

# РАСПОЗНАВАНИЕ ЭМОЦИЙ ЧЕЛОВЕКА ПО АУДИОЗАПИСИ ГОЛОСА

ПОПОВА А. С.

РАССАДИНА А. Г.

ПОНОМАРЕНКО А. А.

НИУ ВШЭ НН

2018

# ПОСТАНОВКА ЗАДАЧИ



# МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

$$y: X \rightarrow Y$$

X - это множество всевозможных описаний объектов  
Y - конечное множество классов

Существует неизвестная целевая зависимость  $y$  – отображение, значения которой известны только на объектах конечной обучающей выборки

$$X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}.$$

# МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Требуется построить алгоритм А: X->Y, способный классифицировать произвольный объект x из выборки X.

Для изображений искомое отображение:

$$y: R^n \rightarrow Y$$

где n - общее число признаков или в случае распознавания аудио сигнала, число сэмплов в окне распознавания.

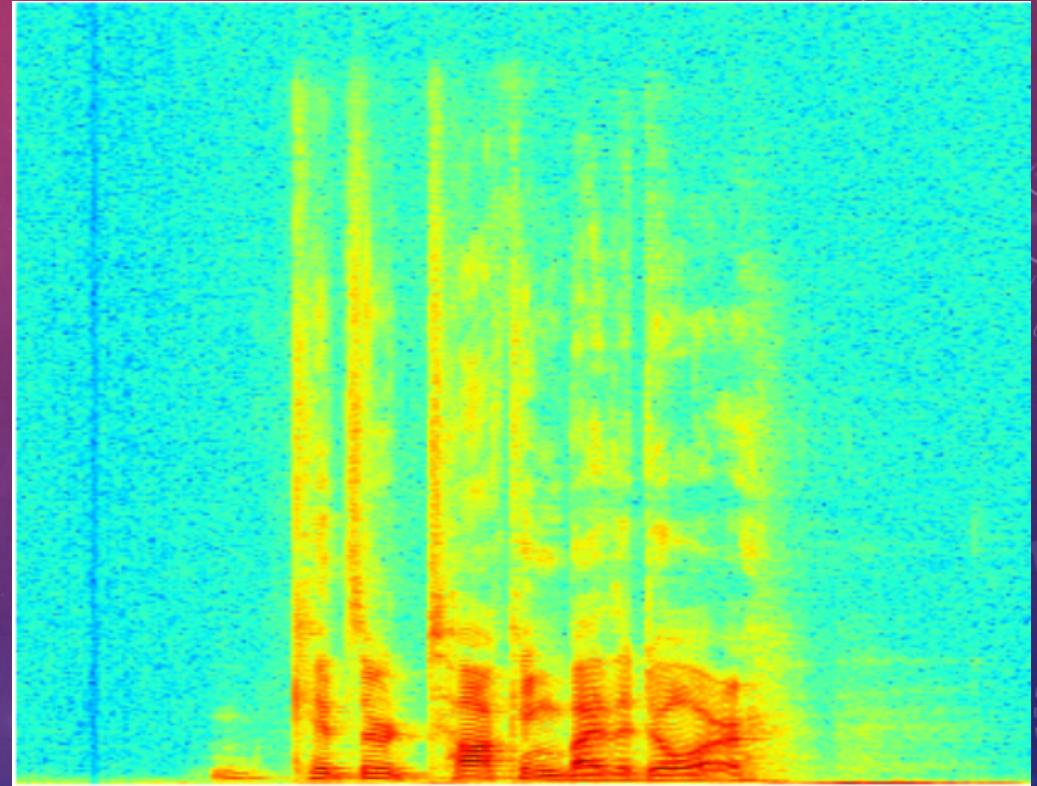
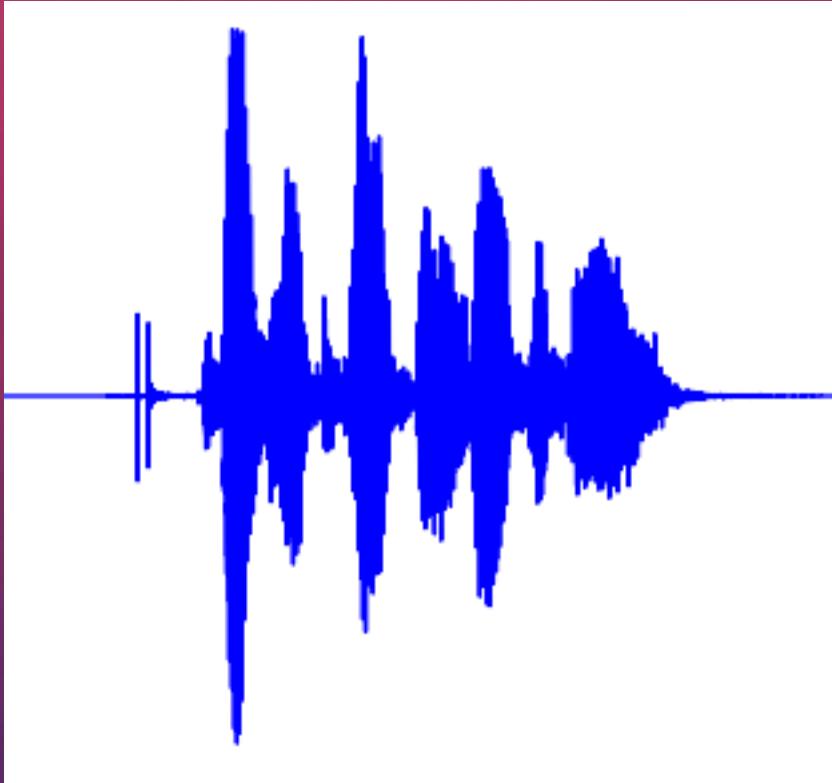
# СБОР ДАННЫХ (RAVDESS DATABASE)

►	1neutral
►	2calm
►	3happy
►	4sad
►	5angry
►	6fearful
►	7disgust
►	8surprised

	03-01-01-01-01-01.wav
	03-01-01-01-01-02.wav
	03-01-01-01-01-03.wav
	03-01-01-01-01-04.wav
	03-01-01-01-01-05.wav
	03-01-01-01-01-06.wav
	03-01-01-01-01-07.wav
	03-01-01-01-01-08.wav

