Системы распознавания речи: базовые принципы и алгоритмы

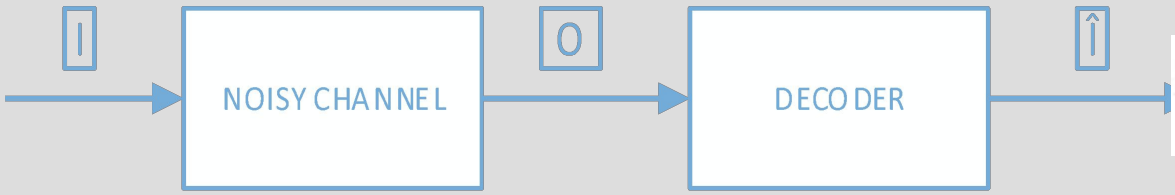
Demystifying

Осень, 2013

Задача распознавания речи

- Система распознавания речи — устройство, которое осуществляет автоматическую трансляцию речи в текст.
- Оно может мыслиться как «печатная машинка», которая осуществляет трансляцию, после чего транслированный текст отображается на экране рабочей станции [1]
- Системы различаются по возможности распознавания слитной речи — слитная речь vs. изолированные слова
- По объему доступного словаря — большой vs. ограниченный
- Способности к распознаванию разговорной речи — разговорная vs. дикторская
- Устойчивости к шуму
- Количеству и качеству дикторов — зависимые от диктора vs. независимые от диктора [2]
- Large-vocabulary continuous speech recognition

Зашумленный канал



$$\check{I} = \arg\max_i P(I|O) = \arg\max_i P(I)P(O|I)$$

Применение	Входная последовательность	Выходная последовательность
Машинный перевод	Последовательность *w* на языке **L1**	Последовательность слов *w′* на языке **L2**
Оптическое распознавание символов	Сканированный текст — цепочка символов *w*	Текст с ошибками распознавания
POS-тэггинг	Последовательность частеречных тегов	Последовательность слов
Распознавание речи	Последовательность слов	Речевой сигнал

Sequence-labeling и обработка естественного языка

Рассмотрим предложение:

Secretariat/NNP is/BEZ expected/VBN to/TO race/?? tomorrow/

Определим часть речи слова race: NN vs. VB

Какие признаки позволяют нам сделать вывод о части речи слова race?

- Лексема race
- Часть речи предыдущего слова — TO
- В написании слова отсутствует -ING
- В написании слова отсутствует большая буква

Наличие/отсутствие каждого признака положительно либо отрицательно сказывается на каждом из решений.
Кроме того, признаки не одинаково информативны

Простейшая модель вывода

- Выберем набор релевантных признаков
- Припишем им числовые значения — «веса». Веса могут быть отрицательными, если соответствующие признаки снижают шансы кандидата на победу
- Для каждого из кандидатов найдем суммарный вес признаков. Побеждает кандидат, набравший больший вес

- f1 = 1 iff word = 'race' & NN
- f2 = 1 iff t_{i-1} = TO & VB
- f3 = 1 iff suffix=-ING & VB
- f4 = 1 iff lower_case(word) = 'race' & VB
- f5 = 1 1 iff word = 'race' & VB
- f6 = 1 iff t_{i-1} = TO & NN

