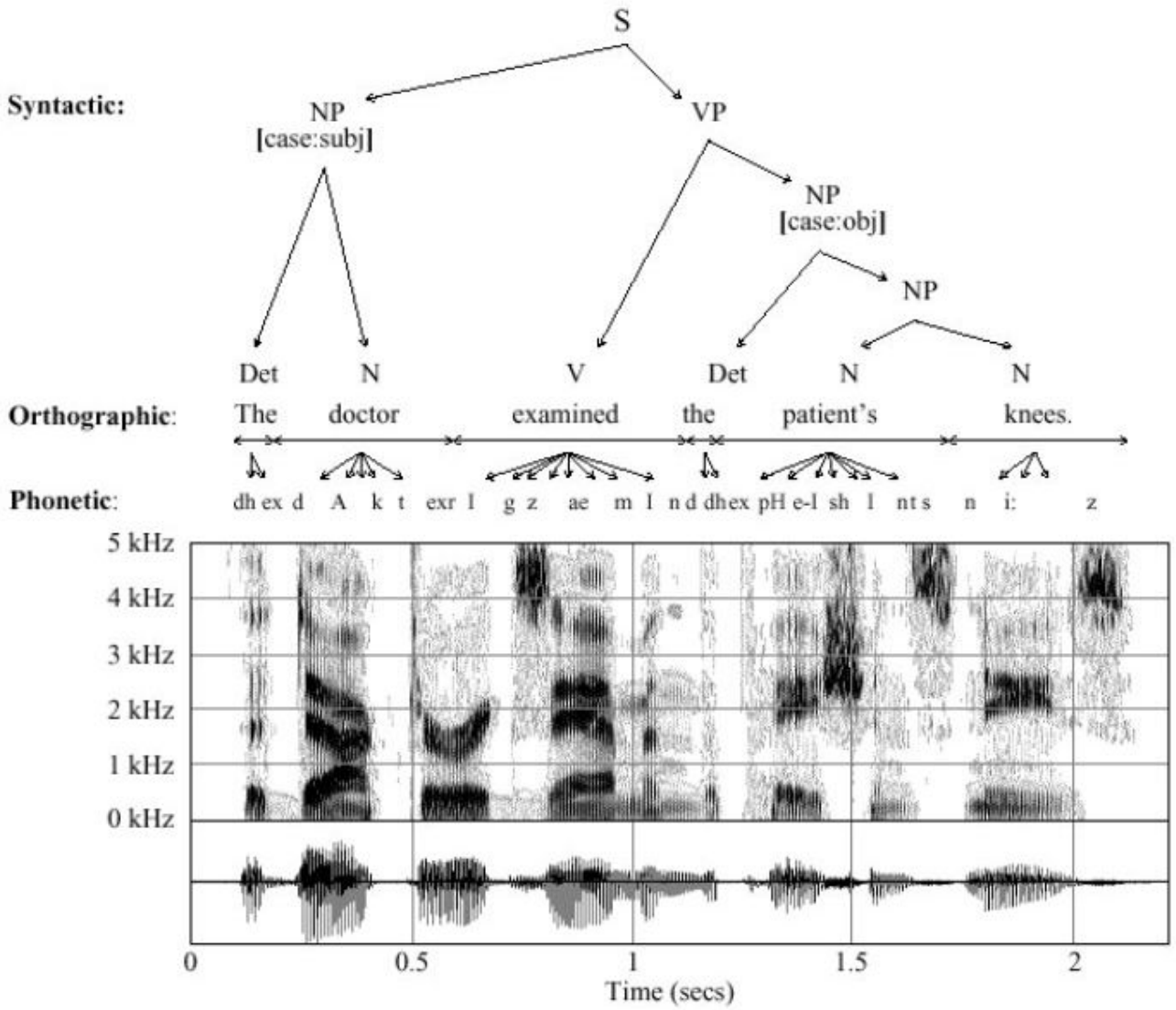
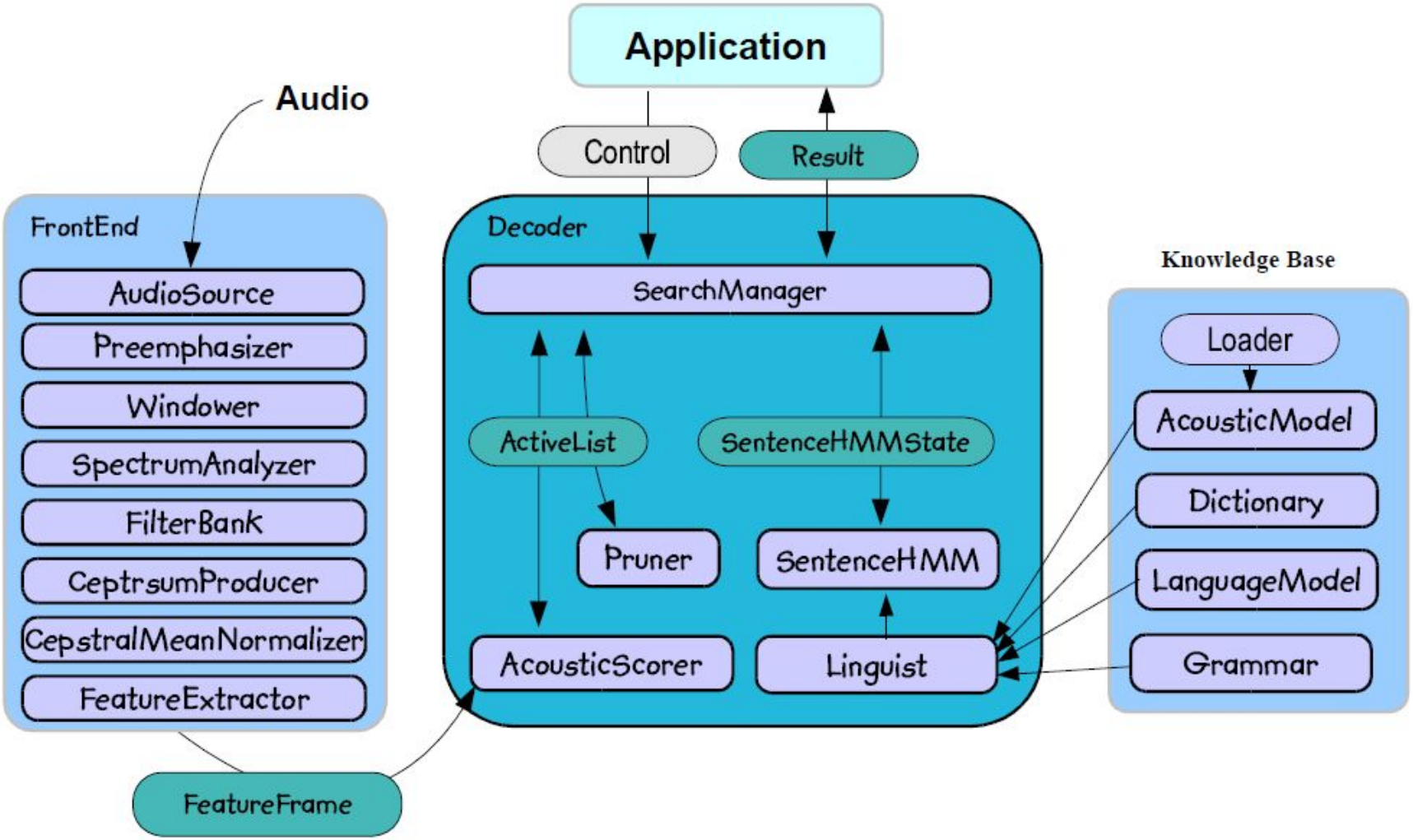
Распознавание речи

А. Р. Нехаев
А. А. Жижелев

Пример спектрограммы



Архитектура Sphinx



Органы речи

- дыхательные органы (лёгкие, бронхи, дыхательное горло)
  - проход воздушной струи через произносительный аппарат
- гортань (голосовые связки)
  - образование голоса
- полости глотки, рта и носа
  - образование специфических свойств отдельных звуков речи