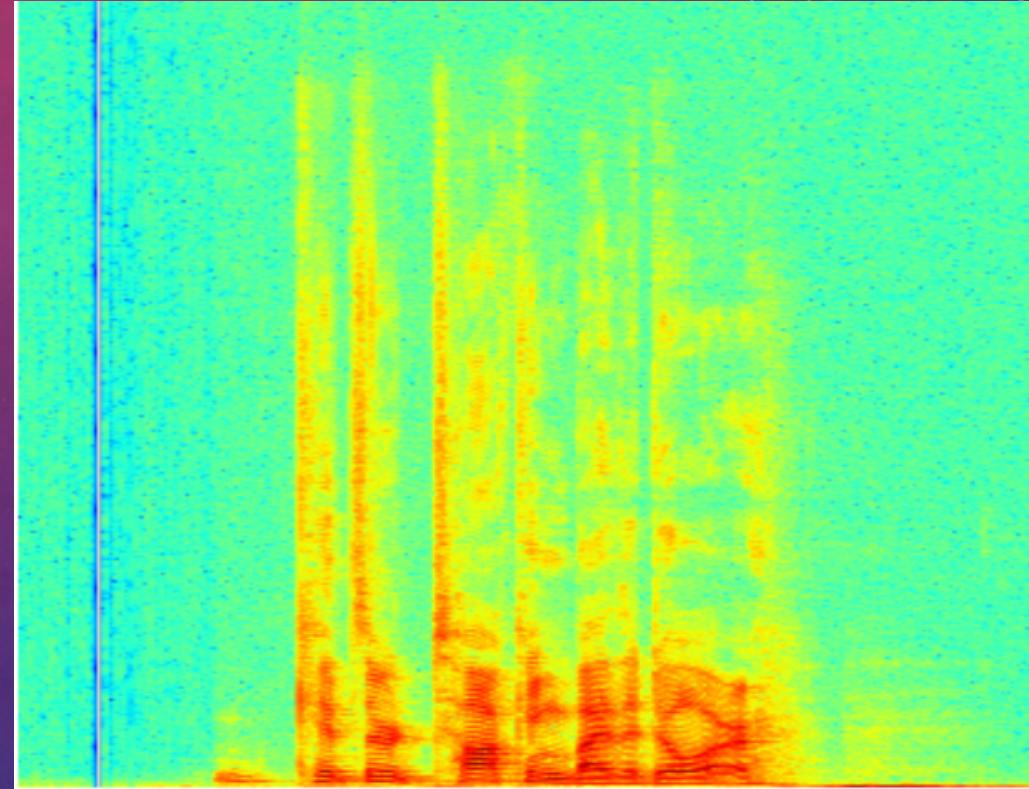
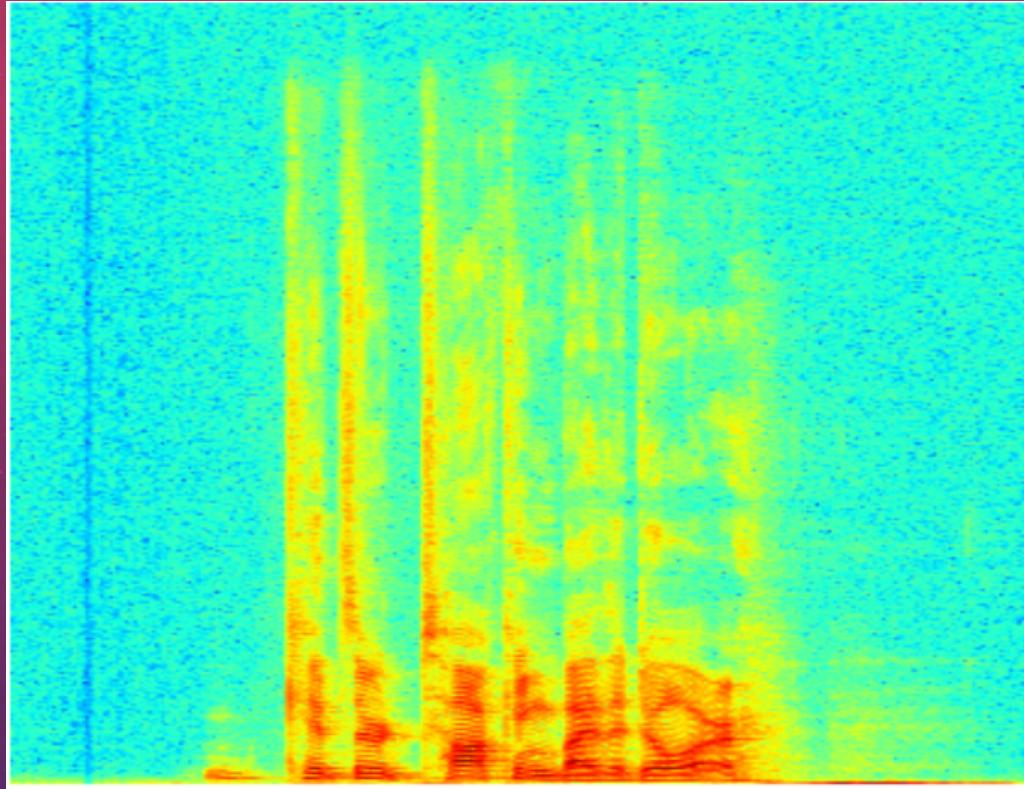
<http://neuron.arts.ryerson.ca/ravdess/?f=3>

