

Распознавание речи — набор технологий, позволяющих управлять компьютером (или любым другим неодушевлённым предметом), используя человеческий голос. Эта давняя мечта человека отражена в команде: «Сезам откройся!», произносимой героем одной из сказок Шехеризады перед входом в пещеру с сокровищами.

Из всех живых существ только человек обладает речью, и во многом благодаря которой ему удалось столь значительно развить свои интеллектуальные способности и стать человеком. Естественно предположить, что нечто подобное происходит на наших глазах и с «интеллектом» компьютера, интенсивно овладевающим широким спектром речевых технологий от синтеза до распознавания и понимания речи.

За столь привычным, лёгким и естественным для человека процессом речевого общения скрываются огромные трудности на пути овладения речью компьютером, для разрешения которых требуется очень большая работа ученых многих специальностей - лингвистов, логиков, математиков, программистов. Чтобы компьютер научился понимать человеческую речь и отвечать на вопросы, исследователи затрачивают много сил и времени, начиняя его гигантской информацией даже для того только, чтобы распознавать отдельные звуки.

У каждого звука сложная волновая структура, включающая колебания с различными частотами и, к тому же, одно и то же слово разные люди произносят по-своему: разный тембр голоса (звуковая окраска), разные интонации, разный стиль произношения. Сколько людей, столько и голосов. Голос - индивидуальный признак личности, такой, как почерк и отпечатки пальцев.

Чтобы научить машину узнавать речь, ее заставляют прослушивать слова, произнесенные разными людьми. Задача компьютера - прослушав, учесть особенности произношения, свести на нет индивидуальность, чтобы потом, услышав слово, не ошибиться.

Речевой способ общения между человеком и компьютером имеет принципиальные преимущества, главными из которых являются следующие:

- удобство, естественность и простота общения, не требующая специальной подготовки. А это существенно расширяет круг потенциальных пользователей компьютерных систем;
- разгрузка зрительного канала при выводе информации и устранение ручных манипуляций при вводе, что увеличивает оперативность взаимодействия с компьютерной системой и уменьшает число ошибок оператора, позволяет ему находиться в движении и даже в темноте;

- возможность использования в качестве терминалов обычных телефонных аппаратов и существующей сети телефонной связи. Это позволяет создать человеко-машинные системы массового обслуживания по телефону, включая Интернет.

В 21 веке голосовое управление становится все более востребованным. Голосовой набор номера поддерживается всеми современными мобильными телефонами. Сейчас повсеместному распространению голосового управления мешает недостаточная вычислительная мощность процессоров и проблема наличия посторонних (внешних) шумов.

Голосовое управление основано на технологии распознавания речи: система получает информацию о колебаниях воздуха через микрофон, сравнивает полученные данные с командами, которые записаны в системе и, в случае совпадения, выполняет предписанное действие. Чем больше слогов в записанной команде, тем больше шанс, что система ее распознает без ошибок. К примеру, слова «Июнь» и «Июль» настолько схожи, что очень велика вероятность ошибки. Разработчики систем голосового управления решают эту проблему по-разному. Так, некоторые программы могут «учиться» интонациям пользователя и «привыкать» к его манере говорить. А одна из программ голосового переводчика просто учитывает, что пользователь запросил перевод слова «Вишня» в рамках раздела «Еда», поэтому программа не перепутает его с божеством «Вишну», который находится в разделе «Религия».

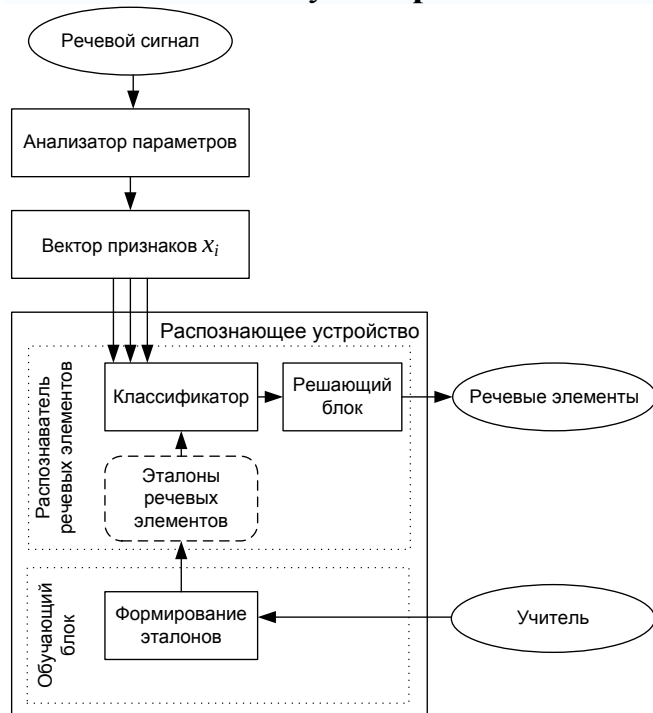


- Она, видно, не расслышала?! Я заказывала сосиски...

Голосовое управление берется на вооружение автомобильными гигантами: производители автомобилей стремятся повысить безопасность и комфорт водителя, поэтому дают возможность управлять бортовой электроникой при помощи голоса (наподобие «Включи радио, станция четыре»). Это позволяет не отводить глаза от дороги и не занимать руки водителя лишними манипуляциями.

Рассмотрим обобщенную структурную схему распознавания речи (см. рисунок). В соответствии с указанной схемой, в ходе **анализа речевого сигнала** формируется **вектор фонетических признаков** - x_i , которые поступают на **распознающее устройство**. Это устройство состоит из двух подсистем: классификатора (распознавателя) и обучающего блока. **Классификатор** сравнивает поступающие признаки с эталонами и передает

работу **решающему блоку**, которые вырабатывает решения по выводу распознанных речевых элементов. **Эталоны** формируются заранее **обучающим блоком** с участием учителя (пользователя), либо на основе подготовленной заранее обучающей выборки большого объема. На выходе такой системы получаем **распознанные речевые элементы**.



Р и с у н о к . Структурная схема распознавания речи

Рассмотрим отдельные элементы схемы распознавания речи более подробно.

Сравнение текущей реализации речевого сигнала и эталона в основном происходит по следующей схеме. Сначала вычисляется либо расстояние от эталона до реализации в некоторой метрике, либо мера подобия речевых элементов. В результате до каждого эталонного элемента получается свое расстояние или своя мера подобия. Тот элемент, до которого расстояние оказывается наименьшим (или мера подобия наибольшей), считается распознанным.

Обучение происходит следующим образом. Известные реализации речевых элементов задаются учителем или считываются с базы данных и по определенным алгоритмам строятся их эталоны. Проще всего взять в качестве эталонов усредненные реализации речевых сигналов, произнесенных диктором. Как правило, учитель произносит каждое слово (или, в некоторых случаях, фразу) несколько раз, а система вырабатывает его усредненный эталон. Существуют специальные базы речевых данных для различных условий распознавания.