

# Проектирование человеко-машинных интерфейсов

Лекция 10.

Распознавание речи

# Что такое распознавание речи?



Система преобразования  
речевых сигналов в текст  
либо в набор управляющих  
КОМАНД

# Основное назначение систем распознавания речи

- Управление различными устройствами при помощи голосовых команд
- Голосовой набор номеров
- Ввод информации в системы с ограниченным словарем
- Полноценная диктовка текстов



## Основные проблемы распознавания

Акустическая  
изменчивость



В разные моменты времени одни и те же речевые фрагменты имеют отличающиеся характеристики

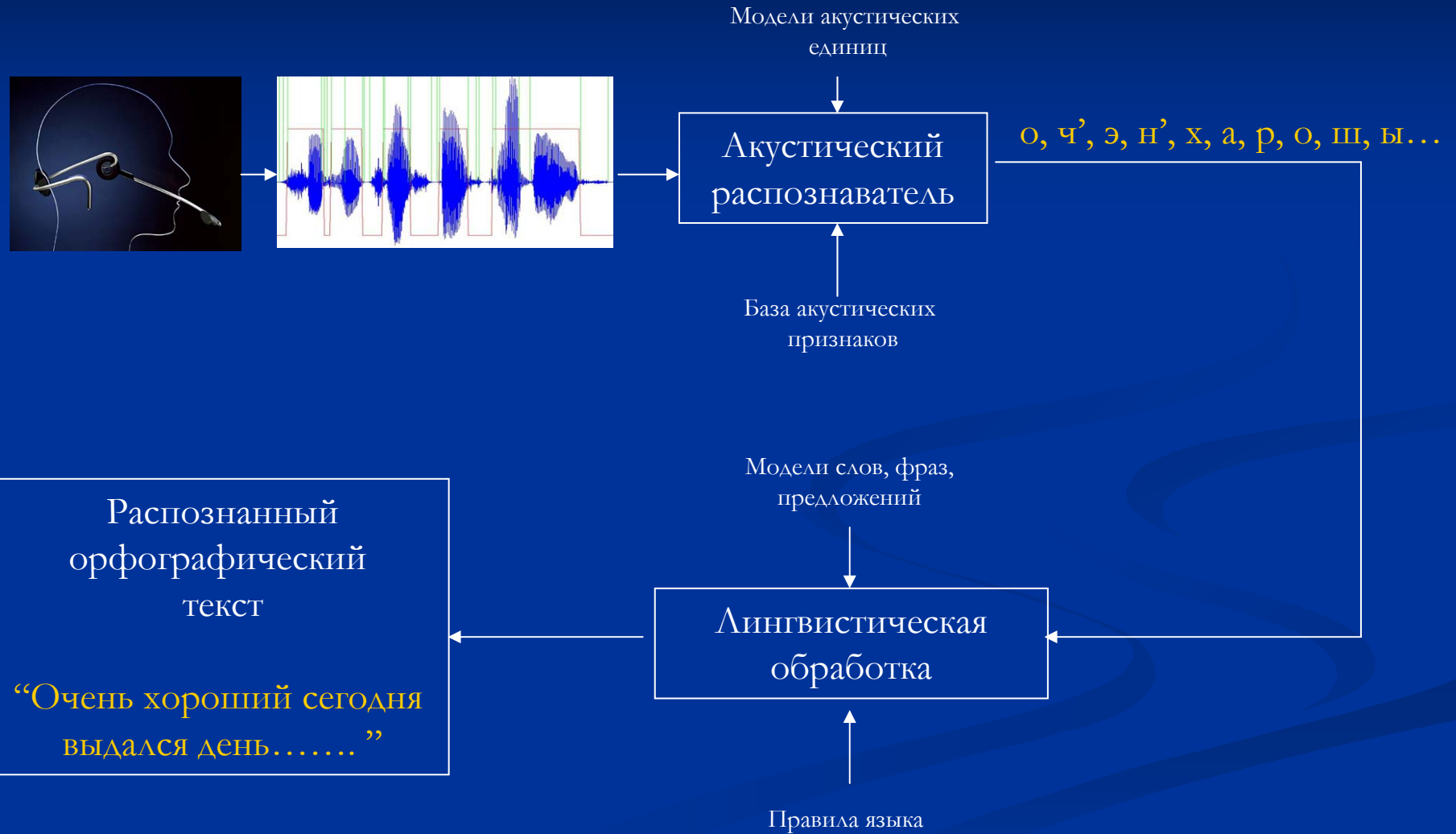
Временная  
изменчивость



В разные моменты времени одни и те же речевые фрагменты имеют различную длительность

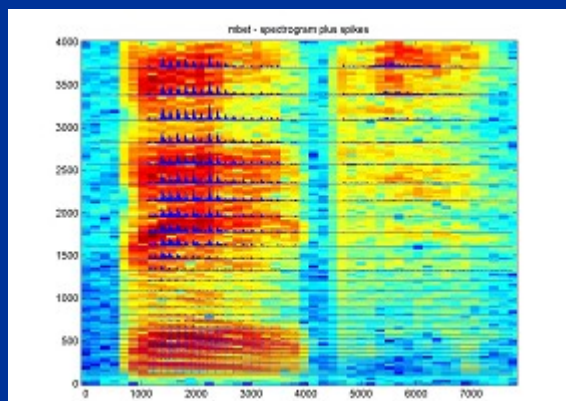
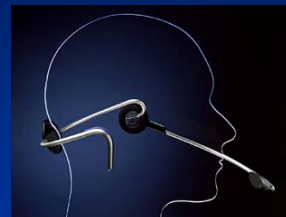
**Нельзя два раза войти в одну и ту же воду!**

# Основная схема систем распознавания речи



# Акустический распознаватель

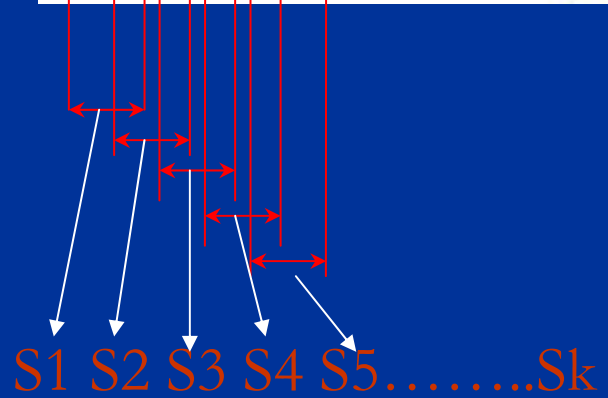
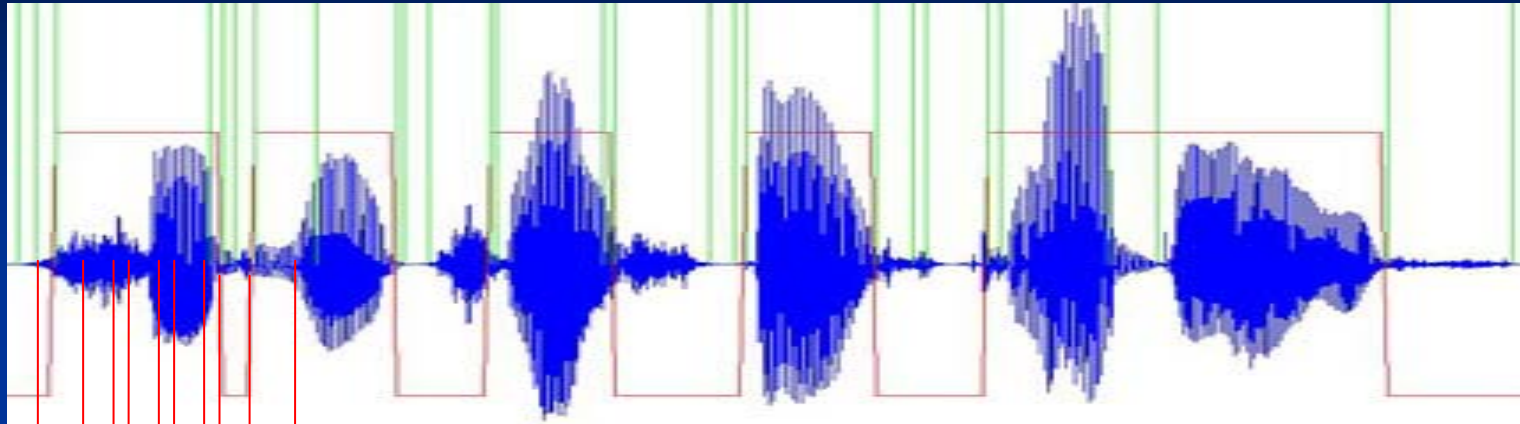
Основная цель- преобразование  
акустического сигнала в  
последовательность акустических  
единиц, соответствующих  
содержанию исходного сигнала