# ПРЕДОБРАБОТКА



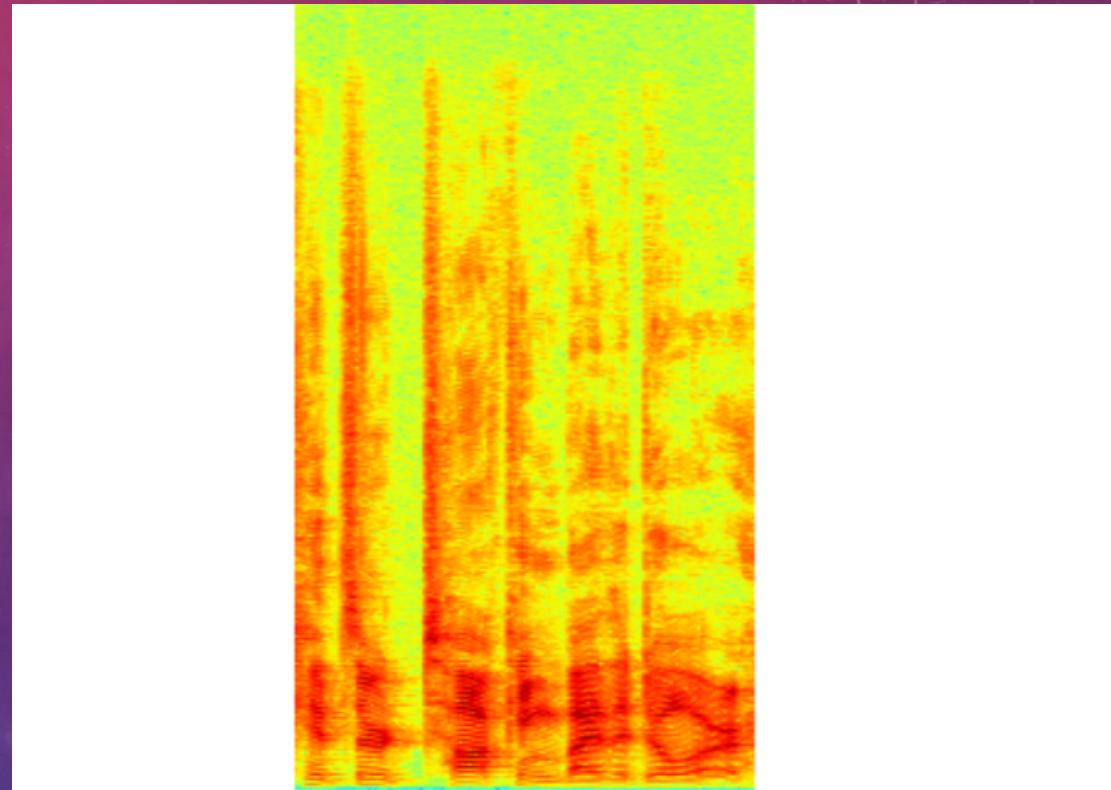
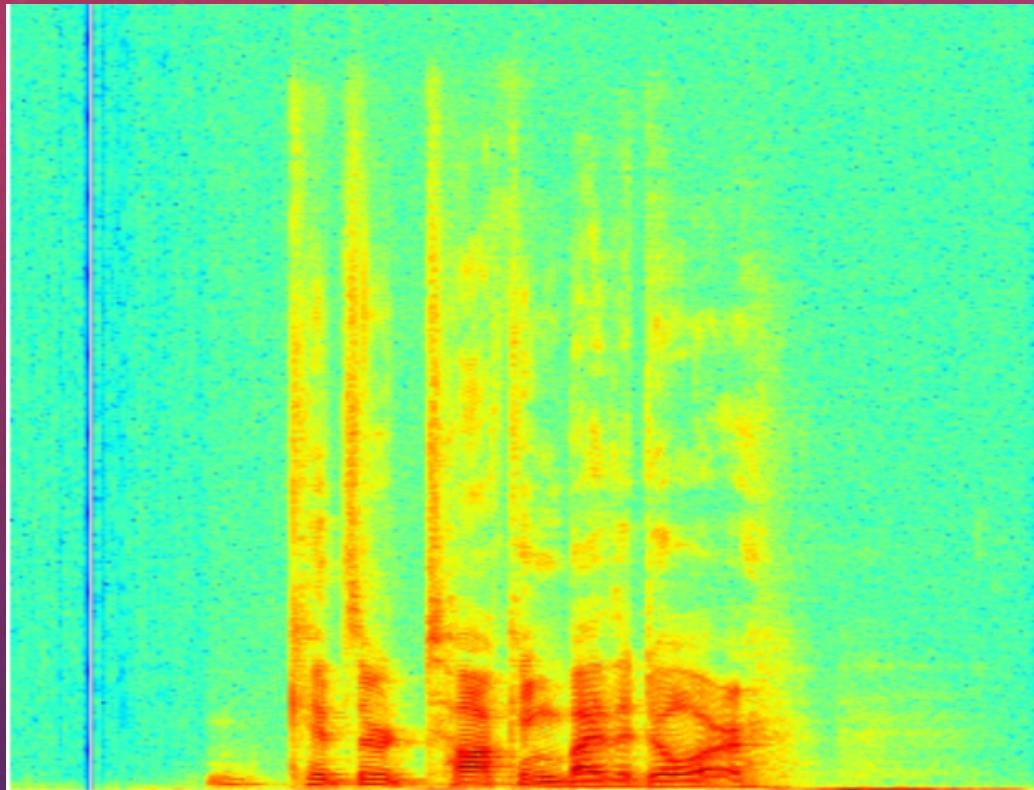
Выравнивание длины

# ПРЕДОБРАБОТКА



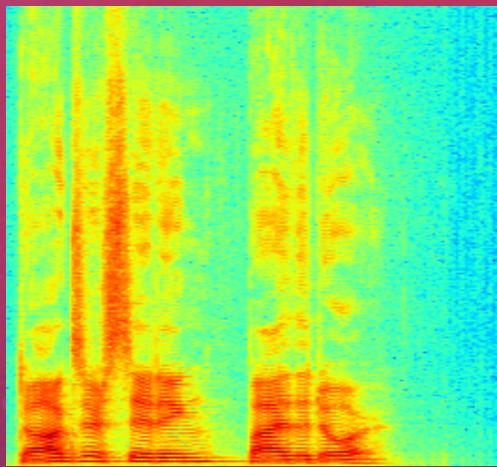
Нормализация громкости

# ПРЕДОБРАБОТКА

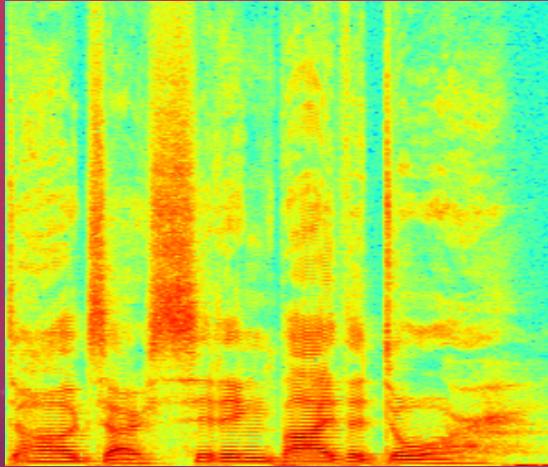


Highpass&Lowpass фильтры, выделение голоса (VAD)

