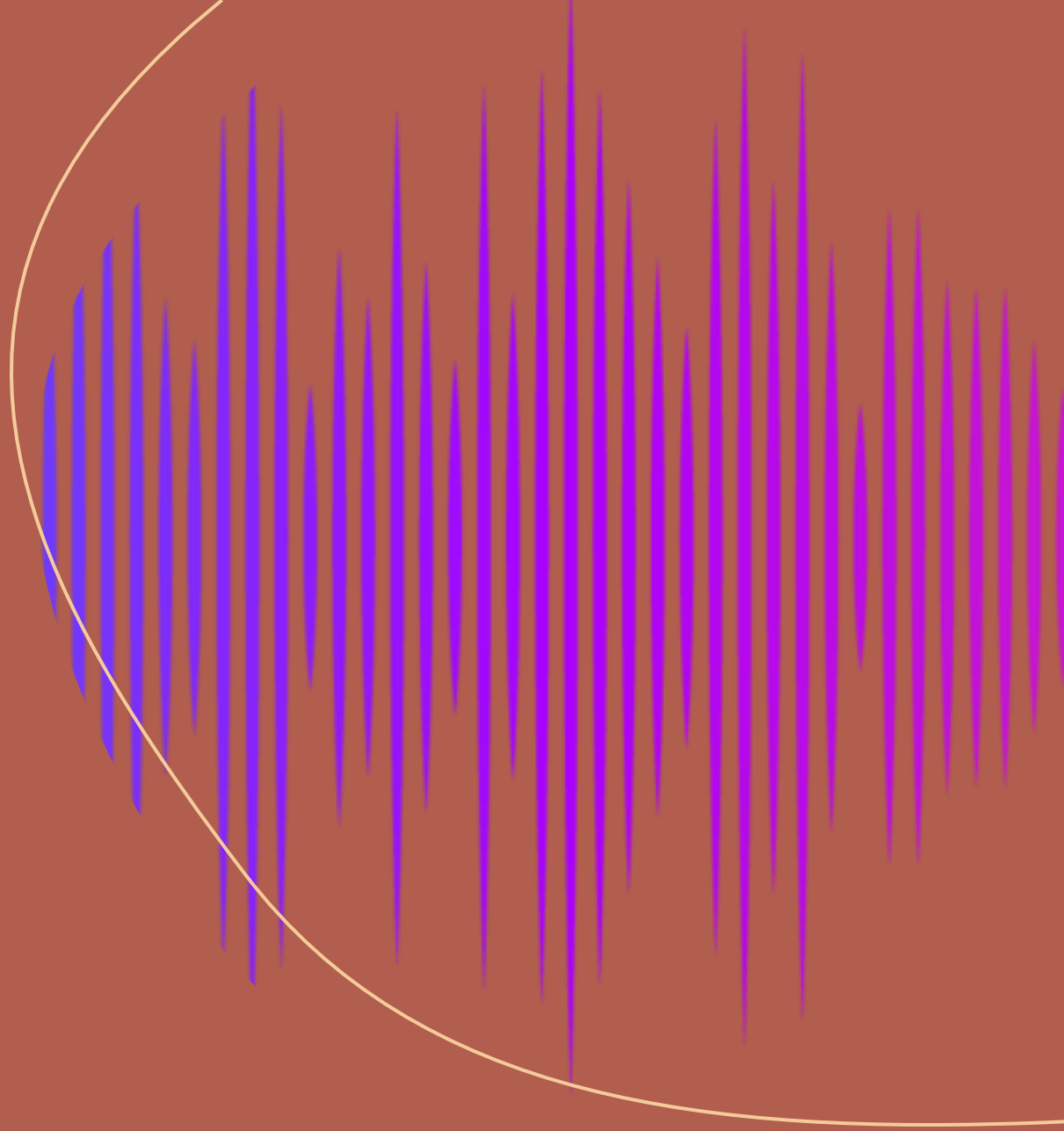
# Natural Language Processing

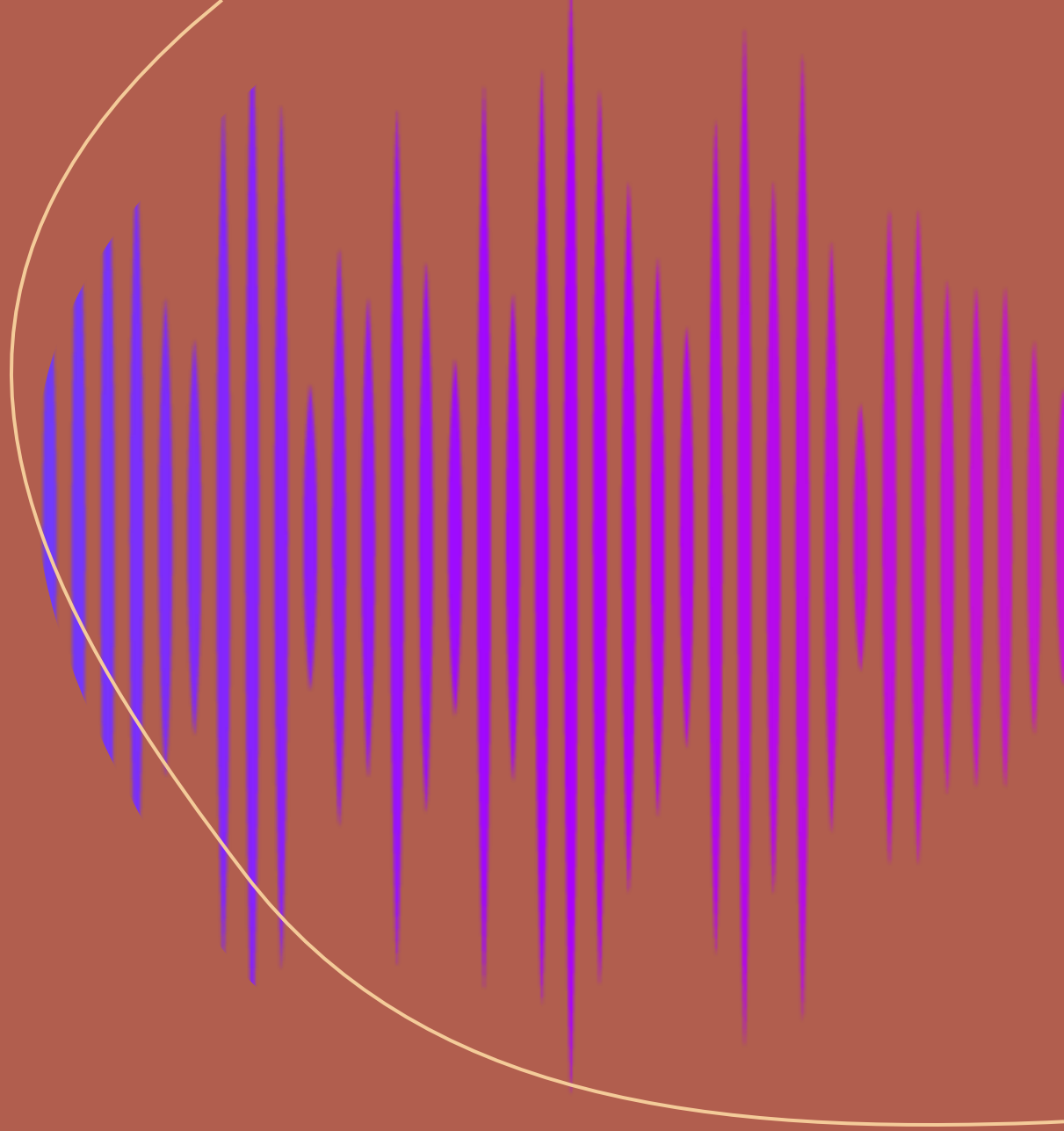
Яковенко Ольга

# Задачи по работе с аудио



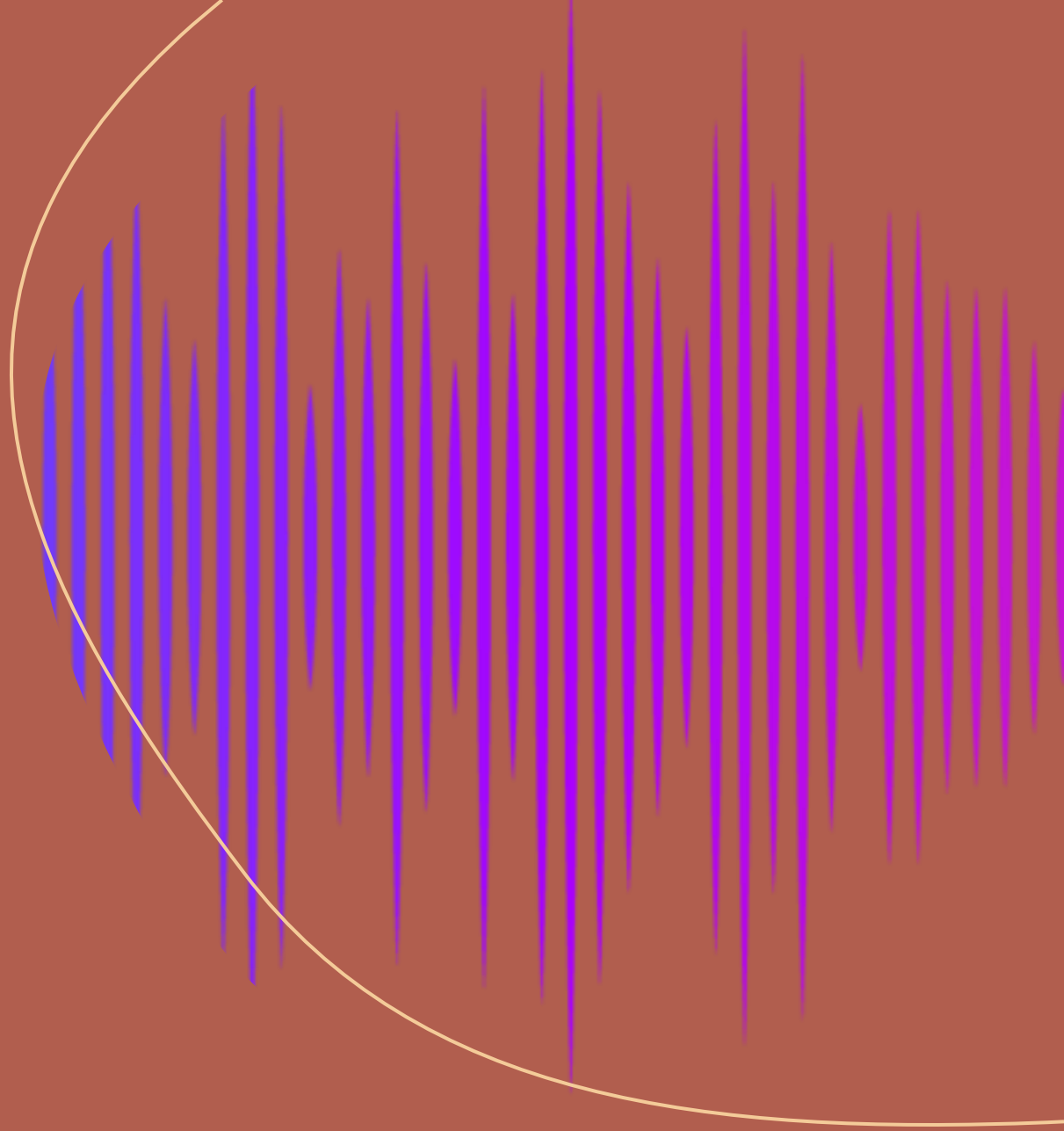