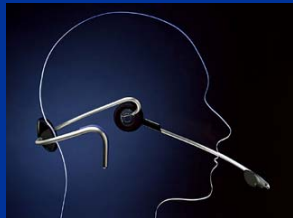
## Этапы акустической обработки

- Сегментация
- Выделение признаков
- Моделирование акустических единиц

# Сегментация



Формируется последовательность  
перекрывающихся  
участков исходного сигнала по  
методике “кадр-за-кадром”

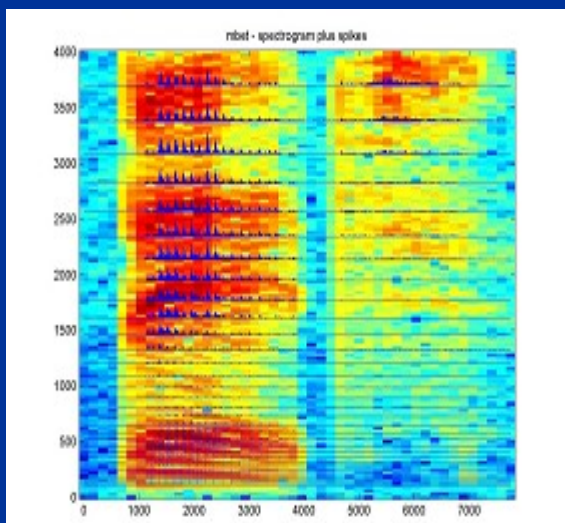


# Выделение признаков

Основная цель- сопоставление каждому речевому сегменту вектора признаков

Требования к вектору признаков:

- информативность
- адекватность
- устойчивость
- доступность вычисления



Сглаживание сигнала

Спектральный анализ

Психоакустический анализ

Кепстральный анализ

Выделение формант

Дельта-параметры

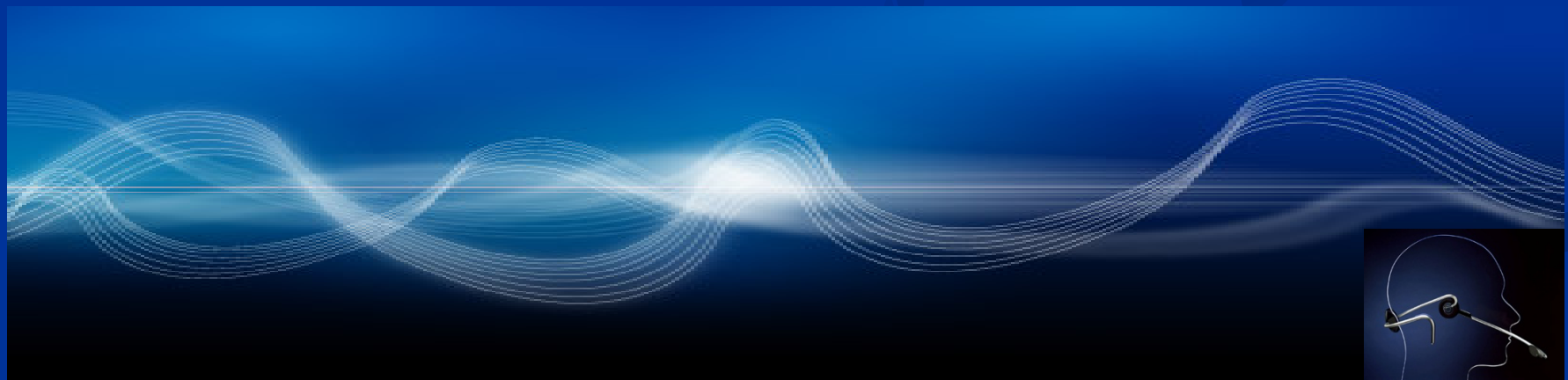
$V_1 V_2 V_3 V_4 V_5 \dots V_k$

**КОНКРЕТНЫЙ ВЫБОР ВЕКТОРА  
ПРИЗНАКОВ ЗАВИСИТ ОТ  
РЕШАЕМОЙ ЗАДАЧИ (ЯЗЫКА,  
УСЛОВИЙ ЗАПИСИ, Т.Д.)**



# Моделирование акустических единиц

Необходимо сопоставить  
последовательности  
векторов признаков  
**V1 V2 V3 V4 V5.....Vk**  
последовательность  
акустических единиц  
**W1, W2, W3.....Wm**



