

Распознавание речи. Акустические модели

П. А. Холявин

p.kholyavin@spbu.ru

26.02.2025





Задача распознавания речи

Если $O = o_1, o_2, \dots, o_n$ – звуковая последовательность,
 $W = w_1, w_2, \dots, w_n$ – последовательность слов, то

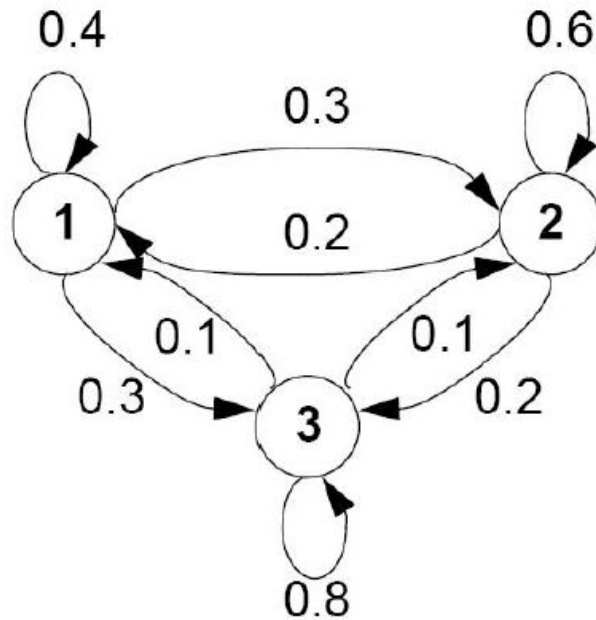
$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} P(W|O)$$

$$\hat{W} = \underset{W \in L}{\operatorname{argmax}} \frac{P(O|W) P(W)}{P(O)} = \underset{W \in L}{\operatorname{argmax}} P(O|W) P(W)$$



Скрытые Марковские модели

1. Цепь Маркова

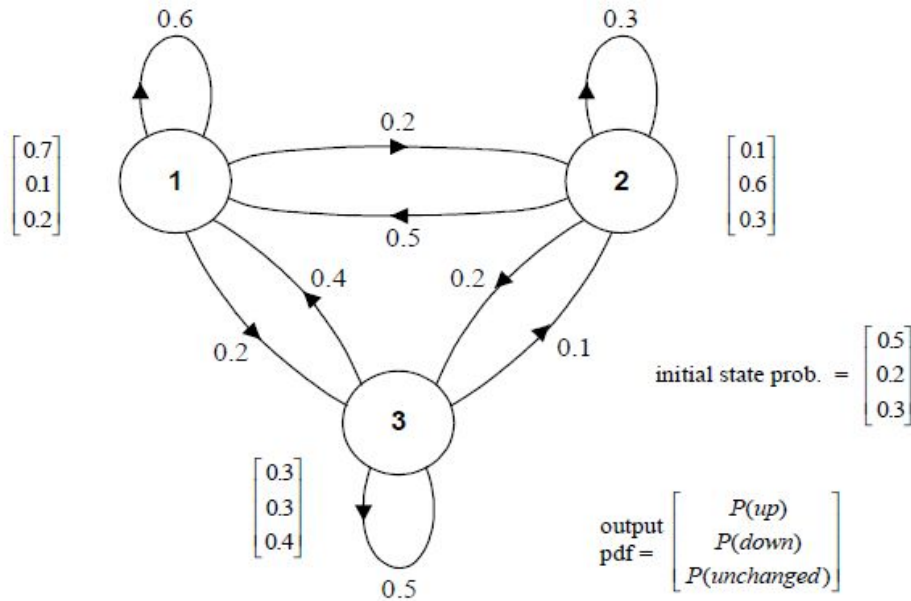


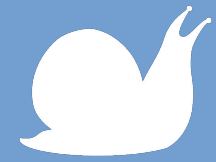
$$A = \{a_{i,j}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$



