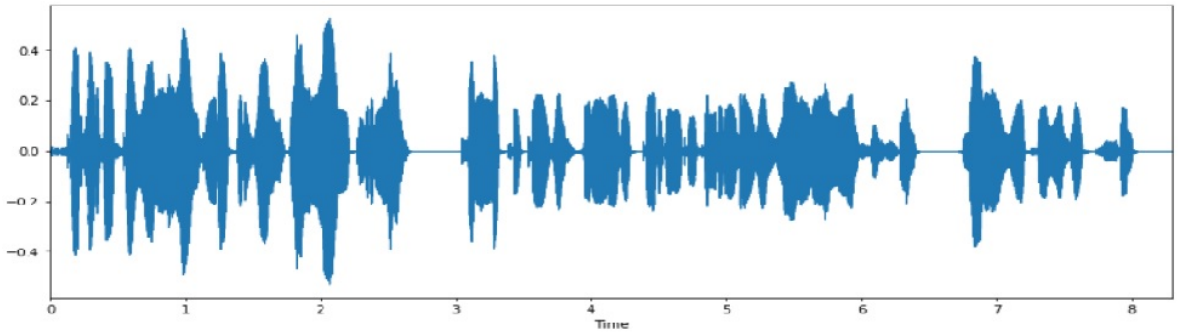


Рисунок 6. Архитектура многомерной сети CNN

При построении систем синтеза речи одной из важнейших задач является сегментация и маркировка баз данных речевых сигналов на семантически и фонетически значимые единицы речи, и в нашем конкретном случае это

фонемы. Полученные сегменты сохраняются в базе данных и используются для машинного обучения акустических моделей в интегрированной системе с последующей генерацией голоса говорящего в системе синтеза текста в речь. Одним из специфических методов работы с обученными системами является настройка параметров системы для определенного языка и выбор алфавита. Поскольку модель синтезатора казахской речи использует фонемы в качестве входных символов алфавита, необходимо было решить задачу транскрибирования текстов по правилам графемы на фонему с учетом фонетических особенностей казахского языка. Задача была решена путем создания модуля фонетической транскрипции, а затем был расшифрован подготовленный текст- обучающий образец.

Таким образом, была создана обучающая экспериментальная база, состоящая из 3500 предложений, и каждое предложение соответствует аудиофайлу в формате wav с частотой дискретизации 22050 Гц. После этого был активирован системный модуль, отвечающий за прием спектрограмм аудио файлов , на основе которых происходит глубокое изучение сетей (рис. 7).



білім беруде, денсаулық сақтау мен көлік саласында болады.