Представление речи

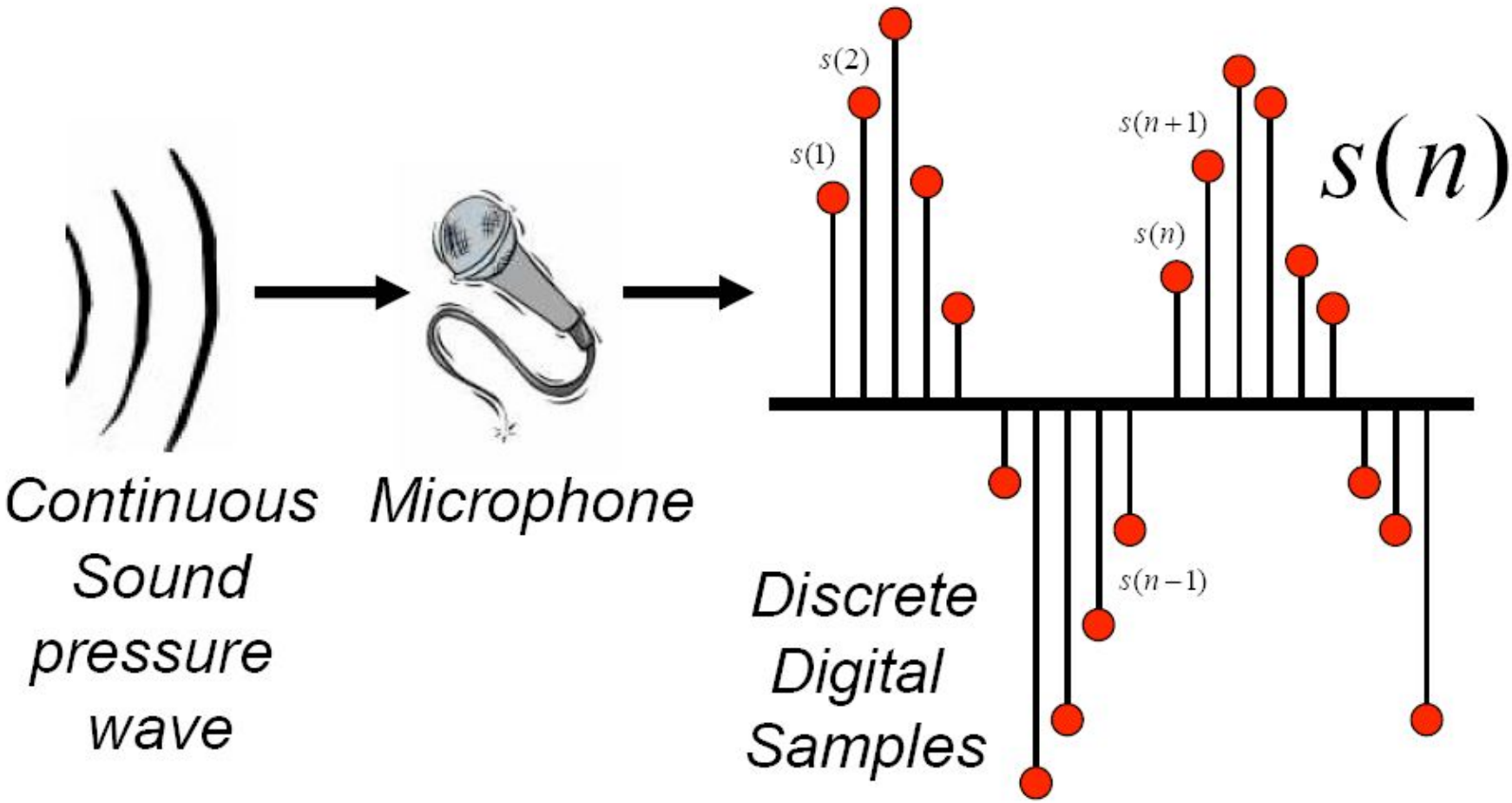
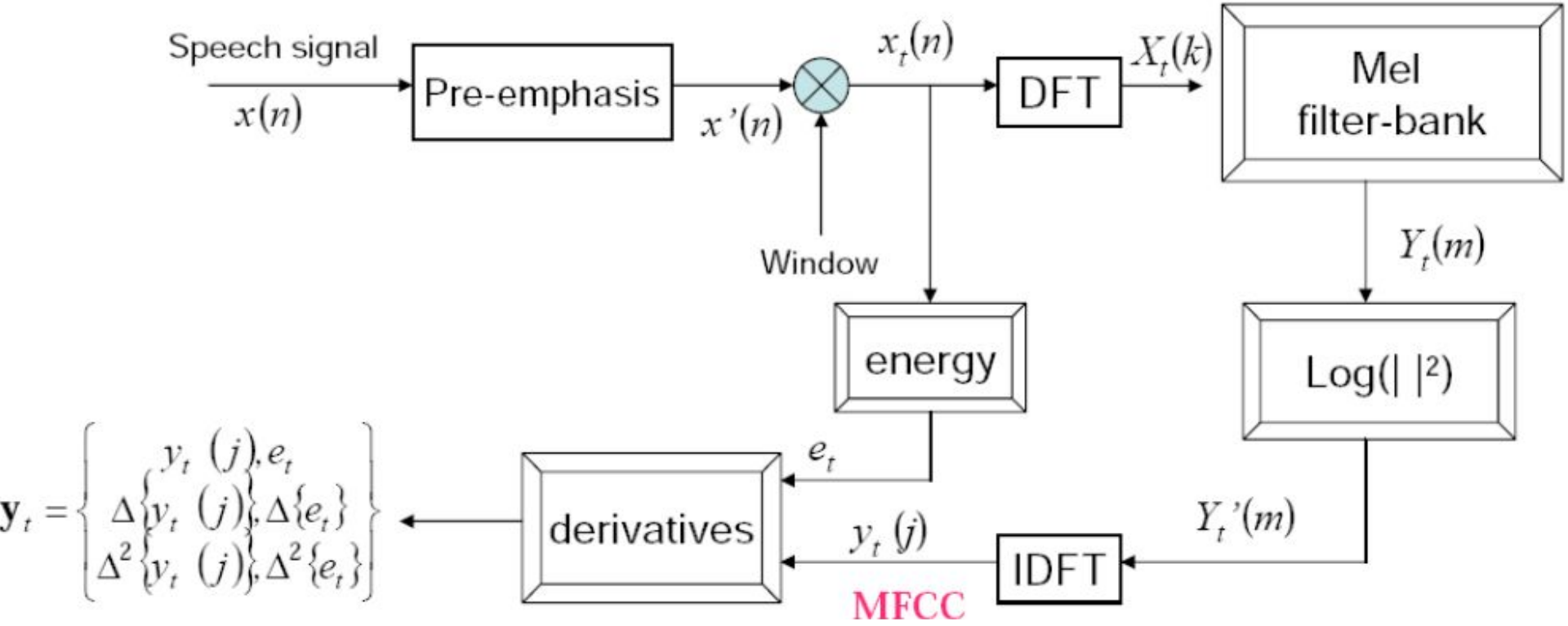
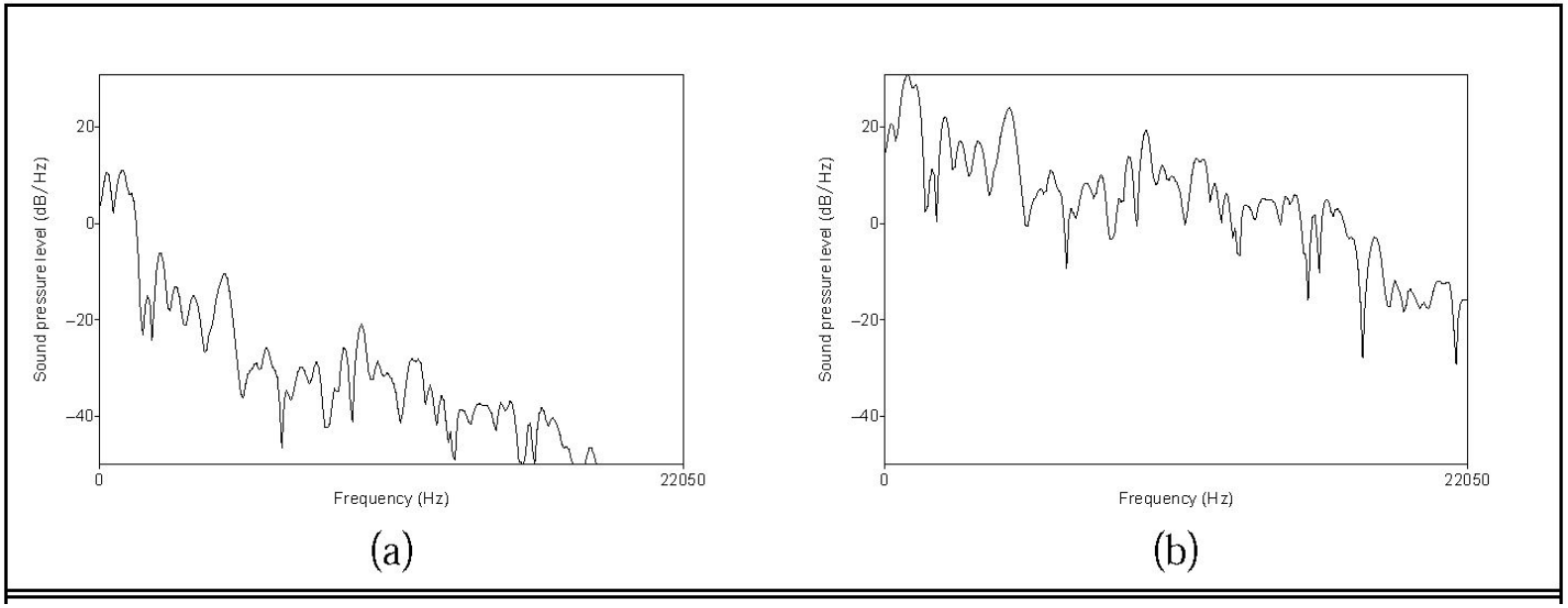