Скрытые Марковские модели

2. Скрытые Марковские модели (СММ, НММ)





Скрытые Марковские модели

3. Марковское допущение:

$$P(s_t | s_1^{t-1}) = P(s_t | s_{t-1})$$

4. Допущение независимости выходных значений:

$$P(X_t | X_1^{t-1}, s_1^t) = P(X_t | s_t)$$



Задачи СММ

1. Оценка: если даны модель Φ и последовательность наблюдений X , какова вероятность $P(X|\Phi)$? – *алгоритм прямого хода*
2. Распознавание: если даны модель Φ и последовательность наблюдений X , какова наиболее вероятная последовательность состояний S , породившая наблюдения? – *алгоритм Витерби*
3. Обучение: если даны модель Φ и несколько последовательностей наблюдений X_i , как изменить параметры модели, чтобы максимизировать вероятность $P(X_i|\Phi)$? – *алгоритм Баума-Уэлша*



Алгоритм прямого хода

$$\alpha_t(i) = P(x_1, x_2, \dots, x_t, q_t = s_i \mid \lambda).$$

Инициализация:

$$\alpha_1(i) = \pi_i B_i(x_1), \quad 1 \leq i \leq N,$$

индукция:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{i,j} \right] B_j(x_{t+1}), \quad 1 \leq t < T, 1 \leq j \leq N,$$

завершение:

$$P(X \mid \lambda) = \sum_{i=1}^N \alpha_T(i).$$



Алгоритм Витерби

ALGORITHM 8.3 THE VITERBI ALGORITHM

Step 1: Initialization

$$V_1(i) = \pi_i b_i(X_1) \quad 1 \leq i \leq N$$

$$B_1(i) = 0$$

Step 2: Induction

$$V_t(j) = \underset{1 \leq i \leq N}{\text{Max}} [V_{t-1}(i) a_{ij}] b_j(X_t) \quad 2 \leq t \leq T; \quad 1 \leq j \leq N \quad (8.25)$$

$$B_t(j) = \underset{1 \leq i \leq N}{\text{Arg max}} [V_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T; \quad 1 \leq j \leq N \quad (8.26)$$

Step 3: Termination

$$\text{The best score} = \underset{1 \leq i \leq N}{\text{Max}} [V_T(i)]$$

$$s_T^* = \underset{1 \leq i \leq N}{\text{Arg max}} [B_T(i)]$$

Step 4: Backtracking

$$s_t^* = B_{t+1}(s_{t+1}^*) \quad t = T-1, T-2, \dots, 1$$

$S^* = (s_1^*, s_2^*, \dots, s_T^*)$ is the best sequence



Алгоритм Баума-Уэлша

ALGORITHM 8.4 THE FORWARD-BACKWARD ALGORITHM

Step 1: Initialization: Choose an initial estimate Φ .

