

## СИСТЕМА РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ АНГЛИЙСКОГО ЯЗЫКА

© 2018 Д. М. Жуманиязов, В. Н. Кострова, Н. А. Коростелева

*Воронежский институт высоких технологий (г. Воронеж, Россия)  
Воронежский государственный технический университет (г. Воронеж, Россия)  
Открытое акционерное общество «Воронежский завод  
полупроводниковых приборов-сборка» (г. Воронеж, Россия)*

*Настоящее исследование направлено на разработку подходящего распознавания речи для языка, для людей, которые физически затруднены и не могут управлять компьютером с помощью клавиатуры и мыши. В данной статье используется НММ (скрытая Марковская модель) для распознавания образцов речи, чтобы дать отличные результаты для изолированных слов.*

*Ключевые слова: речевой сигнал, распознавание, информация.*

Речь является основным средством общения между людьми. Распознавание речи, генерация речевых сигналов, разрабатывается уже несколько десятилетий.<sup>1</sup>

Автоматическое распознавание речи - это процесс, при котором компьютер принимает речевой сигнал и преобразует его в слова.

Это процесс, с помощью которого компьютер распознает то, что сказал человек. Клавиатура, хотя и является популярным средством, не очень удобна, так как требует определенного количества навыков для эффективного использования.

С другой стороны, мышь 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 человеческий голос, хранящийся в компьютере был декодирован 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, стоящий

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 компьютеру, что делать, когда эта команда распознается.

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. Положение системы распознавания речи в этих измерениях определяет, какой алгоритм должен использоваться.

А. Существует два основных типа речи:

1. Слитной речи
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.

В. 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 признаны. Небольшой словарный запас, тот который со-

Жуманиязов Дмитрий Маратович – Воронежский институт высоких технологий, студент, turor45zhumm@yandex.ru.

Кострова Вера Николаевна – Воронежский государственный технический университет, профессор, kost\_3\_veranikol@yandex.ru.

Коростелева Наталья Александровна – Открытое акционерное общество «Воронежский завод полупроводниковых приборов-сборка», сотрудник, koerolostor99\_34@yandex.ru.

держит менее 30 слов. Словарный запас в 500 слов – это средний размер.

Словари с более чем 25000 слов обычно могут рассматриваться как очень большие, хотя это, как правило, зависит от области применения.

С. Динамика зависимости

- динамик зависит от системы;
- дикторо-независимой системы;
- динамик гибкая система.

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

Система, не зависящая от динамика, не требует никакой записи.

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

Разработана адаптивная акустическая система для адаптации ее работы к характеристикам новых динамиков.

Созданная система, если ее представить в виде блок-схемы будет иметь следующие компоненты: Запись звука и компонента распознавания слов, компонента извлечения признаков, компонент распознавания речи, акустическую и языковую модель.

1. Компонента звукозаписи и обнаружения слова отвечает за ввод с микрофона и выявление наличия слов. Обнаружение слова проводится на основе энергетических характеристик и сравнении с эталонными сигналами.

Выходные данные этого компонента могут быть файлом формата \*.wav или прямым образом подаваться на модуль распознавания объектов.

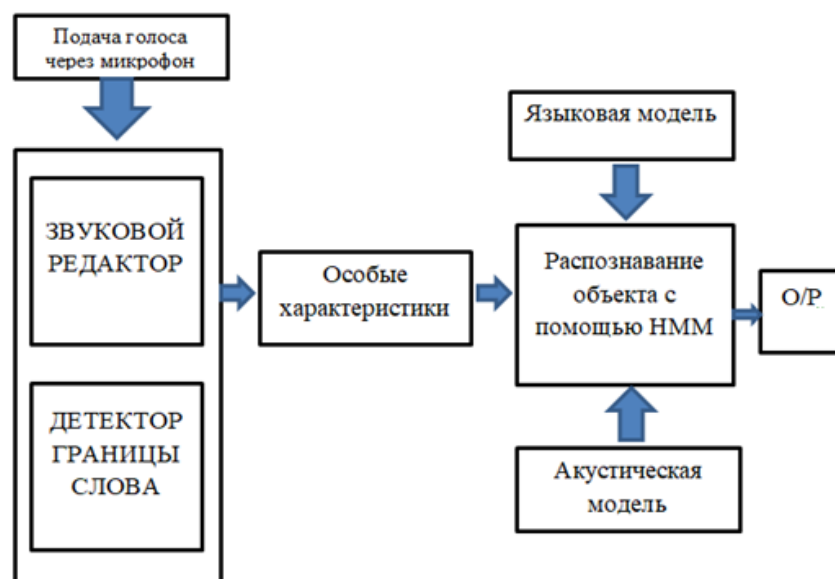


Рисунок 1. Структурная схема системы обучения

2. Компонент распознавания базируется на непрерывной, многомерной скрытой Марковской модели.

Это самый важный компонент системы и отвечает за поиск наилучшего соответствия в базе знаний, за входящие векторы признаков.

3. Модель знания, основывается на том, каким образом происходит звучание слова.