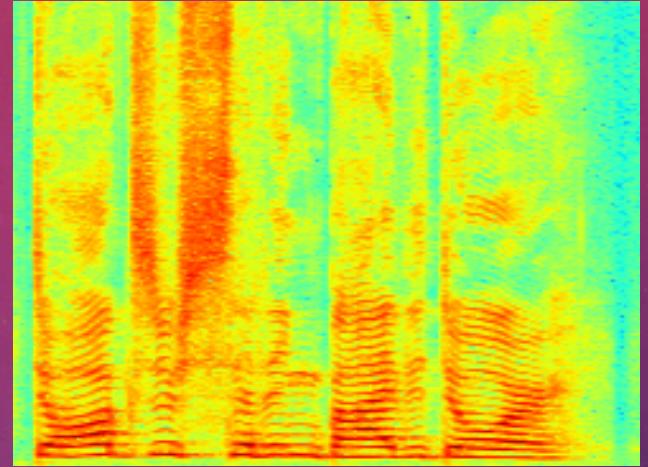
# ОБУЧАЮЩАЯ ВЫБОРКА



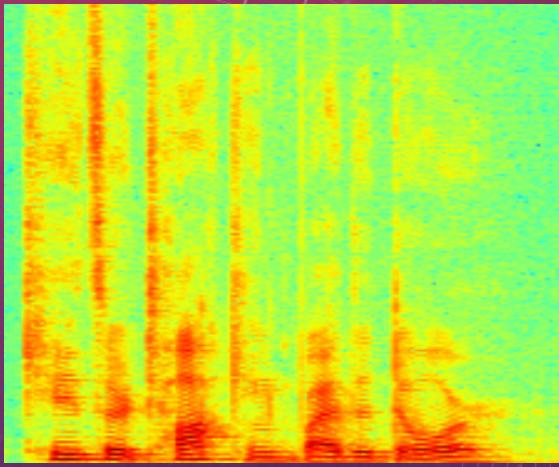
neutral



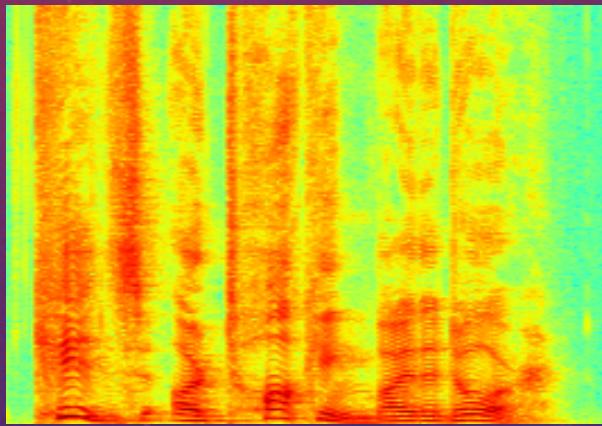
calm



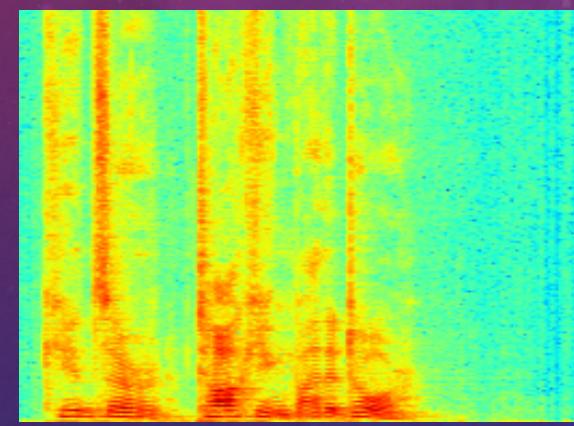