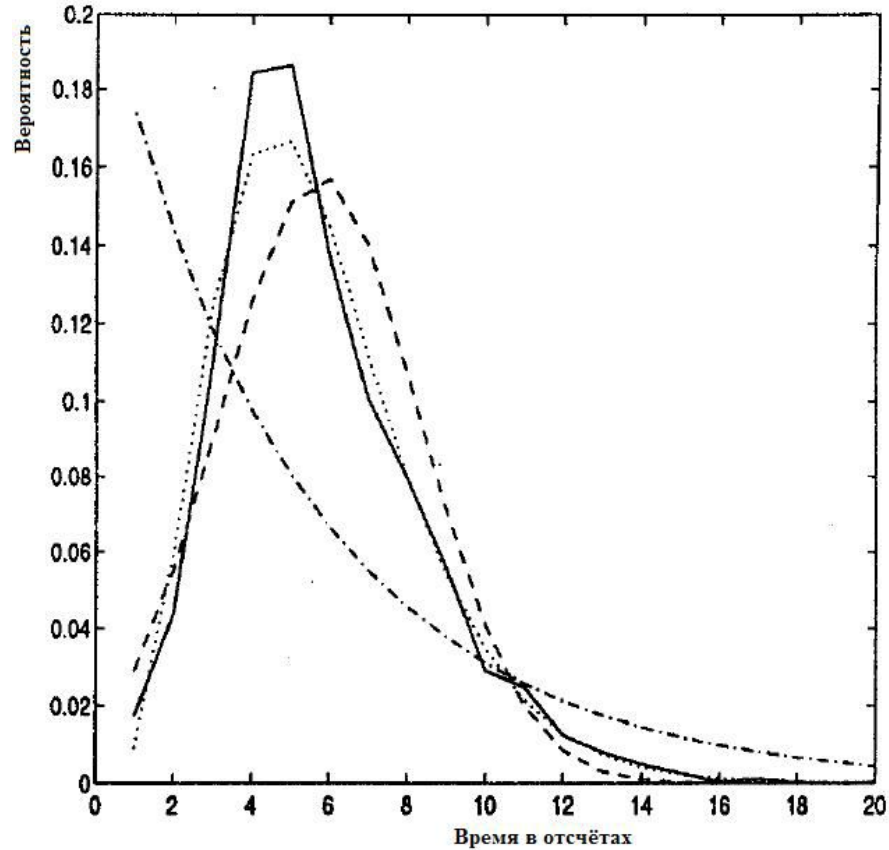
Step 2: E-step: Compute auxiliary function $Q(\Phi, \hat{\Phi})$ based on Φ .

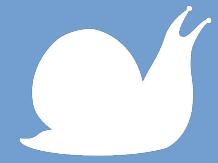
Step 3: M-step: Compute $\hat{\Phi}$ according to the re-estimation Eqs. (8.40) and (8.41) to maximize the auxiliary Q-function.

Step 4: Iteration: Set $\Phi = \hat{\Phi}$, repeat from step 2 until convergence.

