УДК 004.934

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РАСПОЗНАНИЯ ГОЛОСА

Муравьев М.О.

МИРЭА - Российский технологический университет, 119454, Россия, г. Москва, проспект Вернадского, 78, e-mail: maxim_mour@mail.ru

На сегодняшний день наблюдается активная разработка и интеграция инновационных решений из области информационных технологий. При этом особую актуальность получают технологии и методы распознавания голоса, используемые во многих областях жизнедеятельности современного человека. Цель текущей статьи заключается в рассмотрении методов распознавания голоса в голосовых помощниках. Автором предпринимается попытка систематизации материалов относительно существующих методов распознавания голоса.

Ключевые слова: Распознавание голоса, информационные технологии, голосовой помощник, распознавание речи.

COMPARATIVE ANALYSIS OF VOICE RECOGNITION METHODS

Muravyov M.O.

MIREA - Russian Technological University, 119454, Moscow, 78 Vernadskogo Avenue, Russia, e-mail: maxim_mour@mail.ru

Today, there is an active development and integration of innovative solutions from the field of information technology. At the same time, technologies and methods of voice recognition, used in many areas of the life of a modern person, are of particular relevance. The purpose of the current article is to review voice recognition methods in voice assistants. The author makes an attempt to systematize the materials on the existing methods of voice recognition.

Keywords: Voice recognition, information technology, voice assistant, speech recognition.

Введение

В современном мире наблюдаются устойчивые тенденции, связанные с развитием и интеграцией в различных бытовых и профессиональных сферах жизнедеятельности технологий распознавания голоса. На основе данных инструментов представляется возможность взаимодействия между человеком и цифровым устройством исключительно на основе использования голоса. Использование методов распознавания голоса является основой работы современных голосовых помощников, способных упростить и качественно улучшить жизнь современного человека. Наиболее распространёнными из примеров использования данных технологий являются индивидуальные голосовые помощники, телефонные роботы и ряд иных задач, при выполнении которых происходит экономия временных ресурсов человека [1].

Помимо этого, в зависимости от используемых методов распознавания голоса, активно разрабатываются интеллектуальные системы обучения, способные принимать экзамен и помогать в изучении иностранных языков. При этом использование голосовых помощников получает свое развитие и в решении наиболее сложных задач из профессиональных сфер человеческой жизнедеятельности. Примерами данного применения являются – идентификация личности, судебная экспертиза, помощь людям с ограниченными возможностями и ряд иных задач.

Основные направления распознавания голоса

Широкое распространение голосовых помощников стало возможным благодаря развитию методов распознавания голоса. На сегодняшний день существует два основных направления развития по распознаванию речи – традиционное направление автоматического распознавания речи и направление комплексного (end-to-end) подхода к распознаванию речи [9]. При этом необходимо отметить, что каждый из данных методов имеет уникальные особенности, преимущества и недостатки.

Традиционный метод автоматического распознавания речи представляет собой совокупность компьютерного оборудования и программных технологий, выполняющих прямую идентификацию и обработку человеческого голоса. Принцип работы данной технологии может быть определена в качестве автоматической транскрипции разговорного языка в читаемый текст. При этом распознавание голоса происходит в режиме реального времени на основе заранее заданных звуковых шаблонов. Таким образом, данный метод ASR представляет компьютеру возможность выявить слова из человеческой речи и перевести их в электронный текст [2].

Основное использование данного метода наблюдается в задачах распознавания слов для идентификации речи человека. Данный метод подразумевает разделение ответственности за распознание различных компонентов

языка – последовательное преобразование звука в фонемы, фонем в слова, слов – в связные предложения. Подобная архитектура определяет с одной стороны преимущества данного подхода – модульность и возможность повторного использования некоторых данных компонентов для разных языков, имеющих похожие фонемы, словоформы или синтаксис. В то же время, из этого выходят и недостатки подобного подхода – потеря ряда данных, на каждом из преобразований.

