

УДК 519.668

В.И. Свиридов
**ОСНОВНЫЕ ХАРАКТЕРИСТИКИ
МЕТОДОВ РАСПОЗНАВАНИЯ ГОЛОСА**
Воронежский институт высоких технологий

Статья посвящена анализу основных характеристик распознавания голоса. Рассмотрены основные проблемы, связанные с обработкой речи.

Ключевые слова: речь, распознавание.

Распознавание речи это автоматическое отнесение единицы речи к заранее выделенным в ней признакам, близким к классификации, осуществляемой человеком. В настоящее время одним из перспективным приложением речевого управления объектами является автоматизация производства, взаимодействие между пользователями и персональными компьютерами. Речевой контакт облегчает ввод данных в машину, помогает работать человеку и компьютеру в реальном масштабе времени: человек сказал - машина сделала [1].

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

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.

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.

Что же происходит в машине во время прослушивания? То же, что при разговоре по телефону или радио: звуковые колебания преобразуются в электрические. В специальных «решетах» - 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 картинками. Потом картины сравнили по частоте появления. Оказалось, из 14 вариантов чаще всего встречается одна картина. Именно этот узор и записали в памяти машины.

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.

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

3. Основные методы распознавания голоса [1, 2].

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 [3-6]:

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.

4. 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. Например, еще не решили в полной мере задачу отделения речевого сигнала от шума, который идет фоном.

Выводы. Распознавание речи имеет большие перспективы с точки зрения практических приложений. Необходимо учитывать при построении современных систем, какие методы распознавания голоса используются, как проводится фиксация голосового сигнала. Перспективным направлением исследований является алгоритмов и методов понимания речи.

ЛИТЕРАТУРА

1. Кленяева Г.В., Преображенский А.П. Современные проблемы речевой акустики и построения систем автоматического распознавания речи / Вестник Воронежского института высоких технологий. 2007. Т. 1. № 2-1. С. 71-74.
2. Головинов С.О., Миронченко С.Г., Щепилов Е.В., Преображенский А.П. Цифровая обработка сигналов / Вестник Воронежского института высоких технологий. 2009. № 4. С. 64-65.
3. <http://www.russika.ru/t.php?t=4850>
4. <http://www.russika.ru/ef.php?s=5021>
5. <http://chernykh.net/content/view/146/>
6. <http://freepapers.ru/22/raspoznavanie-rechi-kompterom/205188.1284820.list3.html>

V.I. Sviridov

THE KEY FEATURES OF METHODS FOR VOICE RECOGNITION

Voronezh Institute of High Technologies

The paper analyzes the main characteristics of voice recognition. The main problems associated with the processing of speech are considered.

Keywords: speech, recognition.