Модуль распознавания объектов генерирует векторы объектов.

Компонент распознавания с помощью модели знаний выдает результат.

Система протестирована для различных параметров, полученный результат представлен ниже.

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

Для тренировки системы мы использовали 10 пользователей. Каждый пользователь обучал систему в течение некоторого времени.

Распознавание было опробовано на двух видах звуков: известный пользователь, пользователь, голос которого используется для обучения и пользователь, голос которого не использовался для обучения.

Результат эксперимента приведен в таблице 1.

Таблица 1

## Результаты распознавания

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

Таблица 2

## Вероятность распознавания слов

Слова	Количество тестов	Вероятность распознавания
Station	10	100 %
Student	10	100 %
Try	10	100 %
Lemon	10	100 %
Mobile	10	90 %
Biology	10	80 %

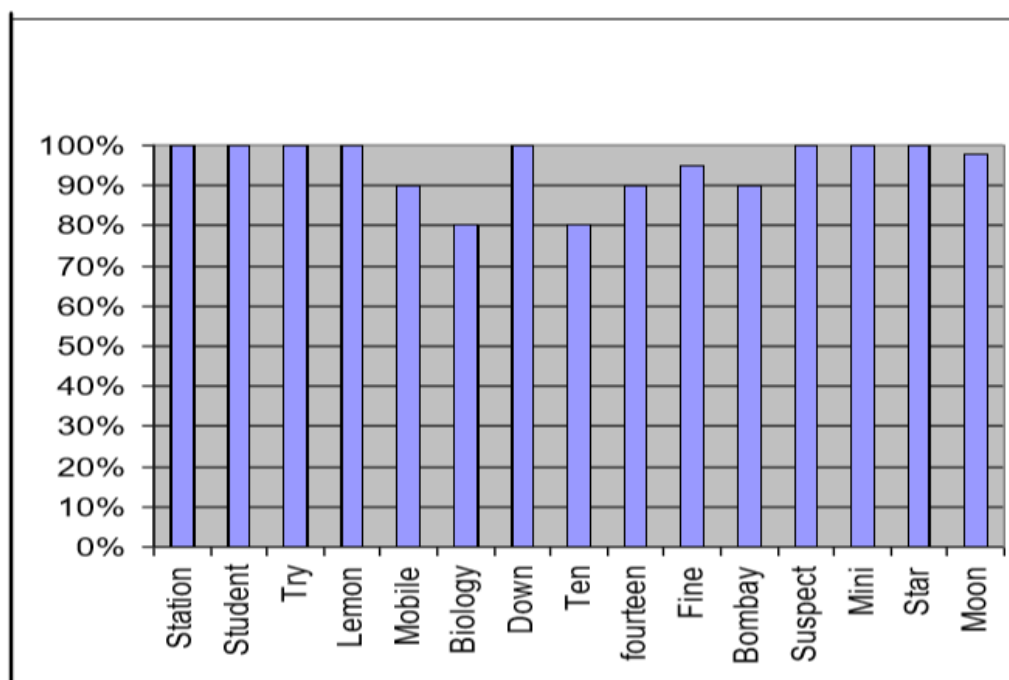


Рисунок 2. Вероятность распознавания слов для 10 пользователей.

Таблица 3

## Результаты, показывающие характеристики динамиков и точность

Down	10	100 %
Ten	10	80 %
Fourteen	10	90 %
Fine	10	95 %
Bombay	10	90 %
Suspect	10	100 %
Mini	10	100 %
Star	10	100 %
Moon	10	98 %

Нет диктора	Точность
First	60 %
Second	70 %
Third	75 %
Fourth	50 %
Fifth	65 %
Sixth	70 %
Seventh	80 %
Eighth	85 %
Ninth	90 %
Tenth	93 %

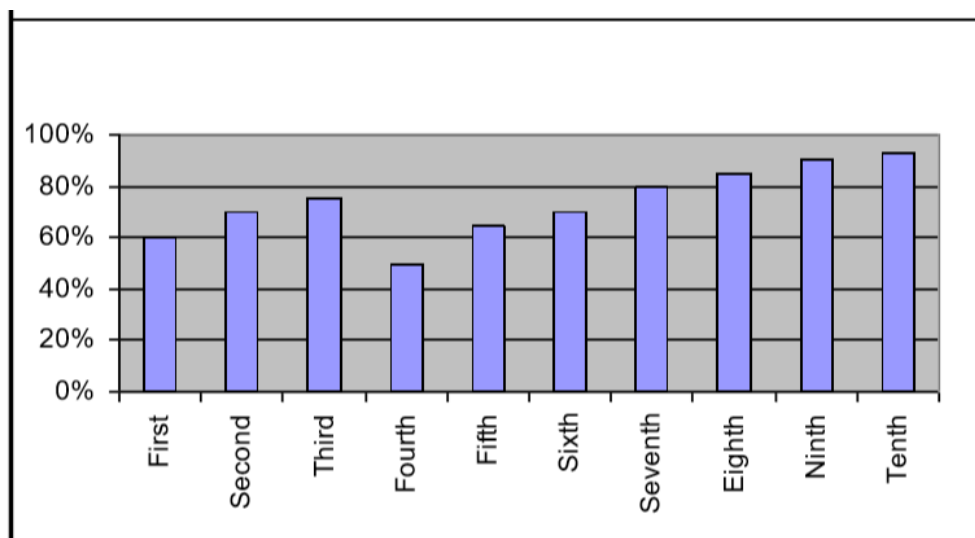


Рисунок 3. Точность и количество динамиков.