happy



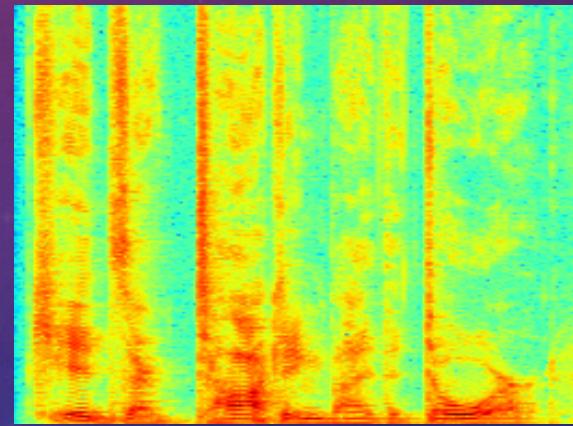
sad



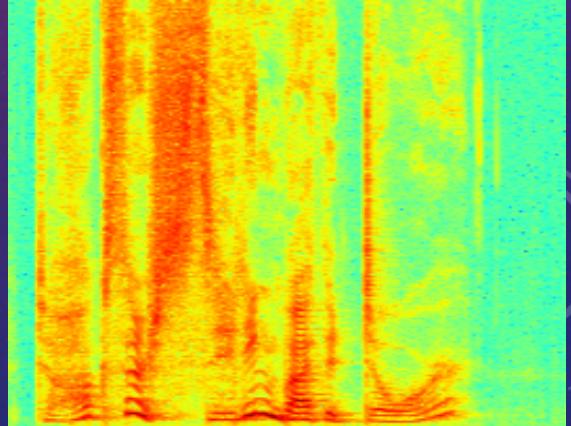
angry



fearful

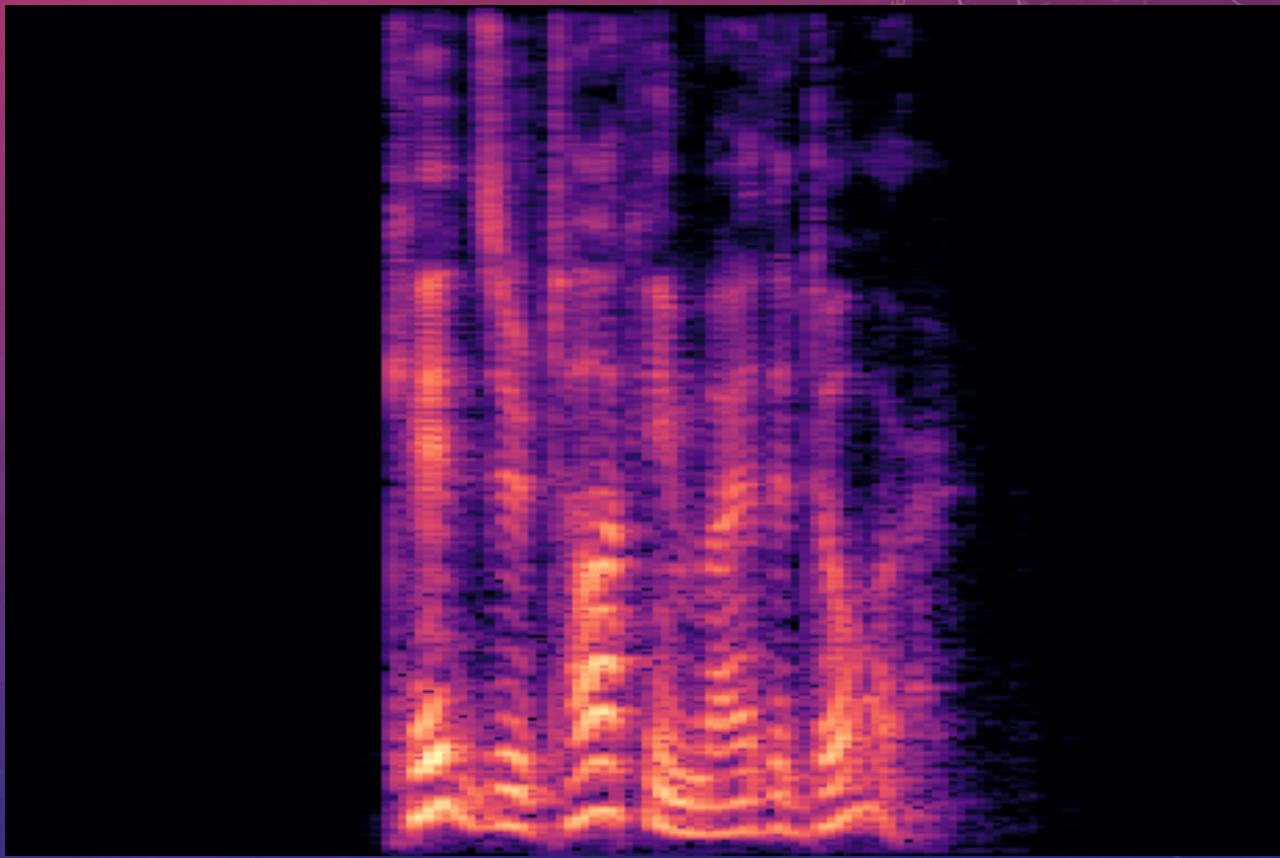
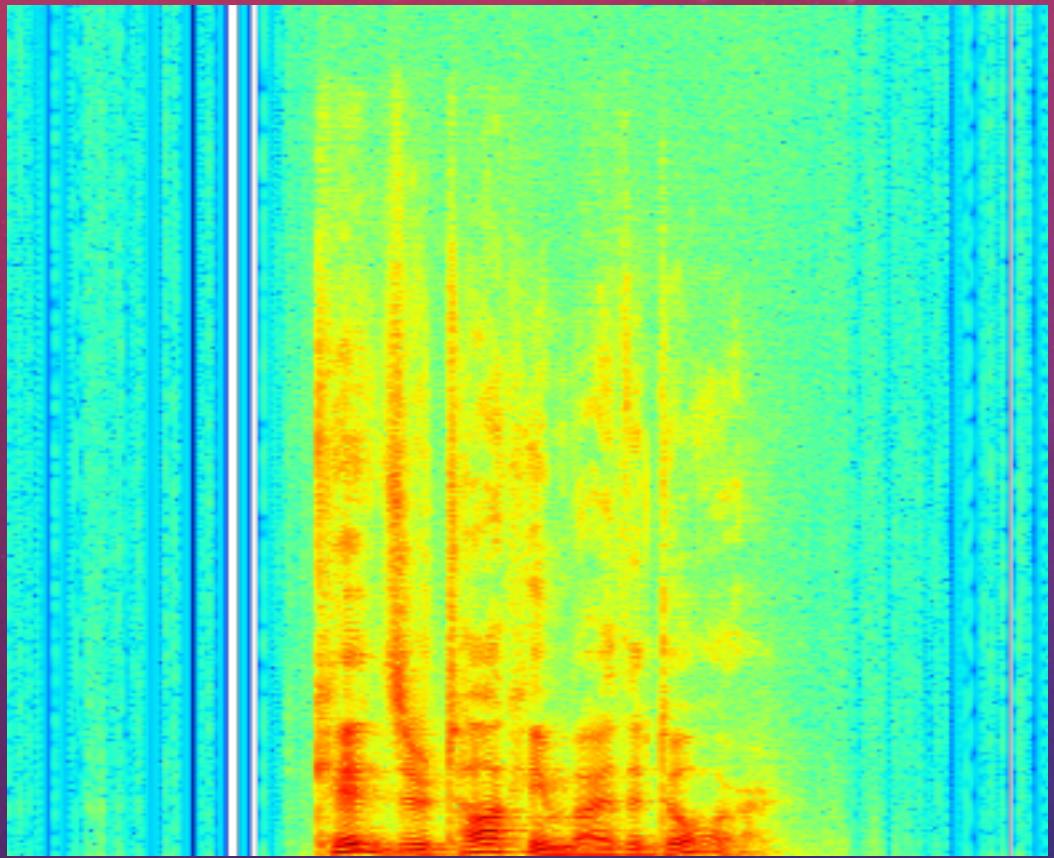