# Задачи по работе с аудио

- Распознавание речи



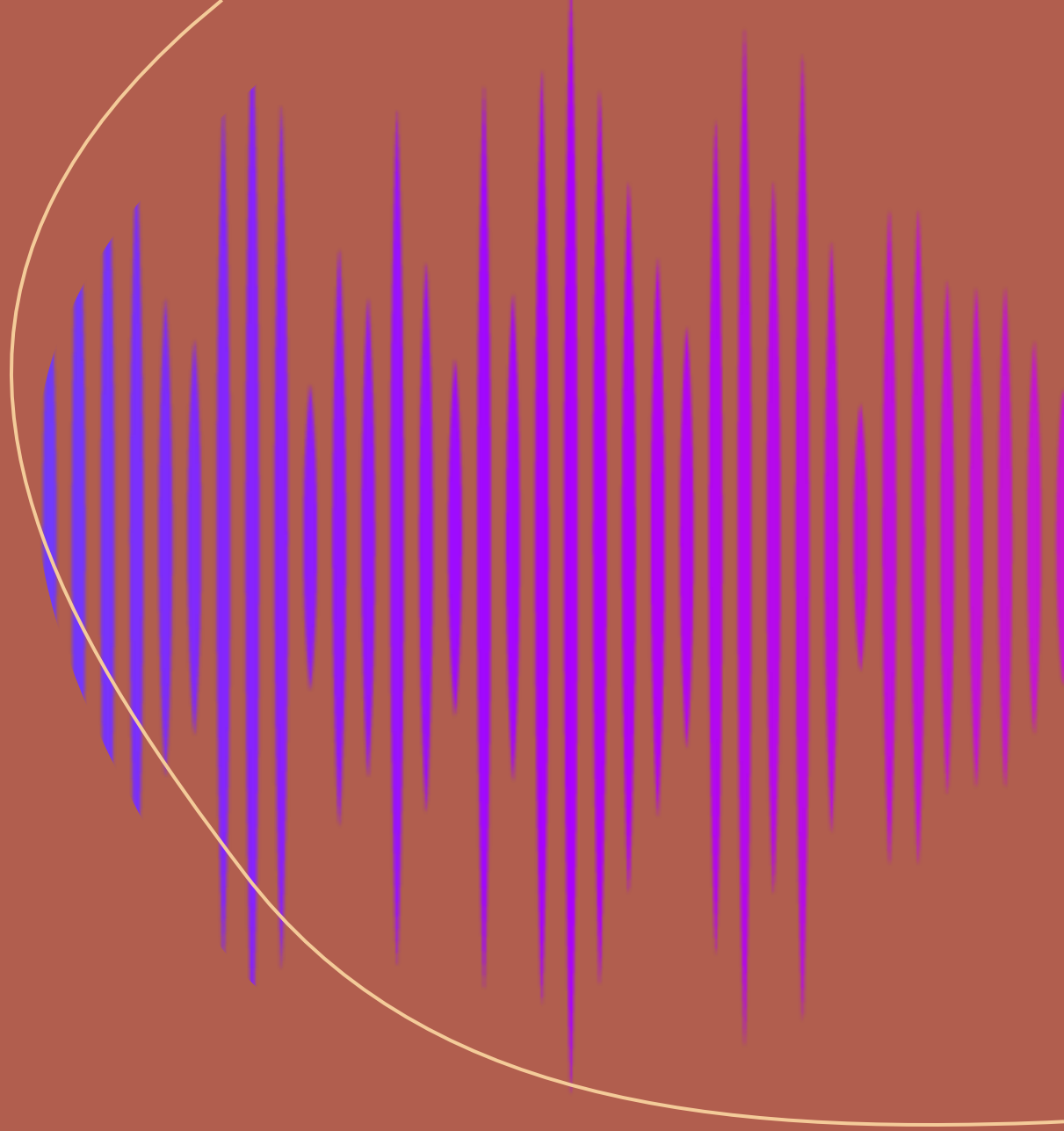
# Задачи по работе с аудио

- Распознавание речи
- Синтез речи



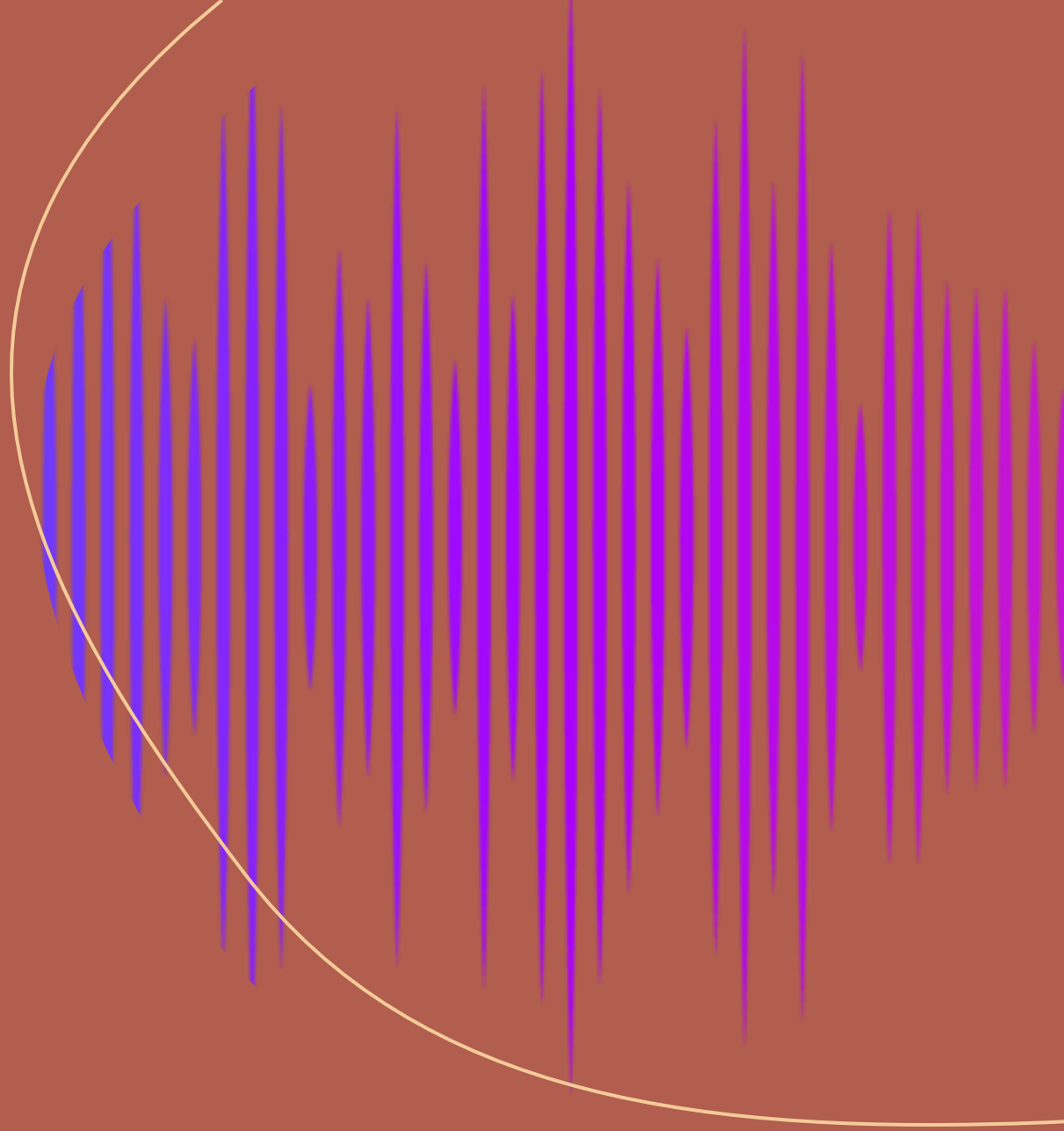
# Задачи по работе с аудио

- Распознавание речи