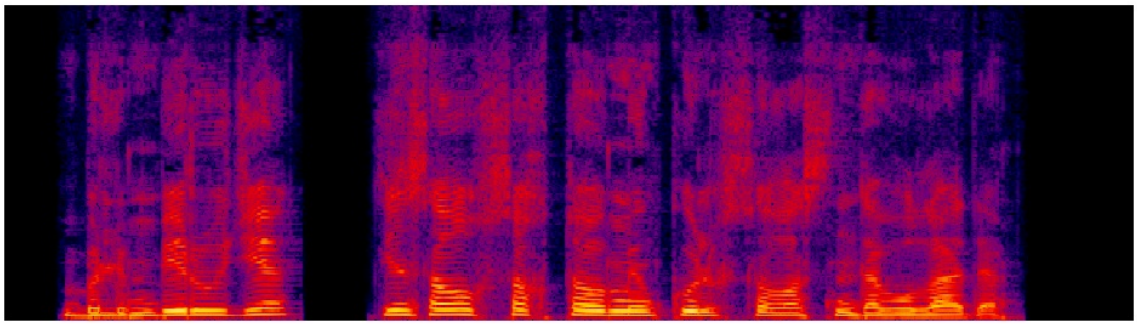
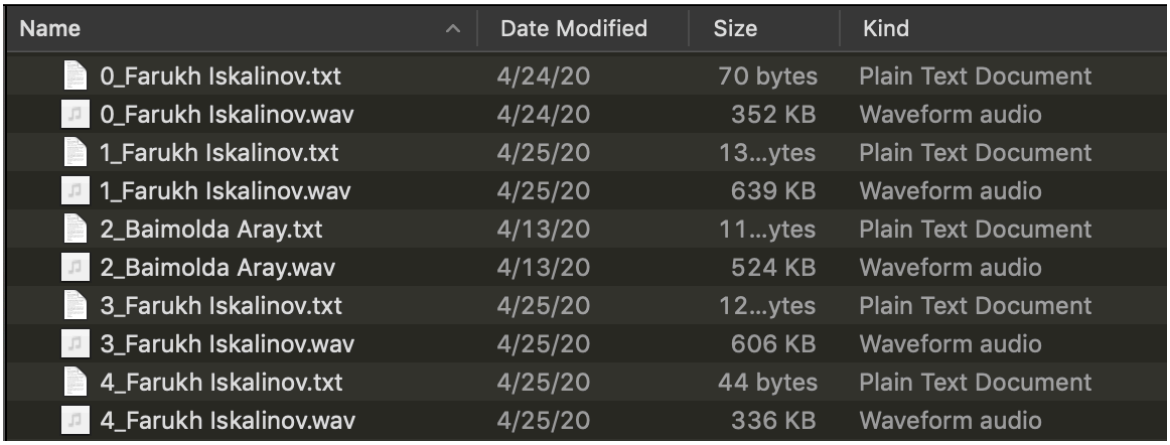


Рисунок 7. Пример предложения из учебной базы

В следующем разделе представлен обзор самого веб-приложения с описанием всех его функций и функций.

Система состоит из двух важных функций для завершения всего процесса сбора данных. Это пользователь-администратор (всего 1-2 пользователя) и большое количество клиентов. Администратор имеет право добавлять любое количество книг в любом формате, удалять и редактировать. Администратор может просматривать процент полноты определенной книги и может прослушивать все записи клиентов. Он может перечислить записи любой книги, а также любого оратора. У него есть возможность удалить нежелательные или поврежденные записи и отправить по электронной почте докладчикам, допустившим ошибку. После завершения всего процесса записи администратор может загрузить все записи. Загруженный zip-файл уже структурирован с папками динамиков и всей необходимой транскрипцией с соответствующими аудиофайлами.



Name	Date Modified	Size	Kind
0_Farukh Iskalinov.txt	4/24/20	70 bytes	Plain Text Document
0_Farukh Iskalinov.wav	4/24/20	352 KB	Waveform audio
1_Farukh Iskalinov.txt	4/25/20	13...ytes	Plain Text Document
1_Farukh Iskalinov.wav	4/25/20	639 KB	Waveform audio
2_Baimolda Aray.txt	4/13/20	11...ytes	Plain Text Document
2_Baimolda Aray.wav	4/13/20	524 KB	Waveform audio
3_Farukh Iskalinov.txt	4/25/20	12...ytes	Plain Text Document
3_Farukh Iskalinov.wav	4/25/20	606 KB	Waveform audio
4_Farukh Iskalinov.txt	4/25/20	44 bytes	Plain Text Document
4_Farukh Iskalinov.wav	4/25/20	336 KB	Waveform audio

Рисунок 8. Структура файлов

Клиент может зарегистрироваться на веб-сайте, только введя имя, пол и возраст, и выбрать любую книгу, указанную администратором, для дальнейшего процесса записи. Клиент не имеет права добавлять, удалять или редактировать какую-либо книгу, за исключением прослушивания своих собственных записей в целях улучшения. В процессе записи клиент имеет возможность повторно записать текущее предложение после многократного прослушивания его записи.

