«Сейфуллин оқулары – 16: Жаңа формациядағы жастар ғылыми – Қазақстанның болашағы» атты халықаралық ғылыми-теориялық конференциясының материалдары = Материалы Международной научно-теоретической конференции «Сейфуллинские чтения – 16: Молодежная наука новой формации – будущее Казахстана. - 2020. - Т.II. - С. 343-346

ПРЕОБРАЗОВАНИЕ РЕЧИ В ТЕКСТ

Ерхасанова А. К.

С появлением компьютеров перед человеком встал целый ряд новых проблем, связанных с передачей и хранением информации. Ввод данных всегда требовал значительных затрат времени и сил, а стремление свести эти затраты к минимуму заставляет постоянно работать над способами перевода знаковой системы, которой пользуется человек, на тот язык, который понятен машине. Наиболее естественным для человека являются речевые интерфейсы взаимодействия. В их реализации используются технологии синтеза и распознавания речи. Внедрению данных технологий препятствует недостаточное качество и эффективность реализующих их алгоритмов. Задача распознавания речи является актуальной проблемой на сегодняшний день. Т.к. большинство методов требуют больших вычислительных ресурсов. Невозможность применения многих алгоритмов сегодня заставляет искать более эффективные методы [1]. Существующие методы распознавания речи не отвечают всем заявленным требованиям. Это обстоятельство определяет актуальность исследований в этом направлении.

В ходе развития компьютерных систем становится очевидным, что эффективность использования этих систем может быть повышена в случае использования естественного и распространенного для человека инструмента общения – речи. В частности это позволит ускорить ввод информации и управление компьютерными, и особенно, мобильными системами. Многие разработчики уже добились некоторых успехов в области распознавания речи, но массового использования таких технологий на сегодняшний день пока не наблюдается, что причиной чего могут являться зависимость от диктора, недостаточная точность распознавания для непрерывной речи и высокая чувствительностью к наличию различного вида помех.

Как сказал наш Елбасы: «Цифровые преобразования меняют облик привычной нам экономики. Глобальный экономический ландшафт сегодня определяется высокотехнологичными компаниями. К примеру, ТОП-5 компаний, по рыночной стоимости, представляют технологические гиганты (Apple, Google, Microsoft, Amazon и Tencent), которые вытесняют традиционные промышленные и инвестиционные корпорации. Их совокупная стоимость составляет порядка 3,5 трлн долларов, что сопоставимо с ВВП некоторых развитых государств» [2].

Специалистами Института информационных и вычислительных технологий Комитета науки МОН РК было создано мобильное приложение KazVoice для распознавания речи, работающее с казахским языком.

Программа позволяет обмениваться информацией с электронными гаджетами звуковыми командами на казахском языке. Разработка представляет собой экспериментальную технологию, позволяющую компьютеру понимать устные команды благодаря технологиям распознавания речи и считывания мимики, движений мышц лица при произнесении слов. Мобильное устройство может получать информацию не только с микрофона, но и с видеокамеры, в буквальном смысле читая слова по губам [3].

В русском сегменте лидирует компания Яндекс с новой разработкой Yandex.SpeechKit которая представляет собой мультиплатформенную библиотеку, предоставляющая разработчикам мобильных приложений доступ к технологии распознавания речи Яндекс [4]. Неплохих успехов добилась корпорация Google Inc., предлагающая речевой ввод при осуществлении поиска в сети Интернет. Системы Google Inc. используют широкую базу образцов речевых шаблонов сотен и даже тысяч дикторов, что позволяет им добиваться уверенного распознавания многих слов неизвестных дикторов [5].