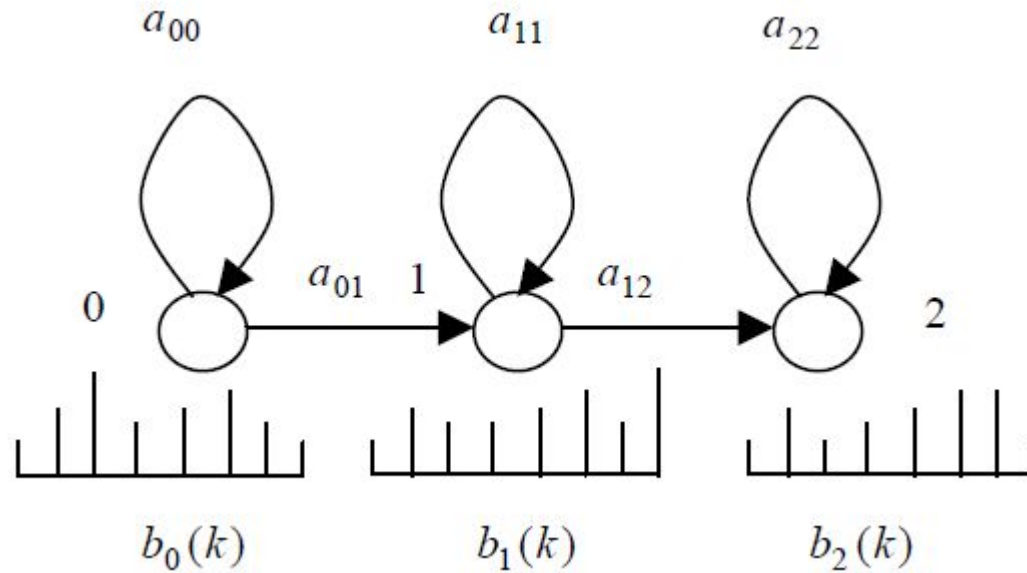


Неоднородная Марковская модель



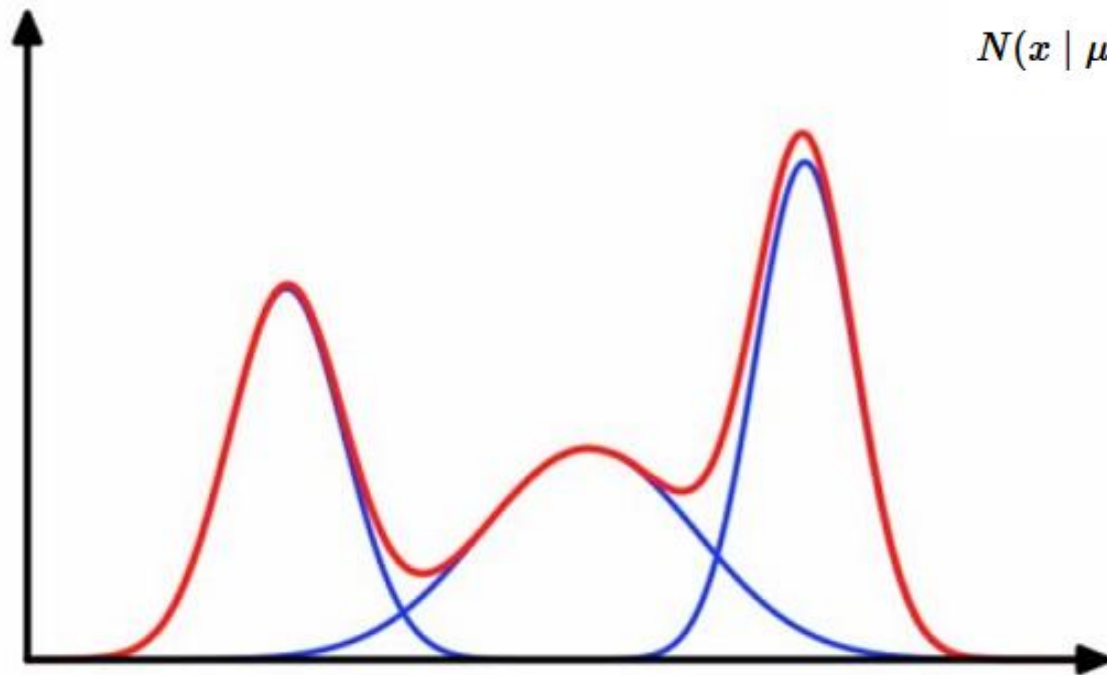


Практическое применение



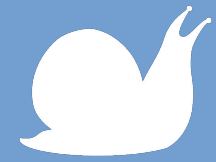


Гауссовские смеси (GMM)