# Модели акустических единиц

## Непараметрические модели

Динамическое  
программирование



## Параметрические модели

Скрытые марковские модели

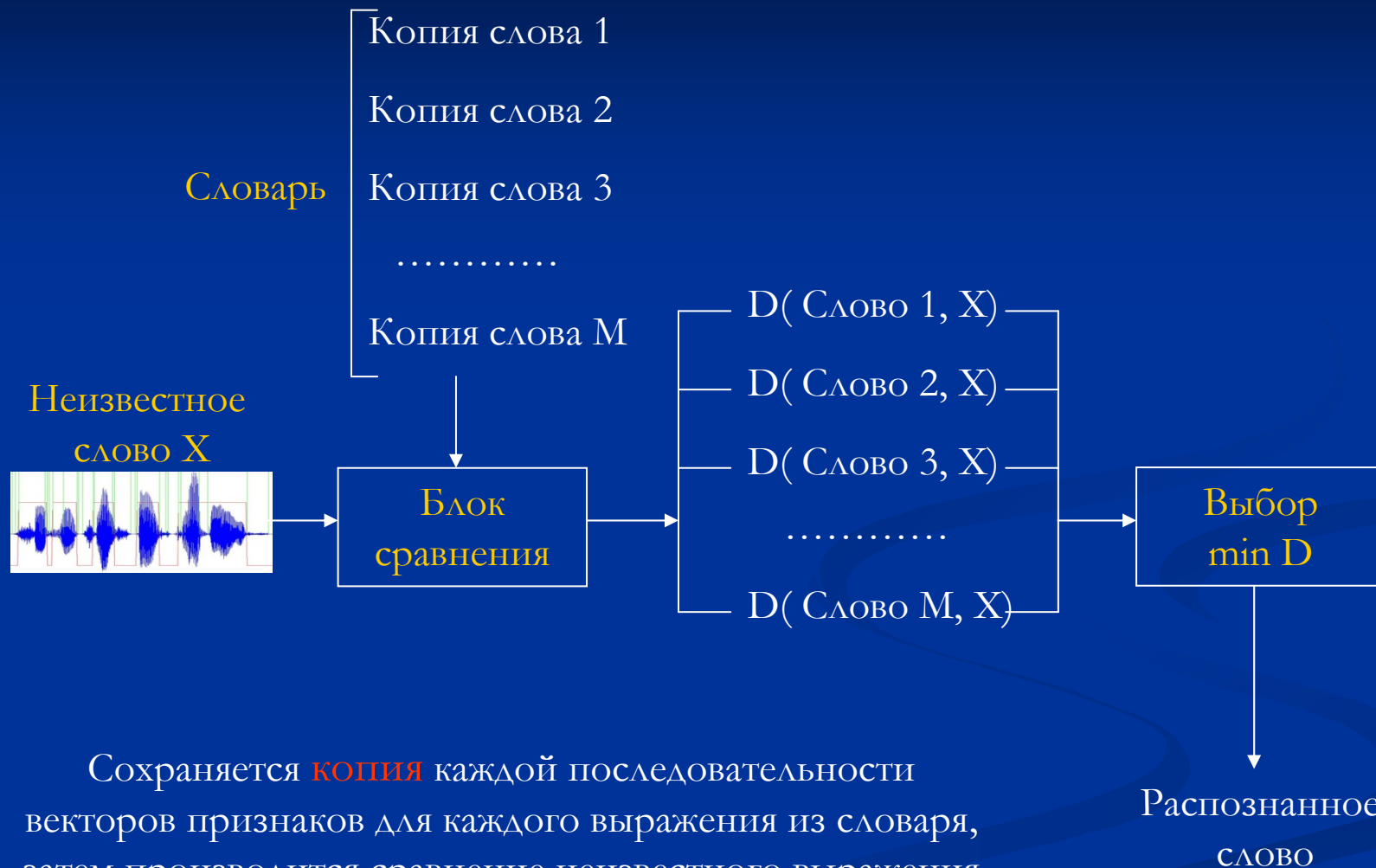
Нейронные сети

Машина на опорных векторах

Генетические алгоритмы

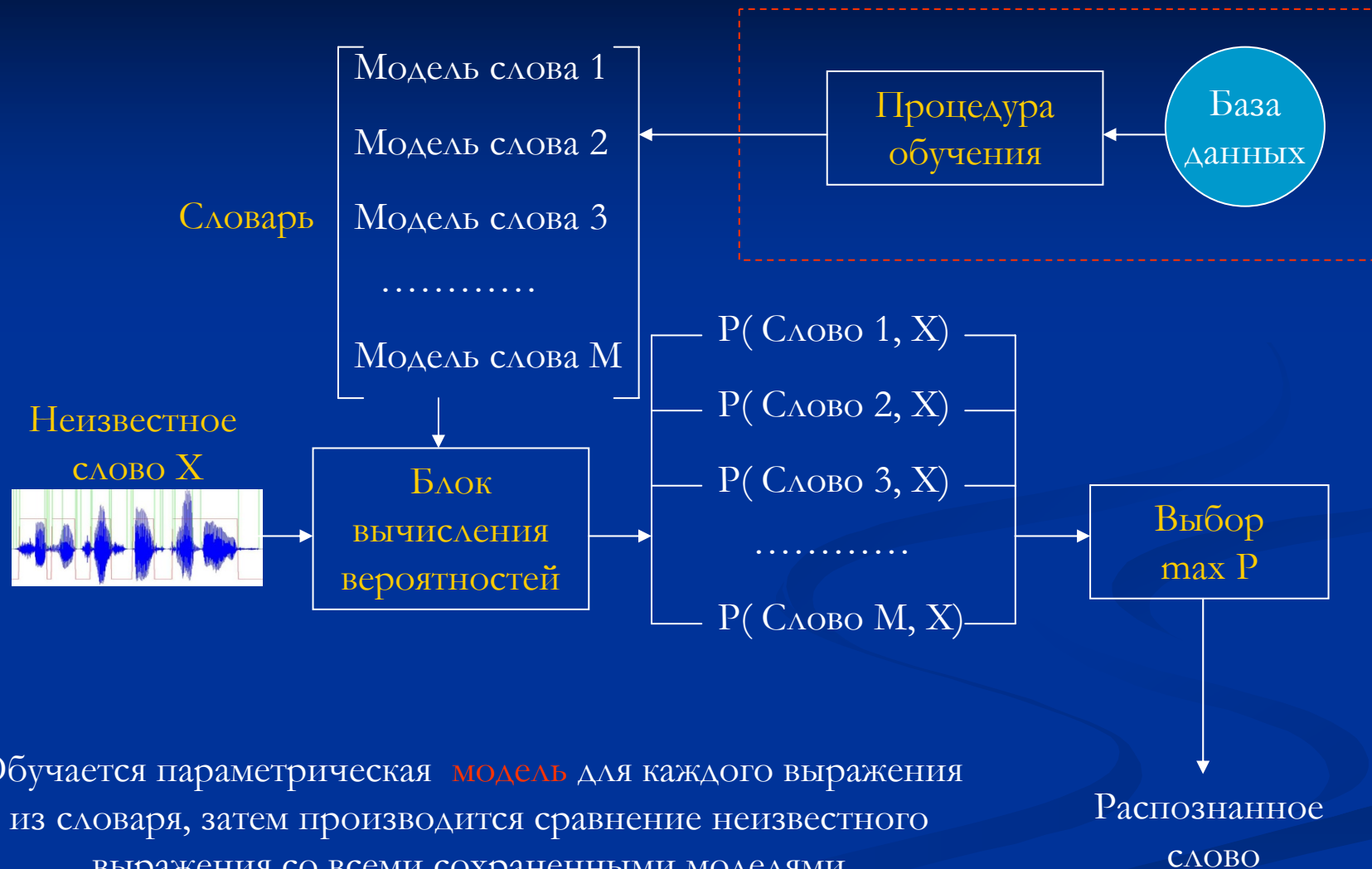
.....