disgust

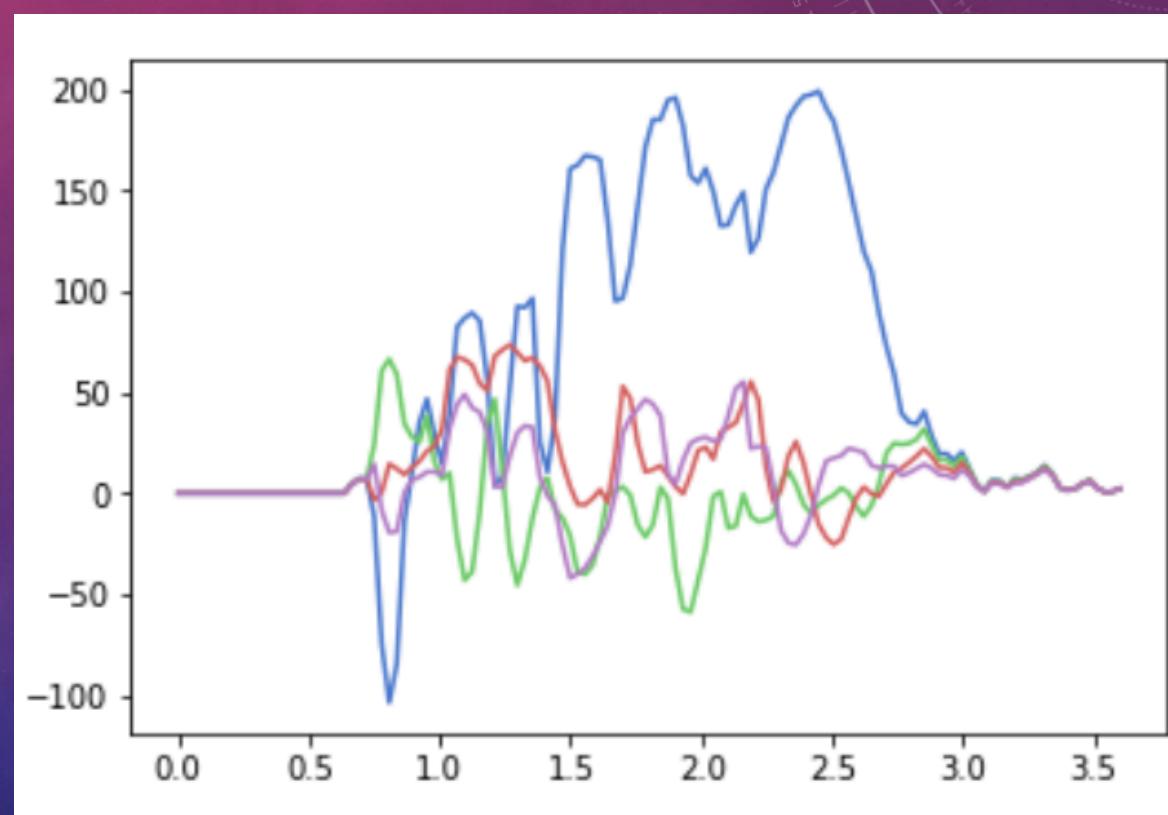
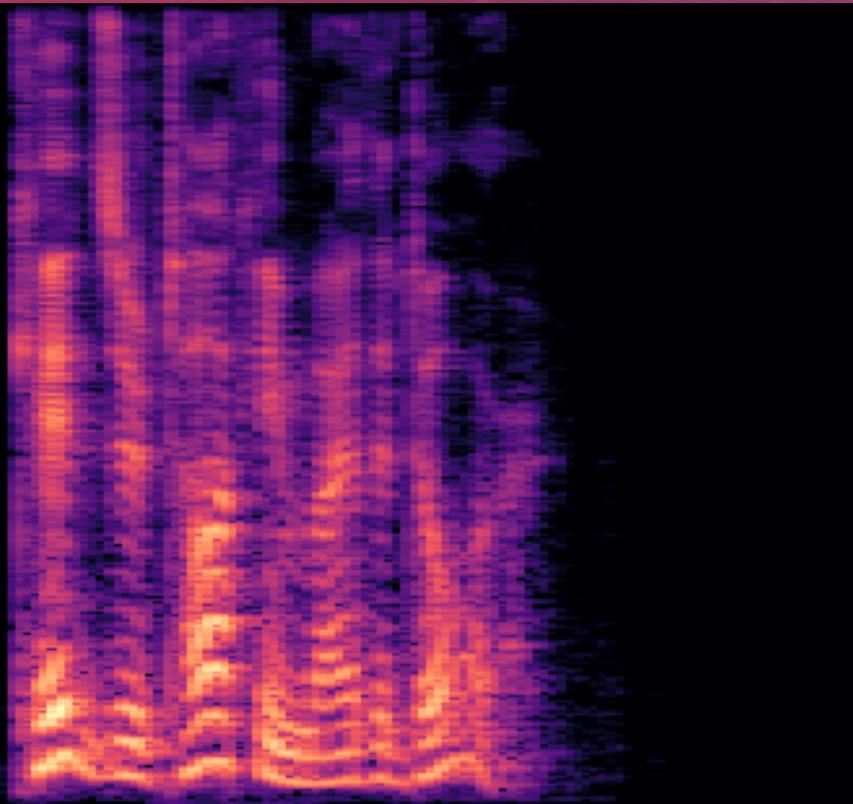


