



УДК 381.3

**MAIN CHARACTERISTICS OF VOICE RECOGNITION METHODS  
ОСНОВНЫЕ ХАРАКТЕРИСТИКИ МЕТОДОВ РАСПОЗНАВАНИЯ ГОЛОСА****Kostrova V.N. / Кострова В.Н.***d.t.s., prof. / д.т.н., проф.*

SPIN: 1330-4645

*Voronezh State Technical University, Voronezh, Moskowskiy, 14, 394026**Воронежский государственный технический университет,**Воронеж, Московский пр, 14, 394026***Tserovscaя T.A. / Цепковская Т.А.***senior lecturer / старший преподаватель*

SPIN: 2567-8407

*Voronezh Institute of High Technologies, Voronezh, Lenina, 73a, 394043**Воронежский институт высоких технологий,**Воронеж, ул.Ленина, 73а, 394043*

**Аннотация.** В работе рассматриваются некоторые особенности задач, направленных на распознавание голосовых сигналов. Показаны процессы, проходящие в момент прослушивания голоса. Указаны основные шаги процедуры распознавания. Обсуждается механизм распознавания фонем. Показаны проблемы, которые связаны с распознаванием голосовых сигналов.

**Ключевые слова:** речь, кодирование, голос, распознавание, сигнал.

**Вступление.**

Распознавание речи рассматривается в виде автоматическое отнесение единицы речи к заранее выделенным в ней признакам, близким к классификации, осуществляемой человеком.

Речевое, столь привычное, естественное для человека управление объектами открыло бы широкие перспективы перед автоматизацией производства; раздвинуло бы границы возможностей общения с машинами, особенно пользователей персональных компьютеров, не знающих языков программирования. Речевой контакт облегчает ввод данных в машину, помогает работать человеку и компьютеру в реальном масштабе времени: человек сказал - машина сделала.

Целью данной работы является анализ основных проблем, связанных с распознаванием речевых сигналов.

**Основной текст.**

Чтобы машина научилась понимать человеческий язык, отвечать на вопросы, исследователи затрачивают много сил и времени, начиная ее гигантской информацией даже для того, чтобы распознавать отдельные звуки [1, 2].

У каждого звука сложная волновая структура, включающая различные частоты и колебания, к тому же, естественно, одно и то же слово разные люди произносят по-своему: разный тембр голоса (звуковая окраска), разные интонации, разная чистота произношения. Сколько людей, столько и голосов. Голос - индивидуальный признак личности, такой, как почерк и отпечатки пальцев [3].



Что же происходит в машине во время прослушивания? То же, что при разговоре по телефону или радио: звуковые колебания преобразуются в электрические. В специальных "решетках" - фильтрах - они "просеиваются" по частоте. Затем в машинной памяти по строго определенному узору сигналов сравниваются с хранящимися там эталонами. Этот узор картина звука - и есть тот усредненный звук, который научилась узнавать машина.

Анализ узора - работа довольно утомительная и однообразная. Например, при одном эксперименте испытуемый 100 раз произнес букву "А". Она фиксировалась все 100 раз 14 вариантами произношения, или картинами. Потом картины сравнили по частоте появления. Оказалось, из 14 вариантов чаще всего встречается одна картина. Именно этот узор и записали в памяти машины [4, 5].

Машина, распознающая речь, сама остается немой: она не говорит, а пишет ответы. Устраивает ли это пользователя? Да, безусловно, и все-таки общение с компьютером в полном смысле слова - это диалог. Тем более что в создании определенного звука не обязательно участие голосовых связок, звук можно искусственно синтезировать, научить говорить машину.

Отвечать на вопросы голосом учат машину давно. О степени трудности создания говорящей машины свидетельствует такой пример.

Чтобы машина произнесла, например, на венгерском языке короткое "добрый день", нужно заложить в ее память 500 различных характеристик. Еще больше усилий стоило японским специалистам заставить говорить наручные часы. При нажатии кнопки они объявляют часы и минуты [7, 8].

Можно использовать компьютер в редактировании текстов, в совершенствовании обучающих машин.

### ***Основные методы распознавания голоса.***

Процедура распознавания голоса осуществляется в несколько шагов:

1. фиксация голосового сигнала и проведение предварительной обработки речи.
2. проведение распознавания фонем и слов.
3. осуществление понимания речи.