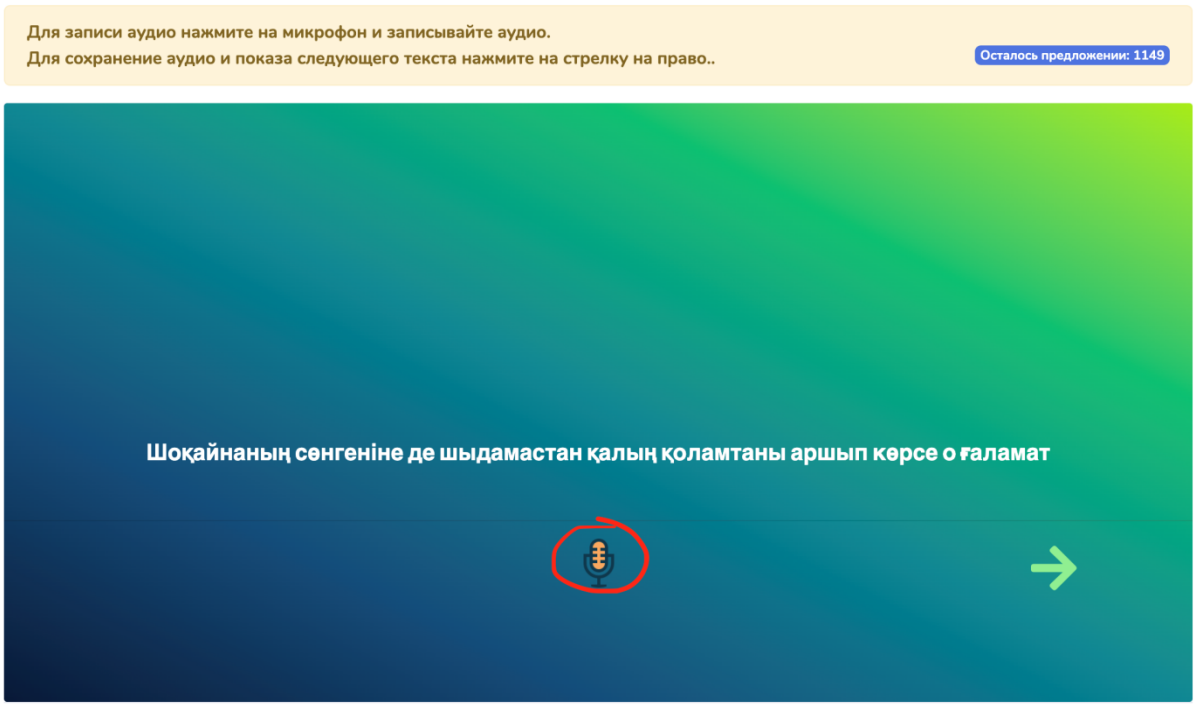


Рисунок 9. Страница для записи

В следующих подразделах описываются функции динамиков, такие как запись предложения (рисунок 9), повторная запись предложения, проверка вручную на наличие искажений, список записей (рисунок 10), структура сохраненных файлов (Рисунок 8) и т. д.

Исследователи могут извлечь выгоду из такой технологии, поскольку она обеспечивает большое количество людей и короткое время сбора данных. Кроме того, он обеспечивает хорошее качество записи звуковой речи, благодаря возможности мониторинга и контроля всего процесса записи. Сейчас все инструкции и правила представлены на казахском языке, но он может быть адаптирован к любому языку с ограниченными ресурсами. Таким образом, этот инструмент приложит огромные усилия для повышения популярности известных языков, помогая собирать речи удобным и безболезненным способом.