- Синтез речи
- Верификация/идентификация диктора



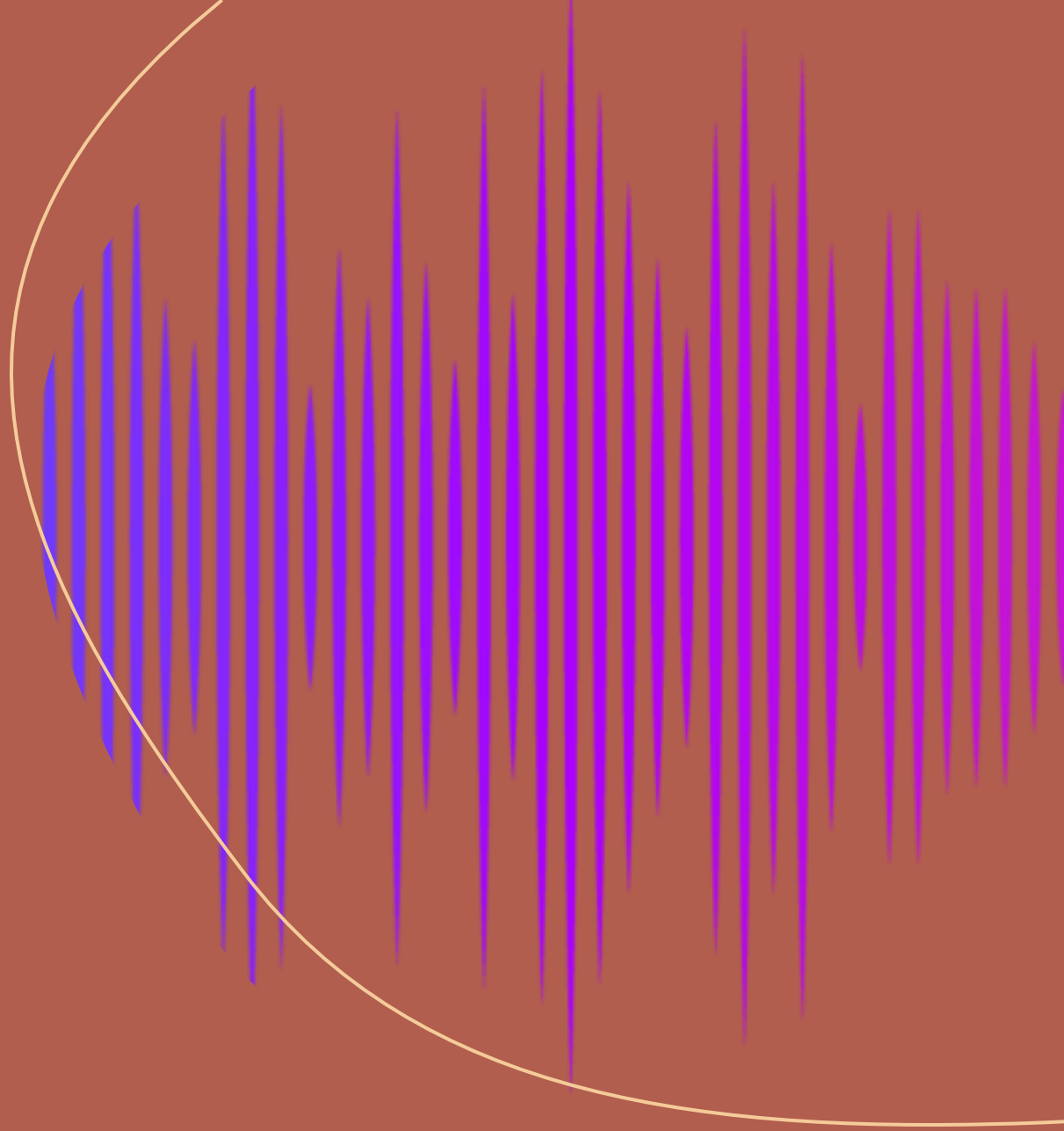
# Задачи по работе с аудио

- Распознавание речи
- Синтез речи
- Верификация/идентификация диктора
- Поиск похожих композиций



