

**ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ПРОГРАММ С ФУНКЦИЯМИ
РАСПОЗНАВАНИЯ ГОЛОСА В РАЗЛИЧНЫХ ОБЛАСТЯХ
ЖИЗНЕДЕЯТЕЛЬНОСТИ ЧЕЛОВЕКА**

Базаров Кирилл Александрович

студент 2 курса института математики,
физики и информационных технологий

Тамбовский государственный университет им. Г.В.Державина

г. Тамбов, Россия

e-mail: bazarov.kirill@mail.ru

Аннотация: в данной работе перечислены преимущества и недостатки голосовых технологий, которые используются в наиболее перспективных и быстроразвивающихся областях жизнедеятельности человека. Также голосовые программы имеют большие перспективы в будущем в связи с развитием технологий в области создания голосового интерфейса визуальных объектов.

Ключевые слова: фоторобот, голосовые интерфейсы, нейронные сети, программы с применением систем распознавания речи.

Общеизвестно, что возможности применения голосовых команд все больше расширяются и становятся частью повседневного комфорта. Голосовое управление — это взаимосвязь вербального общения с устройствами. Считается, что это следующая стадия управления техникой после сенсорного ввода информации.

Голосовое управление происходит по следующим этапам — сегментация, распознавание речи, отклик на требуемую задачу. Переработка речи начинается с определения уровня помех и искажений, полученного речевого сигнала. В сигнале имеются участки, содержащие речь. В них происходит оценка информации о формах слов. Этот этап называется сегментацией.

Далее параметры речи поступают в декодер, где сопоставляются входные речевые потоки информации с потоками, хранящимися в акустических и языковых моделях устройства. Там же определяется некоторая наиболее вероятная последовательность слов, являющаяся в конечном итоге результирующим. Это этап распознавания речи [1,2].

Последним этапом в голосовом управлении является отклик на требуемую задачу. Процессор, встроенный в используемое устройство, обрабатывает полученную команду. Если она совпадает с той, что хранится в его встроенной базе данных, то устройство следует по необходимым условиям. При отсутствии команд, оборудование оповестит о невозможности выполнения.

Опишем некоторые уже существующие программы с функциями распознавания голоса.

Существуют программы с применением систем распознавания речи, которые независимы от диктора, то есть программы, которые не требуют предварительного обучения и могут проводить анализ речи любого диктора. Такие программы уже весьма успешно используются на практике. Например, «Dragon NaturallySpeaking» от компании «Nuance» уже предоставляет огромный функционал для управления, таких как: использование браузеров, отправление электронной почты, создание новых файлов. Точность

распознавания голоса и обработки команд крайне высока[3]. Интересен также факт, что «ViaVoice» от компании «IBM» позволила компании ProVox Technologies создать систему для создания отчетов врачей-радиологов. Она позволяет распознавать специфичные термины и достигает точности распознавания 95-98%.

В 2016 году Google I/O был представлен умный персональный помощник «Google Assistant». В 2017 году российская компания представила голосового помощника «Алиса». Голосовые помощники, такие как «Google Assistant» и «Алиса», уже тесно вошли в нашу жизнь и стали частью повседневности. Они научились вызывать такси, ставить будильник, прокладывать маршрут, общаться с пользователем на отвлеченные темы и многое другое.

С помощью голосового управления можно распознать состояние человека. Например, если дрожит голос, скорее всего, он волнуется. Это вводится в специальную базу данных, где выводится своеобразный «фоторобот».

Технологии голосового управления уже широко используются в медицине, например, для коррекционной развивающей работы с людьми, у которых имеются нарушения произношения и/или слуха.

Речевые системы имеют широкое применение в банковских организациях. Они используют технологии голосового самообслуживания и обеспечивают быстрый доступ абонентов к необходимой для запроса информации.

Голосовые системы самообслуживания помогают операторам телекоммуникационной связи улучшить процесс обслуживания абонентов, поскольку ускорилась обработка звонков, уменьшилось количество ошибочных соединений, усовершенствовалась защита персональных данных клиентов.

Многие современные исследователи считают, что разработка интерфейса голосового управления является очень востребованной, но одновременно и трудоемкой задачей. Реализация занимает много времени и

ресурсов. Существует огромное количество особенностей создания этой среды, например, диалектика, иностранные языки, дефекты речи и так далее[4]. Известно, что прогресс распространяется на все сферы жизнедеятельности общества. Со времен, когда художники стали не только прислугой аристократов для написания автопортретов, но и верными помощниками криминалистов в расследовании обширного спектра дел, связанных с опознанием преступника, прошло уже несколько столетий. Мастера художественного искусства, прислушиваясь к словам жертвы, могли очень точно нарисовать лицо предполагаемого мошенника или грабителя.

В конце 20-го века на смену им пришли технические средства визуального предоставления информации – криминалистика сделала огромных технологический шаг. Однако тогда они не могли научить ЭВМ слушать и понимать услышанное, что и являлось существенным недостатком данного метода по сравнению с человеком. Сейчас такая возможность есть, и можно использовать ее на благо общества. Когда компьютер сам научится составлять фоторобот по голосовым командам, тексту или жестам, необходимость в затрате времени и наличии специальных технических навыков отпадет. Данная технология может пригодится не только в сфере правоохранительной деятельности, но и в наиболее перспективных и быстроразвивающихся областях, например, социальных сетей, рекламной деятельности, кино индустрии, театрального искусства, HR-менеджмента и многих других.