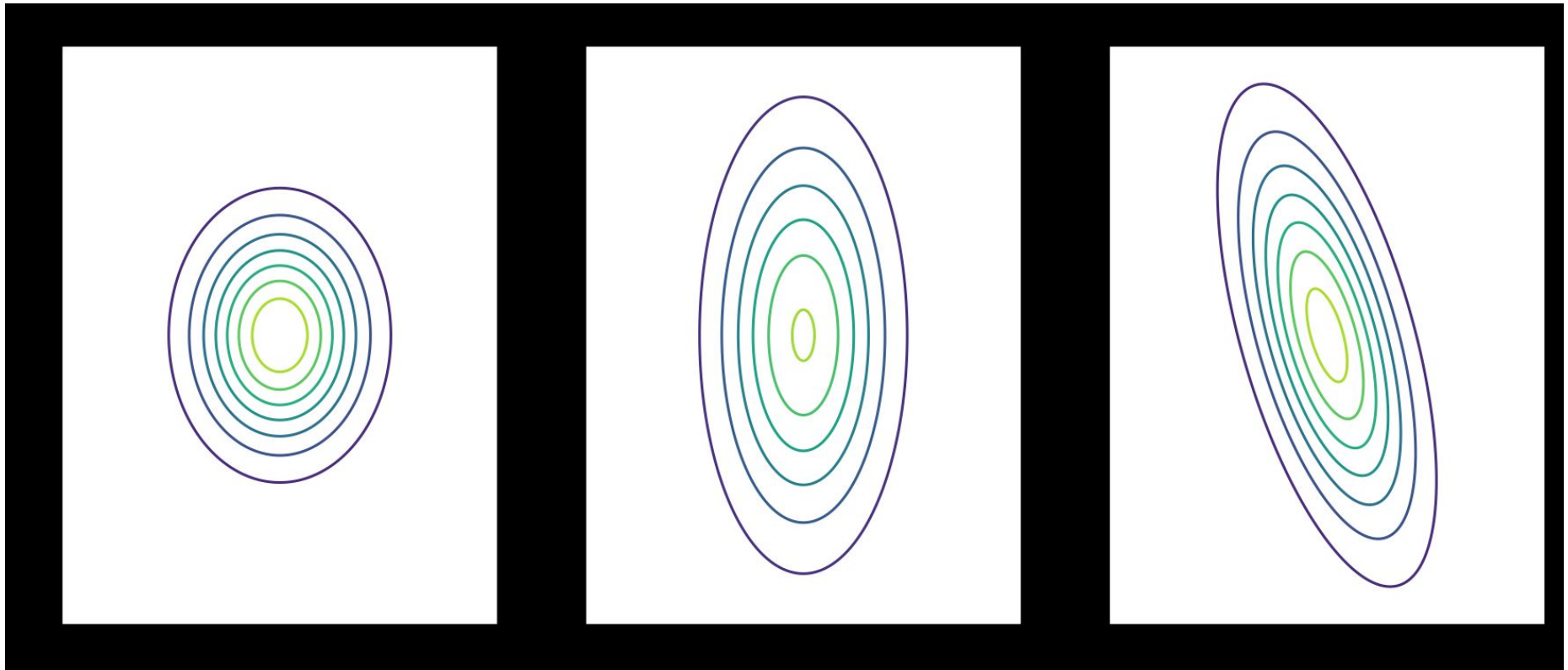


$$N(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

$$N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right]$$



Гауссовские смеси (GMM)

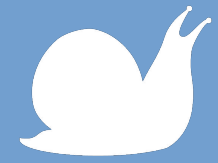




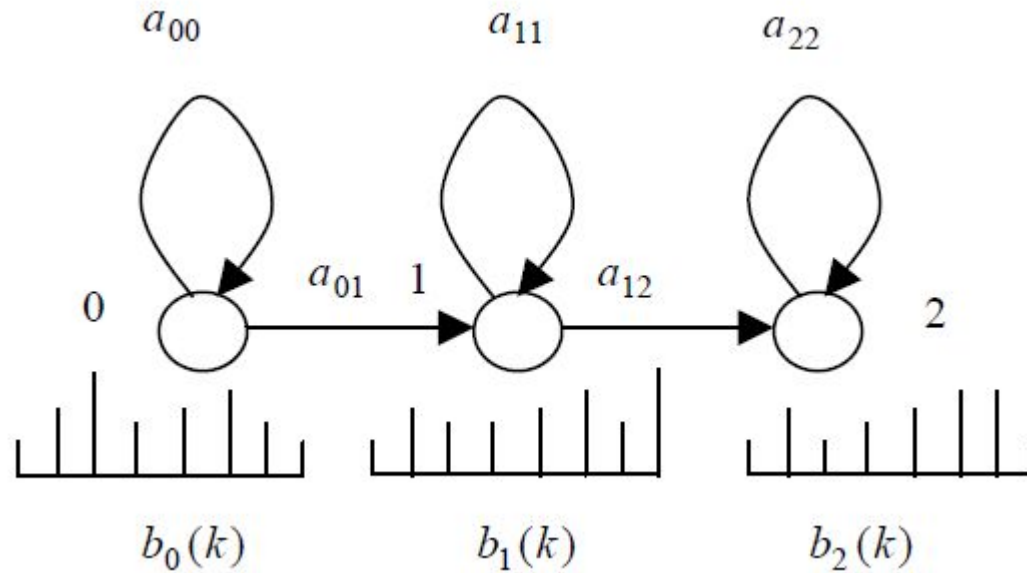
Проблема выбора единиц

Требования к единицам:

1. Точность
2. Обучаемость
3. Обобщаемость

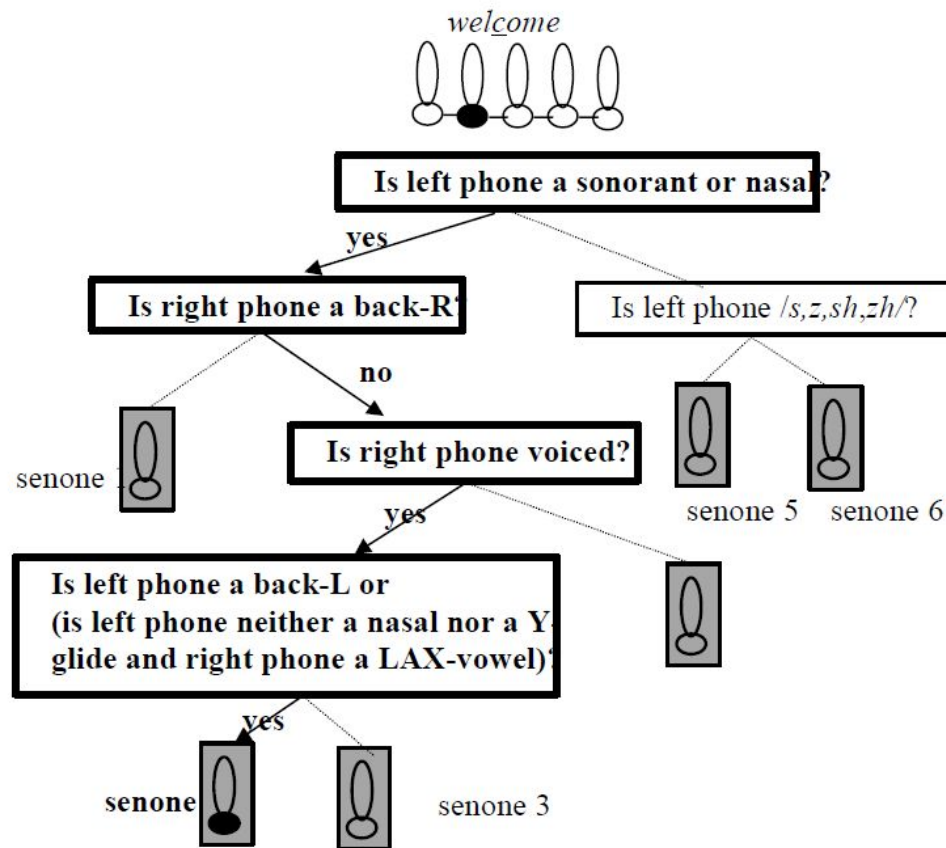


Архитектура СММ: трифоны



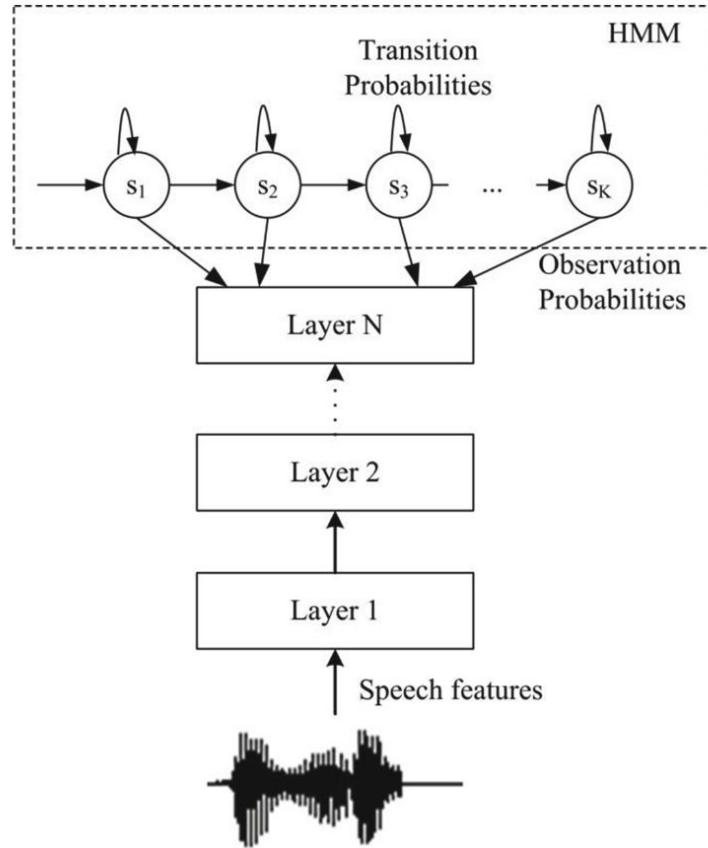