- w1 = 8
- w2 = 8
- w3 = 7
- w4 = 0.1
- w5 = 1
- w6 = -13

P(NN) = 8 - 13 = -5

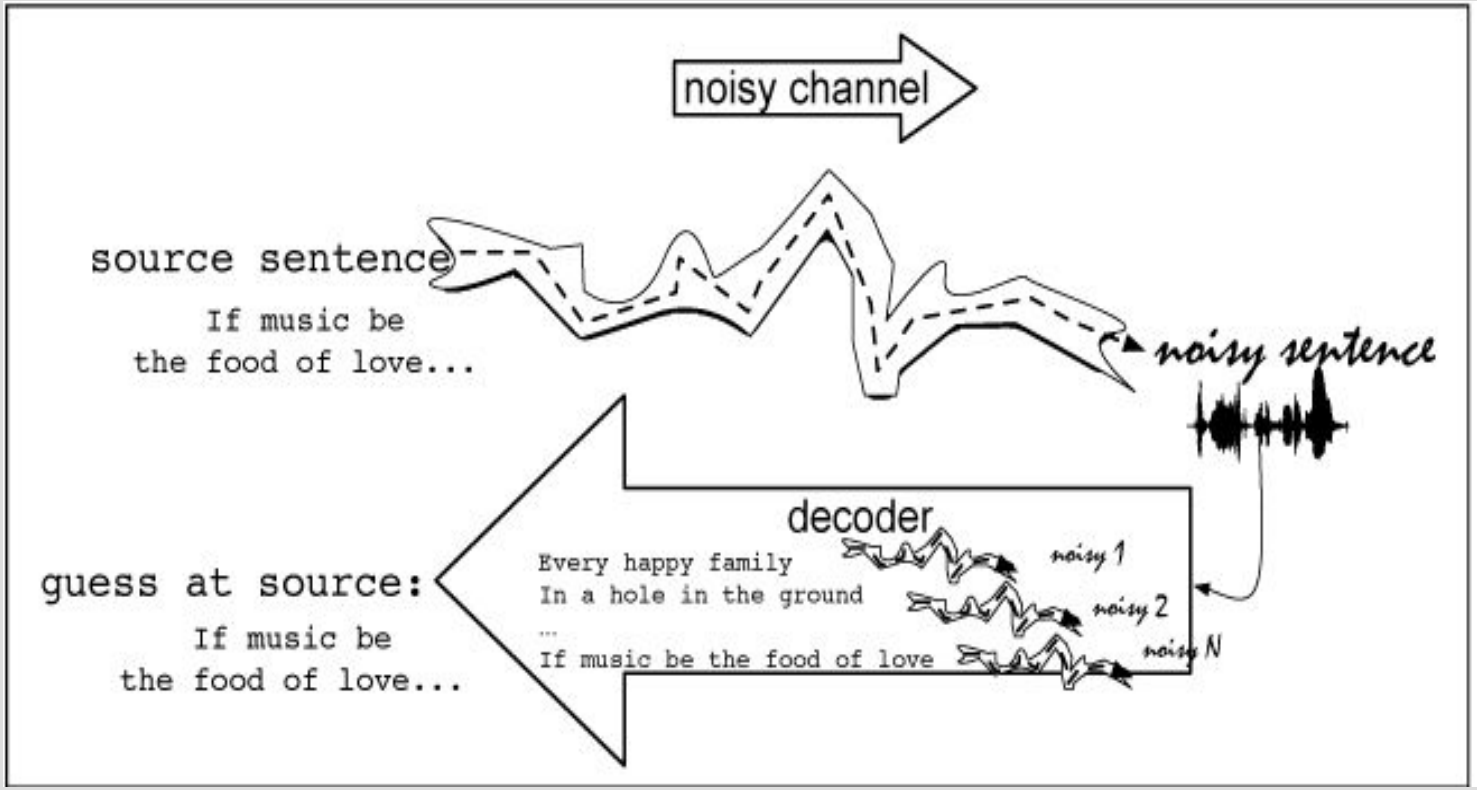
P(VB) = 8 + 0.1 + 1 = 9.1

Но откуда брать значения весов?!

Обработка естественного языка и декодирование

- Большое количество задач NLP может быть так или иначе сведено к проставлению меток
- Морфологический анализ
- Выделение именованных сущностей
- Разрешение анафоры
- Поверхностный синтаксический разбор
- И даже...
- …
- ???
- Распознавание речи
- Ключевой момент - выбор релевантных признаков и инвентаря скрытых и наблюдаемых состояний