Языковая модель содержит список слов. При распознавании системы необходимо знать, где найти модель для каждого слова и какое слово соответствует модели.

Эта информация хранится в файле models в каталоге.

Эти модели в каталоге известны как HMMs. Если пользователю нужно распознать новое слово, пользователь должен будет обучить систему этому слову.

В процессе распознавания звук сопоставляется с каждой моделью, чтобы найти наилучшее соответствие.

Построение модели для слова требует от нас сбора звуковых файлов слов от различных пользователей. Эти звуковые файлы затем используются для обучения модели НММ.

Точность распознавания речи оценивается по количеству ошибок изолированного слова, всего используется пять сотен отдельных слов. Тестирование случайно выбранных пятидесяти слов производится различными говорящими.

Система распознавания речи является независимой от акустической системы и тес-

тируется совокупностью динамиков, и каждый раз результат меняется.

На рисунке 3 показана точность, рассчитанная при тестировании системы десятью различными динамиками.

Точность эксперимента зависит от времени обучения.

Если время тренировки системы увеличено, то точность автоматически увеличится. время прямо пропорционально точности.

**Вывод.** В статье используется акустическая модель на основе слова. Эта модель может быть использована только для ограниченного размера словаря. По мере увеличения размера словаря производительность системы снижается.

Система не может правильно различать похожие слова, так как у них похожие звуковые фонемы.

В ходе эксперимента была использована система, базирующаяся на среднем словаре. Система может быть расширена до непрерывного распознавания слов при большом размере словаря, за счет использования техники НММ или с использованием других

перспективных методов, таких как искусственная нейронная сеть.

#### ЛИТЕРАТУРА

1. Головинов С. О. Цифровая обработка сигналов / С. О. Головинов, С. Г. Миронченко, Е. В. Щепилов, А. П. Преображенский // Вестник Воронежского института высоких технологий. 2009. № 4. С. 064-065.

2. Зазулин А. В. Особенности построения семантических моделей предметной области / А. В. Зазулин, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. 2008. № 3. С. 026-028.

3. Зяблов Е. Л. Разработка лингвистических средств интеллектуальной поддержки на основе имитационно-семантического моделирования / Е. Л. Зяблов, Ю. П. Преображенский // Вестник Воронежского института высоких технологий. 2009. № 5. С. 024-026.

4. Кленяева Г. В. Современные проблемы речевой акустики и построения систем автоматического распознавания речи / Г. В. Кленяева, А. П. Преображенский // Вестник Воронежского института высоких технологий. 2007. Т. 1. № 2-1. С. 071-074.

5. Lvovich I. Ya. Analysis of potential of error-correcting capabilities of codes / I. Ya.

Lvovich, A. P. Preobrazhensky, O. N. Choporov // Life Science Journal. 2013. Т. 10. № 4. С. 830-834.

6. Маркел Дж. Линейное предсказание речи / Дж. Маркел, А. Грей. – М.: Связь, 1980. – 308 с.

7. Методы и стандарты кодирования и сжатия речи в цифровой телефонии: Интернет-публикации. – Режим доступа: <http://dox.sbnnet.ru:8082/~serge/speech.ru/coder/>.

8. Преображенский А. П. Анализ методов кодирования разных видов информации / А. П. Преображенский // Вестник Воронежского института высоких технологий. 2013. № 10. С. 74-77.

9. Рабинер Л. Р. Цифровая обработка речевых сигналов: пер. с англ. / Л. Р. Рабинер, Р. В. Шафер; под ред. М. В. Назарова и Ю. Н. Прохорова. – М.: Радио и связь, 1981. – 496 с.

10. Смит, С. В. Научно-техническое руководство по цифровой обработке сигналов / С. В. Смит; пер. с англ. В. Н. Покровского, В. И. Силантьева. – СПб.: АВТЭКС, 2001. – 630 с.

#### THE SPEECH RECOGNITION SYSTEM FOR ENGLISH

© 2018 D. M. Zhumaniyazov, V. N. Kostrova, N. A. Korosteleva

*Voronezh Institute of high technologies (Voronezh, Russia)  
Voronezh state technical University (Voronezh, Russia)  
Open Joint-Stock Company «Voronezh Plant of Semiconductor  
Devices-Assembly» (Voronezh, Russia)*

*This paper presents a technique to improve the efficiency of the algorithm of compression of speech signals, without much loss of speech signals. Compression in this case means removing from the digital signal those quantization steps that can be predicted. We propose to reduce the number of these quantization steps using a modified variable order linear predication algorithm. This allows to reduce the time compression.*

*Key words: speech signal, identification, information.*