# Непараметрические модели



Сохраняется **копия** каждой последовательности векторов признаков для каждого выражения из словаря, затем производится сравнение неизвестного выражения со всеми сохраненными копиями

# Параметрические модели



Обучается параметрическая **модель** для каждого выражения из словаря, затем производится сравнение неизвестного выражения со всеми сохраненными моделями

# Требования к базе данных

Множество дикторов, участвующих в записи, разного пола и возраста (не менее 50-ти)

Фонетически полный набор выражений для записи

Множество записей для одного диктора (не менее 10-ти для каждого выражения )

Качественная запись (хороший микрофон, звуковая карта, студия)

Качественная сегментация на акустические единицы (с точностью не ниже 3 мс)



# Использование скрытых моделей Маркова (СММ или НММ)

- При анализе естественного языка первым шагом необходимо определить: к какой части речи относится каждое из слов в предложении. В английском языке задача на этом этапе называется Part-Of-Speech tagging.
- Рассмотрим предложение на английском языке: «The can will rust».
  - the –определенный артикль или частица «тем»;
  - can – может одновременно являться и модальным глаголом, и существительным, и глаголом;
  - will – модальный глагол, существительное и глагол;
  - rust – существительное или глагол.
- В статистическом подходе необходимо построить таблицу вероятностей использования слов в каждом грамматическом значении. Эту задачу можно решить на основе тестовых текстов, проанализированных вручную. И сразу можно выделить одну из проблем: слово «can» в большинстве случаев используется в качестве глагола, но иногда оно может являться и существительным.

# Использование скрытых моделей Маркова (2)

- Учитывая этот факт, была создана модель, принимающая во внимание тот факт, что после артикля следует прилагательное или существительное:

$$\operatorname{argmax}_{t1..tn} \prod_{t=1}^n p(w_i | t_i) p(t_i | t_{i-1})$$

- Где:

$t$  – тэг (существительное, прилагательное и т.д.)

$w$  – слово в тексте (rust, can ...)