Теперь более подробно о каждом:

*1. Фиксация голосового сигнала и проведение предварительной обработки речи.* Фиксация голосового сигнала или, другими словами, проведение дискретизации голоса представляет собой процесс измерения и преобразования акустического сигнала. При этом голос рассматривается как явление колебаний акустического давления в микрофоне, которое характеризуется в общем, низкочастотными сигналами, которые лежат в диапазоне от 0 до 4 кГц. Существуют два вида звуков: глухие и звонкие. Звонкие возникают за счет вибраций голосовых связок, когда через них проходит воздух. При этом поток воздуха, формирующий звук, называют "волной, которая идет из голосовой щели". Такой сигнал получается квазипериодический, и его период относится к основному тону. В свою очередь, резонансный сигнал, относящийся к звонким звукам, как правило, формируется из 4 частотных составляющих, которые называют формантами. Эти форманты представляют собой "голосовую печать"



для разных звуков, которые делаются речевым аппаратом человека. В свою очередь, глухие звуки получаются когда проходит воздух через речевой канал, в том случае, если нет колебаний голосовых связок. Эти два типа звуков могут анализироваться как временные ряды, которые отсчитаны в течение регулярных интервалов времени. С целью проведения изоляции требуемого интервала применяют пространственные окна. Определенные оконные функции вычисляют среднюю амплитуду, а также число нулевых пересечений и преобразование Фурье сигнала в течение интервала. С целью того, чтобы убрать шум применяют разные методы, позволяющие фильтровать сигнал.

2. *Распознавание фонем и слов.* При распознавании фонем, а также групп фонем и слов применяют такие способы, как скрытую марковскую модель или НММ (hidden Markov modelling), нейронные сети (ИНС).

Достаточно часто и хорошо при осуществлении распознавания фонем и слов применяется скрытая марковская модель (НММ). НММ можно применять для того, чтобы представить любую составляющую речевого сигнала, к которым относятся слова или фонемы.

Так как для потока речи можно отметить строгую временную направленность, то используется специальная топологическая процедура для направленного потока (слева направо). Исследователями было продемонстрировано, что ИНС можно применять для того, чтобы повысить мощность тех распознавателей, у которых структура базируется на скрытой марковской модели. Например, ИНС с несложной уровневой структурой дает возможность обеспечения получения оценок вероятностей для НММ моделей.

3. *Понимание речи.* "Понять" речь - это довольно трудная задача. На таком шаге обработки сигнала предложения необходимо преобразовать в определенные представления о том, что желал сказать тот, кто говорил. Проблема, которая решается при распознавании голоса состоит в распознавании говорящего. При этом можно говорить о идентификации или о верификации того, кто говорит. Идентификация состоит в определении в известной совокупности контрольных фраз экземпляра, который соответствует данной манере говорящего. Способ распознавания диктора дает возможность применять голос для того, чтобы обеспечить контроль доступа.

*Нерешенные задачи и определенные проблемы в будущем.*

На данный момент стоят три основных проблемы на пути создания эффективных систем распознавания речи:

1. довольно большие объемы словарей;
2. сложности с построением шаблонов непрерывной речи;
3. существование различных акцентов и произношений.

Это базовые препятствия для создания автоматизированных систем распознавания голоса. Помимо них есть и другие.

### **Заключение и выводы.**

Таким образом, в статье проведен анализ основных проблем, связанных с распознаванием речевых сигналов. Кратко описаны характеристики способов распознавания голоса.



## Литература:

1. Преображенский Ю.П., Мясников О.А. Анализ перспектив информационных технологий в сфере интернет вещей // Вестник Воронежского института высоких технологий. 2020. № 1 (32). С. 43-45.
2. Мельникова Т.В., Преображенский А.П. О перспективах передачи информации в информационно-телекоммуникационных системах к 2100 году // Вестник Воронежского института высоких технологий. 2020. № 2 (33). С. 28-30.
3. Львович Э.М., Холодков А.М. Проблемы передачи информации в автоматизированных системах управления // Вестник Воронежского института высоких технологий. 2020. № 3 (34). С. 30-33.
4. Львович Я.Е., Преображенский А.П., Преображенский Ю.П. Анализ возможностей применения технологий обработки информации в образовательной сфере // Вестник Воронежского института высоких технологий. 2020. № 3 (34). С. 41-43.
5. Клименко Ю.А., Преображенский А.П. О способах устранения шумов на базе числовой обработки сигналов // Вестник Воронежского института высоких технологий. 2020. № 4 (35). С. 41-43.
6. Клименко Ю.А., Преображенский А.П. Особенности методов многомерного анализа // Вестник Воронежского института высоких технологий. 2020. № 4 (35). С. 44-46.
7. Львович И.Я., Чупринская Ю.Л., Кравцова Н.Е. О разделении и анализе перекрещивающихся сигналов // Вестник Воронежского института высоких технологий. 2020. № 4 (35). С. 65-67.

***Abstract.** The paper discusses some of the features of tasks aimed at recognizing voice signals. The processes taking place at the moment of listening to the voice are shown. The main steps of the recognition procedure are indicated. The phoneme recognition mechanism is discussed. Problems related to the recognition of voice signals are shown.*

***Key words:** speech, coding, voice, recognition, signal.*

Статья отправлена: 07.06.2021 г.  
© Кострова В.Н., Цепковская Т.А.