

# **Задачи в NLP и распознавание речи**

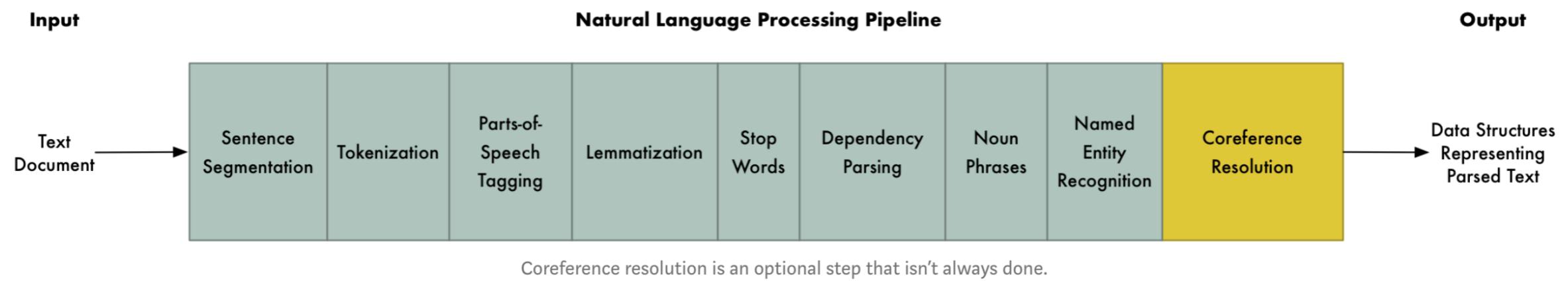
**Титизян Армине, БПМИ172**

# Содержание

- Что такое «обработка естественного языка» (NLP)?
- Какие задачи решают в NLP?
- Задача на стыке: распознавание речи. История
- Задача на стыке: распознавание речи. Вариант решения и проблемы
- Connectionist Temporal Classification (CTC)

**NLP =**  
**компьютерные науки +**  
**искусственный интеллект +**  
**математическая лингвистика**

# Задачи



# Синтаксис

- Индукция грамматики: восстановление формальной грамматики на основе наблюдений
- Лемматизация
- Морфологическая сегментация: разбить слово на морфемы (*корень, приставка, суффикс, окончание и тп.*)
- Тэггинг частей речи
- Разбивка сырого текста на предложения
- Стэмминг

# Семантика

- Лексическая семантика: какое значение имеет слово?
- Дистрибутивная семантика: насколько два слова близки по значению, судя по их распределению в некотором корпусе? Эмбеддинги :)
- Машинный перевод
- Генерация текста
- Вопросно-ответные системы
- Разрешение лингвистической многозначности: какое значение имеет слово в данном контексте? Например, «замок двери» и «замок принцессы» (омонимичная двузначность)

