

# **Системы распознавания речи**

# Что такое распознавание речи?

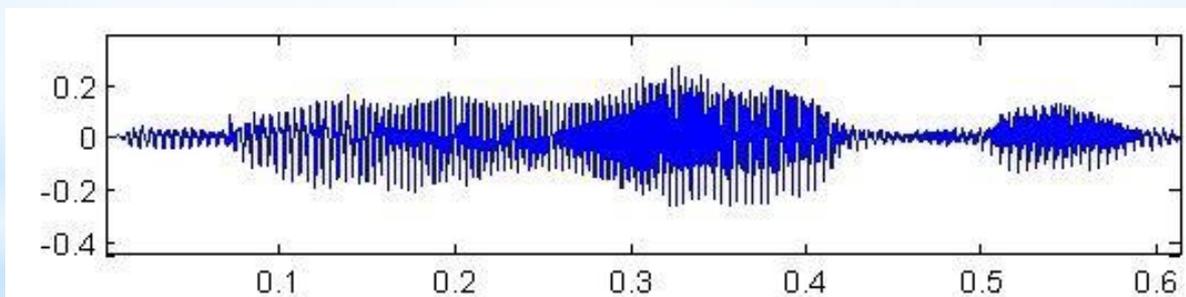
Распознавание речи - это многоуровневая задача распознавания образов, в которой акустические сигналы анализируются и структурируются в иерархию структурных элементов (например, фонем), слов, фраз и предложений

# Структура стандартной системы распознавания речи



# Необработанная речь

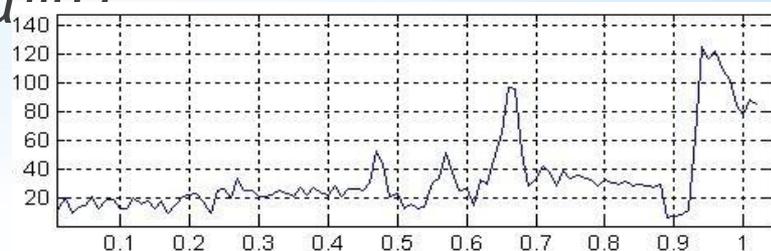
Обычно, поток звуковых данных, записанный с высокой дискретизацией (20 КГц при записи с микрофона либо 8 КГц при записи с телефонной линии)



# Анализ сигнала

Поступающий сигнал должен быть изначально трансформирован и сжат, для облегчения последующей обработки. Есть различные методы для извлечения полезных параметров и сжатия исходных данных в десятки раз без потери полезной информации. Наиболее используемые методы:

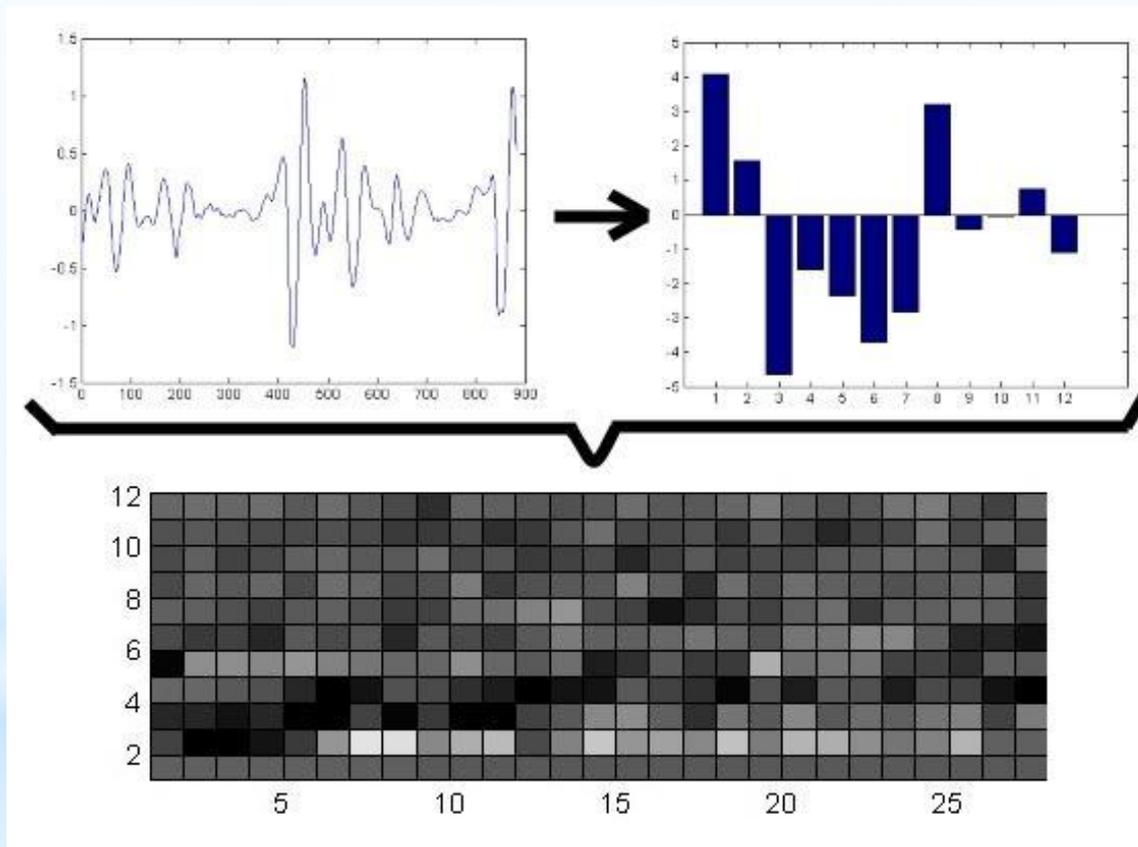
1. анализ Фурье;
2. линейное предсказание речи;
3. кепстральный анализ