Для начала необходим сбор примеров голосовых команд для создания голосовых меток. Звук должен пройти дальнейшую обработку и запись в определённый формат, например, в WAV, это PCM формат записи без потерь. Аудио файлу обязательно необходимо иметь некое значение для обеспечения следующего шага.

Далее необходимо провести генерацию вариаций звуковых сигналов. Нужно генерировать новые аудио файлы, дополняющие программу изменёнными данными. Изменение высоты, громкости, постороннего шума и тембра речи должно увеличить процент распознавания, так как система

нацелена на разных дикторов, в том числе с дефектами речи и специфичными диалектами в лексиконе. Создание всех возможных случаев и учет непредвиденных ситуаций обязательно должны присутствовать в процессе обучения нейронной сети. Сеть должна быть самообучающаяся. Пользователь будет дополнять базу обучающими голосовыми метками. Этот этап в большей степени определит успешность проекта по созданию программ в будущем [5] .

Самым важным этапом работы можно считать перевод речи человека в программные команды для выборки правильных визуальных элементов или исправления неправильных. Для этого необходимы принципиально отличные способы преобразования голосовых меток в тезаурус программного кода.

После создания приложения его необходимо протестировать на смысловые ошибки, неточности, ограниченность словарного запаса, возможность создания новых голосовых меток в процессе работы.

Основная функция программы – составление наиболее точной визуальной модели с помощью голосового ввода данных. Такой функционал полностью подходит для применения в правоохранительной сфере, в криминалистике и расследовании преступлений, в частности. В настоящее время полиция и другие службы, ведущие оперативно-розыскные действия, активно используют фотороботы для расследования преступлений. Составляются они экспертами, но их штат крайне мал и жертва нападения может забыть некоторые особенности лица преступника. Время будет потеряно.

Применение системы создания фоторобота по голосовому вводу позволит намного точнее и эффективнее использовать данные свидетелей, так как в руки сотрудников правоохранительных органов попадет не примерное текстовое сообщение, а готовое графическое изображение правонарушителя. Так же пострадавший может составить фоторобот самостоятельно и прикрепить к заявлению в полицию.

Однако это не единственная область применения голосового интерфейса визуальных объектов. Эта программа поможет и в других сферах жизни общества.

Часто случается, что требуется найти человека в социальных сетях по приблизительным воспоминаниям. Использование голосовых команд поможет с легкостью сузить диапазон поисков и найти человека.

Данная система также поможет решить кадровый вопрос в кино индустрии и театральном искусстве. Она самостоятельно подготовит подходящие анкеты, не прибегая к ручному перебору и рутинной работе.

Описанные в настоящей статье возможности применения программ с функциями голосовых команд в различных областях жизнедеятельности человека имеют широкие перспективы использования и в будущем, например, в создании голосовых интерфейсов.

Список литературы:

1. Петрушенко А. А., Петрушенко Р. В. Речевые технологии — следующий уровень сервиса [Текст] /А.А.Петрушко, Р.В.Петрушко. - Технические науки в России и за рубежом: материалы V Междунар. науч. Конф.: Москва, январь 2016. — М.: Буки-Веди, 2016. — С. 6-8. — Режим доступа: <https://moluch.ru/conf/tech/archive/164/9303/> (дата обращения: 25.10.2019).

2. Алимуратов А. К., Чураков П. П. Выбор оптимального набора информативных параметров речевых сигналов для систем голосового управления[Текст] /А.Ю. Тычков-режим доступа: <https://cyberleninka.ru/article/v/vybor-optimalnogo-nabora-informativnyh-parametrov-rechevyh-signalov-dlya-sistem-golosovogo-upravleniya>

3. Алимуратов А. К. Обработка речевых команд в системах голосового управления[Текст]/А.К.Алимуратов-Режимдоступа: <https://cyberleninka.ru/article/v/obrabotka-rechevyh-komand-v-sistemah-golosovogo-upravleniya>

4. Алимуратов А.К., Чураков П.П. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи/А.К. Алимуратов,

П.П.Чураков - Режим доступа: <https://cyberleninka.ru/article/v/obzor-i-klassifikatsiya-metodov-obrabotki-rechevyh-signalov-v-sistemah-raspoznavaniya-rechi>

5. Проблемы интерпретации голосового ввода — как это работает у нас – режим доступа: <https://habr.com/ru/post/235763/>

6. Горшенин В.И. Особенности профессиональной социализации будущего специалиста среднего звена / В.И. Горшенин // Современные проблемы науки и образования. - 2016. - № 6. - С. 446.

POSSIBILITIES OF USING PROGRAMS WITH VOICE RECOGNITION FUNCTIONS IN VARIOUS AREAS OF HUMAN LIFE

Bazarov Kirill Alexandrovich

2nd year student of the Institute of Mathematics,
Physics and Information

Technology Tambov State University G.V. Derzhavina
Tambov, Russia

e-mail: bazarov.kirill@mail.ru

Abstract: this paper lists the advantages and disadvantages of voice technologies that are used in the most promising and rapidly developing areas of human life. Also, voice programs have great prospects in the future in connection with the development of technologies in the field of creating a voice interface for visual objects.

Keywords: image robot, voice interfaces, neural networks, programs using speech recognition systems.