

О РАЗРАБОТКЕ СИСТЕМЫ ДЛЯ РАСПОЗНАВАНИЯ РЕЧИ

© 2018 Д. М. Жуманиязов, И. Я. Львович

Воронежский институт высоких технологий (г. Воронеж, Россия)

В статье описываются основные особенности системы, связанной с распознаванием голосовых сообщений. Приведена структурная схема системы обучения. Даны результаты тестового распознавания.

Ключевые слова: распознавание звука, обработка сигналов, информационная система, микрофон, голосовое управление.

Речь – это основное средство коммуникации среди людей. Работы по распознаванию речи и генерации сигналов речи ведутся уже ни одно десятилетие.

Автоматическое распознавание речи – это процесс, при котором компьютер принимает речевой сигнал и преобразует его в слова [1].

Компьютер распознает то, что сказал человек. Клавиатура, хотя и является популярным средством, не очень удобна, так как требует владение определенными навыками.

С другой стороны, мышь (манипулятор) требует хорошей координации глаз. Людям с ограниченными физическими возможностями трудно пользоваться компьютером. Частично слепым людям трудно читать с монитора.

Все эти ограничения должны быть устранены. Речевой интерфейс помогает нам решать эти проблемы. Цель состоит в том, чтобы зафиксировать человеческий голос в цифровом компьютере и декодировать его в соответствующий текст.

Распознавание речи может быть определено как процесс преобразования акустического сигнала, захваченного микрофоном или телефоном, в набор слов.

Когда два человека говорят друг с другом, они оба узнают слова и смысл, который стоит за ними. Компьютеры, с другой стороны, способны только на первое: они могут распознавать отдельные слова и фразы, но на самом деле они не понимают речь так же, как люди. Компьютер распознает команду и программное обеспечение сообщает компь-

ютеру, что делать, когда эта команда будет распознана.

1. Классификация системы распознавания речи.

Существует большое разнообразие в технологии распознавания речи и важно понимать различия. Системы распознавания речи можно классифицировать по типу речи, размеру лексики, основным единицам и зависимости от говорящего [2]. Особенности системы распознавания речи в этих измерениях определяет, какой алгоритм должен использоваться.

A. Существует в основном два типа речи:

1. Слитная речь.

2. Дискретная речь.

Дискретная речь состоит из отдельных слов, разделенных паузами. Преимущество дискретной речи заключается в том, что границы слов могут быть установлены точно в то время как с непрерывной речью; слова будут произноситься без молчания.

B. Размер словаря.

Размер словаря является вторым типичным аспектом технологии распознавания речи. Словарь представляет собой набор слов, которые должны быть признаны.

Небольшой словарный запас – тот, который содержит менее 30 слов. Словарный запас в 500 слов – это средний размер.

Размер словаря является вторым типичным аспектом технологии распознавания речи. Словарь представляет собой набор слов, которые должны быть распознаны.

Словари с более чем 25000 слов обычно рассматриваются как очень большие, хотя эти определения, как правило, зависят от области применения.

C. Динамика зависимости параметров

- звучание зависит от системы;
- дикторнезависимая система;

Жуманиязов Дмитрий Маратович – Воронежский институт высоких технологий, магистрант, yZhumman4RTOTE@yandex.ru

Львович Игорь Яковлевич – Воронежский институт высоких технологий, д. т. н., профессор, office@vvt.ru.

- звучание рассматривается как гибкая система.

Некоторые системы, зависящие от звучания, требуют только, чтобы пользователь записал подмножество системного словаря, чтобы сделать весь словарь узнаваемым.

Система, не зависящая от звучания, не требует никакой записи перед системой, зависящей от звучания, требует, чтобы пользователь записал пример использования слова, предложения или фразы.

Независимая акустическая система разработана для работы с любым произношением определенного типа (например, американский английский язык).

Нами разработана адаптивная акустическая система для того, чтобы она была готова к характеристикам новых говорящих людей.

2. Проектирование системы.

Сформированная система при визуализации в виде блок-схемы будет иметь следующие компоненты: Запись звука и компонент распознавания слов, компонент извлечения признаков, компонент распознавания речи, акустическую и языковую модель.