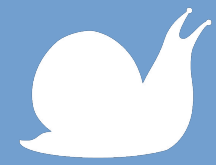
Кластеризация



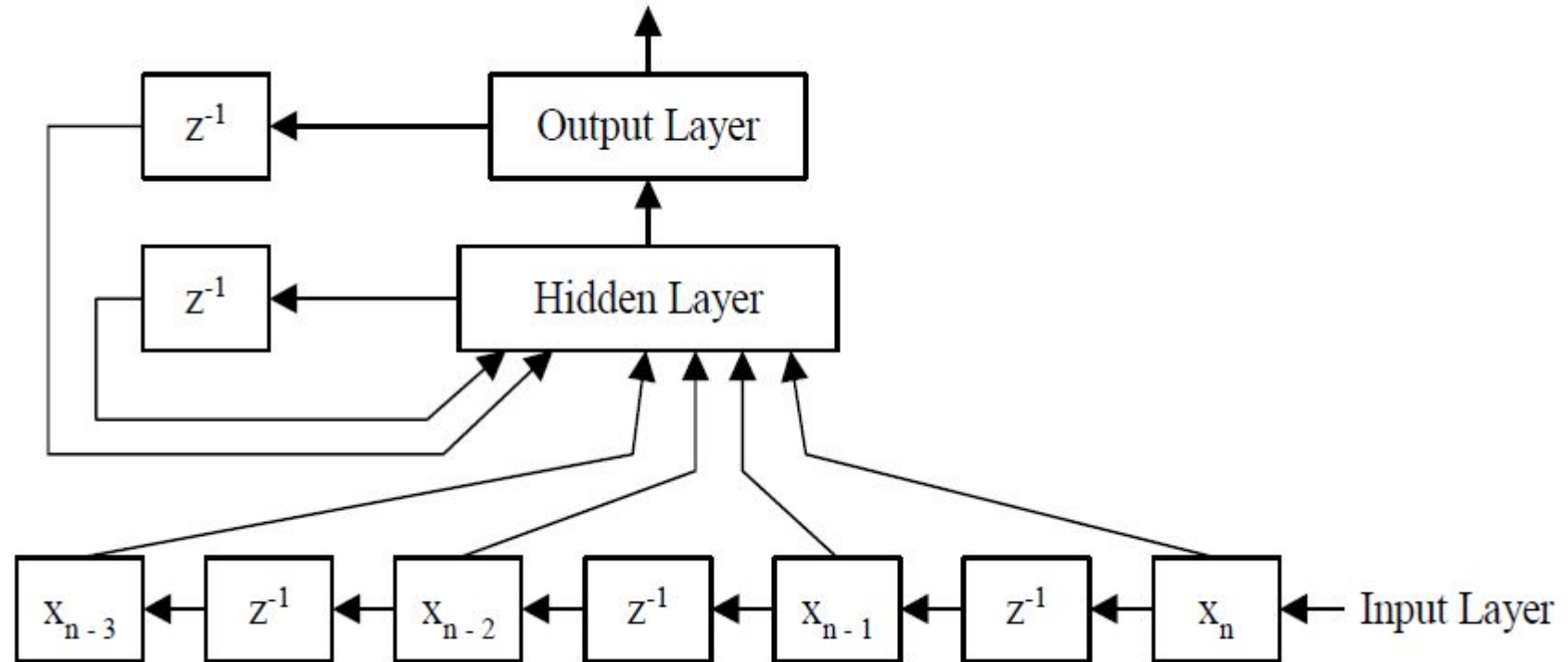