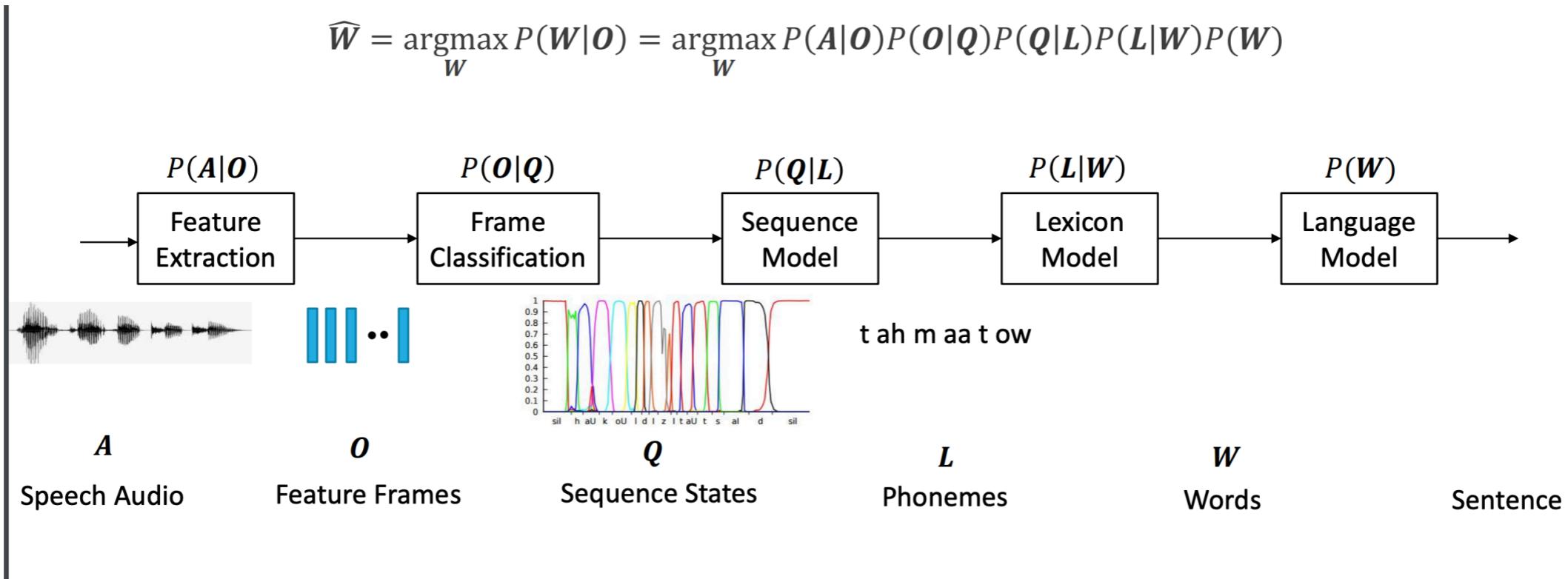
# Дискурс

- Автоматическое обобщение: генерация текста меньше, чем данный, о данном тексте
- Разрешение кореферентности: как понять, что два выражения ссылаются на один объект?  
*Например, имя «средство для чистки» кореферентно имени «чистящее средство».*

# Речь

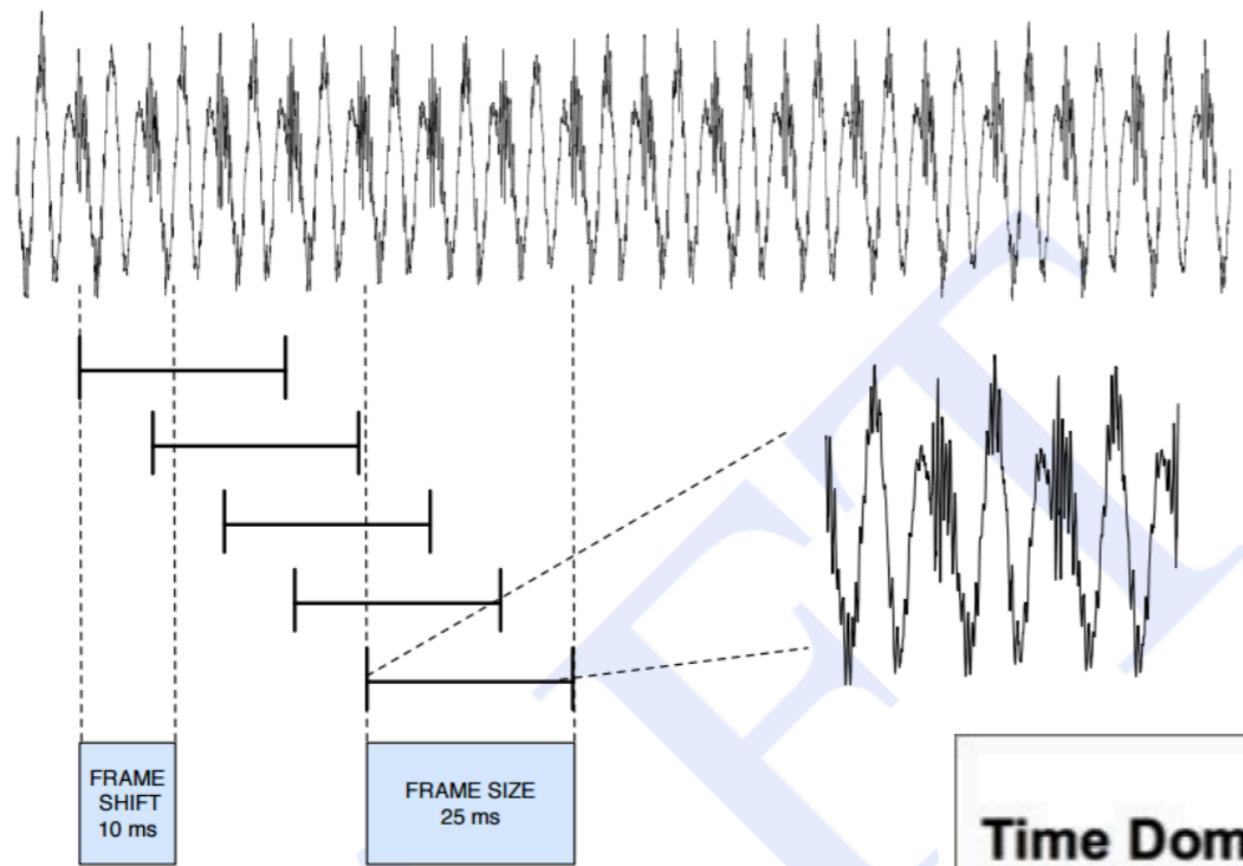
- Распознавание речи  
(speech2text и text2speech)

# Варианты



- Speech2text: глубокие модели на основе скрытых марковских моделей, Connectionist Temporal Classification (CTC), модели на основе архитектуры «attention»
- Text2Speech: WaveNet, DeepVoice, Tacotron

