

СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ (ASR)

В.Д. Косых

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. В работе рассматриваются принципы функционирования и ключевые области применения систем автоматического распознавания речи (Automatic Speech Recognition, ASR). Анализируются основные технологические этапы преобразования аудио в текст. Так же рассматривается практическое применение данной системы в разных сферах. Освещаются перспективные направления развития.

Ключевые слова: автоматическое распознавание речи, ASR, преобразование речи в текст, транскрибация аудио, голосовые помощники, искусственный интеллект (ИИ), голосовые помощники, нейронные сети, субтитрирование, обработка языка, языковая модель.

AUTOMATIC SPEECH RECOGNITION SYSTEMS (ASR)

V.D. Kosykh

Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The paper discusses the principles of functioning and key areas of application of Automatic Speech Recognition (ASR) systems. The main technological stages of audio-to-text conversion are analyzed. The practical application of this system in various fields is also being considered. Promising areas of development are highlighted.

Keywords: automatic speech recognition, ASR, speech-to-text conversion, audio transcription, voice assistants, artificial intelligence (AI), voice assistants, neural networks, subtitling, language processing, language model.

Наши предшественники не могли и думать о том, что в будущем появится голосовое управление и расшифровка речи с аудио. Сегодня широко

используется данная функция как обыденность, мы произносим: «Привет, Алиса, расскажи о погоде» - и устройство выполняет запрос. За это отвечает технология – автоматическое распознавание речи (Automatic Speech Recognition, ASR). Это область искусственного интеллекта, которая есть на всех современных устройствах.

Процесс превращения звука в текст – многоуровневый алгоритм. Рассмотрим его подробнее:

- Акустический анализ

Здесь происходит оцифровка аналогового звукового сигнала. Отделяются фоновые шумы (гул, треск, посторонние звуки). Далее запись разбивается на отрезки (семплы). Для каждого отрезка выделяются свои акустические характеристики;

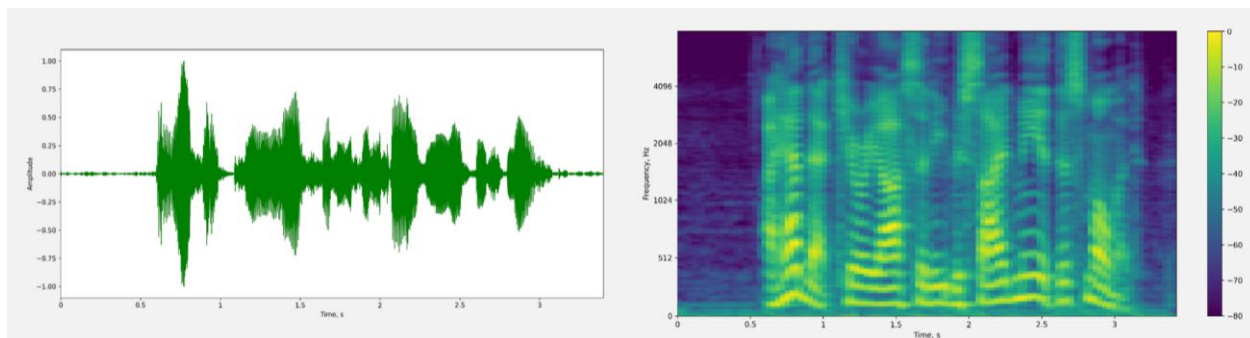


Рисунок 1 – Необработанная аудиозапись (слева) и мел-спектограмма (справа)

- Акустическое моделирование

На данном этапе сравниваются характеристики с фонемами – минимальные единицы звука. Современные модели делают это с высокой точностью;

- Языковое моделирование

Система так же анализирует контекст. Языковая модель предполагает последовательность слов, которая по смыслу может понять какое слово следует за другим;

- Декодирование

Заключительный этап, на котором подбирается наиболее оптимальный вариант между звуками и словами, результатом этого этапа является готовый текст.



Рисунок 2 – Алгоритм системы ASR

Система распознавания применяется в разных сферах:

- Субтитрование и транскрибация

Самое массовое применение, увидеть вы можете при просмотре видео на YouTube, новостная лента в телевизоре, а также к любому видео в сети Интернет можно применить транскрибацию. Очень удобная система, которая помогает расшифровать невнятную речь или экономить время просмотра, сделав импорт текста;

- Голосовые помощники и умный дом

Алиса и Siri являются результатом развития ASR. Они не только выполняют команды, но и обучены быть собеседником, навигатором, что создает голосовое взаимодействие между системой и человеком. Так же эти помощники обеспечивают доступность людям с ограниченными возможностями;

- Клиентский сервис и колл-центры

Роботы, которые отвечают на звонки и ведут с Вами диалог, это тоже ASR, которая распознает речь и направляет в нужный отдел. Так же они используются для анализа записи разговоров, для того чтобы оценить качество работы и сервиса;

- Профессиональная сфера

Подходит для тех, кто использует платформы Zoom, Microsoft Teams, где предлагается функция live-транскрибация. То есть, весь деловой звонок превращается в текстовый документ.