surprised

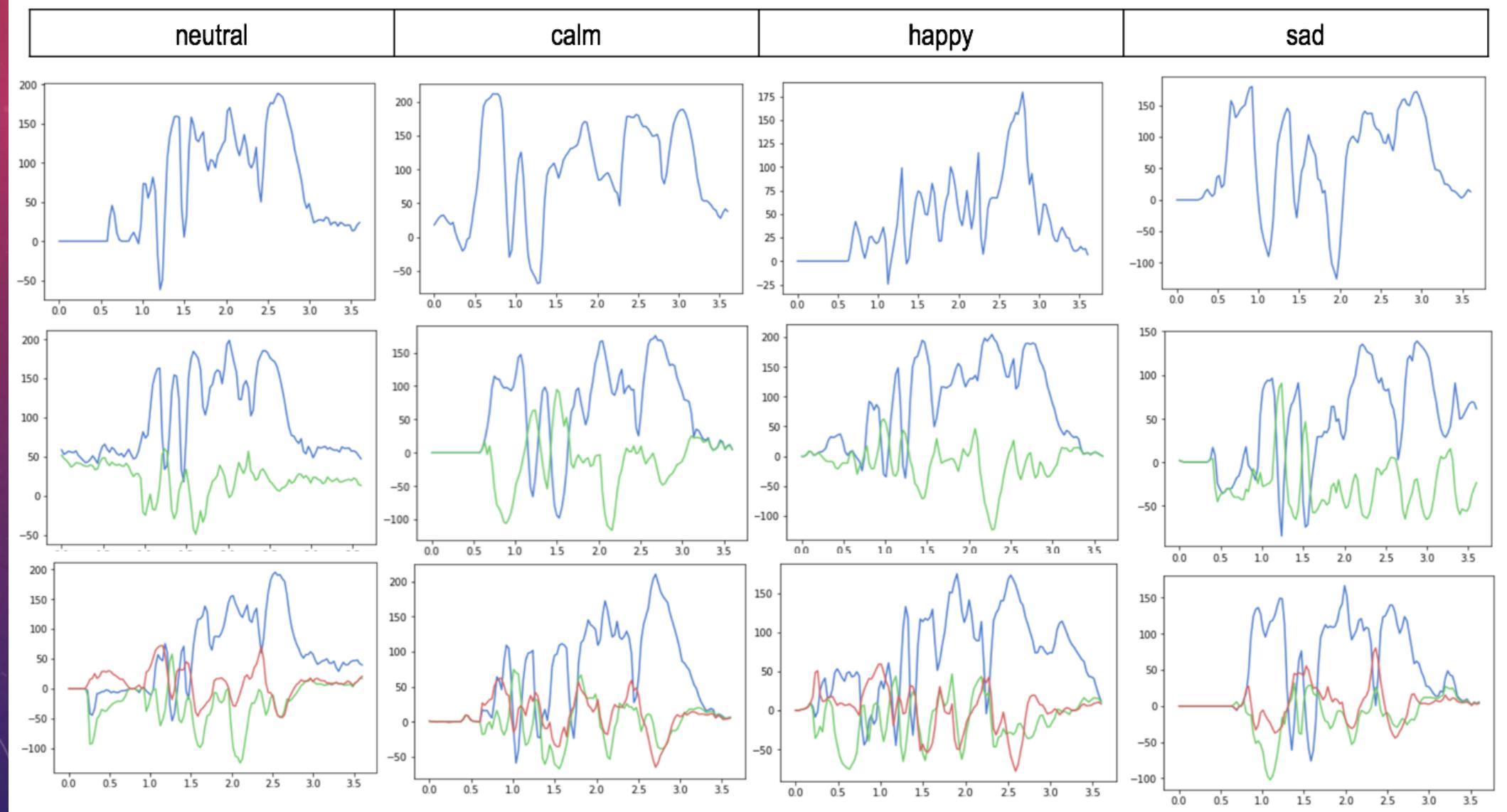
# МЕЛСПЕКТРОГРАММА



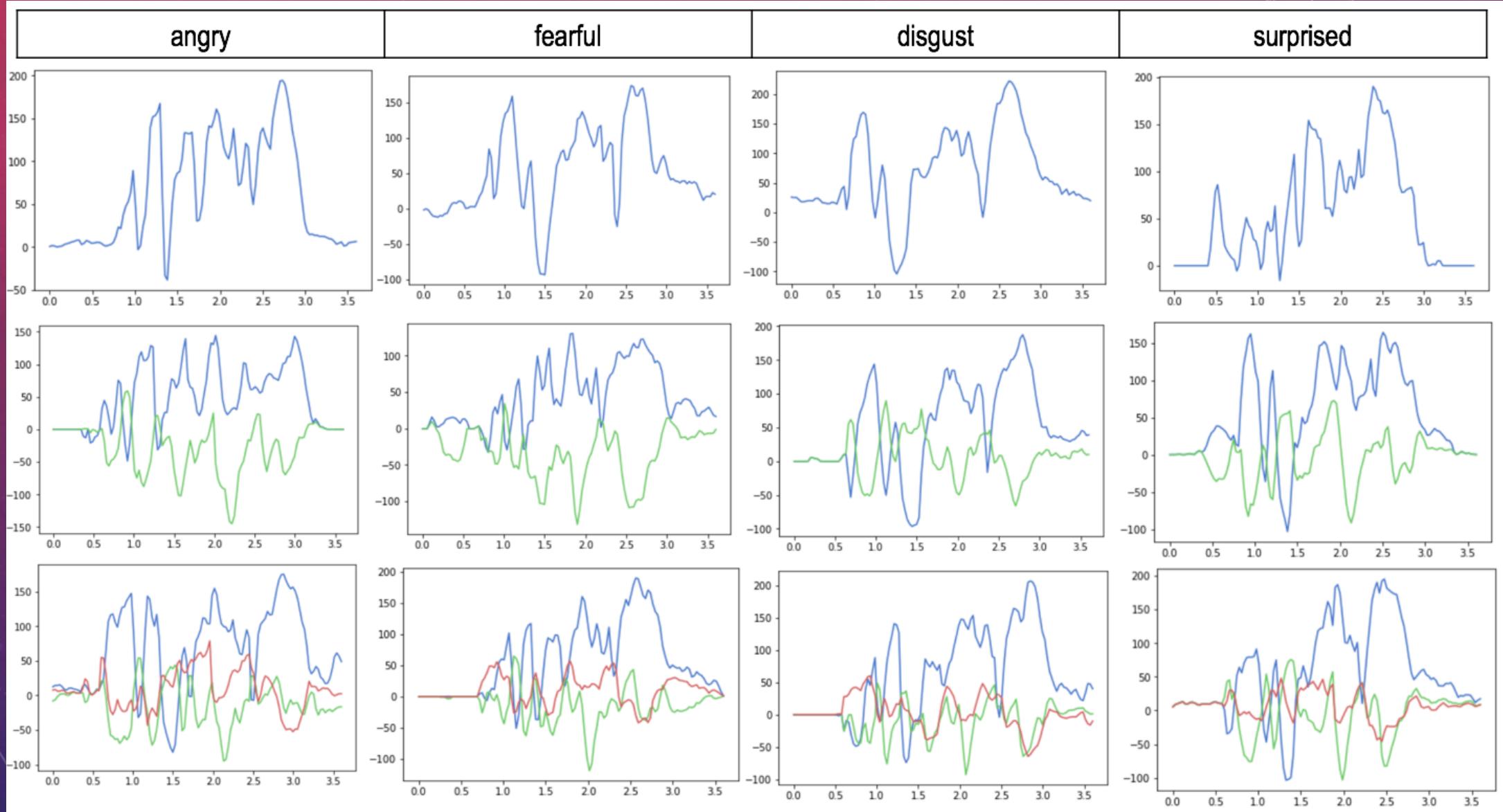
# МЕЛ-КЕПСТРАЛЬНЫЕ КОЭФФИЦИЕНТЫ