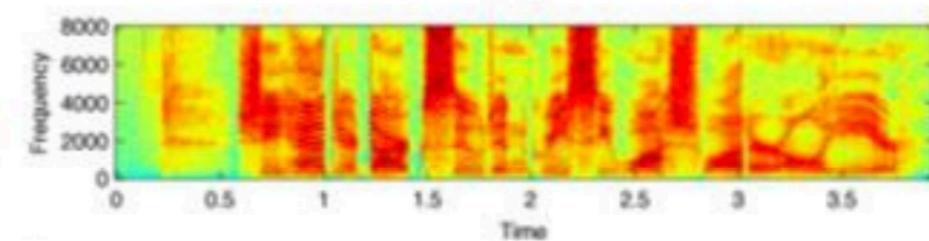
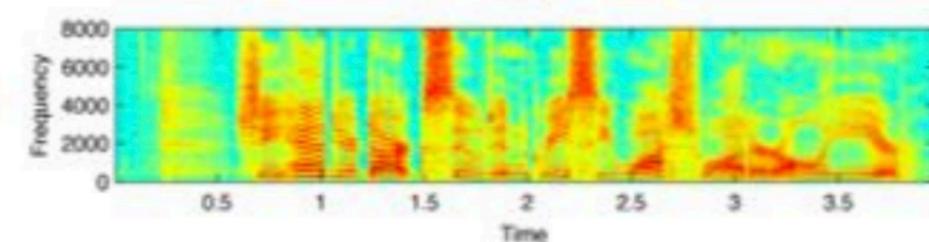
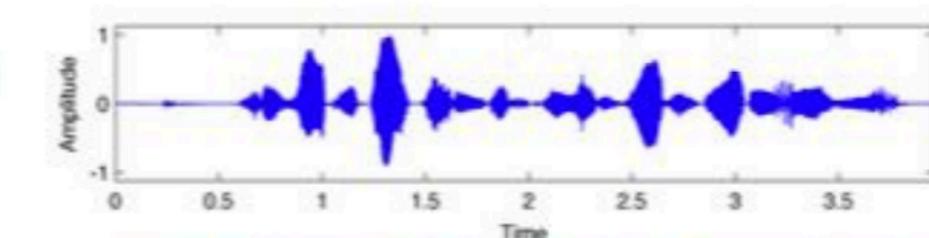
# Предобработка аудио



**Time Domain  
Waveform**

**Spectrogram**

**MFCC  
Spectrogram**



# Лексическая модель

Deterministic

Word	Pronunciation
ТОМАТО	t ah m aa t ow
	t ah m ey t ow
COVERAGE	k ah v er ah jh
	k ah v r ah jh