Особое внимание заслуживает инновационный метод комплексного подхода к распознаванию речи. Данный метод представляет собой актуальное направление из области машинного обучения, в котором обработка голоса производится на основе интеллектуальных алгоритмов. Данный подход отличается от традиционного тем, что фонемы преобразуются напрямую в текст, минуя промежуточные формы представления. При этом один из вариантов процесса распознавания голоса в данном методе, примененного в голосовом помощнике, может выглядеть таким образом: запись речи человека; преобразование машиной слов из аудио в электронный текст; разбор текста на основные составляющие для понимания контекста беседы и целей человека; После определения контекста, алгоритм может возвратиться к этапу анализа аудио для более точного преобразования в текст и повторить цикл; по результатам работы система определяет команду на выполнение [3].

В таблице 1 сведены основные особенности, сферы использования, преимущества и недостатки рассмотренных методов распознавания голоса:

Таблица 1. Сравнение методов традиционного и комплексного распознания речи

	Традиционное распознание речи	End-to-end распознание речи
Преимущества	- модульность архитектуры ПО на основе этой архитектуры - повторное использование одних и тех же решений для схожих языков - высокая производительность, работа в реальном времени	- отсутствие потери данных из-за преобразования данных в различные промежуточные абстракции - возможность учета контекста на уровне распознания речи - высокое качество распознания
Недостатки	- нет возможности учета контекста в ходе распознания - низкое качество распознания сленговых выражений, имен собственных, аббревиатур	- более высокие требования к производительности - высокие требования к качеству и размеру корпуса текстов для обучения
Особенности	Сочетание высокой производительности и модульности позволяет использовать даже на портативных устройствах.	Высокое качество распознания при высоких требованиях к производительности на сегодняшний день ограничивают применение облачными технологиями.
Области использования	Широко применяется в случаях необходимости распознания речи в реальном времени или распознания на локальных устройствах – голосовые помощники, «умные» устройства с голосовым управлением.	Является перспективным решением для извлечения текстовых стенограмм, системах корпоративного и государственного управления, перспективных голосовых помощниках с обработкой голоса в облаке.

Для экспериментальной оценки данных методов распознания речи было проведено тестирование на примере двух программных продуктов, использующих данные два метода распознания речи соответственно.

Проведение тестирования существующих программных продуктов

В рамках данной статьи будут рассмотрены следующие программные продукты, которые применяют данные технологии для конвертации аудио дорожки в текстовой формат:

– Vosk-Browser Demo [5];
– RealSpeaker Transcriber [4].

Vosk-Browser Demo использует технологию традиционного распознания речи и применяет библиотеку Vosk — практическую библиотеку распознавания речи в реальном времени, которая поставляется с набором точных моделей, сценариев, практик и обеспечивает готовый к использованию сервис распознавания речи для различных платформ. Библиотека поставляется со множеством различных словарей на различных языках.

RealSpeaker Transcriber является результатом российского стартапа «РеалСпикер» и представляет приложение по превращению аудио- и видеофайлов в текстовые документы. RealSpeaker Transcriber основан на технологии по распознаванию речи на базе интеллектуальной системы, оценивающей не только аудиодорожку, но и видео с человеческой мимикой, и контекст распознаваемого текста, который позволяет точнее определять частоту встречаемости слов и их взаимодействия между собой, позволяя уточнять в том числе уже распознанные части предложения [4], что говорит о использовании комплексного подхода к распознанию речи.

Оба этих приложения используются для конвертации в текст записи голоса на естественных языках и имеют широкий круг использования, то есть не имеет ярко выраженную специфику речи. В рамках этой статьи будет рассмотрено применение технологии распознавания голоса на примере данных двух приложений. Для тестирования был подготовлен аудио файл на основании судебного заседания, представленного в открытой сети Интернет – заседание арбитражного суда по делу А12-6556/2020 [6].

По результатам работы приложений были получены следующие результаты (таблица 2).

Таблица 2. Результаты распознавания голоса. Vosk-Browser Demo и RealSpeaker Transcriber