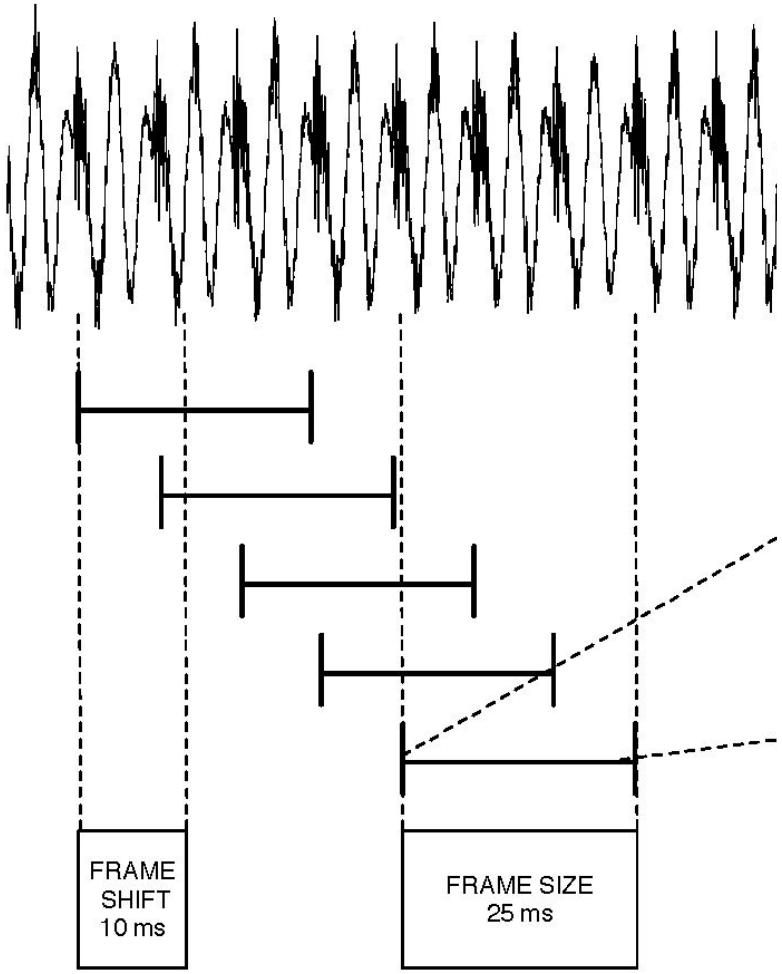
Схема формирования вектора признаков

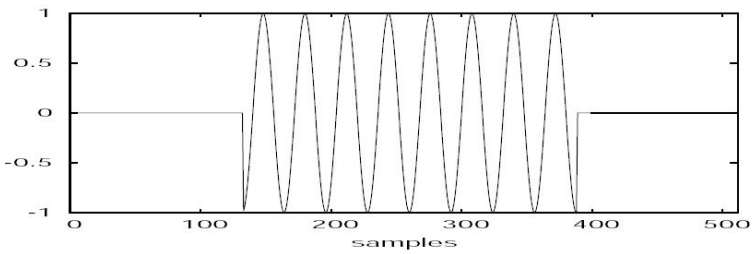


Предусиление

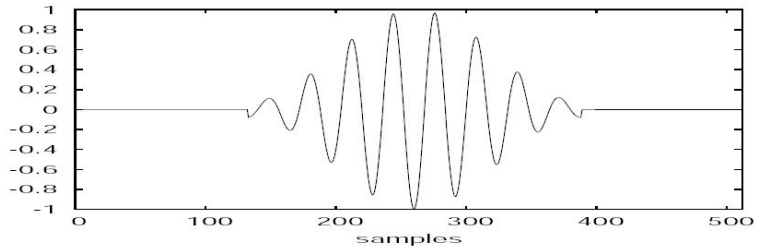


Разбиение на фреймы



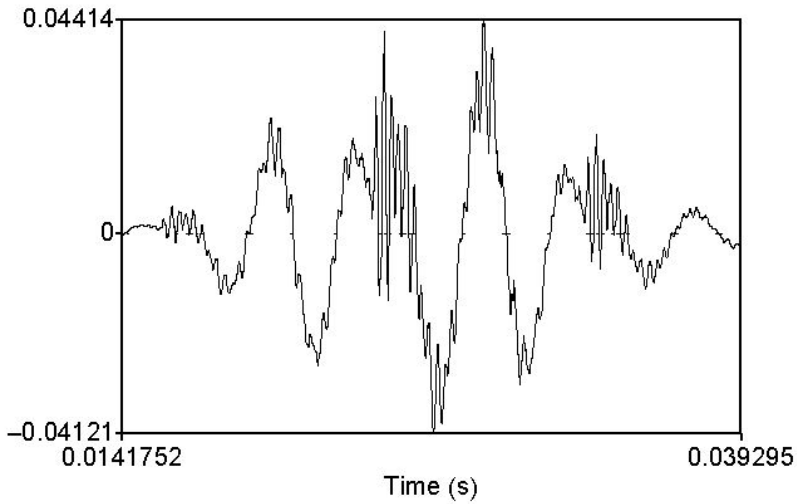


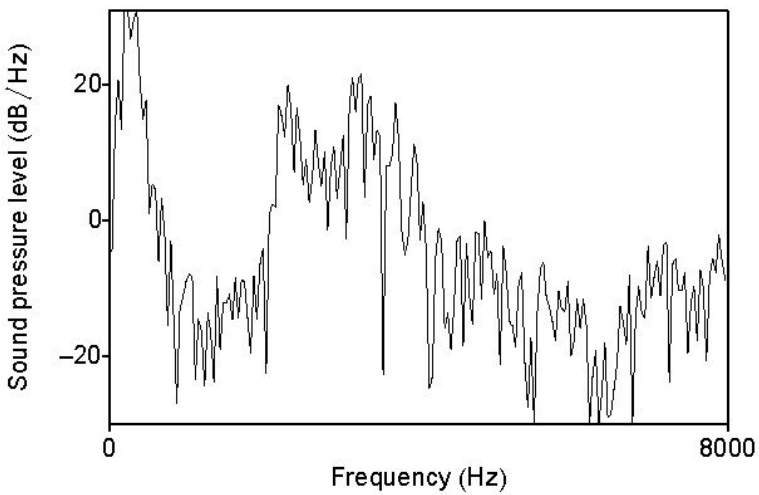
(a) Rectangular window



(c) Hamming window

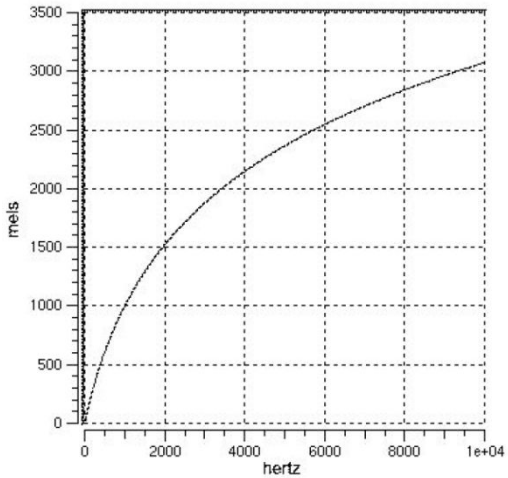
Преобразование Фурье

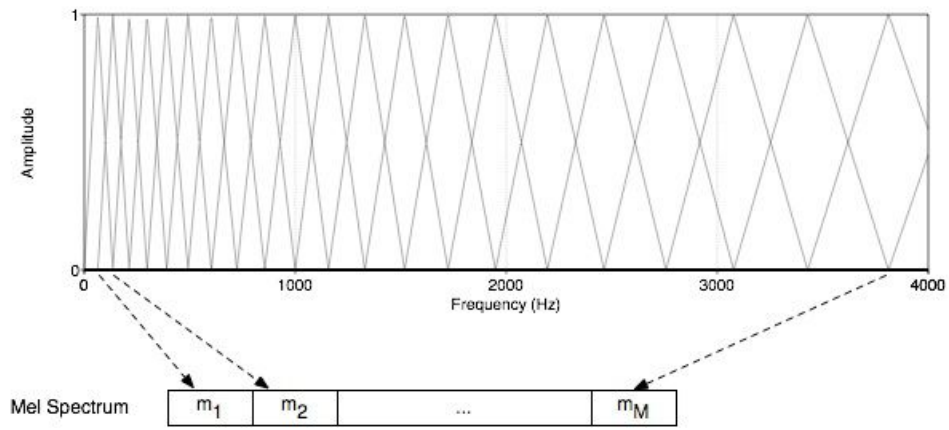


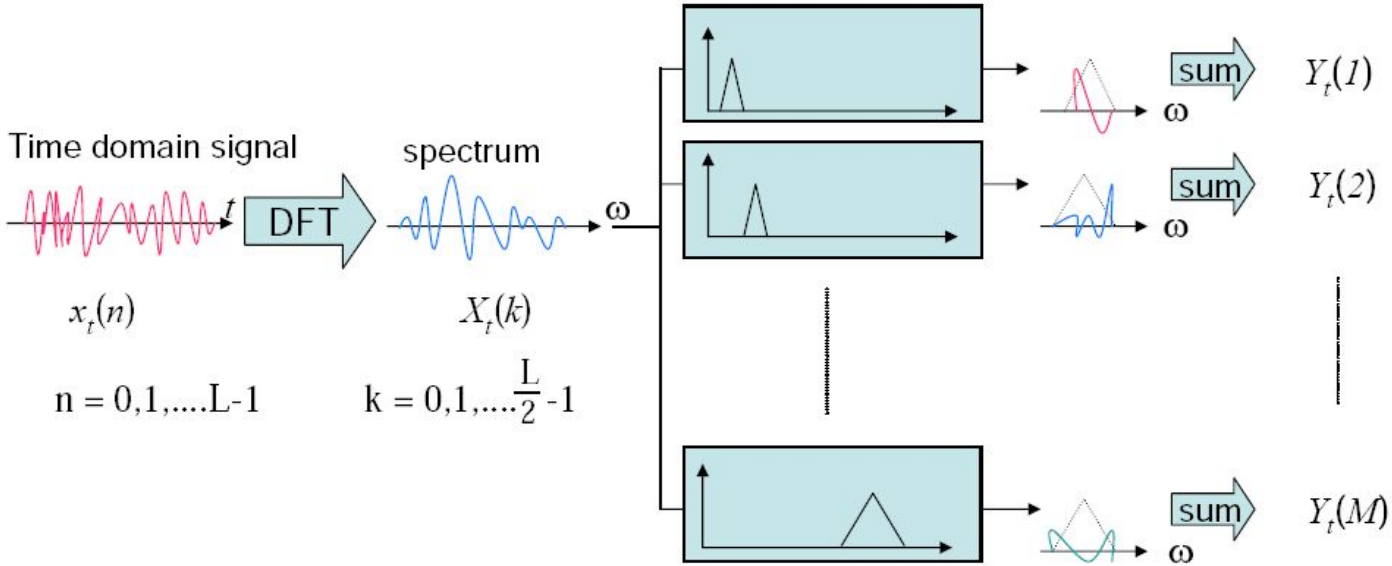


$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\frac{\pi}{N}kn}$$