В работе ученых Хисаёши К., Донсук Ю. «Классификация систем распознавания речи» [6] идет речь о таких понятиях, как: *размера словаря* (частота ошибок системы распознавания напрямую зависит от количества слов в словаре системы распознавания); *дикторозависимость* (т.е. настраивается под индивидуальные характеристики речи, в то время как дикторонезависимая система предназначена для работы с любым диктором и не учитывает индивидуальных особенностей произношения); *тип речи* (для ввода используются либо отдельные слова и словосочетания, либо требуется найти слова маркеры в слитной речи [7]); *тип лексической структурной единицы* (при анализе речи, в качестве базовой единицы анализа могут быть выбраны отдельные слова и словосочетания, слоги, а также такие элементы как фонемы, аллофоны, дифоны, трифоны); *механизм работы.*

В работе Титова Ю.Н. «Современные технологии распознавания речи» [7] описывается многообразие существующих систем распознавания речи, которое можно условно разделить на следующие группы:

1. Программные ядра для аппаратных реализаций систем распознавания речи;
2. Наборы библиотек, утилит для разработки приложений, использующих речевое распознавание;
3. Независимые пользовательские приложения, осуществляющие речевое управление и/или преобразование речи в текст;
4. Специализированные приложения, использующие распознавание речи;
5. Устройства, выполняющие распознавание на аппаратном уровне;
6. Теоретические исследования и разработки.

Рассмотрим каждую из этих групп подробнее.

1. *Программные ядра для аппаратных реализаций*. В основе любой речевой технологии лежит так называемый «engine» или ядро программы – набор данных и правил, по которым осуществляется обработка данных. В зависимости от назначения этого ядра различают TTS и ASR engine. TTS (Text-to-Speech) engine предоставляет возможность синтеза речи по тексту, а ASR (Automatic Speech Recognition) engine – для распознавания речи [8]. Существует несколько крупных производителей, занимающихся созданием ASR ядер и среди них такие компании, как SPIRIT, Advanced Recognition Technologies, IBM.

Корпорация IBM уже более 30 лет занимается вопросами автоматического распознавания речи и достигла в этой области больших успехов. Так компания ProVox Technologies на основе программного ядра ViaVoice от IBM создала систему для диктовки отчетов врачей-радиологов VoxReports [9]. Технология распознавания речи все больше применяется в средствах подвижной связи. Так компания Advanced Recognition Technologies создала систему smARTspeak NG, встраиваемую в мобильные телефоны. Сейчас система smARTspeak NG применяется в бесклавиатурных телефонах от Siemens [10], телефонах Panasonic стандарта TDMA в США и других.

2. *Наборы библиотек для разработки приложений.* С развитием речевых технологий и все большим внедрением мобильных устройств возникла идея применения речевого управления при создании сетевых приложений. Для этого было необходимо разработать стандарт для интеграции речевых технологий. Один из открытых стандартов на основе XML языка — VoiceXML (Voice eXtensib1e Markup Language) [11].

3. *Независимые пользовательские приложения.* В настоящее время рынок программных распознавателей речи представлен множеством приложений. Рассмотрим наиболее известные из них.Dragon NaturallySpeaking Preferred фирмы Dragon Systems [13] – единственная программа, приблизившаяся к тому, чтобы соответствовать заявленным характеристикам. В целом она очень близко подходит к достижению заявленной безошибочности распознавания - 95%.

4. *Специализированные приложения.* Распознавание речи может применяться не только для ввода текста или подачи команд, но и для более специфичных целей. Например, российская компания «Центр Речевых Технологий» разрабатывает и производит программные продукты, технологии и образцы техники для подразделений полиции, служб экстренной помощи, центров обработки вызовов и для других пользователей, в деятельности которых особое значение придается регистрации и обработке речевой информации [14].

5. *Устройства, выполняющие распознавание на аппаратном уровне.* Для использования функций речевого распознавания в различных устройствах, роботах, игрушках, разрабатываются аппаратные методы решения данной проблемы. Так американская компания Sensory Inc. разработала интегральную схему Voice Direct 364, осуществляющую

дикторозависимое распознавание небольшого числа команд (около 60) после предварительного обучения [15].