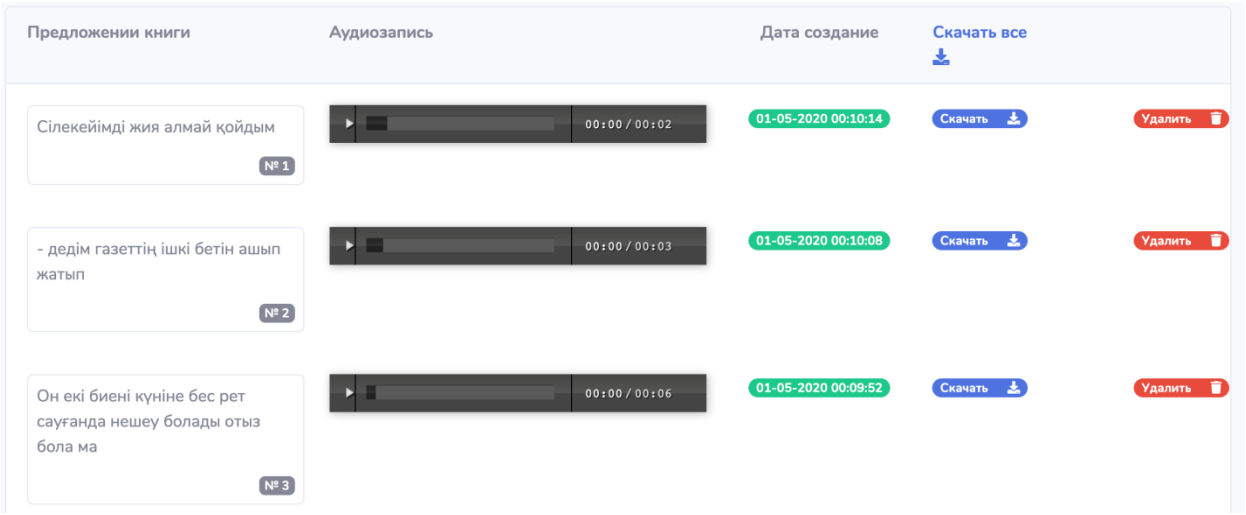


Рисунок 10. Список записей

В пятой главе реализована система автоматического распознавания речи (ASR) с использованием речевых данных, собранных в предыдущей главе. Общие данные состоят из четких 100 часов выступлений. Основной идеей построения нейронной сети является подход к обучению передаче (рис. 11).

Используя предварительно обученную модель распознавания русской речи, все веса были перенесены в нашу собственную нейронную сеть. Модель распознавания русской речи была обучена на наборе данных VoxForge, содержащем 100 часов речевых данных.

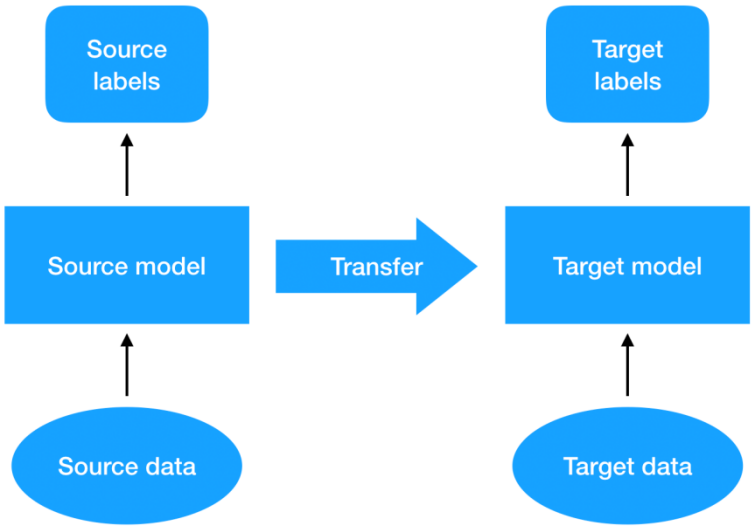


Рисунок 11. Высокоуровневое представление трансфертного обучения

Мы провели в нашем эксперименте 2 различных типа RNN: Долговременная кратковременная память (LSTM) и двунаправленная LSTM. Поскольку BiLSTM в недавних исследованиях показал многообещающие результаты и очень хорошо понимает контекст, мы решили сравнить его с обычным LSTM.Как мы уже упоминали выше, в качестве среды для этого эксперимента мы использовали ноутбук Jupiter на языке программирования Python. В частности, мы использовали библиотеку Tensorflow на основе Python для создания и обучения модели ASR. Более того, для построения самой нейронной сети было использовано расширение библиотеки Keras Tensorflow. Набор данных речи клонирован из репозитория GitHub. Хранилище состоит из трех папок: набор для обучения, набор для проверки и набор для тестирования. Обучение проводилось на нескольких графических процессорах (GPU) Tesla K80. Сервер с этими процессорами был арендован на три месяца использования. Список параметров, которые использовались в нашем эксперименте, выглядит следующим образом:

- два слоя LSTM и BiLSTM отдельно
- 128 единиц нейронов в каждом слое
- 500 эпох обучения
- Слой отсева после каждого слоя LSTM с вероятностью 50%
- Размер партии составляет 4
- Значение импульса в оптимизаторе импульса равно 0,9
- Скорость обучения составляет 0,0005
- Функция потерь- CTC
- Метрикой является Частота ошибок меток (LER)

Мы рассмотрели 4 различных сценария: 1) Нейронная сеть LSTM без использования модели русского языка; 2) Нейронная сеть LSTM с использованием модели русского языка; 3) Двунаправленный LSTM без использования модели русского языка; 4) Двунаправленный LSTM с использованием модели русского языка. Они использовали одинаковое количество слоев и одинаковое количество нейронов в каждом слое. Результаты процесса обучения можно оценить в таблице 2.Результат каждой рекуррентной нейронной сети на самом деле очень близок, но мы видим, что архитектуры с трансфертным обучением явно улучшают все.

Многоуровневая нейронная сеть LSTM с внешней моделью повысила стоимость обучения до 8%, в то время как частота ошибок меток увеличилась до 4%. Двунаправленный LSTM показал очень многообещающие результаты, повысив стоимость обучения до 24% и снизив частоту ошибок маркировки до 32% (рис. 12,13).

Эксперимент показал, что использование внешней модели системы ASR на русском языке для передачи своих знаний в систему казахского языка значительно повышает производительность.

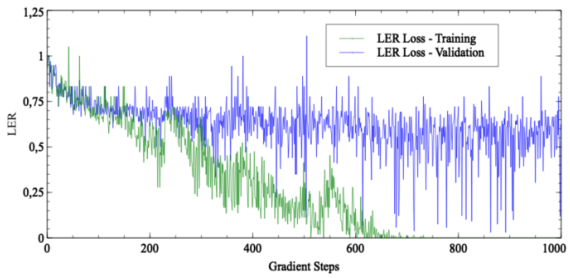


Рисунок 12. Иллюстрация LER

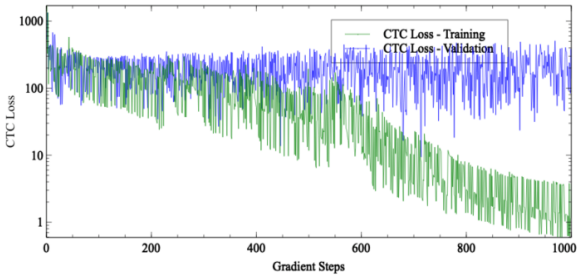


Рисунок 13. Иллюстрация CTC потери

Заключение.

В процессе исследования области распознавания речи были получены все поставленные цели и задачи:

- Была построена платформа с хорошо продуманной средой и системой мониторинга для автоматического сбора речевых данных. Он основан на веб-приложении , содержащем модель синтеза речи, обученную с использованием сверточной нейронной сети.

- С помощью инструмента для сбора речевых данных было собрано более 50 часов данных. В процессе сбора данных из общей численности говорящих 83% составляли мужчины и 17% - женщины.

- Подход к обучению передаче был применен к казахской системе ASR с использованием предварительно построенной модели распознавания русской речи. Путем построения нейронной сети на основе долговременной кратковременной памяти и переноса всех весов с модели распознавания русской речи была достигнута многоязычная модель распознавания речи.

Используя среду сбора речи, было задействовано 65 носителей языка, и менее чем за 1,5 месяца было получено более 50 часов чистых речевых данных.

Система распознавания казахской речи была обучена с использованием нейронной сети на основе слоев LSTM, BiLSTM. Применение многоязычного подхода с использованием трансфертного обучения (русская готовая модель) улучшило производительность казахской модели ASR на 24% с точки зрения частоты ошибок в метках.