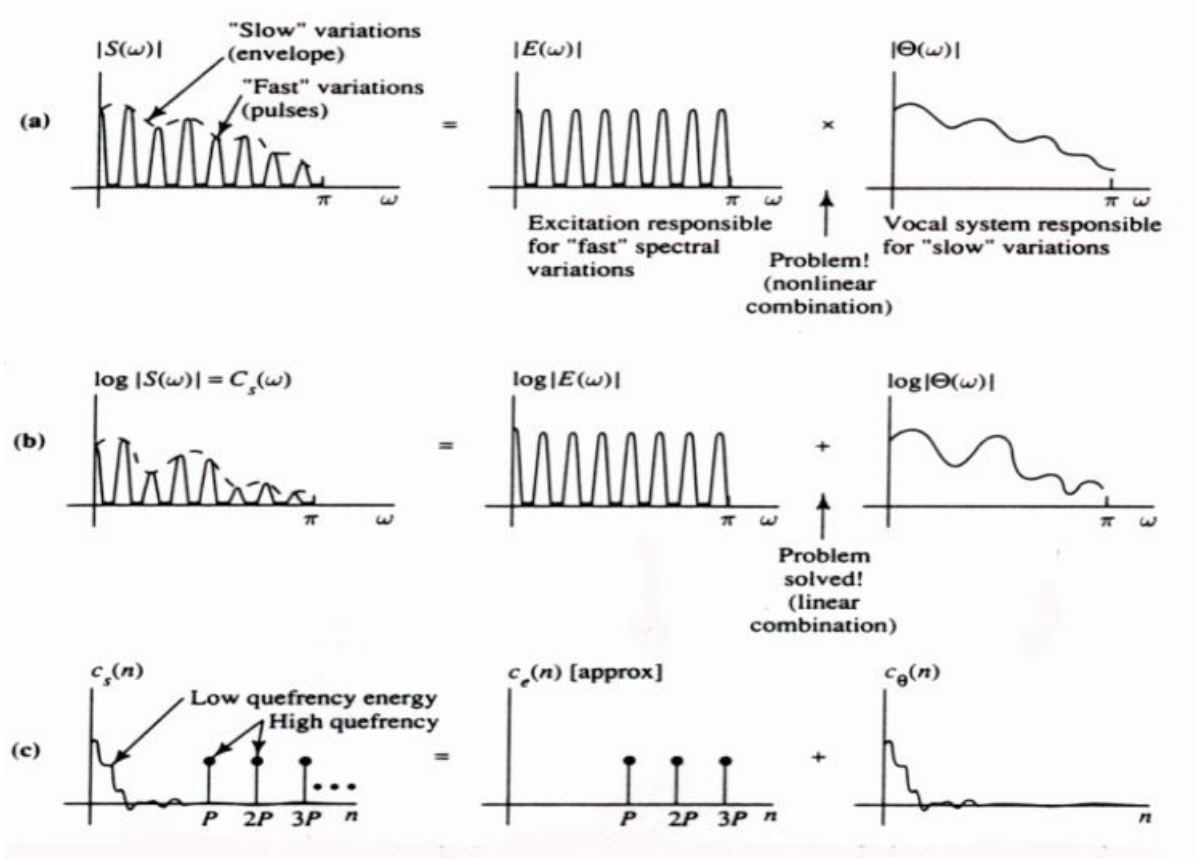
Mel-filter Bank



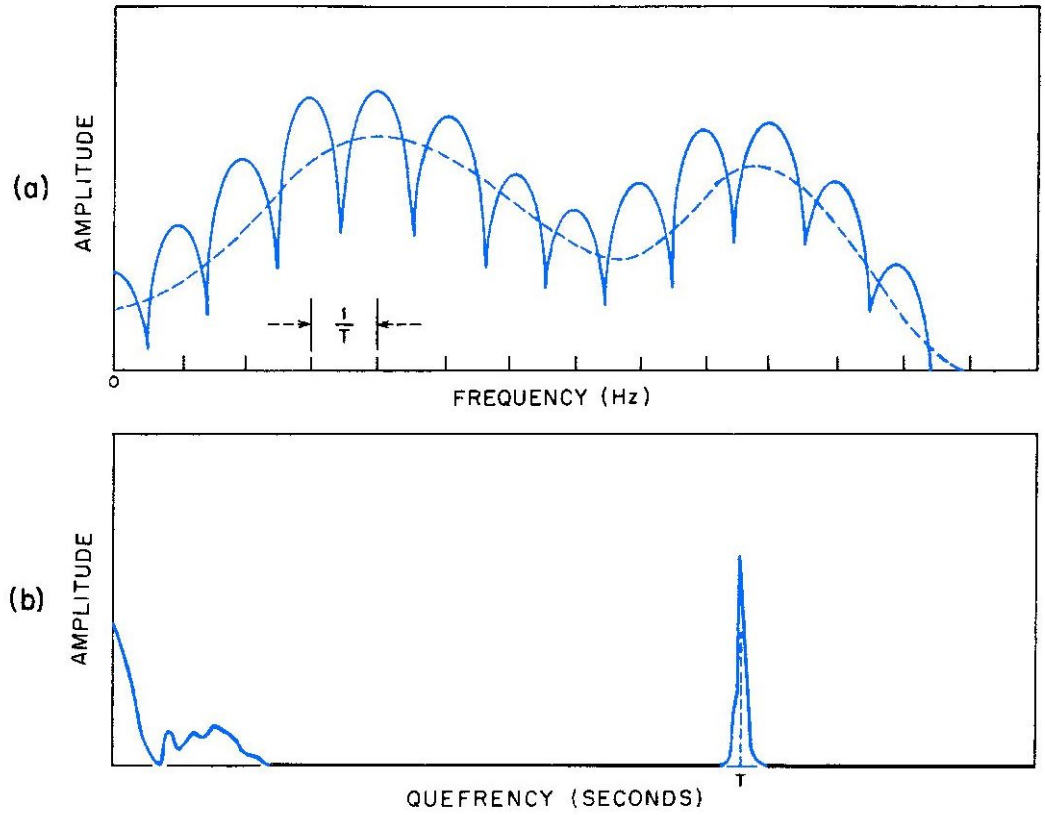




Кепстр

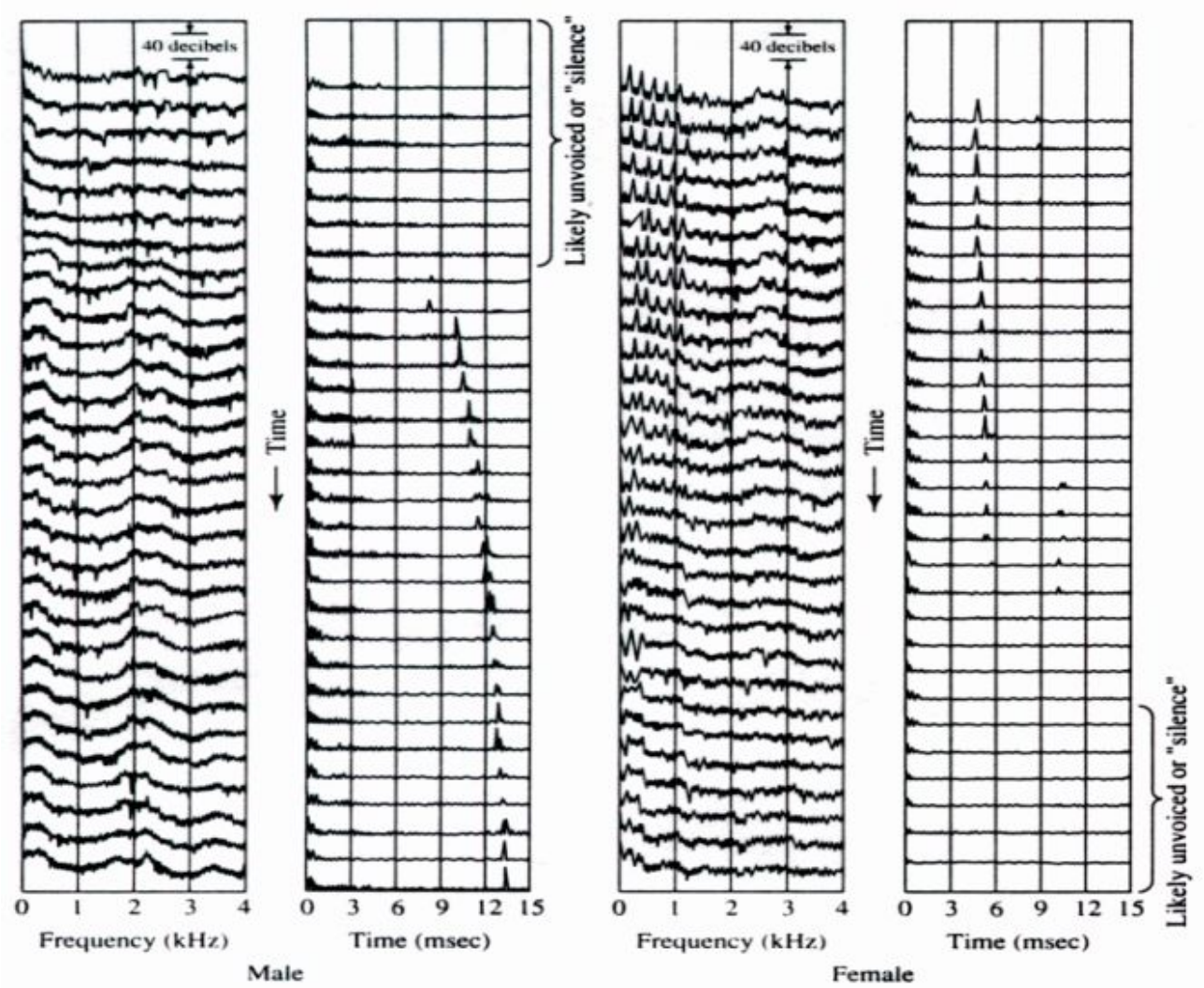

(a)
|S(ω)|
"Slow" variations (envelope)
"Fast" variations (pulses)
π ω
=
|E(ω)|
π ω
×
|Θ(ω)|
π ω
Excitation responsible for "fast" spectral variations
Problem! (nonlinear combination)
Vocal system responsible for "slow" variations
(b)
log |S(ω)| = C_s(ω)
π ω
=
log|E(ω)|
π ω
+
log|Θ(ω)|
π ω
Problem solved! (linear combination)
(c)
c_s(n)