6. *Теоретические исследования и разработки*. Из всего разнообразия научных разработок подробно рассмотрим работы отечественных исследовательских групп, а именно разрабатываемую в Институте проблем информатики и управления (ИПИУ) систему синтеза казахской речи по тексту [2]. Данная система является электронным казахскоязычным диктором. В систему синтеза речи загружается произвольный текст на казахском языке. После завершения процесса синтеза можно услышать, как компьютер читает данный текст естественным человеческим голосом, соблюдая все знаки препинания, правильно расставляя ударения, делая паузы в нужных местах и акцентируя интонацией значимые фрагменты текста.

Технологии распознавания речи считаются одними из наиболее перспективных в мире. Однако если сравнить показатели современных систем распознавания с показателями систем времен начала зарождения этой области науки, то можно сказать, что за прошедшие десятки лет исследователи недалеко продвинулись. Это заставляет некоторых специалистов сомневаться относительно возможности реализации речевого интерфейса в ближайшем будущем с использованием существующих подходов.

Проблема, о которой говорит большинство исследователей, называется «неустойчивость систем распознавания, к внешним условиям» и заключается в том, что методы реализации механизма распознавания речи не совсем согласуются с тем, как реальные люди распознают и понимают речь друг друга. Поэтому, думаю, что стоит развивать эти системы, заостряя внимание на том, как можно улучшить данное восприятие, какие технологии стоит применить. Это могут быть различные высокочувствительные микрофоны, технологии, связанные с постоянно обучающимся ИИ, развитие уже существующих алгоритмов, увеличение базы шаблонов речи с диалектами, акцентами, с учетом определенных особенностей произношения и т.д.

Я считаю, что использование этих систем намного расширится, если станет возможным управление машиной обычным голосом в реальном времени, а также ввод и вывод информации в виде обычной человеческой речи. Тем самым это могло бы позволить людей с ограниченными возможностями лучше взаимодействовать с остальными людьми нашего общества.

Список использованной литературы

1. Асхатова Г. (2019). Главные цитаты Назарбаева о вызовах современности. https://365info.kz/2019/05/glavnye-tsitaty-nazarbaeva-o-vyzovah-sovremennosti
2. Институт Информационных и Вычислительных технологий. Интеллектуальные облачные технологии (2019). https://iict.kz/cloud-technologies/

3. Косенов А. (2014). Статья «Мобильное приложение для распознавания речи научили казахскому языку». https://tengrinews.kz/gadgets/mobilnoe-prilojenie-raspoznavaniya-rechi-nauchili-kazahskomu-264272/
4. Яндекс. SpeechKit API. http://api.yandex.ru/speechkit/.
5. Google. How Google uses pattern recognition to make sense of images. https://policies.google.com/technologies/pattern-recognition?hl=en
6. Титов Ю.Н. (2006). Современные технологии распознавания.
7. речи SantoshK.Gaikwad,BhartiW.Gawali, PravinYannawar (2010). “A Review on Speech Recognition Technique”,International Journal of Computer Applications (0975 – 8887).
8. Хабрахабр. Интерактивное голосовое редактирование текста с помощью новых речевых технологий от Яндекса. https://habr.com/ru/company/yandex/blog/243813/
9. Компания SPIRIT DSP. https://www.spiritdsp.com/company/
10. CMU Sphinx Open Source инструментарий для распознавания речи оценки. http://cmusphinx.sourceforge.net/ .
11. Wikipedia. Opera. https://ru.wikipedia.org/wiki/Opera
12. Wikipedia. VoiceXML. https://ru.wikipedia.org/wiki/VoiceXML
13. MS Technology. About. http://www.mstechnology.ru
14. Группа компаний «Центр речевых технологий». https://www.speechpro.ru/
15. Wikipedia. Sensory, Inc. https://en.wikipedia.org/wiki/Sensory,_Inc.