Преимущества технологии ASR:

- Эффективность

Скорость преобразования преобладает над распознаванием человеком;

- Масштабируемость

Данная система одновременно может обработать тысячи часов записи;

- Автоматизация;

- Доступность.

Недостатки:

- Точность

Диалекты, незнакомые слова, а также шум, могут дезориентировать алгоритм;

- Контекст и специализация

Распознавание слов, которые относятся к определенной специальности, требует переобучение модели;

- Конфиденциальность.

Современные системы не стоят на месте, а постоянно совершенствуют свой алгоритм. Следующим шагом будет распознавание эмоций в голосе, то есть система будет определять злобу, радость, грусть и другие эмоции. В врачебной практике думаю будет помогать определять какие-то психологические проблемы или болезни.

Система в фоновом режиме может обучаться и набираться лексики в определенной сфере (инженеры, юристы), чтобы правильно интерпретировать речь.

Так же нельзя опустить возможность, что можно воссоздать копию голоса, которая будет полезна в озвучке, а также людям с ограниченными возможностями, они смогут говорить с помощью синтезатора своим голосом.

Будущее ASR – переход от транскрибации к аудио-интерфейсу, который будет понимать пользователя, как человека или даже лучше. Данная система станет незаменимым посредником между человеком и цифровой средой.

Список литературы

1. Иванов, А. С., Петровский, А. В. Обработка речевых сигналов в системах автоматического распознавания речи: современные подходы / А. С. Иванов, А. В. Петровский // Вестник Воронежского государственного лесотехнического университета. – 2021. – Т. 15, № 2. – С. 45-58.
2. Козлов, М. Ю., Иванова, Т. Н. Глубокие нейронные сети и их применение в речевых технологиях / М. Ю. Козлов, Т. Н. Иванова // Информационные технологии и вычислительные системы. – 2022. – № 3. – С. 15-24.
3. Морозова, Е. П., Григорьев, К. А. Использование сверточных нейронных сетей для повышения точности распознавания речи / Е. П. Морозова, К. А. Григорьев // Журнал прикладной информатики. – 2023. – Т. 29, № 2. – С. 88-95.
4. Николаев, П. С., Орлова, М. В. Современные подходы к интеграции языковых моделей в системы распознавания речи / П. С. Николаев, М. В. Орлова // Информационные системы и технологии. – 2022. – Т. 14, № 3. – С. 55-64.
5. Рэдфорд, А., Ким, Дж. В., Сюй, Т., Брокман, Г., Макливи, К., Сутскевер, И. Надёжное распознавание речи с помощью масштабного слабого обучения // Международная конференция по машинному обучению (ICML). – 2023.
6. Соколова, Е. А., Кузнецов, В. П. Применение машинного обучения для анализа лингвистических данных в автоматическом распознавании речи / Е. А. Соколова, В. П. Кузнецов // Известия Воронежского государственного лесотехнического университета. – 2019. – Т. 12, № 1. – С. 112-125.
7. Юрафский, Д., Мартин, Дж. Х. Обработка речи и языка (3-е изд.) / Д. Юрафский, Дж. Х. Мартин. – Пирсон, 2023. – 1200 с.

References

1. Ivanov, A. S., Petrovsky, A.V. Speech signal processing in automatic speech recognition systems: modern approaches / A. S. Ivanov, A.V. Petrovsky // Bulletin of the Voronezh State Forestry University. – 2021. – Vol. 15, No. 2. – pp. 45-58.
2. Kozlov, M. Yu., Ivanova, T. N. Deep neural networks and their application in speech technologies / M. Yu. Kozlov, T. N. Ivanova // Information technologies and computing systems. - 2022. – No. 3. – pp. 15-24.
3. Morozova, E. P., Grigoriev, K. A. Using convolutional neural networks to improve speech recognition accuracy / E. P. Morozova, K. A. Grigoriev // Journal of Applied Informatics. 2023. Vol. 29, No. 2. pp. 88-95.
4. Nikolaev, P. S., Orlova, M. V. Modern approaches to the integration of language models into speech recognition systems / P. S. Nikolaev, M. V. Orlova // Information Systems and Technologies. – 2022. – Vol. 14, No. 3. – pp. 55-64.
5. Radford, A., Kim, J. V., Xu, T., Brockman, G., McLeavy, K., Sutskever, I. Reliable speech recognition using large-scale weak learning // International Conference on Machine Learning (ICML). – 2023.
1. Sokolova, E. A., Kuznetsov, V. P. Application of machine learning for linguistic data analysis in automatic speech recognition / E. A. Sokolova, V. P. Kuznetsov // Proceedings of the Voronezh State Forestry University. – 2019. – Vol. 12, No. 1. – pp. 112-125.
2. Yurafsky, D., Martin, J. H. Speech and Language Processing (3rd ed.) / D. Yurafsky, J. H. Martin. – Pearson, 2023. – 1200 p.