Low quefrency energy
High quefrency
P 2P 3P n
=
c_e(n) [approx]
P 2P 3P n
+
c_θ(n)
n

Кепстр



$$c[n] = \sum_{n=0}^{N-1} log\left(\left|\sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn}\right|\right) e^{j\frac{2\pi}{N}kn}$$

Вектор признаков: пример кепстрального анализа

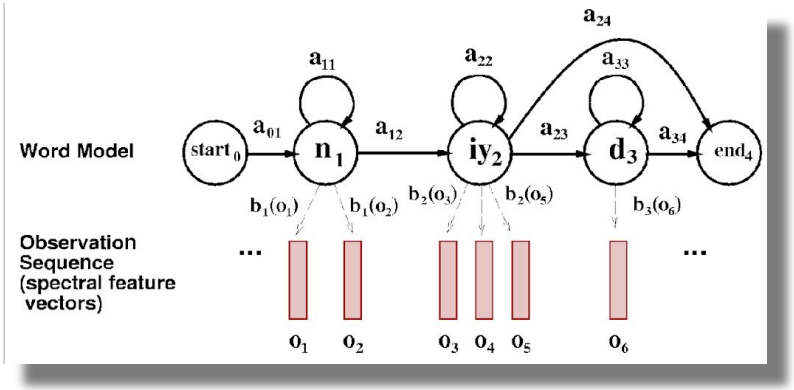


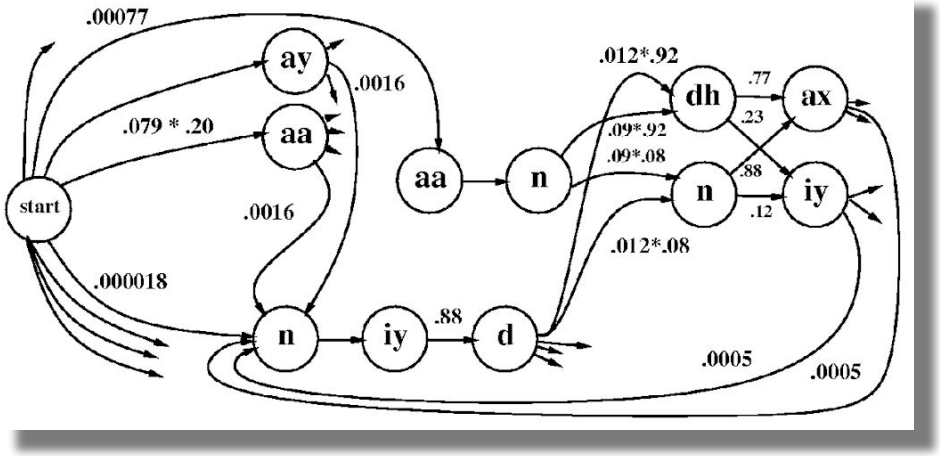
Распознавание по вектору признаков

- Кластеризации недостаточно
- Узнавание основано на цепочке стохастических векторов признаков
- Ожидание играет ключевую роль