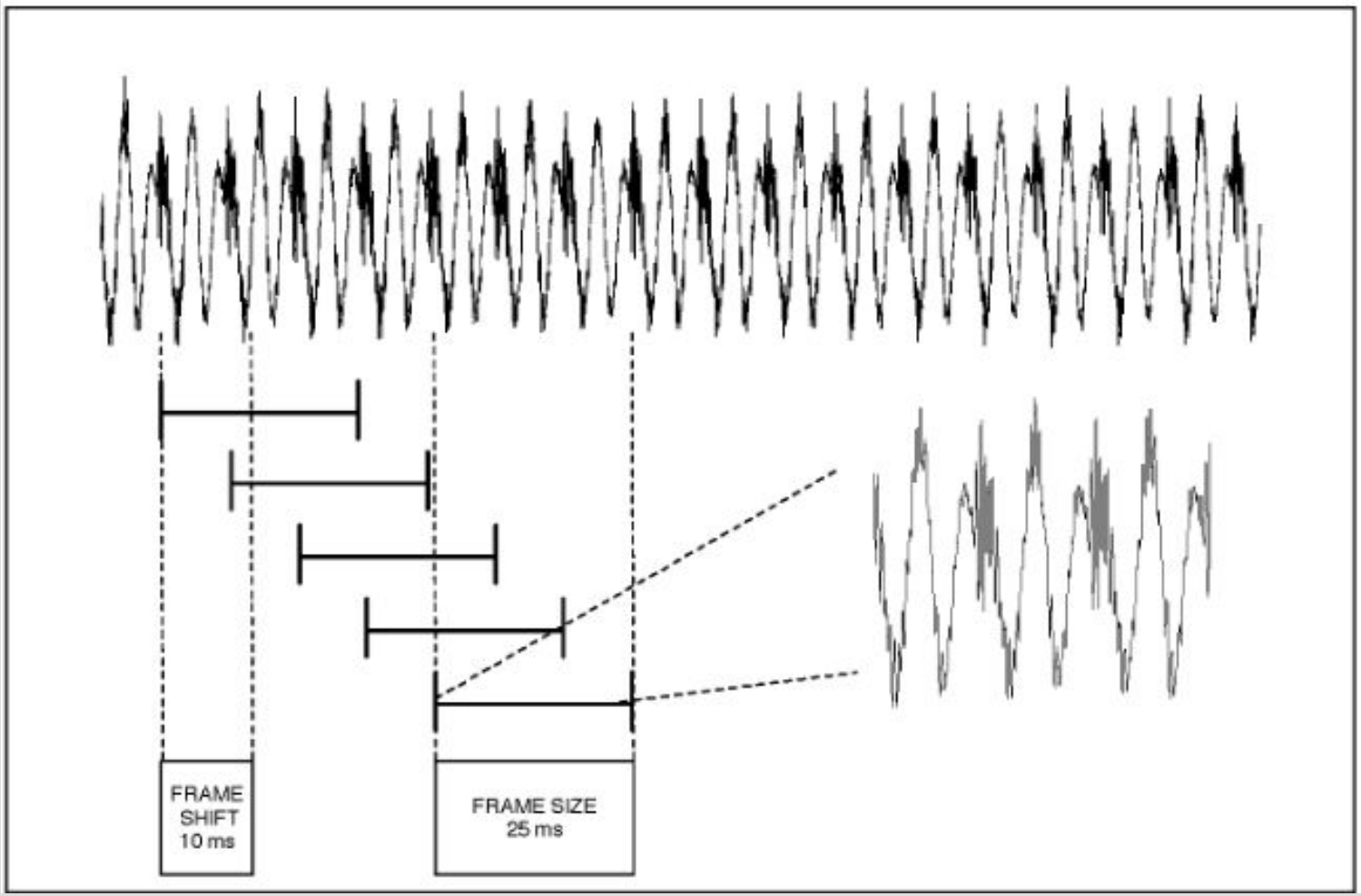
Модель зашумленного канала в распознавании речи