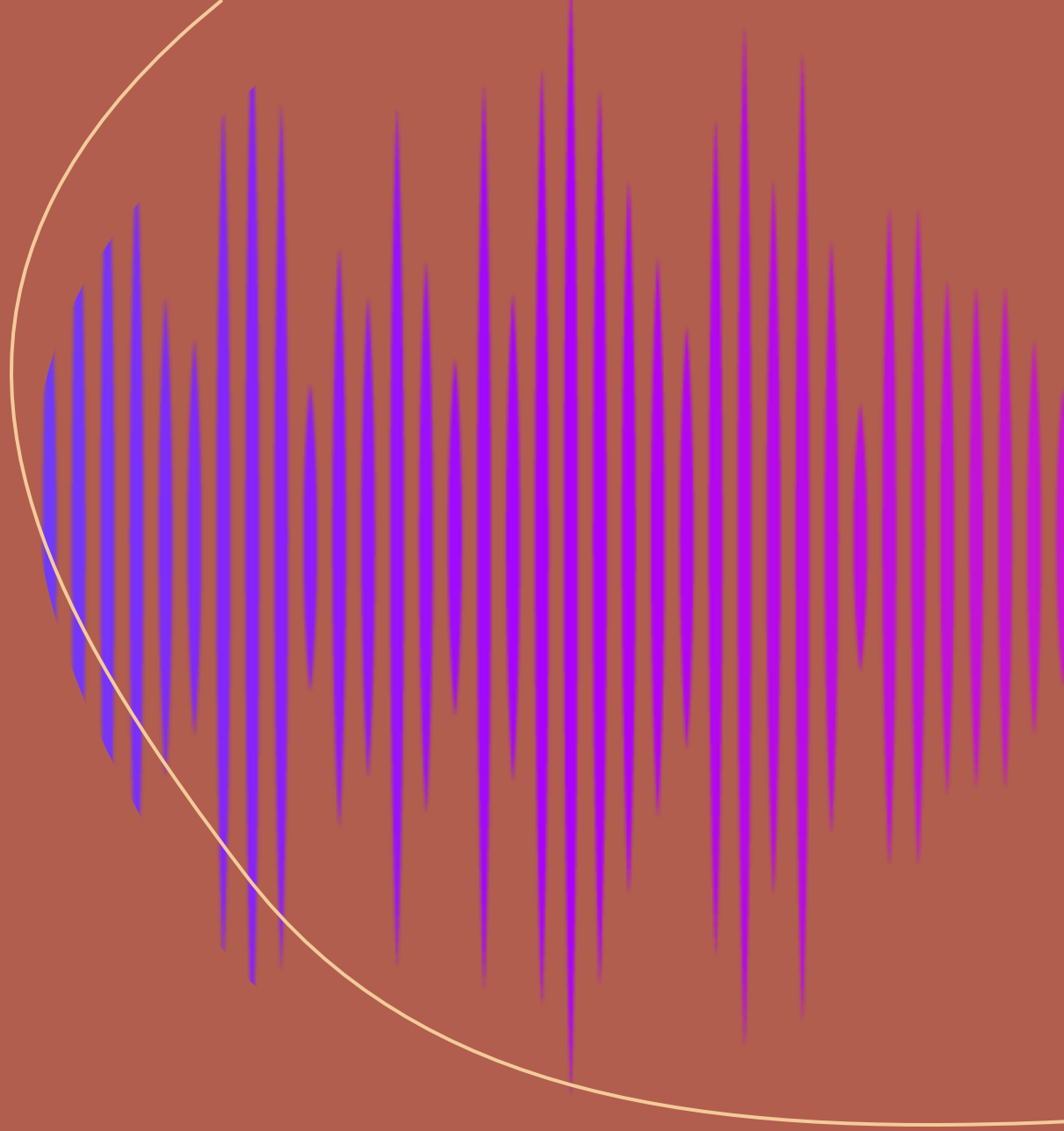
# Задачи по работе с аудио

- Распознавание речи
- Синтез речи
- Верификация/идентификация диктора
- Поиск похожих композиций
- Шумоподавление



# Задачи по работе с аудио

- Распознавание речи
- Синтез речи
- Верификация/идентификация диктора
- Поиск похожих композиций
- Шумоподавление



# Распознавание речи

# Распознавание речи

Автоматическое обнаружение в аудио произносимого человеком текста

# Распознавание речи

## Компонентный подход

## End-to-End подход

# Компонентный подход

## Входной аудио поток

WAV файл

## Извлечение признаков

Мел-частотные кепстральные коэффициенты (MFCC)

I-vectors

Кепстральная нормализация (CMVN)

## Акустический блок

Оконное распознавание звуков

Определение наиболее вероятной цепочки звуков

## Лингвистический блок

Языковая модель

## Выходной текст

Строка

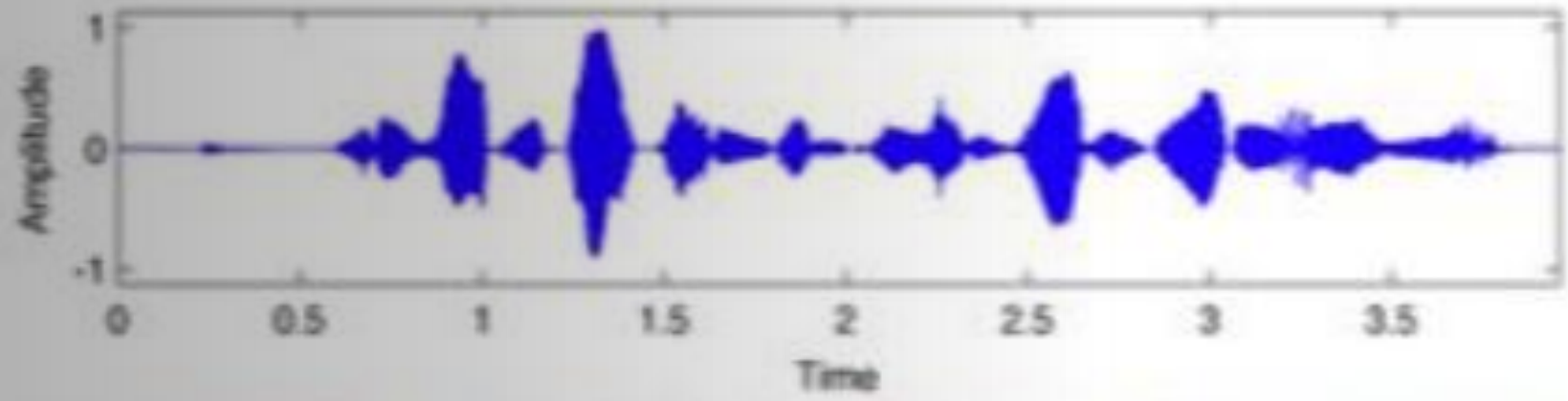
# Входной поток аудио

- WAV формат:
- Частота дискретизации (8 кГц, 16 кГц, 44 кГц)
- Количество каналов
- Битовая глубина

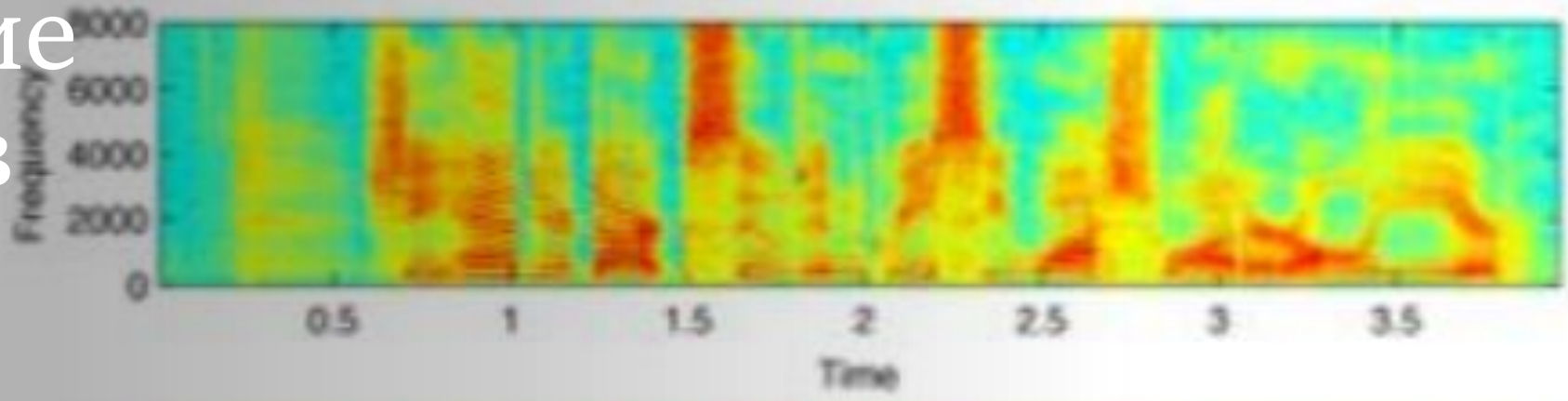
```
-bash-4.2$ soxi audio.wav

Input File     : 'audio.wav'
Channels       : 1
Sample Rate    : 8000
Precision      : 16-bit
Duration       : 01:31:33.97 = 43951752 samples ~ 412048 CDDA sectors
File Size      : 87.9M
Bit Rate       : 128k
Sample Encoding: 16-bit Signed Integer PCM
```