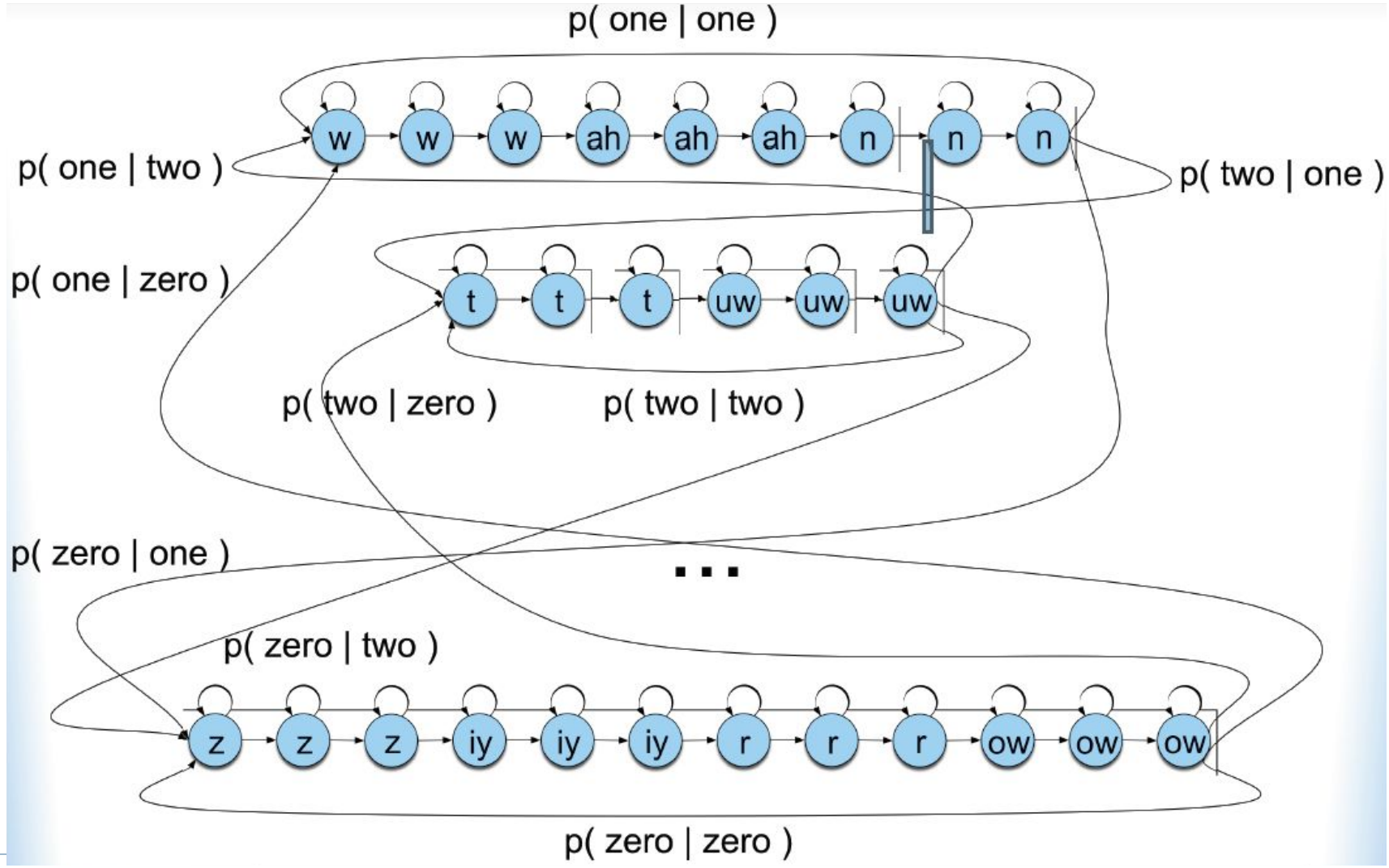
- Подходящий мат.аппарат – скрытые марковские цепи