$p(w | t)$  – вероятность того, что слово  $w$  соответствует тэгу  $t$

$p(t1 | t2)$  – вероятность того, что  $t1$  идет после  $t2$

# Использование скрытых моделей Маркова (3)

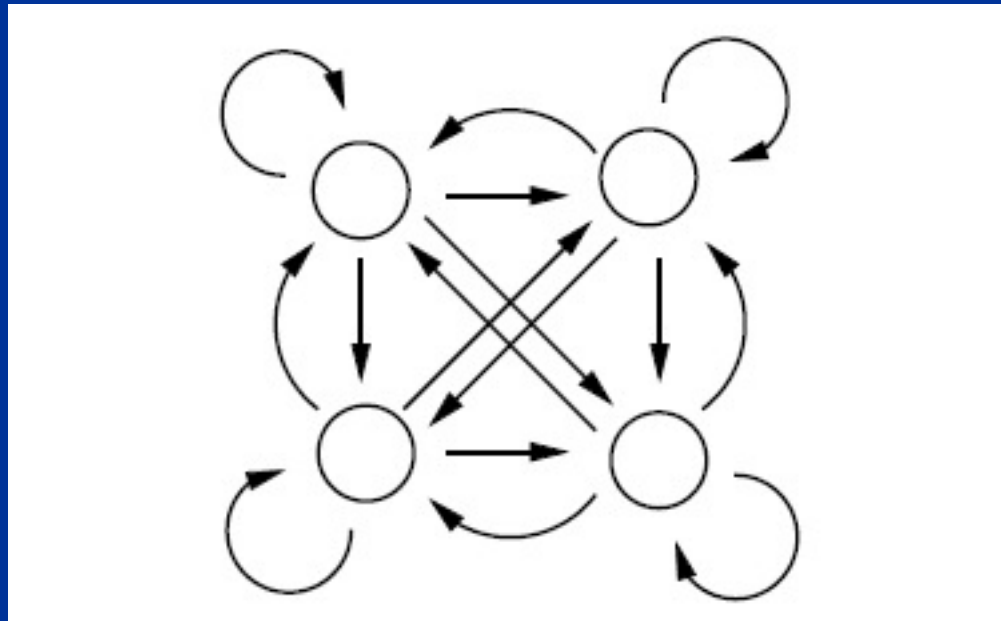
- Из формулы видно, что мы пытаемся подобрать тэги так, чтобы слово подходило тэгу, и тэг подходил предыдущему тэгу. Данный метод позволяет определить, что «can» выступает в роли существительного, а не модального глагола.