# Извлечение признаков

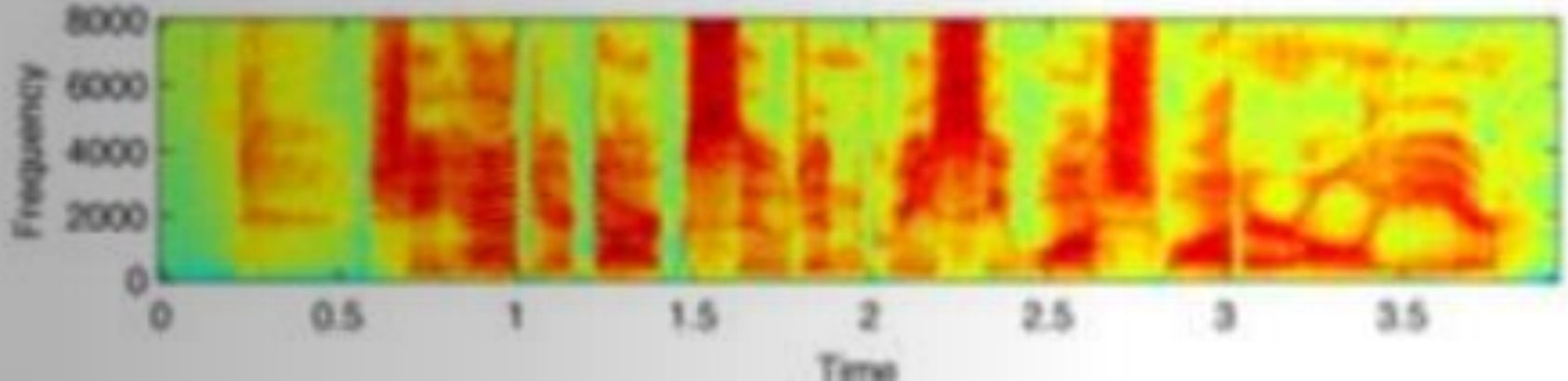
**Time Domain Waveform**

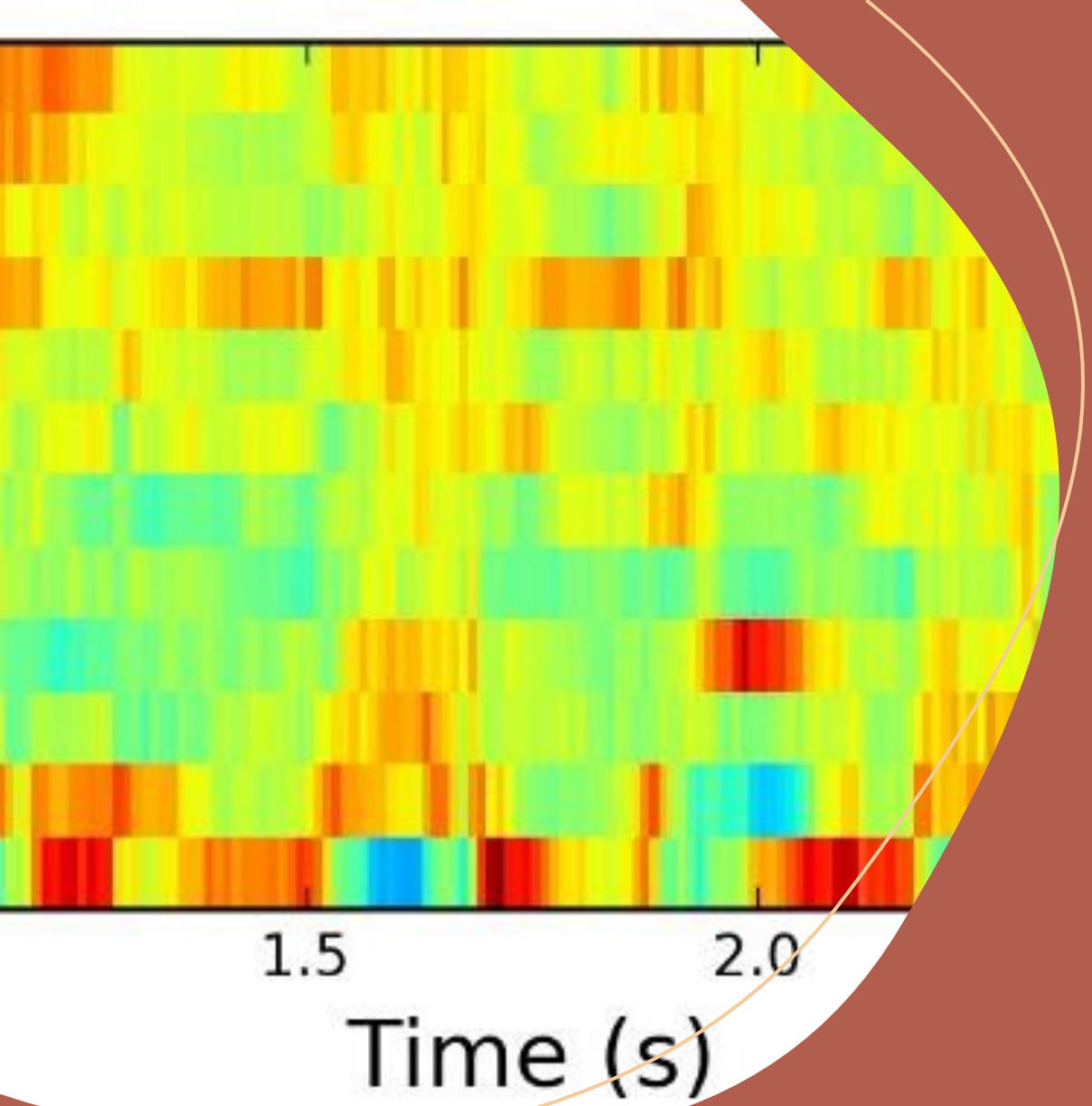



**Spectrogram**




**MFCC Spectrogram**



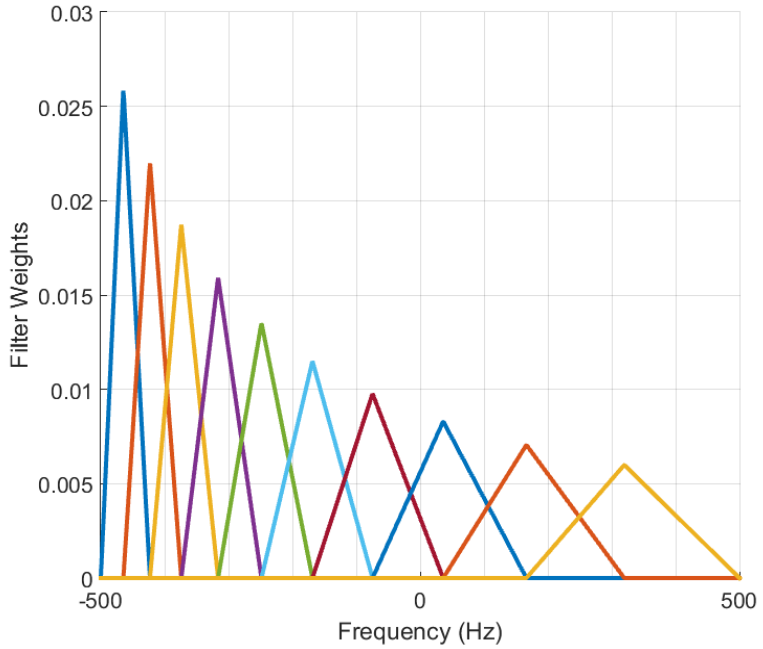


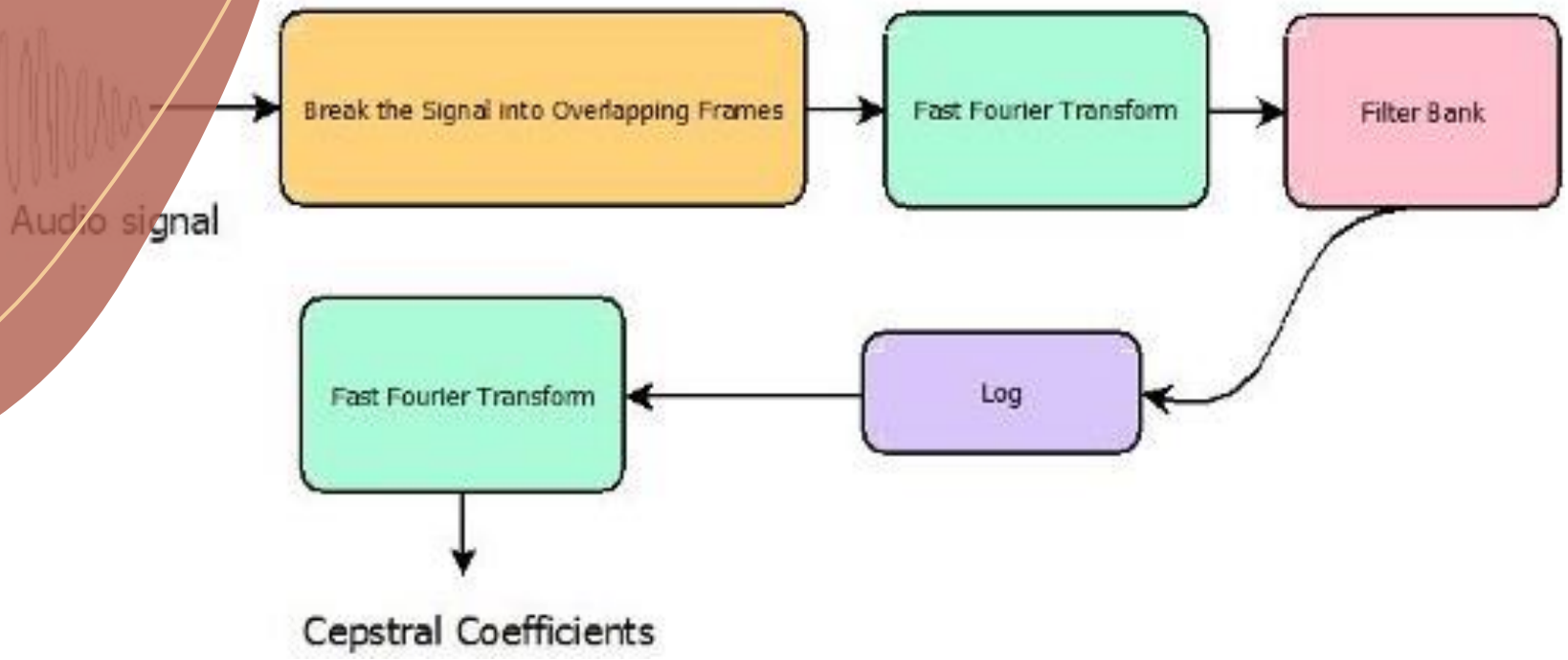

# Извлечение признаков

- Мел-частотные кепстральные коэффициенты
- Mel-frequency cepstral coefficients
- MFCC