Марковские модели

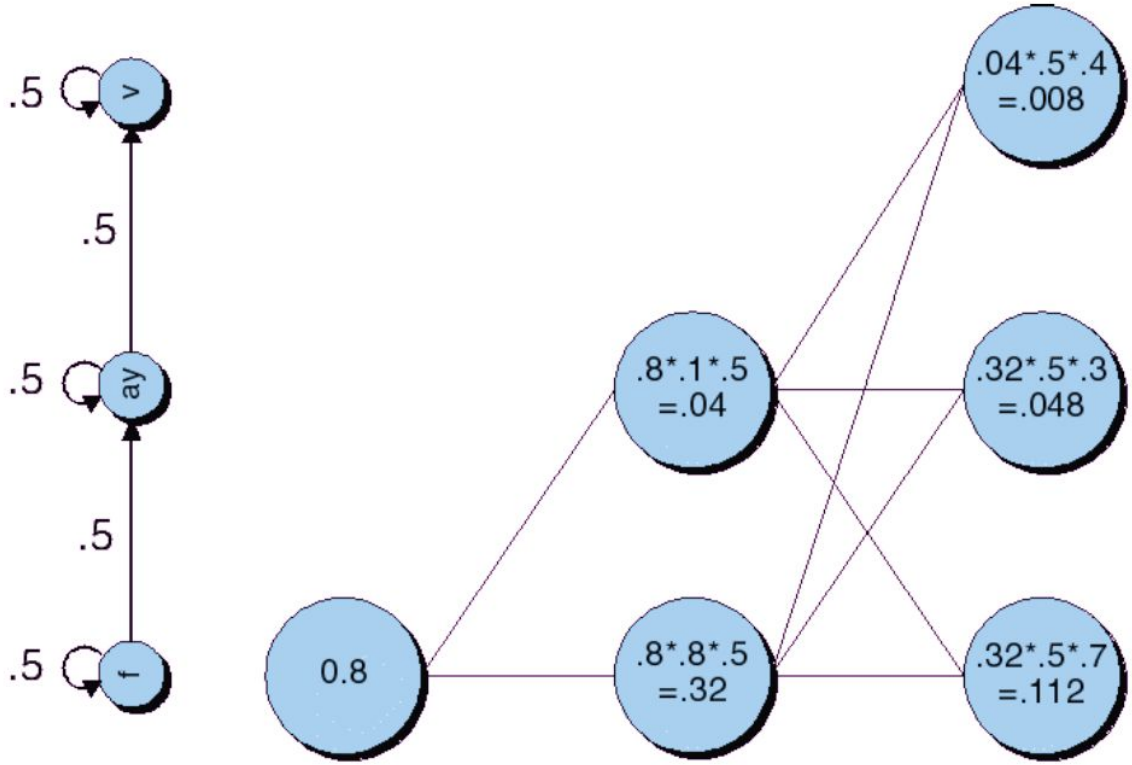


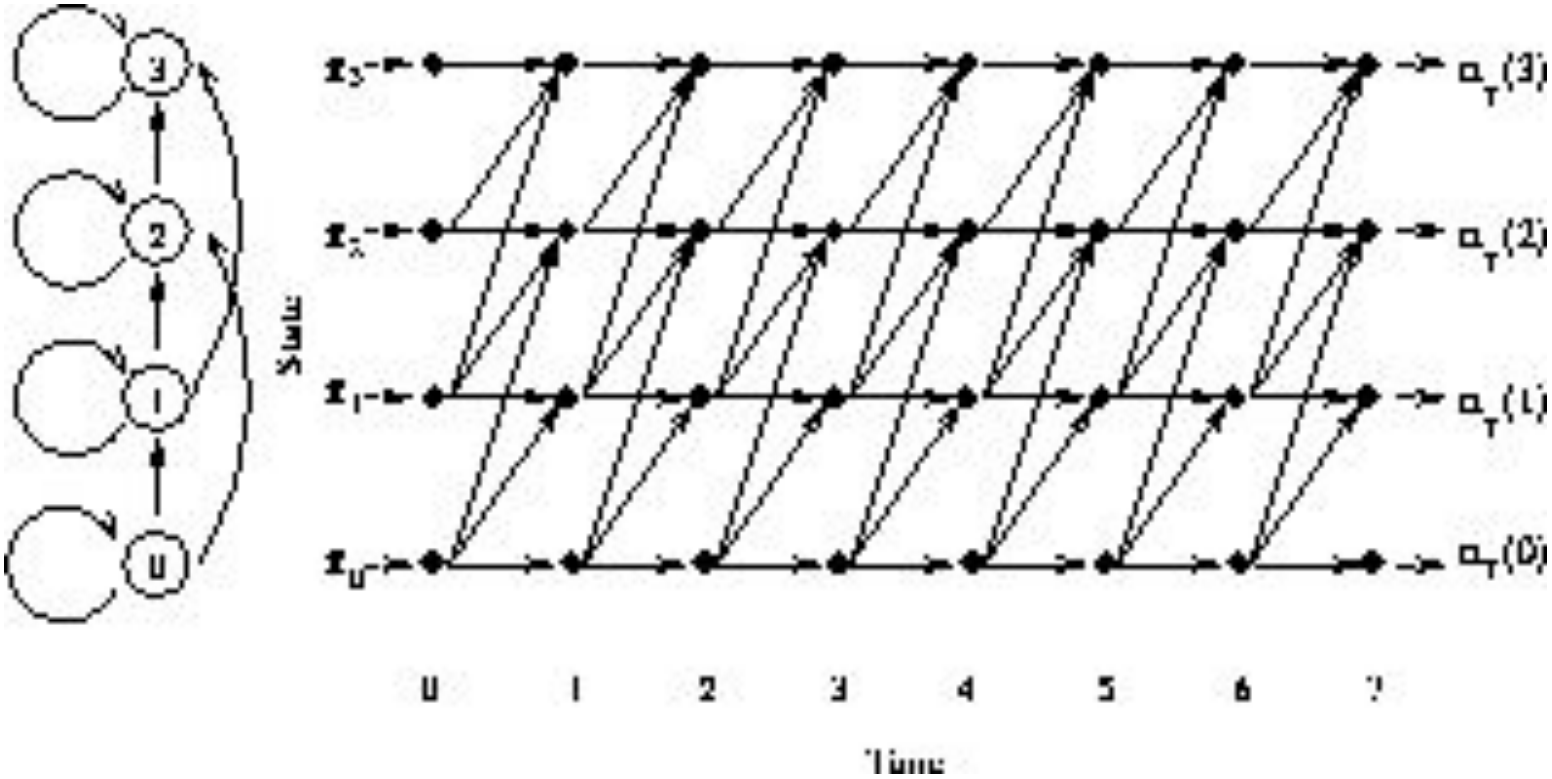


Структура скрытой марковской модели для цифр

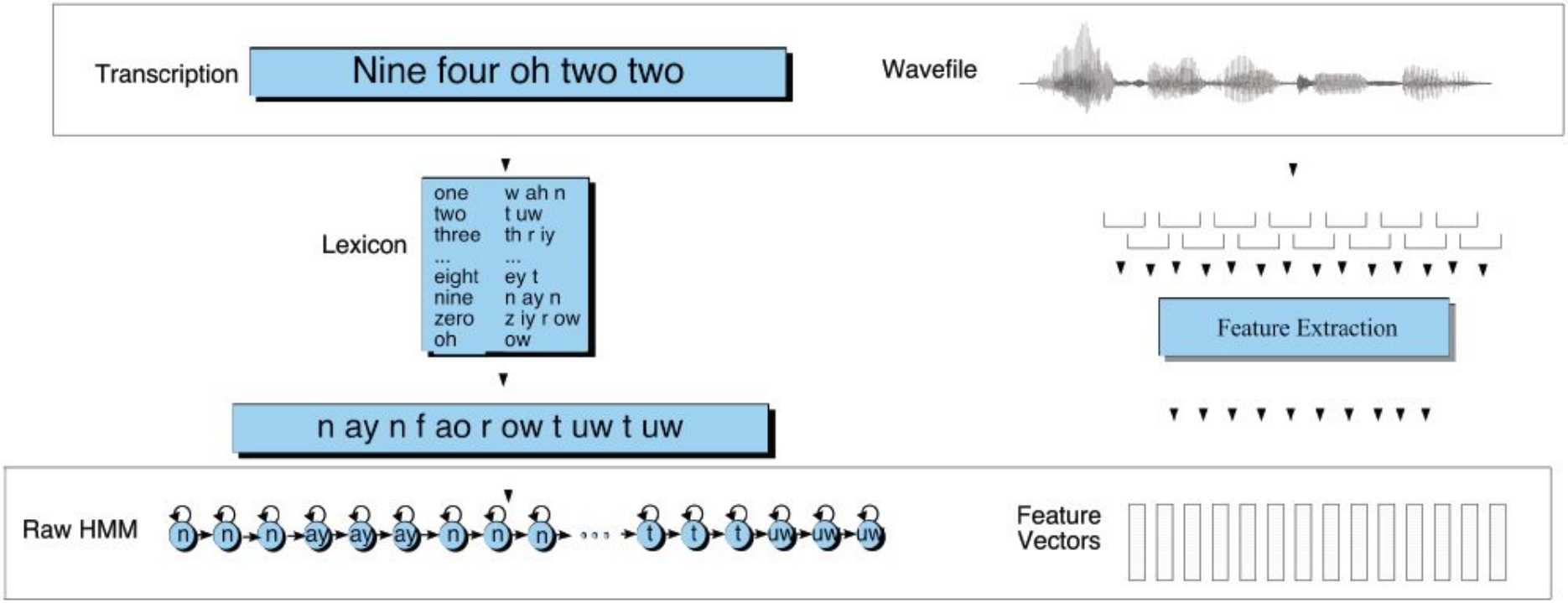


Трелис

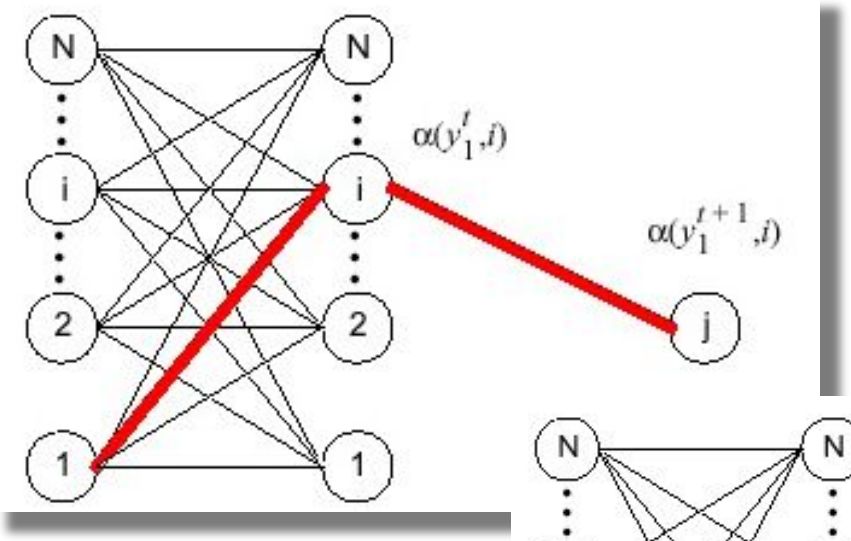


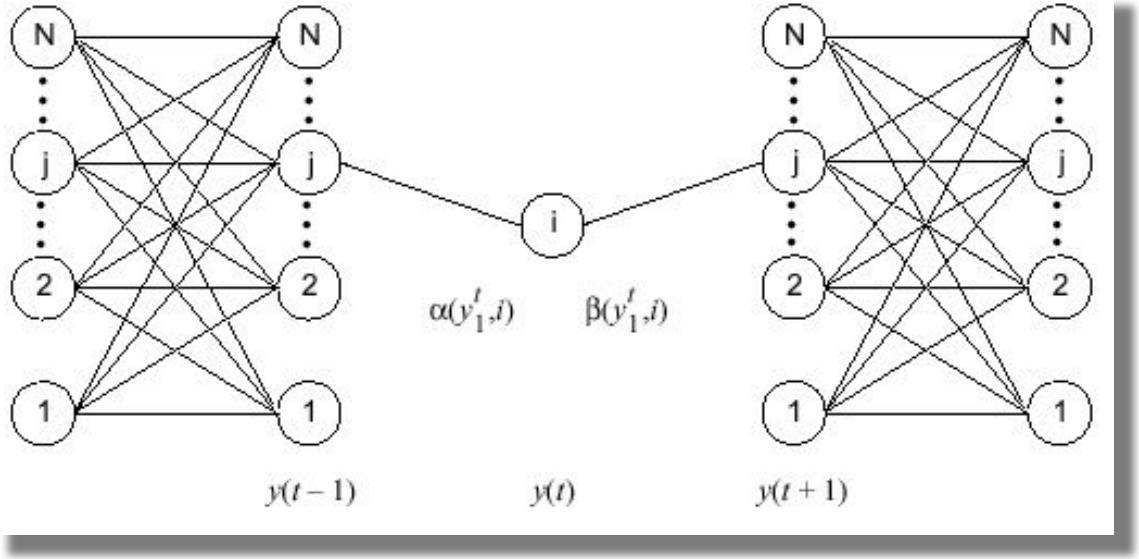

3
2
1
0
State
x_3
x_2
x_1
x_0
$\alpha_T(3)$
$\alpha_T(2)$
$\alpha_T(1)$
$\alpha_T(0)$
0
1
2
3
4
5
6
7
Time