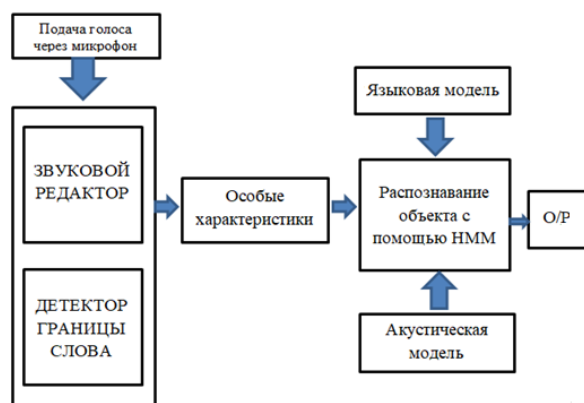


Рисунок 1. Структурная схема системы обучения.

А. звукозапись и компонента обнаружения слов отвечает за ввод с микрофона и выявление наличия слов.

Выходные данные могут быть wav-файлом или напрямую передаваться в устройство обработки [3].

В. компонент извлечения признаков: компонент генерирует векторы признаков для звуковых сигналов, передаваемых ему.

С. компонент распознавания: это непрерывная, многомерная скрытая марковская модель.

Это самый важный компонент системы, он отвечает за поиск наилучшего соответст-

вия в базе знаний, за входящие векторы признаков.

Д. модель знания. Система распознавания использует эту модель при распознавании звукового сигнала.

После того, как обучение будет проведено, модуль извлечения объектов генерирует векторы объектов, которые затем перенаправляются в компонент распознавания.

Компонент распознавания с помощью модели знаний формирует результат.

3. Результаты.

Система протестирована для различных параметров, полученный результат представлен ниже.

В эксперименте скрытая Марковская модель используется для распознавания изолированных английских слов.

А. При обучении системы мы использовали 10 потребителей. Каждый пользователь обучал систему в течение некоторого времени, помехи были исключены. Распознавание было опробовано на двух видах звуков. Известный пользователь: пользователь, голос которого используется для обучения.

Неизвестный: пользователь, голос которого не использовался для обучения.

Результаты эксперимента приведены в таблицах 1, 2.

Система распознавания речи является независимой от акустической системы и тестируется совокупностью говорящих людей

Точность эксперимента зависит от времени обучения. Если время тренировки системы увеличено, то точность автоматически увеличивается, причем, время обучения прямо пропорционально точности. Вывод. Рассмотрены особенности распознавания голосовых сообщений в созданной системе.

Ключевым фактором при проектировании такой системы является целевая аудитория. Система не может правильно различать похожие слова, потому что у них похожие звуковые фонемы. В ходе эксперимента была использована система со средним словарным запасом.

Система может быть расширена до непрерывного распознавания слов с большим словарным запасом на основе акустической модели телефона, используя технику скрытых марковских моделей или с использованием других перспективных методов, таких как искусственная нейронная сеть.

Таблица 1

Результаты распознавания

Вид пользователя	Нет звука	Правильное распознавание	Неправильное распознавание
Известный пользователю	10	8	2
Неизвестный	10	5	5

Таблица 2

Вероятность распознавания слов

Слова	Количество тестов	Вероятность распознавания
Парк	10	100 %
Привет	10	100 %
Баня	10	100 %
Яблоко	10	100 %
Трактор	10	90 %
Пей	10	80 %

ЛИТЕРАТУРА

1. Львович, И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова // Воронеж, Издательство: Воронежский институт высоких технологий (Воронеж). – 2014. – 339 с.

2. Львович, Я. Е. Адаптивное управление марковскими процессами в конфликтной ситуации / Я. Е. Львович, Ю. П. Преоб-

раженский, Р. Ю. Паневин // Вестник Воронежского государственного технического университета. – 2008. – Т. 4. – № 11. – С. 170-171.

3. Завьялов, Д. В. О применении информационных технологий / Д. В. Завьялов // Современные наукоемкие технологии. – 2013. – № 8-1. – С. 71-72.

ABOUT THE DEVELOPING A SYSTEM FOR SPEECH RECOGNITION

© 2018 D. M. Zhumaniyazov, I. Ya. Lvovich

Voronezh Institute of high technologies (Voronezh, Russia)

The paper describes the main features of the system associated with the recognition of voice messages. The block diagram of the training system is given. The results of the test of recognition are shown.

Key words: sound recognition, signal processing, information system, microphone, voice control.