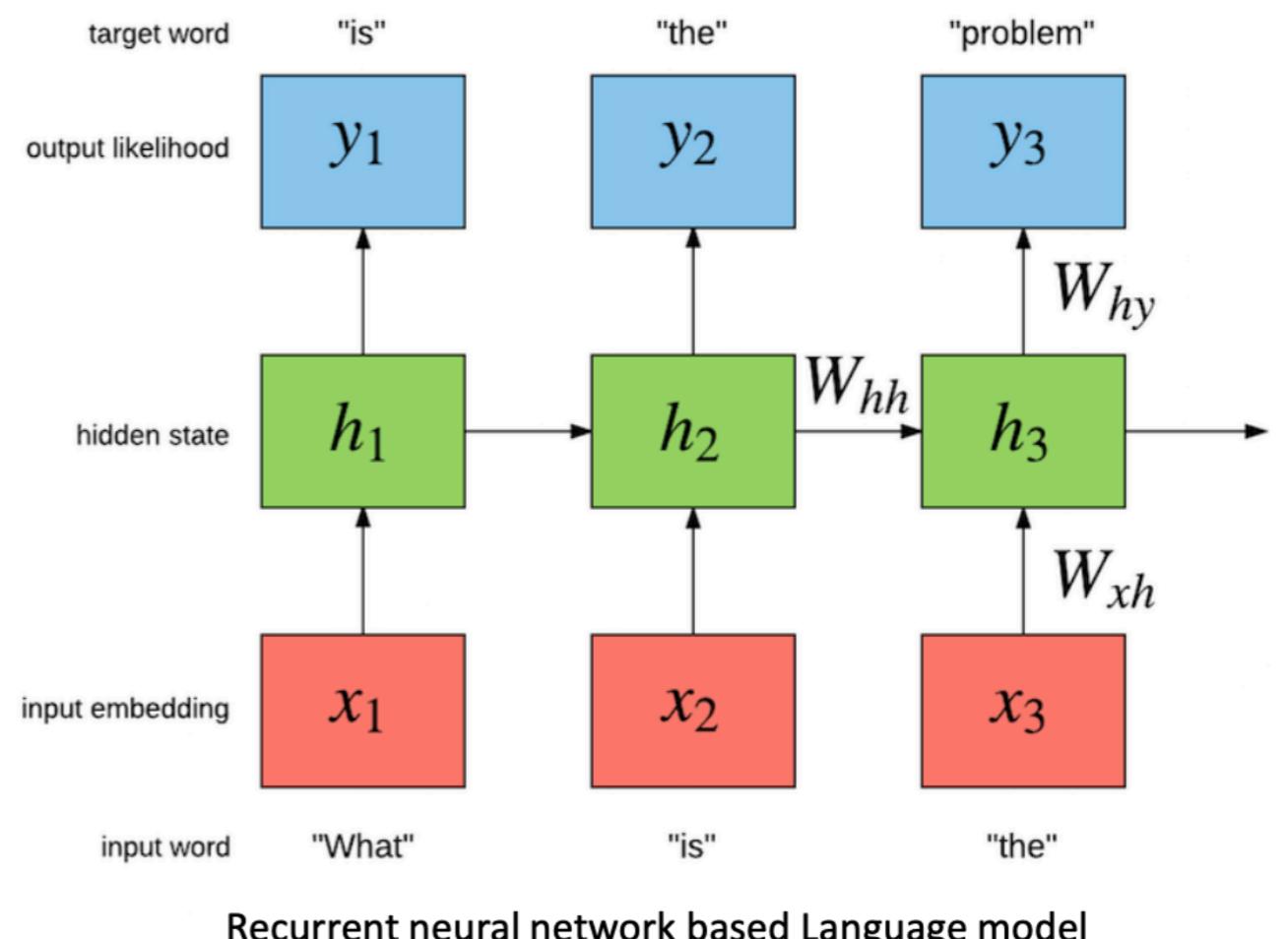
Probabilistic

Word	Pronunciation	Probability
ТОМАТО	t ah m aa t ow	0.45
	t ah m ey t ow	0.55
COVERAGE	k ah v er ah jh	0.65
	k ah v r ah jh	0.35

- Чаще всего, это словарь, который сопоставляет предсказанной последовательности фонем слово. Вероятностная модель возвращает также вероятность данного слова при данной последовательности

# Языковая модель

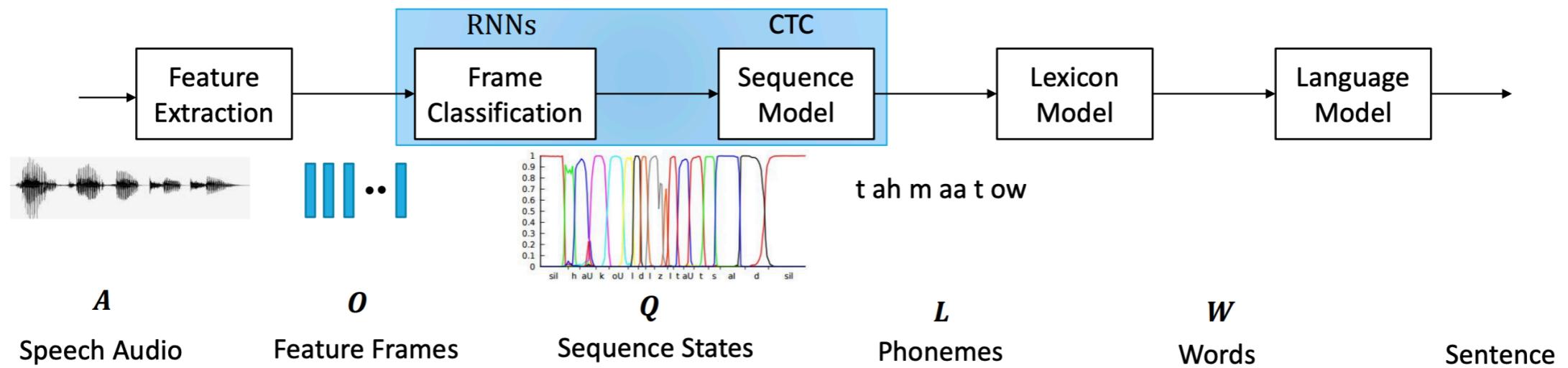
- Вероятностная модель; используется, чтобы **а)** предсказать возможные следующие слова **б)** разрешать проблему неоднозначности для двух последовательностей фонем
- Возвращает вероятность того, что последовательность распознана