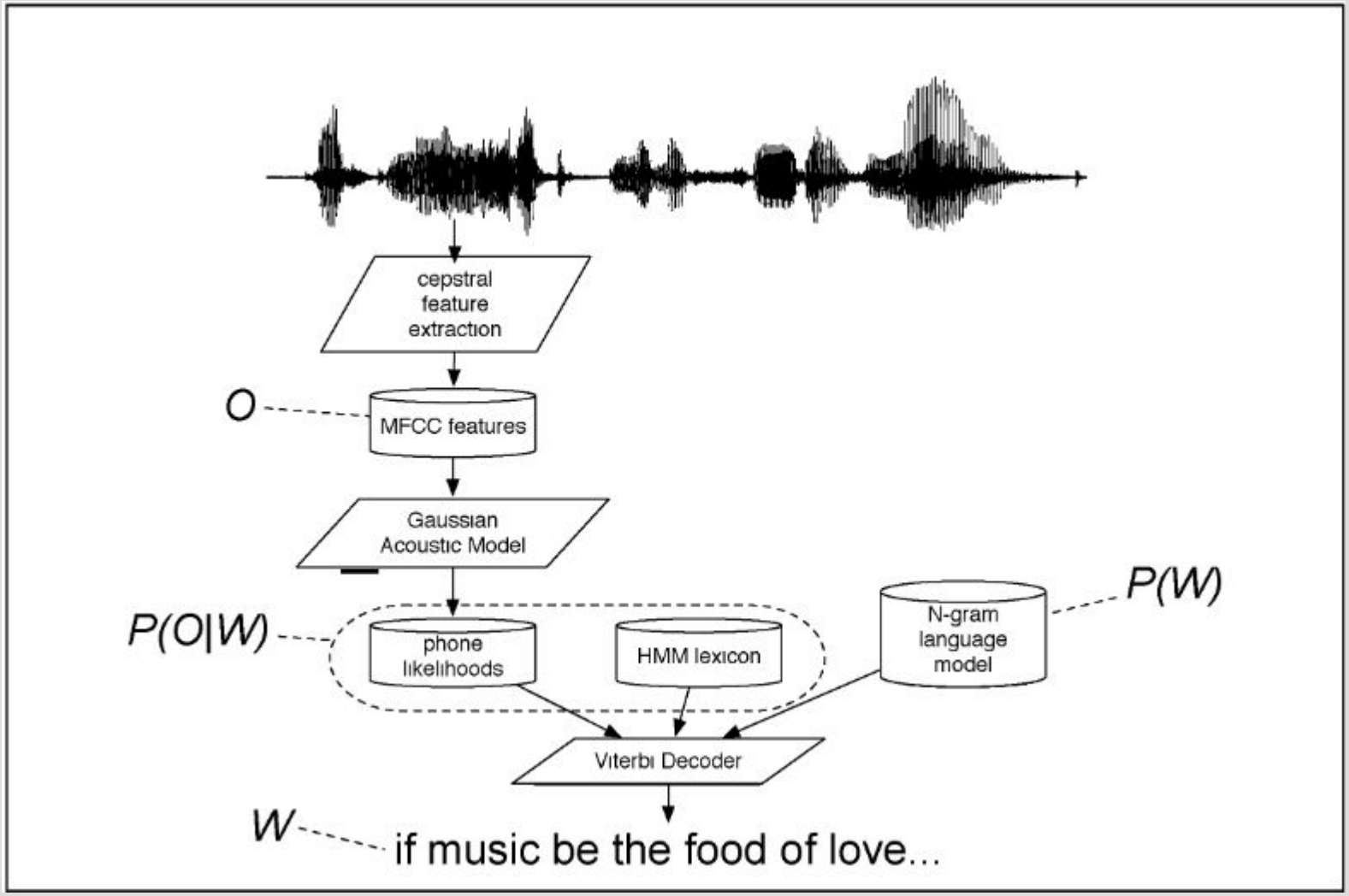
Sequence-labelling и распознавание речи

- Скрытые состояния — фонемы/аллофоны
- Наблюдаемые состояния — векторы акустических признаков
- В основе алгоритмов декодирования лежит аппарат скрытых марковских моделей
- Выделение акустических признаков базируется на дискретном преобразовании Фурье
- Непрерывный речевой сигнал представляется в виде дискретной последовательности векторов акустических признаков
- Для анализа слитной речи необходима также **модель языка** — вероятность появления данной цепочки слов в анализируемом языке
-

Представление речевого сигнала в виде дискретной последовательности



Общая схема системы распознавания речи



Основные разделы курса

- Выделение акустических признаков
  - Дискретное преобразование Фурье и его разновидности
  - MFCC (mel-frequency cepstral coefficients)
- Алгоритмы распознавания
  - Скрытая марковская модель
  - Алгоритм Витерби декодирования в СММ
  - Алгоритм Баума-Велша обучения СММ
- Языковая модель
- Контекстная вариативность звуков и аллофоны
- Распознавание речи за пределами СММ

Коротко о MATLAB

- MATLAB — популярный программный пакет для инженерных вычислений и моделирования
- Поддерживает встроенный язык программирования
- Операции с векторами и матрицами поддерживаются на уровне синтаксиса
- Имеется множество готовых функций и алгоритмов для работы с изображениями и цифровыми сигналами
- Имеет свободный аналог — GNU Octave
- https://class.coursera.org/pgm-003/lecture/index ML-class Ocatve Tutorial

Литература

1. F. Jelinek. Statistical methods for speech recognition
2. D. Jurafsky, J. Martin. Speech and language processing