# Извлечение признаков

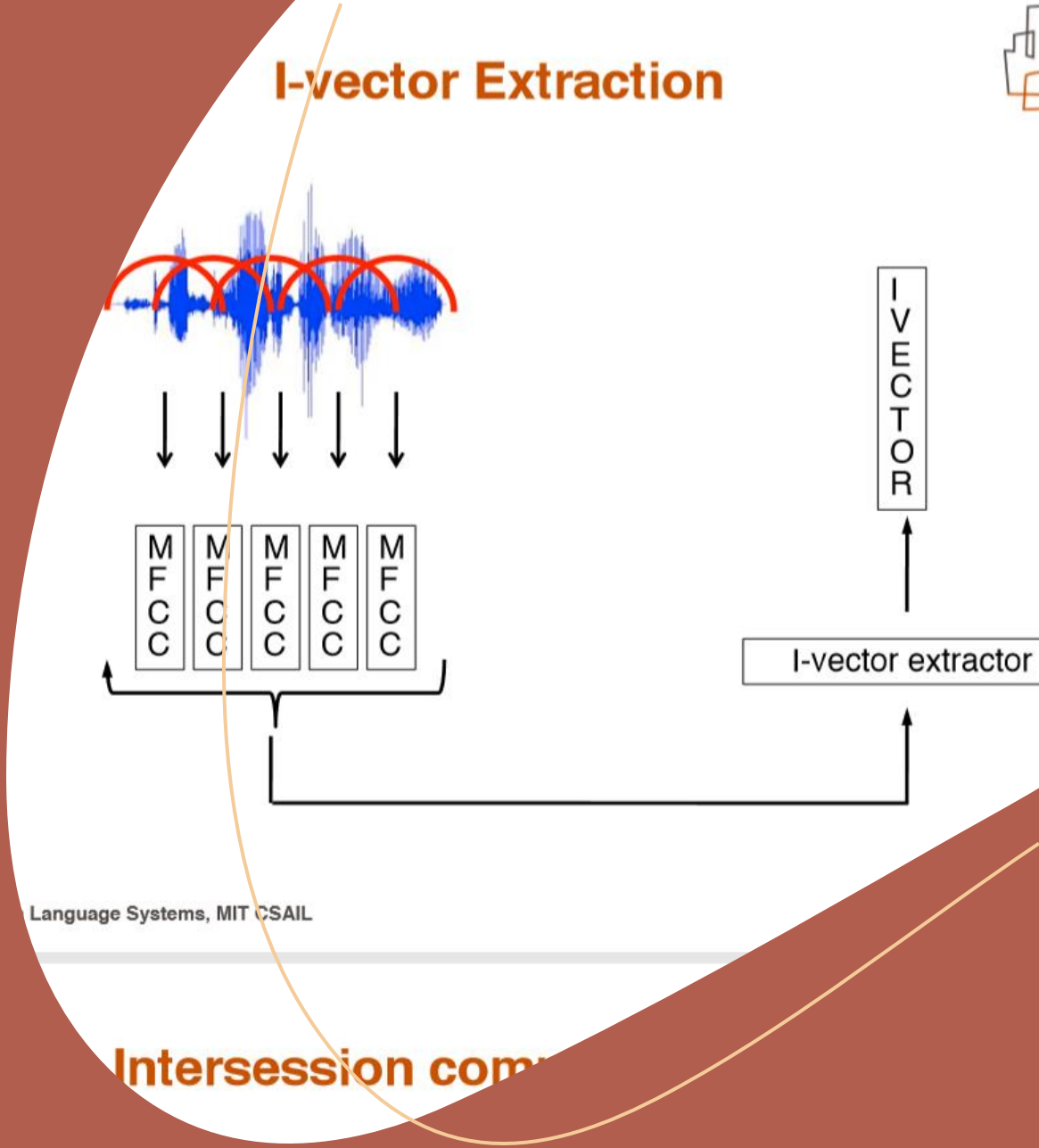
MFCC

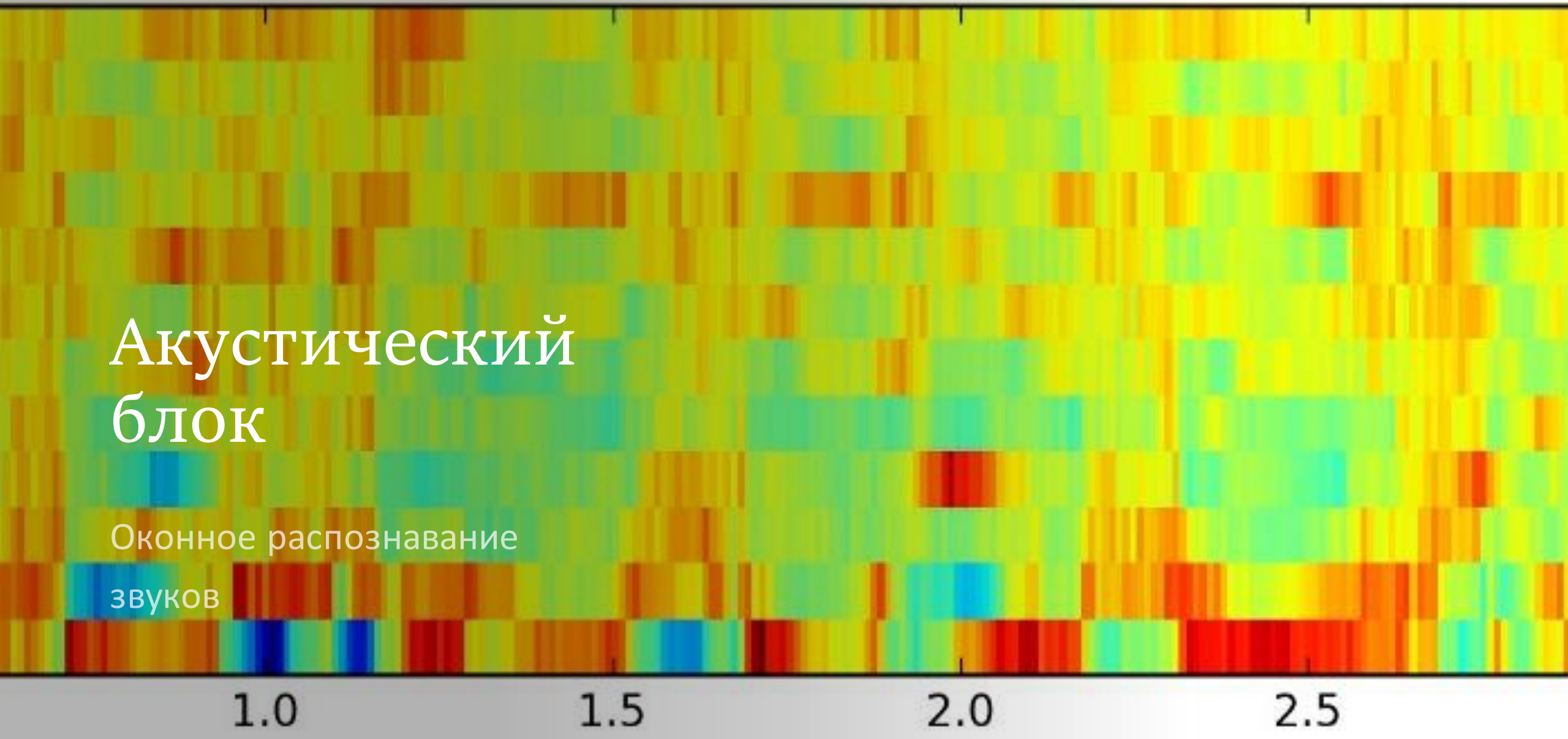


# Извлечение признаков

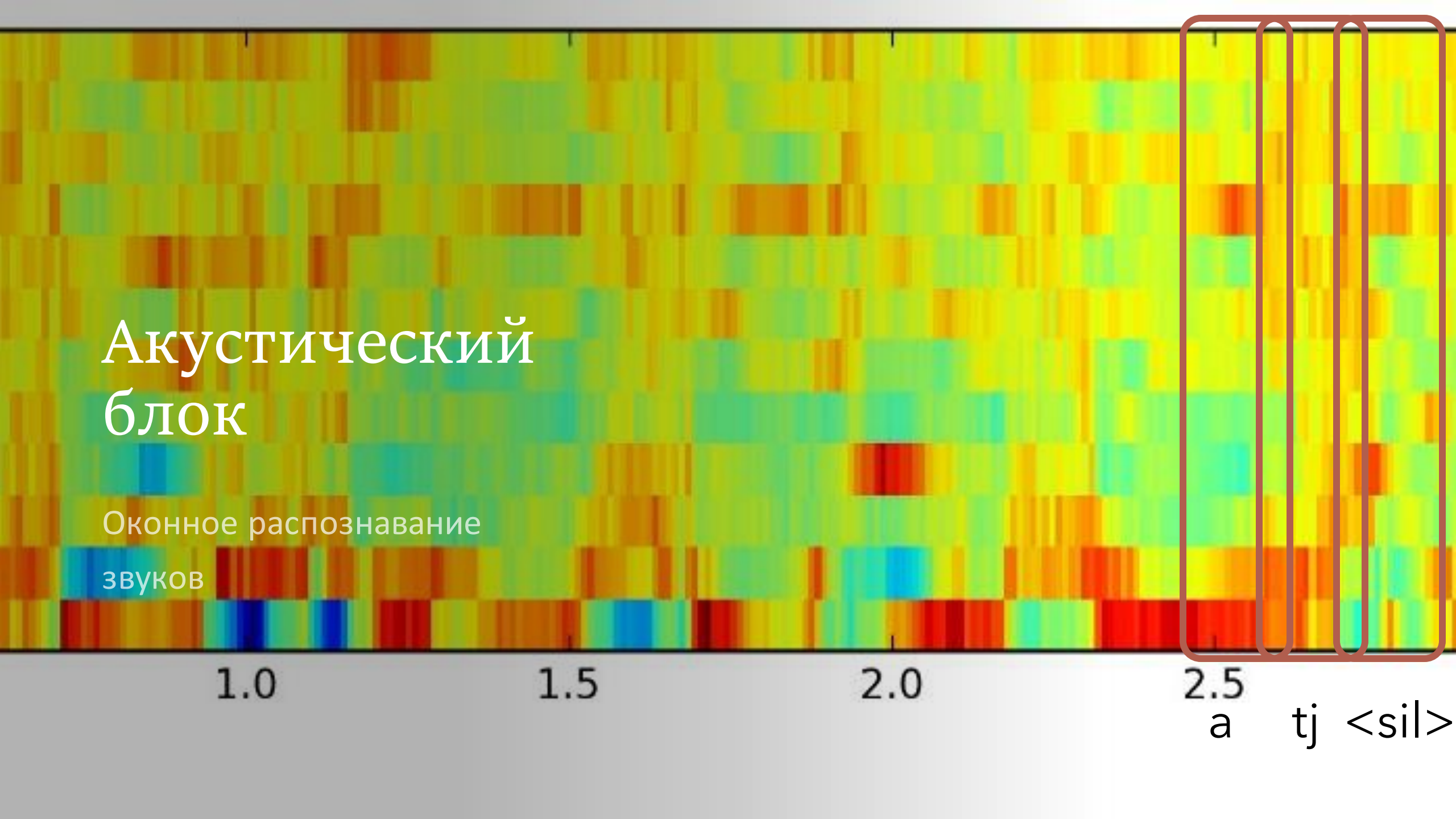
- Identity vectors
- I-vectors






# Акустический блок

Оконное распознавание звуков


Акустический блок
Оконное распознавание звуков
1.0
1.5
2.0
2.5
a
tj
<sil>

# Акустический блок

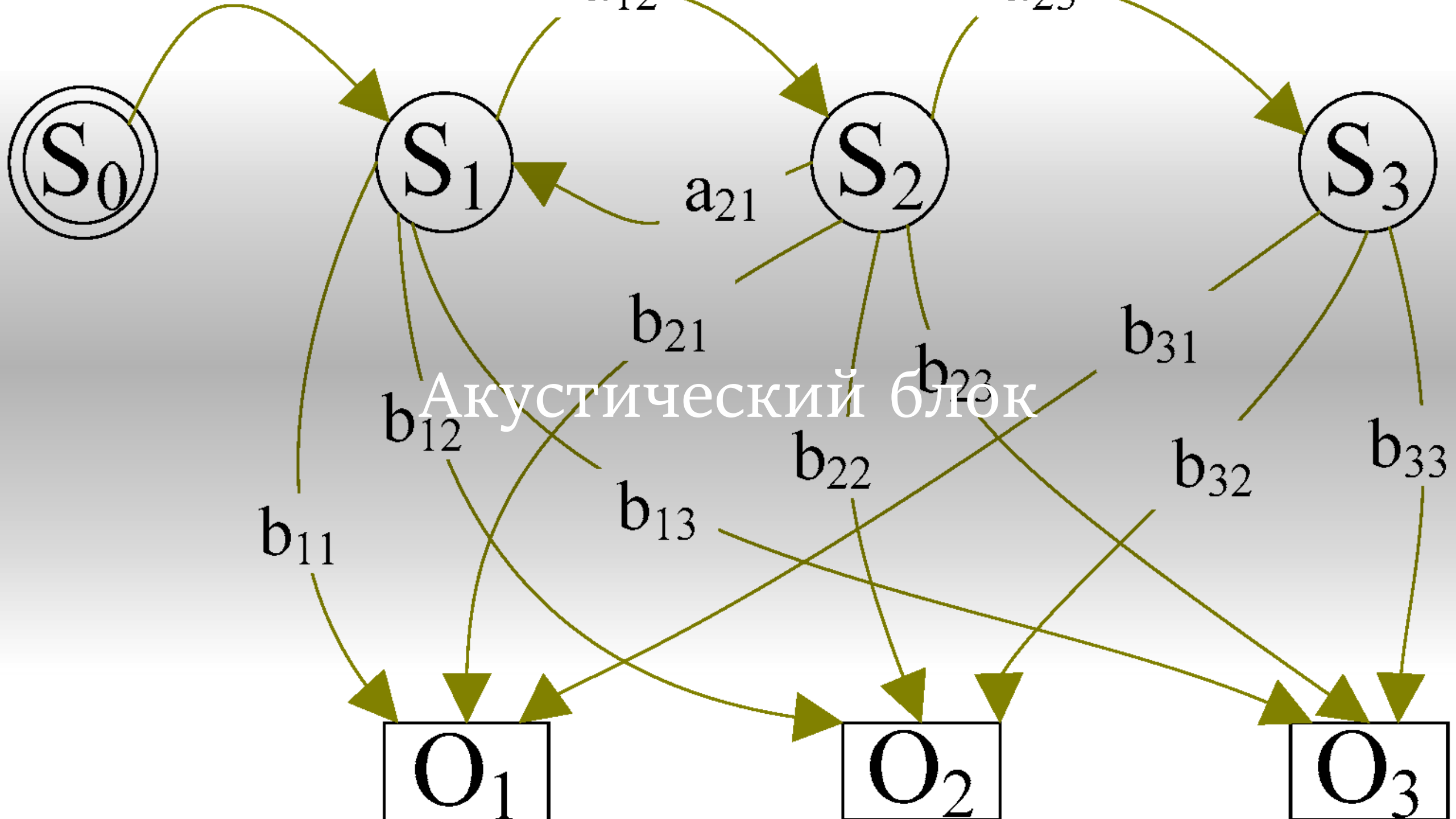
# Акустический блок

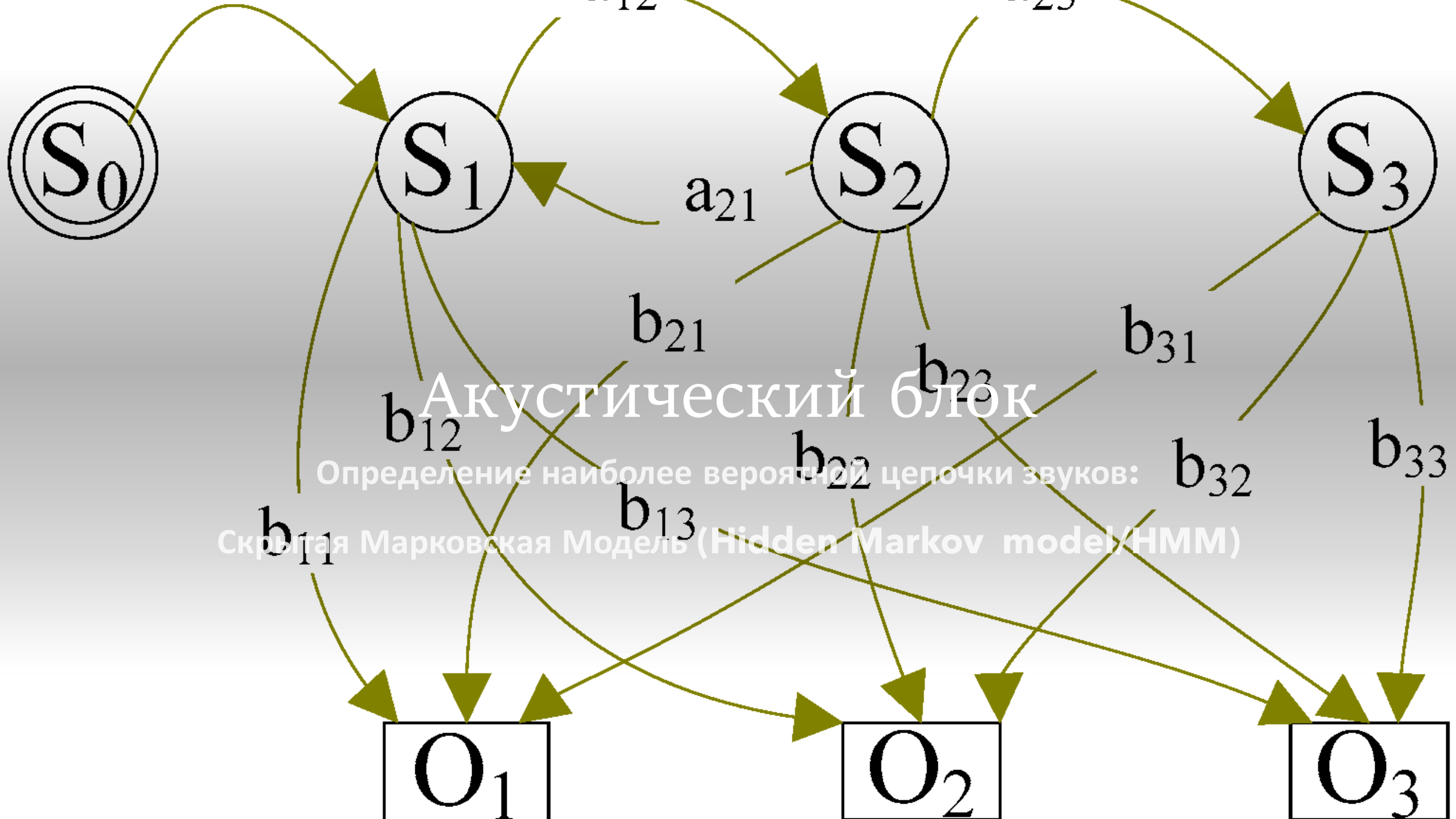
## Оконное распознавание звуков

- Гауссовы смеси (Gaussian mixture models, GMM)
- Глубокие нейронные сети (Deep neural networks, DNN)

# Акустический блок

Определение наиболее вероятной цепочки **звуков**


$S_0$
$S_1$
$S_2$
$S_3$