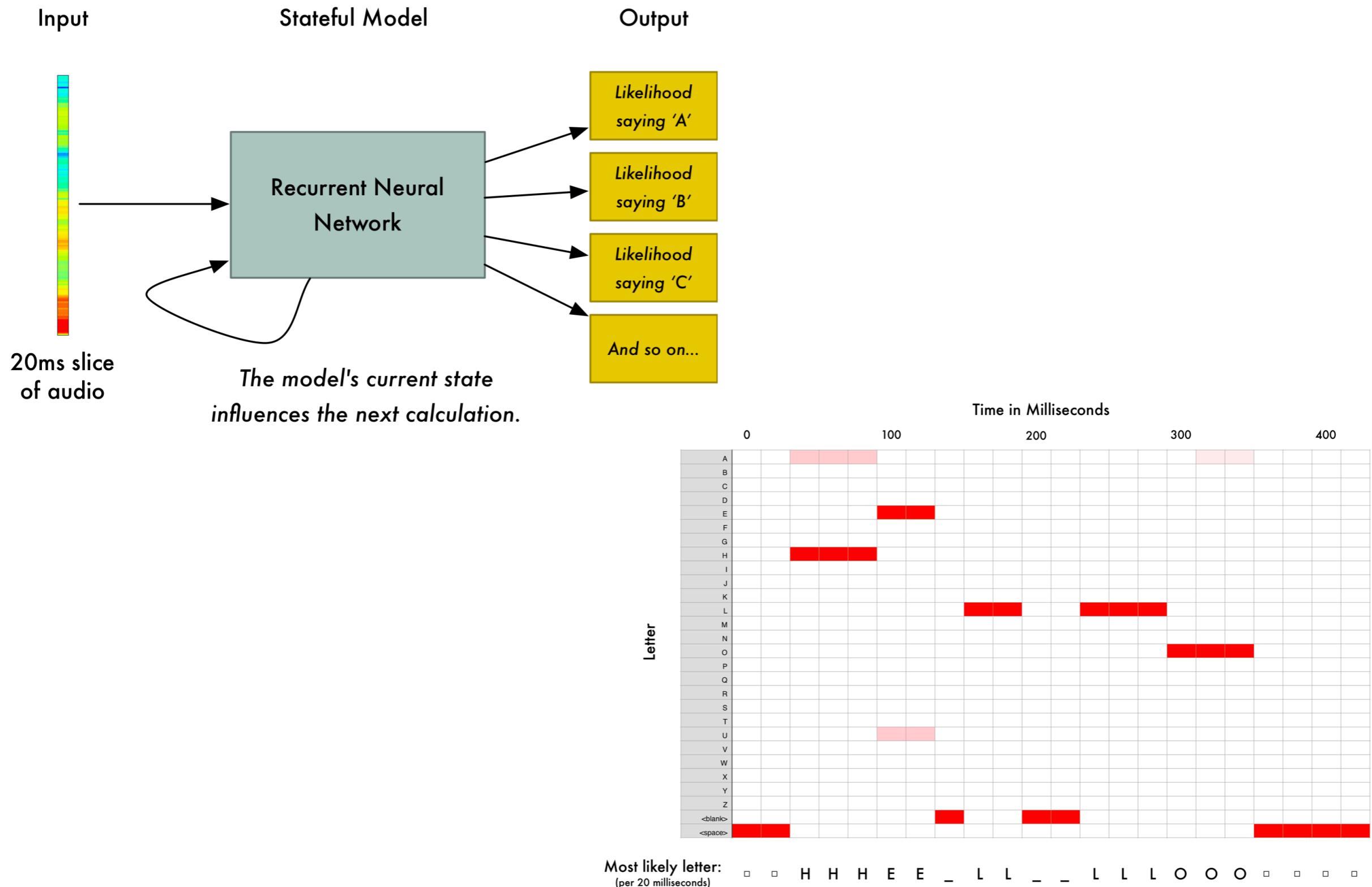
# RNN+CTC

RNN: Recurrent Neural Networks

CTC: Connectionist Temporal Classification