Ручная стенография	**Результат работы Vosk-Browser Demo**	**Результат работы RealSpeaker Transcriber**
Да добрый день. Здравствуйте. Да слышно видно. Да кто-то мы не сумели подключиться сначала 4 часа, так судебное заседание, по делу а12, 6556/2020 объявляется открытым, 12 Арбитражный апелляционный суд рассматривает апелляционную жалобу публичного акционерного общества, «Волгоград Энергосбыт» на решение Арбитражного суда Волгоградской области от 22 июня 2020 года. судебное заседание проводится с использованием системы видеоконференцсвязи обеспечивает, который Арбитражный суд Волгоградской области нашего коллегу просим представиться. Кто из судей обеспечивает видеоконференцсвязь, Кто ведёт в судебном заседании протокол судебного заседания в арбитражном суде Волгоградской области и проверить полномочия про семья лица, обеспечивших явку в судебное заседание. Организация видеоконференцсвязи осуществляет судья арбитражного суда Волгоградской области Репникова Виктория Вячеславовна протокол судебного заседания ведёт помощник судьи Лабутова Надежда Фёдоровна судебные заседания явился представитель публичного акционерного общества, Волгоград Энергосбыт Дундуков Никита Андреевич. Личность подтверждаются полномочия подтверждаются доверенностью номер 220 выдана 27 декабря 2019 года сроком действия до 1 января 2021 года, также представлен диплом, подтверждающий наличие высшего юридического образования. иные представители не явились. Все, спасибо вам большое не задерживаем и со стороной продолжаем судебное заседание. Прошу встать, для вас объявляется заседание суда. Дело слушается в составе, председательствующего судьи Борисова члена судебной коллегии судей в составе дубровиной и цуцковой. Секретарь судебного заседания - Кравцова	сошла добрый день да слышно вина что там и не сумели подключиться сначала четыре часа судебное заседание по делу о двенадцать шесть тысяч пятьсот пятьдесят шесть дробь две тысячи двадцать объявляется открытым двенадцатой арбитражный апелляционный суд рассматривает апелляционной жалобы публичного акционерного общества волгоград энерго сбыт на решение арбитражного суда волгоградской области от двадцать второго июня две тысячи двадцатого года на заседании проводится с использованием системы видеоконференцсвязи обеспечивая которую арбитражный суд волгоградской области нашего коллегу на просим представиться кто есть судьи обеспечивает видеоконференцсвязи кто ведёт судебного протокол судебного заседания в арбитражном суде волгоградской области и проверить полномочия просим я лица обеспечивающих явку судебное заседание организацию видеоконференцсвязи осуществляет судья арбитражного суда городского области реплика викторию вячеславовна протокол судебного заседания ведёт помощник судьи во бутово надежда ещё давно судебное заседание явился представитель публичного общества в обратном ноутбуков никита андреевич точность подтверждается паспортом полномочия подтверждаются доверенностью номер двести двадцать выдана двадцать седьмого декабря две тысячи девятнадцатого года сроком действия до первого января две тысячи двадцать первого года также представлен диплом подтверждающие наличие высшего юридического образования длины представитель тело спасибо вам большое не задерживаем и со стороны продолжаем судебное	Да добрый день. Здравствуйте Да слышно видно. Да кто-то мы не сумели подключиться сначала 4, так судебное заседание, по делу а12, 6550 6/2020 объявляется открытым, 12 Арбитражный апелляционный суд рассматривает апелляционную жалобу публичного акционерного общества, Волгоград Энергосбыт на решение Арбитражного суда Волгоградской области от 22 июня 2020 судебное заседание проводится с использованием системы видеоконференцсвязи обеспечивает, который Арбитражный суд Волгоградской области нашего коллегу просим представиться. Кто из судей обеспечивает видео связь Кто ведёт в судебном заседании протокол судебного заседания в арбитражном суде Волгоградской области и проверить полномочия про семья лица, обеспечивших явку в судебное заседание. Организация видеоконференцсвязи осуществляет судья арбитражного суда Волгоградской области Репникова Виктория Вячеславовна протокол судебного заседания ведёт помощник судьи Лабутены Надежда Фёдоровна судебные заседания явился представитель публичного акционерного общества, Волгоград Энергосбыт под Быков Никита Андреевич подтверждаются полномочия подтверждаются доверенностью номер 220 выдана 27 декабря 2019 года сроком действия до 1 января 2021 года представлены диплом, подтверждающий наличие высшего юридического образования иными представителями явились тела. Спасибо вам большое не задерживаем и со стороной продолжаем судебное заседание. Прошу встать для вас дела. Слушай это составе, председательствующего судьи Борисова члена судебной коллегии судей в составе дубровиной цуцкова. Секретарь судебного заседания. Кравцова составу суда сих Доверяете процессуальное право, предусмотренное статьей 41