# Речевые кадры

Результатом анализа сигнала является последовательность речевых кадров. Обычно, каждый речевой кадр - это результат анализа сигнала на небольшом отрезке времени (порядка 10 мс.), содержащий информацию об этом участке (порядка 20 коэффициентов). Для улучшения качества распознавания, в кадры может быть добавлена информация о первой или второй производной значений их коэффициентов для описания динамики изменения речи.

# Речевые кадры



# Акустические модели

Для анализа состава речевых кадров требуется набор акустических моделей. Рассмотрим две наиболее распространенные из них.

1. *Шаблонная модель.*
2. *Модель состояний.*

# Шаблонная модель

В качестве акустической модели выступает каким-либо образом сохраненный пример распознаваемой структурной единицы (слова, команды). Вариативность распознавания такой моделью достигается путем сохранения различных вариантов произношения одного и того же элемента (множество дикторов много раз повторяют одну и ту же команду). Используется, в основном, для распознавания слов как единого целого (командные системы).

# Модель состояний

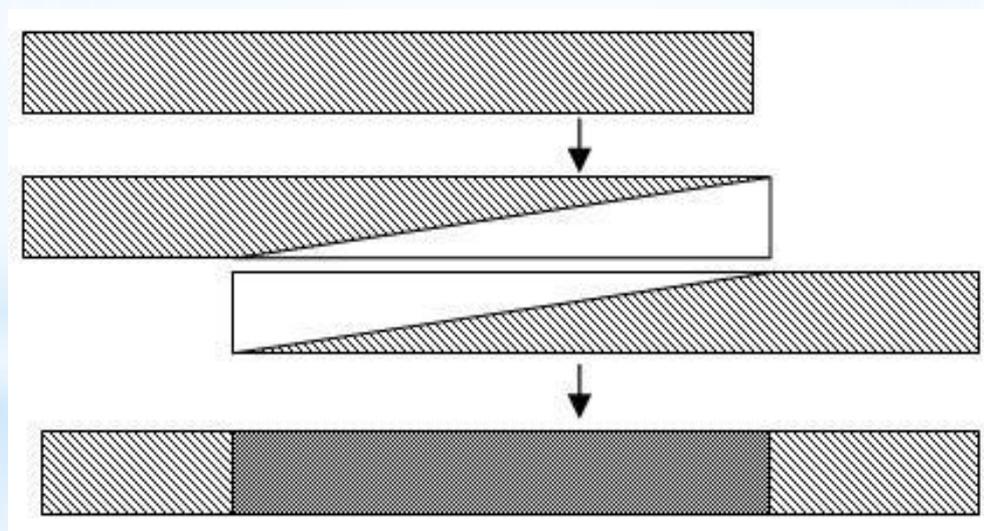
Каждое слово моделируется как последовательность состояний указывающих набор звуков, которые возможно услышать в данном участке слова, основываясь на вероятностных правилах. Этот подход используется в более масштабных системах.

# Акустический анализ

Состоит в сопоставлении различных акустических моделей к каждому кадру речи и выдает матрицу сопоставления последовательности кадров и множества акустических моделей. Для шаблонной модели, эта матрица представляет собой Евклидово расстояние между шаблонным и распознаваемым кадром. Для моделей, основанных на состоянии, матрица состоит из вероятностей того, что данное состояние может сгенерировать данный кадр.

# Корректировка времени

Используется для обработки временной вариативности, возникающей при произношении слов (например, “растягивание” или “съедание” звуков).



# Последовательность слов

В результате работы, система распознавания речи выдает последовательность (или несколько возможных последовательностей) слов, которая, наиболее вероятно, соответствует входному потоку речи.

**Спасибо за внимание!**