# Модель распознавания фрейма



# RNN+CTC

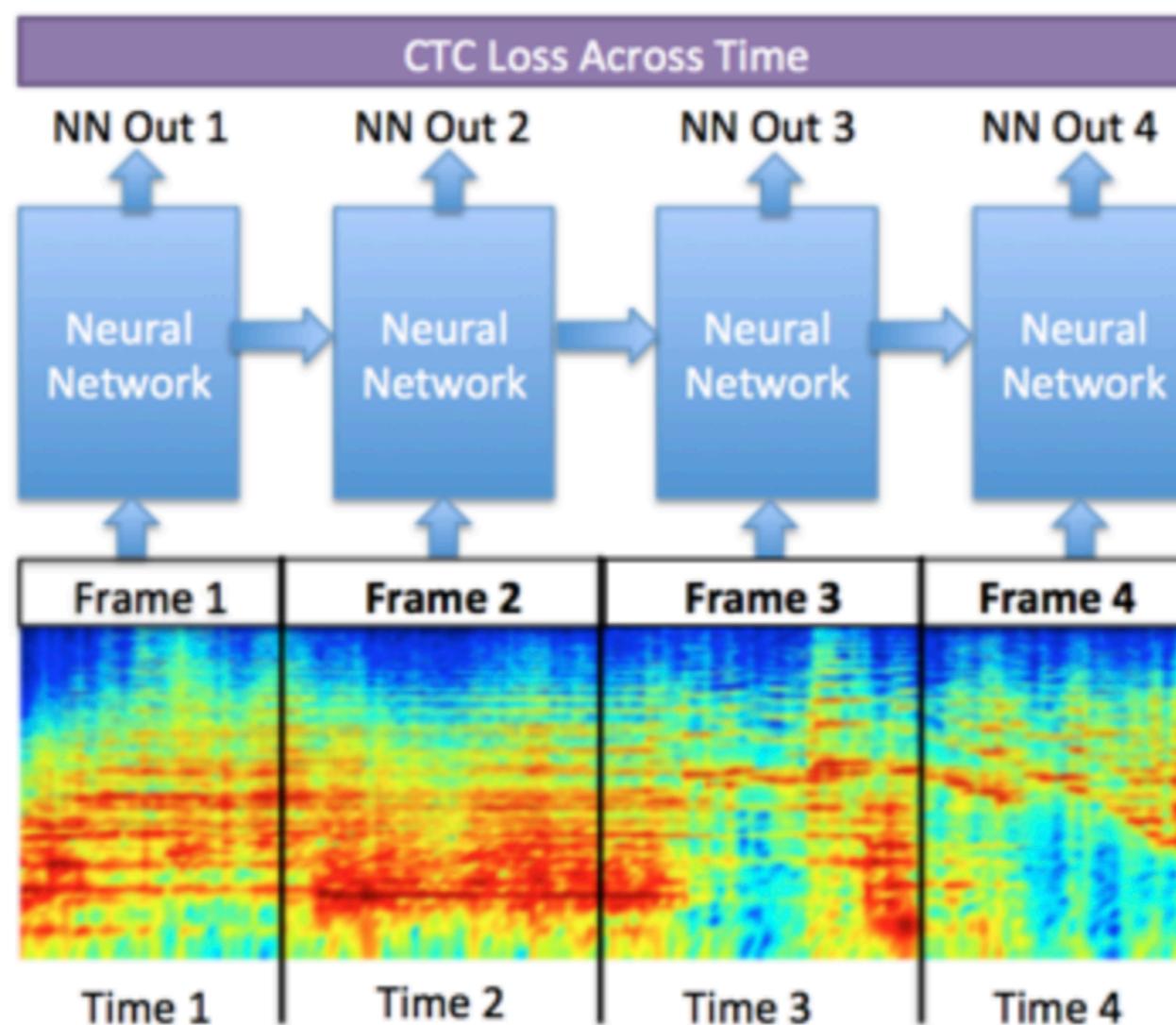
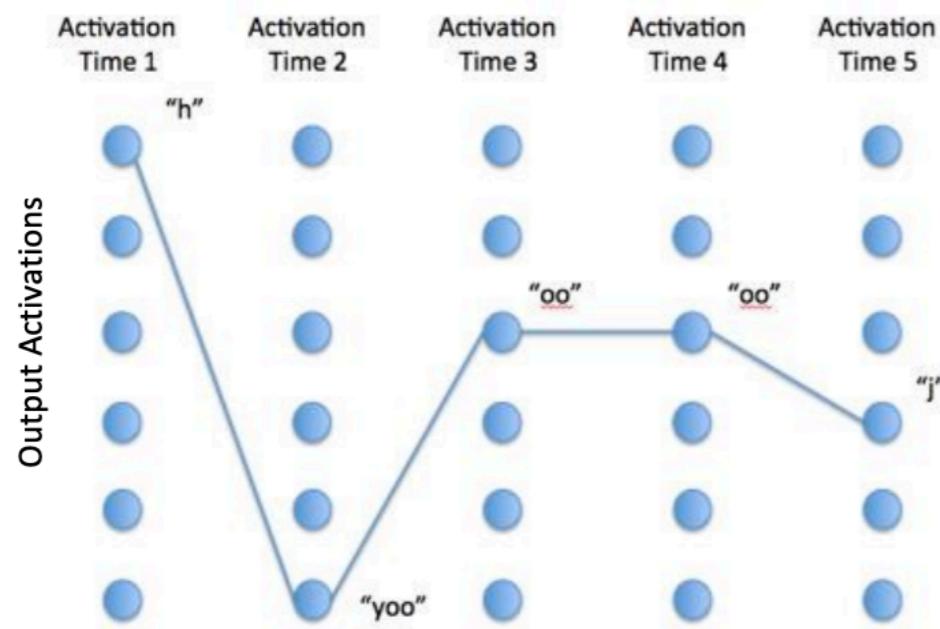