«Встроенное» обучение

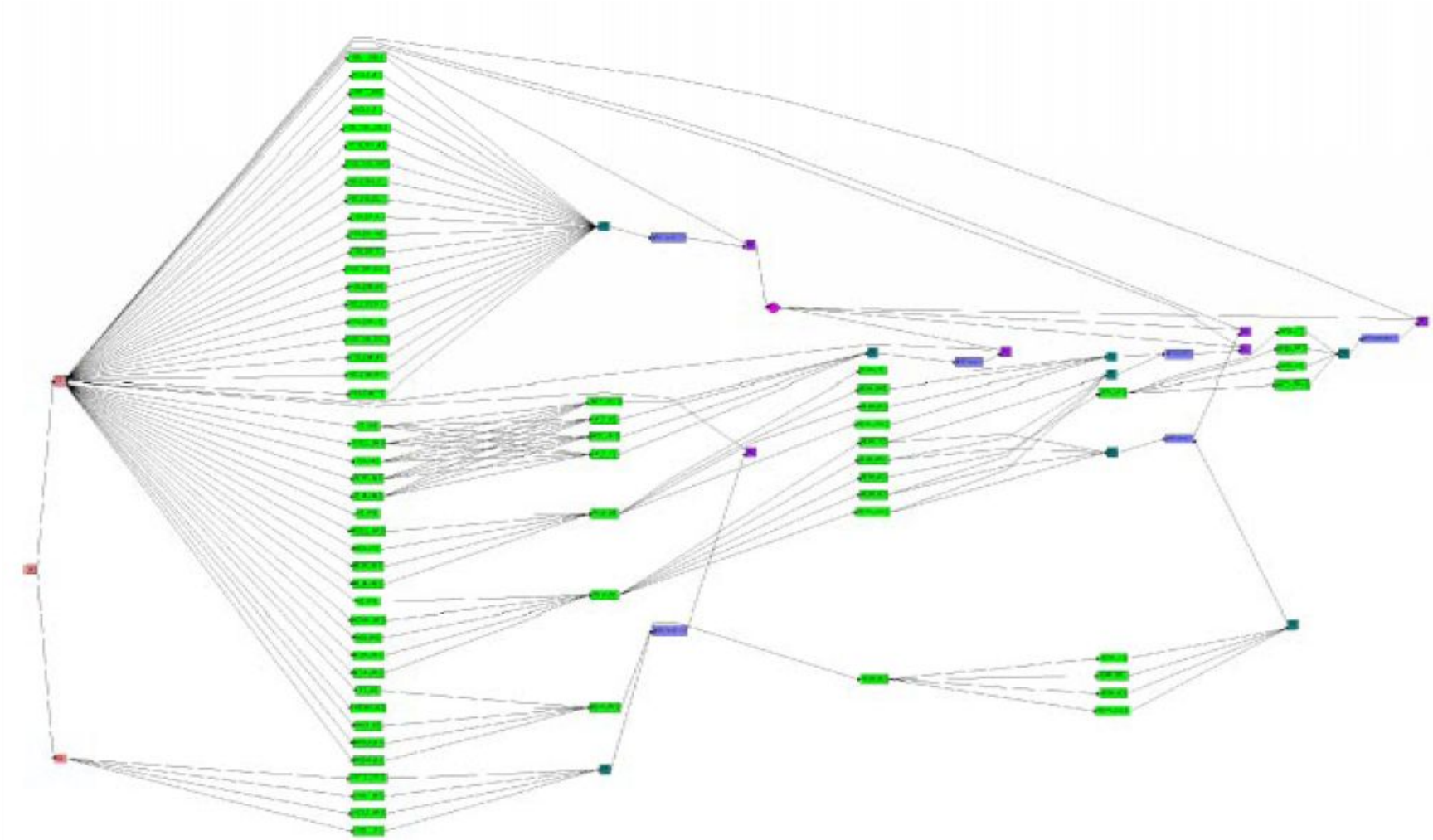
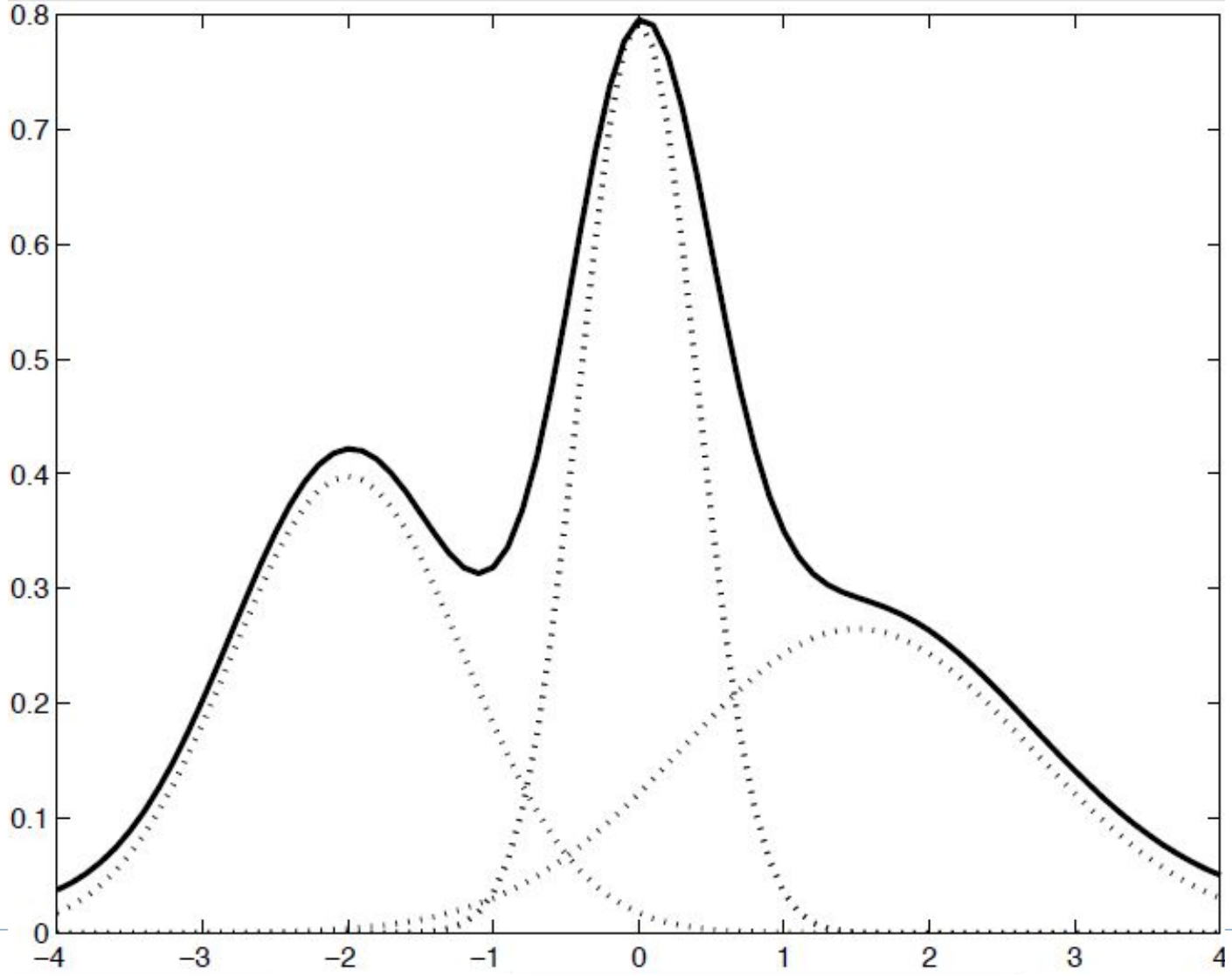


Алгоритм Витерби





Марковская цепь предложения



Моделирование формы кластера с помощью смеси гауссианов