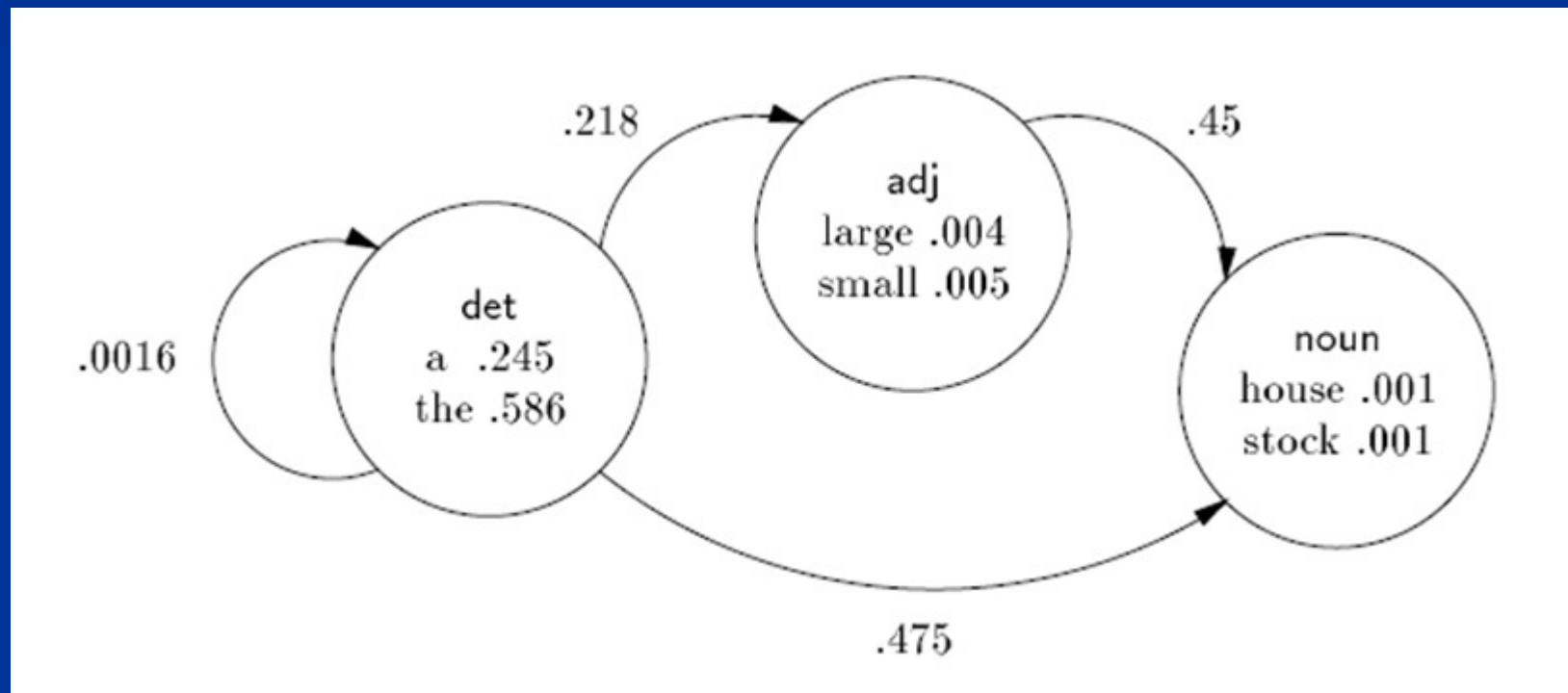


# Использование скрытых моделей Маркова (4)

Эта статистическая модель может быть описана как эргодическая СММ:



# Использование скрытых моделей Маркова (5) Пример СММ



# N-граммные модели

- Основаны на предположении, что вероятность употребления очередного слова в предложении зависит только от  $n-1$  слов.
- Сегодня наиболее популярные биграммные и триграммные модели языка. Поиск в таких моделях происходит по большой таблице (корпусу). Несмотря на быстро работающий алгоритм, такие модели не способны уловить семантические и синтаксические связи, если зависимые слова находятся на расстоянии 5 слов друг от друга. Использование же  $n$ -граммных моделей, где  $n$  больше чем 5, требует огромных мощностей.

# Триграммная модель

- Рассмотрим предложение «I want to go home».
- Вероятность этого предложения можно вычислить от счета частоты n-грамм (в этом примере n=3):

$$P(I, \text{ want, to, go, home}) \approx P(I) * P(\text{want} | I) * P(\text{to} | I, \text{ want}) * P(\text{go} | \text{want, to}) * P(\text{home} | \text{to, go})$$

# Дальнодействующая триграммная модель

- Анализ ведется не только по двум предшествующим словам, а по любой паре слов, находящихся рядом.
- Такая триграммная модель может пропускать малоинформативные слова, тем самым улучшая предсказуемость сочетаемости в модели.

# Основные проблемы распознавания



Очень хороший день

Акустический распознаватель

о, ч', э, і, н', у, х, і, р, у, о, ш, ы, д' | н' ...

Ошибки

Замены

Пропуски

Вставки

Причины ошибок???

## Основные причины ошибок

- Ошибочное и нечеткое произношение
- Плохая дикция говорящего
- Высокий уровень посторонних шумов
- Недостаточное или плохое обучение моделей
- Большая схожесть слов словаря
- Произнесение с разной интонацией
- Акцент и диалект говорящего