$a_{21}$
$b_{11}$
$b_{12}$
$b_{13}$
$b_{21}$
$b_{22}$
$b_{23}$
$b_{31}$
$b_{32}$
$b_{33}$
$O_1$
$O_2$
$O_3$
Акустический блок




# Акустический блок

**Определение наиболее вероятной цепочки звуков:**

**Скрытая Марковская Модель (Hidden Markov model)(HMM)**

# Лингвистический блок

# Лингвистический блок

Языковая модель

- N-граммная языковая модель
- Реккурентная нейронная сеть в качестве языковой модели (Reccurent neural network language model/RNNLM)


# Лингвистический блок

Языковая модель

- N-граммная языковая модель
- Реккурентная нейронная сеть в качестве языковой модели (Reccurent neural network language model/RNNLM)

# Компонентный подход

- CMUSphinx
- HTK
- Kaldi

# End-to-End подход

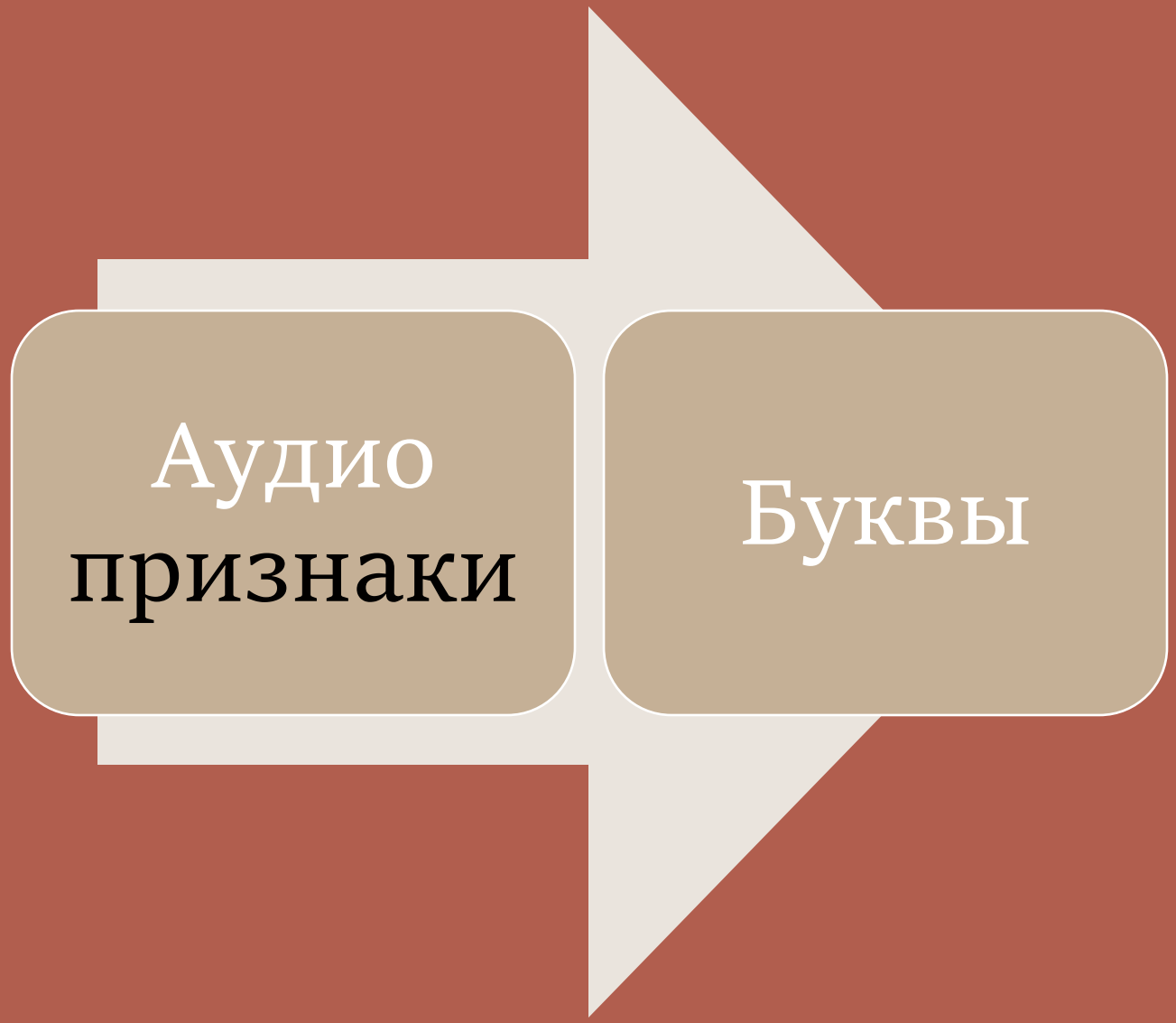
| Входной аудио поток | Извлечение признаков | Акустический блок | Лингвистический блок | Выходной текст |
| --- | --- | --- | --- | --- |
| WAV файл | Мел-частотные кепстральные коэффициенты (MFCC)<br>I-vectors<br>Кепстральная нормализация (CMVN) | Глубокие нейронные сети (DNN) | Статистическая языковая модель (n-gram LM) | Строка |

# Акустический блок

Определение наиболее вероятной цепочки **букв**

# Акустический блок

Определение наиболее вероятной цепочки **букв**



# End-to-End подход

- Kaldi
- DeepSpeech
- Wav2letter

# Практика

https://github.com/DinoTheDinosaur/FocusStart_NLP/blob/master/notebooks/Speech_recognition_and_sentiment.ipynb