УДК 004.8

Гранкин А. М.
Студент 3 курса
факультет «Информационные системы и технологии»
ФГБОУ ВО ПГУТИ
Россия, г. Самара
Пальмов С.В.
к.т.н.
доцент кафедры «Информационные системы и технологии»
ФГБОУ ВО ПГУТИ
Россия, г. Самара
Grankin A. M.
3rd year student
Faculty of "Information systems and technologies"
Volga State University of Telecommunications and Informatics,
Russia, Samara
Palmov S.V.
Ph.D. of Engineering Sciences
associate professor of the department
"Information systems and technologies"
Volga State University of Telecommunications and Informatics,
Russia, Samara

РАСПОЗНАВАНИЕ И СИНТЕЗ РЕЧИ

Аннотация: С каждым годом, человечество всё ближе подходит к созданию компьютера, который сможет общаться с человеком на разных языках. Можно предвидеть приложения-переводчики, которые позволят переводить с одного языка на другой в режиме реального времени. В данной статье рассматривается распознавание и синтез речи, основные алгоритмы распознавания звука и значение распознавания речи в жизни человека.

Ключевые слова: искусственный интеллект, распознавание речи, синтез речи

RECOGNITION AND SYNTHESIS OF SPEECH

Annotation: Every year, mankind comes closer to creating a computer that can communicate with a person in different languages. You can anticipate translation applications that will translate from one language to another in real time. This article deals with recognition and synthesis of speech, the main algorithms of sound recognition and the meaning of speech recognition in human life.

Keywords: artificial intelligence, speech recognition, speech synthesis

Распознавание речи позволяет создавать более простые интерфейсы для пользователей. Синтез речи обеспечивает выход из ситуации, когда человек не может читать, например, при вождении автомобиля, тем самым облегчая деятельность пользователя.

Речевые интерфейсы обычно добавляют в графические интерфейсы пользователя, например, в качестве дополнительной функции для людей с нарушением зрения. Но они также используются в сочетании с другими новыми способами взаимодействия, к примеру, жестикуляция в средах VR, чтобы создать естественные условия.

Распознавание речи – это процесс преобразования речевого сигнала в цифровую информацию. Человек в обычном разговоре произносит от 10 до 15 звуков в секунду. И именно из-за этого, попытки создания компьютерных систем распознавания оказались трудными. Но, несмотря на это, существует множество систем, которые достигли определённого успеха в некоторых аспектах распознавания речи. Несмотря на ограниченность возможностей и, как правило, отсутствие «естественного» языка и жаргона, эти системы теперь являются неотъемлемой частью человеческой жизни.

Синтез текста в речь – это превращение строк текста в речь, который воспроизводится через динамики. Конечным результатом является то, что

компьютер разговаривает с пользователем. Распознавание речи – это возможность компьютера принимать речь человека и интерпретировать её, что делает возможным пользователю управлять компьютером голосом, а не использовать мышь и клавиатуру, например, просто диктовать текст документа.

Распознавание гласных звуков достигается путём идентификации первых двух или трёх формантов. Форманты – это области усиления энергии в спектре звука. Но при некоторых условиях гласные могут быть распознаны из самых высоких формант, когда две низкие отсутствуют. Формантная структура каждого человека отличается друг от друга, у каждого человека свой тембр, скорость речи и высота голоса. Например, речь ребёнка по звуковому ряду значительно отличается от речи взрослого мужчины, но всё же, мы понимаем, что говорит нам ребёнок.

Распознавание согласных звуков осуществляется идентификацией всплесков высокочастотного шума в паре с гласным звуком, например, «т». А более низкая частота всплеска в паре с гласным звуком будет распознаваться как буквы «п» и «к». В обычном разговоре, мы говорим с использованием разных частотных переходов, от 1800 Гц до 300 Гц, и именно эти переходы позволяют распознать какую букву мы произносим. Например, буква «ш» имеет частоту звука в диапазоне 2000-3000 Гц, а буква «с» - более 4000 Гц.

На качество распознавания также влияет зашумленность, невнятная речь, пропуск букв и «проглатывание» окончаний. Такие факторы снижают значение данного показателя от 1800 Гц (понятная речь) до наиболее узкополосной отфильтрованной речи – ниже 800 Гц [2]. Фильтрация шума в подобной ситуации не сто процентов не помогает - он всё равно останется, и не позволит точно определить, что сказал человек. Однако лингвистические и семантические сигналы всё же позволяют разобрать предложения.

Большинство, если не все современные синтезаторы речи, используют библиотеки речевых звуков, которые затем объединяются вместе для формирования слов. Это требует хранения огромных баз данных различных звуков и переходов. Синтезатор, основанный на физической модели голосовой трактовки, самым лучшим образом обеспечит наиболее гибкую систему синтеза речи.

Распознавание речи в настоящее время считается необходимым элементом для обеспечения взаимодействия с современными технологиями, и особенно это важно для слепых людей и людей со слабым зрением. Это обеспечивает или расширяет доступ к печатной или электронной информации, ежедневной или социальной деятельности, а также к частным или общественным объектам.

Одна из первых компаний, которые начало изучение и внедрение речевых технологий – Microsoft. Она занимается развитием своих систем уже много лет, и у них есть своя область. Речевая технология Speech может быть добавлена к синтезу текста в речь и в распознавание речи. И уже в 1995 году компания представила данную технологию как одну из частей служб WindowsOpenServicesArchitecture. Это предназначалось для упрощения и удобной работы с продуктом компании. И уже сейчас группа исследователей компании разработала систему распознавания речи, которая совершает равное или меньшее количество ошибок, чем специалист по распознаванию речи. По данным исследований, частота ошибочных слов составляет 5,9%, что примерно равно количеству ошибок, совершаемых при написании под диктовку людьми того же фрагмента текста. Данный прорыв в этой сфере не означает, что компьютер научился распознавать человеческую речь идеально, а лишь приблизил его к такому же количеству совершаемых ошибок при помощи новейших технологий с применением нейронных моделей языка. Для обучения системы были использованы тренировочные

наборы больших данных, а также собственную систему для обучения, с открытым кодом [3].

Несмотря на достижения в сфере распознавания речи, учёные сейчас работают над точностью распознавания речи в реальных условиях, с фоновым шумом, разнообразием голосов, с другими разговорами на фоне и акцентами. В дальнейшем исследования также будут заключаться не в простом распознавании речи, а в понимании речи компьютером.

Список использованных источников

1. Speech Recognition and Speech Synthesis [Электронный ресурс] [https://xsrv.mm.cs.sunysb.edu/hci/speech/speech.html] / Speech Recognition and Speech Synthesis. -2018. – Режим доступа: https://xsrv.mm.cs.sunysb.edu, свободный. – SUNYSB.
2. Speech Recognition and Synthesis [https://ccrma.stanford.edu/CCRMA/Courses/152/speech_recognition.html] / Speech Recognition and Synthesis. -2018. – Режим доступа: https://ccrma.stanford.edu, свободный. – CCRMA.
3. Историческое достижение — исследователи Microsoft достигли уровня человеческих возможностей при автоматическом распознавании речи [https://news.microsoft.com/ru-ru/microsoft-dostigli-urovnya-chelovecheskih-vozmozhnostej-pri-avtomaticheskom-raspoznavanii-rechi/] / Историческое достижение — исследователи Microsoft достигли уровня человеческих возможностей при автоматическом распознавании речи. - 2016. – Режим доступа: https://www.microsoft.com/ru-ru/, свободный. – Microsoft.