Ручная стенография	Результат работы Vosk-Browser Demo	Результат работы RealSpeaker Transcriber
составу суда секретарю Доверяете? Да процессуальное право, предусмотренное статьей 41 Арбитражного процессуального кодекса Российской Федерации вам известны? Да заявления ходатайства с вашей стороны какие-либо будут нет считаете возможным приступить к рассмотрению апелляционной жалобы в отсутствие, не явившихся представителей ответчика? Да Суд, совещаясь на месте, надлежащим всех лиц, участвующих в деле, определил приступить к рассмотрению апелляционной жалобы. В составе, который Вам был озвучен в отсутствие представителя ответчика отзыв на апелляционную жалобу, которую вы направили в адрес апелляционного суда не поступал, поэтому никаких документов доказательств суд апелляционной инстанции не располагает. Значит, тогда слушаем доводы вашей апелляционной жалобы.	заседание прошу стать для вас объявляет суда дело слушается составе председательствующего судьи борисовой члена судебной коллегии судей составе дубровина секретарь судебного заседания кравцова состава суда секретарю доверяете процессуальные права предусмотренные статьёй сорок первый арбитражного процессуального кодекса российской федерации вам известны заявление ходатайство с вашей стороны какие-либо будут не считаете возможным приступить к рассмотрению апелляционной жалобы в отсутствия имеющуюся представитель ответчика совещаются учитывая сведения надлежащим отречение всех лиц участвующих в деле определил приступить к рассмотрению апелляционной жалобы составе которой вам был озвучен в отсутствия представителя ответчика отзыв на апелляционной жалобы которую вы направили в адрес апелляционного суда не поступал этому никаких новых документов доказательств суд апелляционной инстанции суд не располагает тогда слушаем доводы вашей апелляционной жалобы	Арбитражного процессуального кодекса Российской Федерации вам известны заявления ходатайства с вашей стороны какие-либо будут нет считаете возможным приступить к рассмотрению апелляционной жалобы в отсутствие, не явившихся представители ответчика надлежащим всех лиц, участвующих в деле определил приступить к рассмотрению апелляционной жалобы. В составе, который Вам был озвучен в отсутствие представителя ответчика отзыв на апелляционную жалобу, которую вы направили в адрес апелляционного суда не поступал, поэтому никаких документов доказательств суд апелляционной инстанции не располагает. Значит, тогда слушаем доводы ваши апелляционной жалобы.

Примечание: при сравнении использовалась ограниченная версия RealSpeaker Transcriber, что ограничило фрагмент аудиодорожки.

Технологии подлежали оценке по следующим метрикам:

- Точность;
- Полнота.

Коэффициент полноты – это доля полученных релевантных документов по сравнению с их общим количеством в поисковом массиве. Коэффициент точности – это доля релевантных документов среди документов, выданных в результате.[7]

Полнота и точность результатов приложений представлена в таблице 3.

Таблица 3. Полнота и точность результатов. Vosk-Browser Demo и RealSpeaker Transcriber

Мера	RealSpeaker Transcriber	Vosk-Browser Demo
Полнота	93%	87%
Точность	91%	89%

RealSpeaker Transcriber были некорректно выделены 23 слова, из которых:

- 1 сложносоставное числовое выражение;
- 2 имени собственных;
- 5 крайне зашумлены или сказаны заметно тише на записи.

А также частично неверными (неправильная словоформа, окончание, приставка), оказались 9 слов.

В номере дела допущена только одна ошибка в связи с опущением цифры шесть. Прослеживается работа не

только над выделением слов, но и над пунктуацией, что помогает передать смысловую нагрузку. Приложение различает предложения, включая сложные, а также абзацы, заданные явными паузами в диалоге.

В то же время, орфография не соответствует предметной области (что затрудняет сравнение с оригинальным текстом), а также не различает дикторов.

Vosk-Browser Demo были некорректно выделены 32 слова, среди которых:

– 6 имен собственных;
– 4 крайне зашумлены или сказаны заметно тише на записи.

А также частично неверными (неправильная словоформа, окончание, приставка), оказались 8 слов.

При этом пунктуация и выделение предложений не производилась.