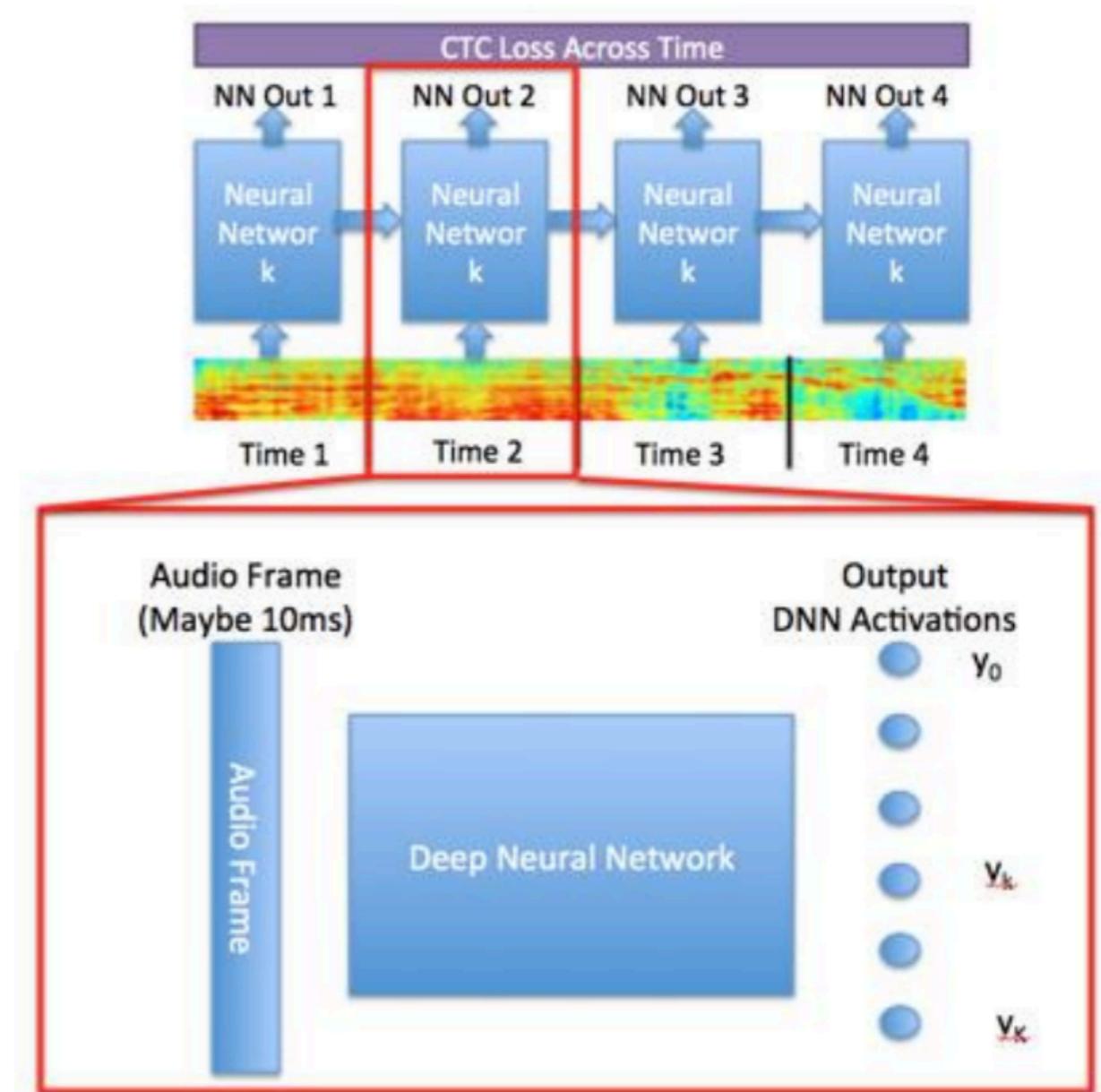
Diagram 1: CTC loss (purple) across several NN (blue) outputs

# Оптимизация путей



$$P(\pi|x) = \prod_{t=1}^T y_{\pi_t}^t, \forall \pi \in L^T$$

$$L^T \triangleq \pi = \{\pi_1, \dots, \pi_T\}$$



# Распознавание последовательностей

- Предсказывать каждую фонему
- Предсказывать последовательностей

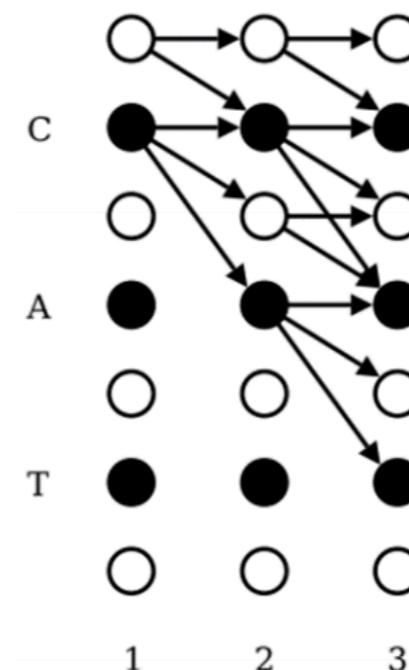
$$P(l|x) = \sum_{s=1}^{|l|} \frac{\alpha_t(s)\beta_t(s)}{y_{l_s}^t}$$

...  
...