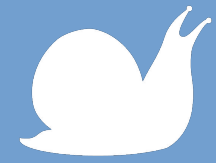
HMM-DNN (гибридные модели)



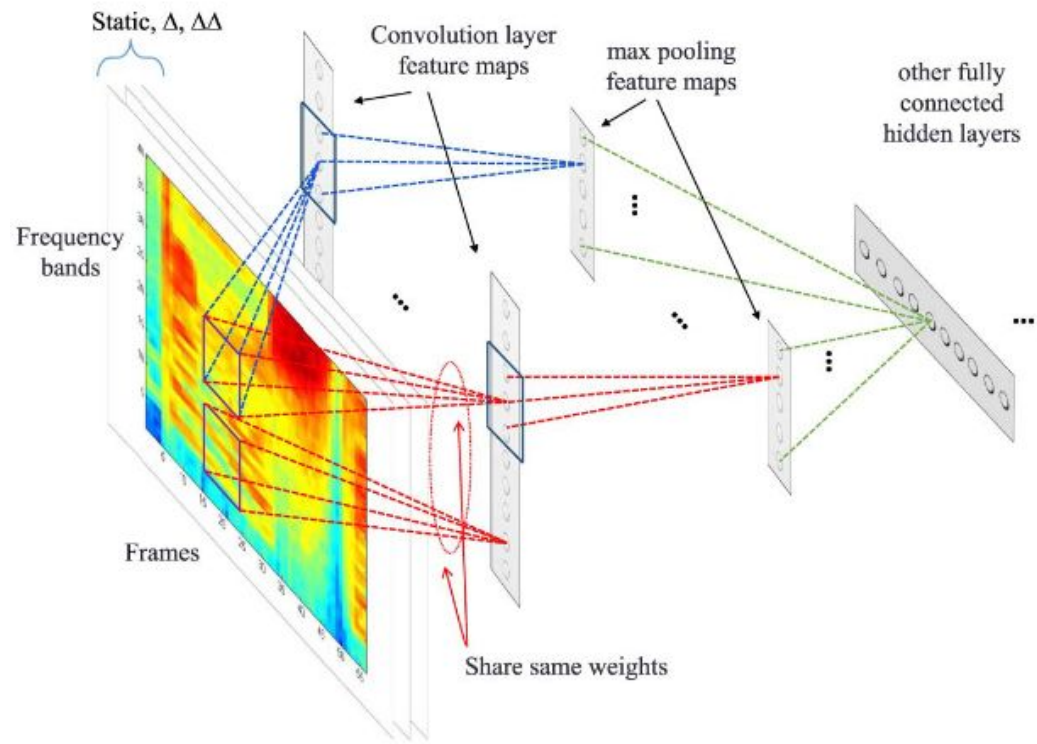


RNN





CNN



Спасибо за внимание!