Таким образом, можно наблюдать, что задачу обработки аудио, насыщенного профессиональной терминологией, оба подхода решают на сравнимом уровне. Но так как основная задача состоит в составлении юридически верного, точного, грамматически и орфографически верного документа – протокола судебного заседания, то возможность выделения речевых оборотов, расставление пунктуационных знаков является неоспоримым преимуществом. Как было отмечено в исследовании, современный подход позволяет распознавать и выделять пунктуацию.

Заключение

В заключение необходимо отметить, что каждый из рассмотренных методов имеет индивидуальные особенности, а также ряд преимуществ и недостатков своего использования в голосовых помощниках. Было определено, что оба подхода имеют недостатки в распознавании специфической информации, такой как номера, шифры и специфичные термины, традиционно записываемые на иных языках. При этом подход на основе понимания естественного языка содержит возможность распознавания пунктуационных знаков по стилю и темпу речи. Таким образом, можно сделать вывод, что развитие подхода с использованием систем понимания естественного языка позволяет расширить спектр задач, где применяются методы распознавания речи на задачи, имеющие существенную профессиональную специфику с точки зрения оформления и терминологии, к примеру на задачи протоколирования судебных заседаний. Однако, текущие реализации технологии ASR все еще обеспечивают недостаточную точность распознания речи для автоматизации многих чувствительных к специфике процессов, что открывает возможности для реализации более специфичных систем на основе представленных методов, которые смогут заполнить данную нишу.

Список литературы

1. Хеин М.З. Современное состояние проблемы обработки, ана-лиза и синтеза речевых сигналов // Computational nanotechnology. 2018
2. Хлопенкова А.Ю., Белов Ю.С. Методы обработки естественно-го языка в виртуальных голосовых помощниках // E-Scio. 2019
3. Abougarair A.J. Design and implementation of smart voice assis-tant and recognizing academic words // International Journal of Robotics and Automation. 2022.
4. RealSpeaker Transcriber. Режим доступа: https://realspeaker.net/, (Дата обращения 01.12.2022)
5. Vosk-Browser Demo. Режим доступа: https://alphacephei.com/vosk/, (Дата обращения 01.12.2022)
6. Судебное заседание по делу А12-6556/2020. Режим доступа: https://www.youtube.com/watch?v=46oPauC9_Aw (Дата обращения 01.12.2022)
7. Зубец В. В., Ильин А. А. О коэффициенте полноты информационного поиска // Вестник российских университетов. Математика. 2004. №1. Режим доступа: https://cyberleninka.ru/article/n/o-koeffitsiente-polnoty-informatsionnogo-poiska (Дата обращения: 03.12.2022).
8. Riqiang W. Automatic Speech Recognition 101: How ASR systems work // Dialpad. Режим доступа: https://www.dialpad.com/blog/automatic-speech-recognition/ (Дата обращения: 04.12.2022)

References

1. Hein M.Z. The current state of the problem of processing, analysis and synthesis of speech signals // Computational nanotechnology. 2018
2. Khlopenkova A.Yu., Belov Yu.S. Natural language processing methods in virtual voice assistants // E-Scio. 2019
3. Abougarair A.J. Design and implementation of smart voice assist-tant and recognizing academic words // International Journal of Robotics and Automation. 2022.
4. Tsitulsky A.M., Ivannikov A.V., Rogov I.S. NLP - natural language processing // StudNet. 2020.
5. RealSpeaker Transcriber. Access mode: https://realspeaker.net/, (Accessed 01.12.2022)
6. Vosk-Browser Demo. Access mode: https://alphacephei.com/vosk/, (Accessed 01.12.2022)
7. Court session in case A12-6556/2020. Access mode: https://www.youtube.com/watch?v=46oPauC9_Aw (Accessed 12/01/2022)
8. Zubets V. V., Ilyin A. A. On the coefficient of completeness of information retrieval // Bulletin of Russian Universities. Mathematics. 2004. No. 1. Access mode: https://cyberleninka.ru/article/n/o-koeffitsiente-polnoty-informatsionnogo-poiska (Date of access: 03.12.2022).
9. Riqiang W. Automatic Speech Recognition 101: How ASR systems work // Dialpad. Access mode: https://www.dialpad.com/blog/automatic-speech-recognition/ (Date of access: 04.12.2022)