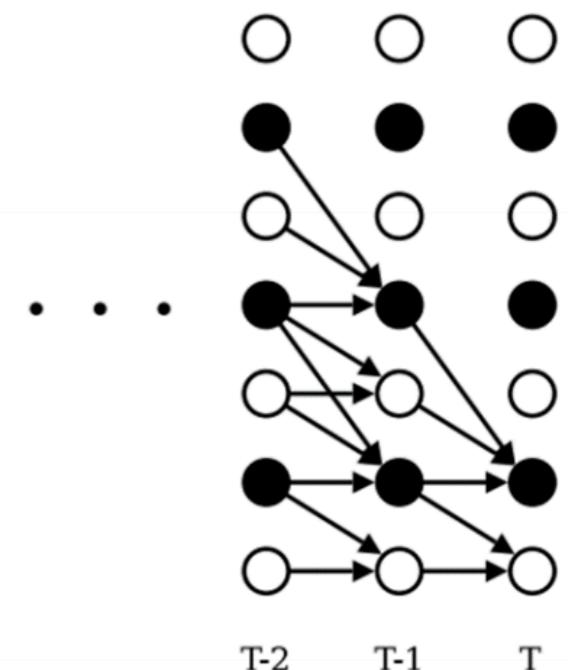
$$P(l|x) = \sum_{\pi} P(l|\pi)P(\pi|x)$$

$$P(l|x) = \sum_{s=1}^{|l|} \frac{\alpha_t(s)\beta_t(s)}{y_{l_s}^t}$$

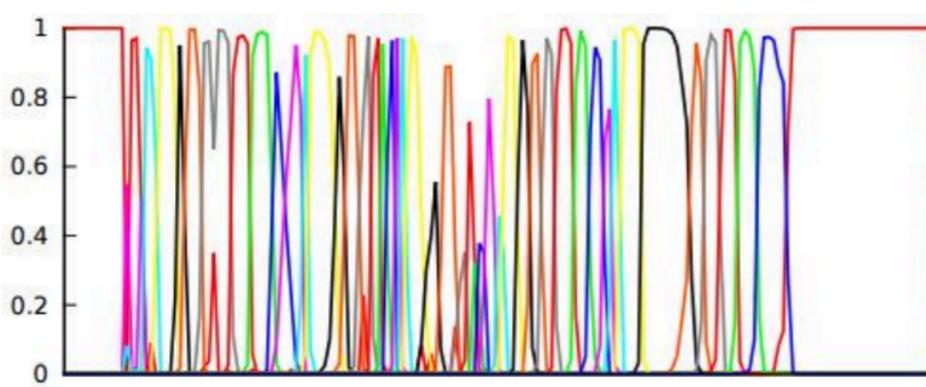
$l = \{a\}$   
aa \_\_\_\_\_  
aaa \_\_\_\_\_  
\_aaa \_\_\_\_\_  
\_\_\_aaa \_\_\_\_\_  
\_aaaaaaa



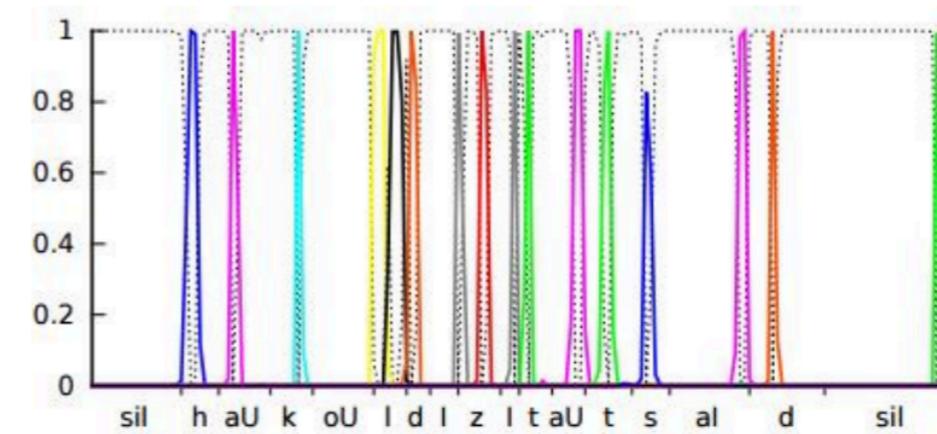
$l = \{bee\}$   
bbbbee\_e\_e  
\_bb\_ee\_\_e  
\_\_bbbe\_e\_



HMM



CTC



**HMM vs CTC**

# Вопросы

- Перечислите три задачи в NLP, кроме распознавания речи.
- Из каких частей состоит распознавание речи?
- Что такое лексическая и языковая модели?
- Зачем нужен СТС? Перечислите два преимущества СТС против скрытых моделей Маркова
- Напишите формулу максимизации пути