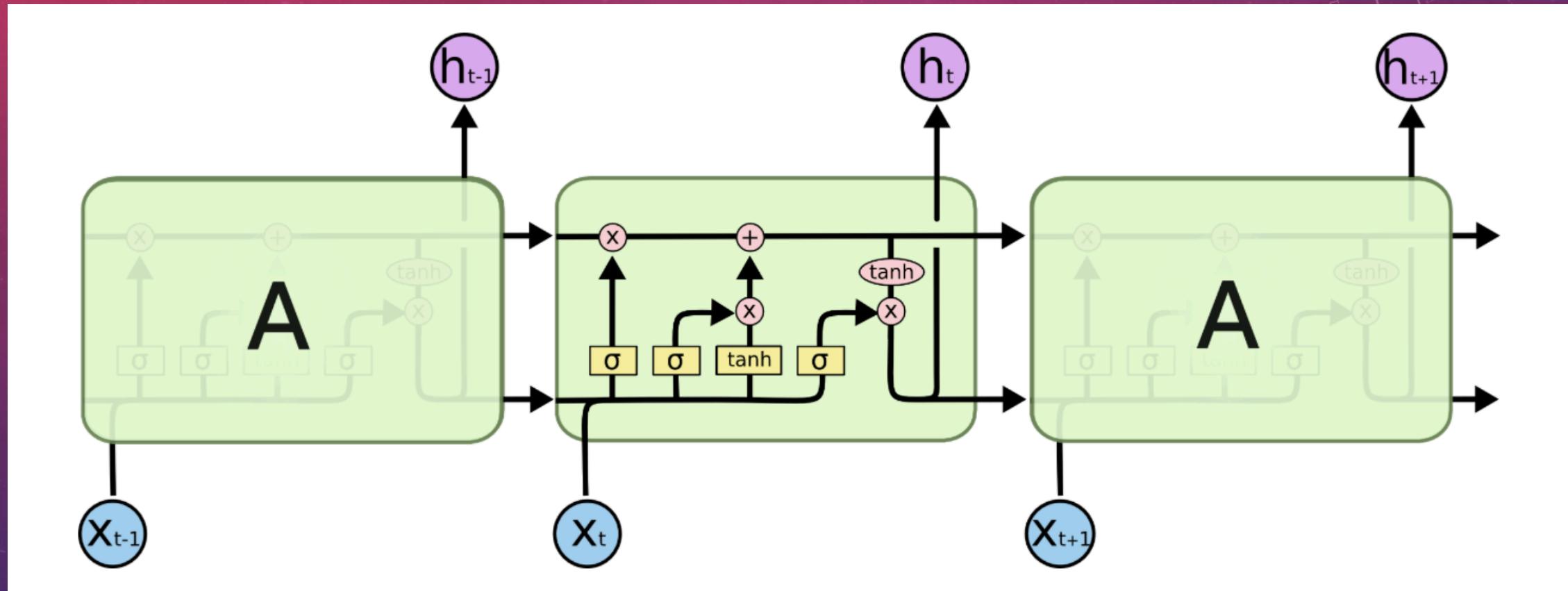
# ПРИМЕРЫ ДЛЯ РАЗНЫХ ЭМОЦИЙ



# ПРИМЕРЫ ДЛЯ РАЗНЫХ ЭМОЦИЙ



# LSTM NETWORK

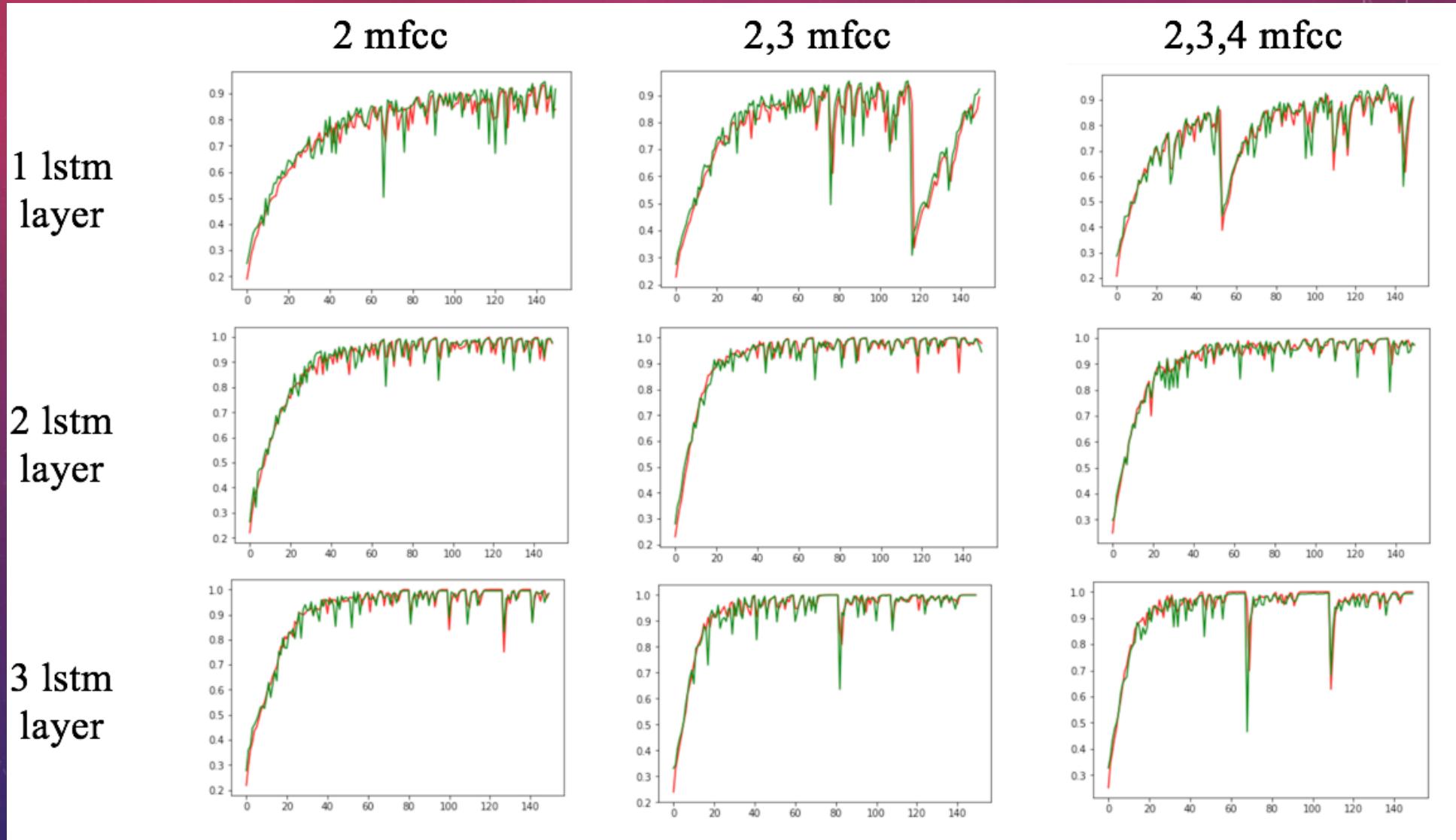


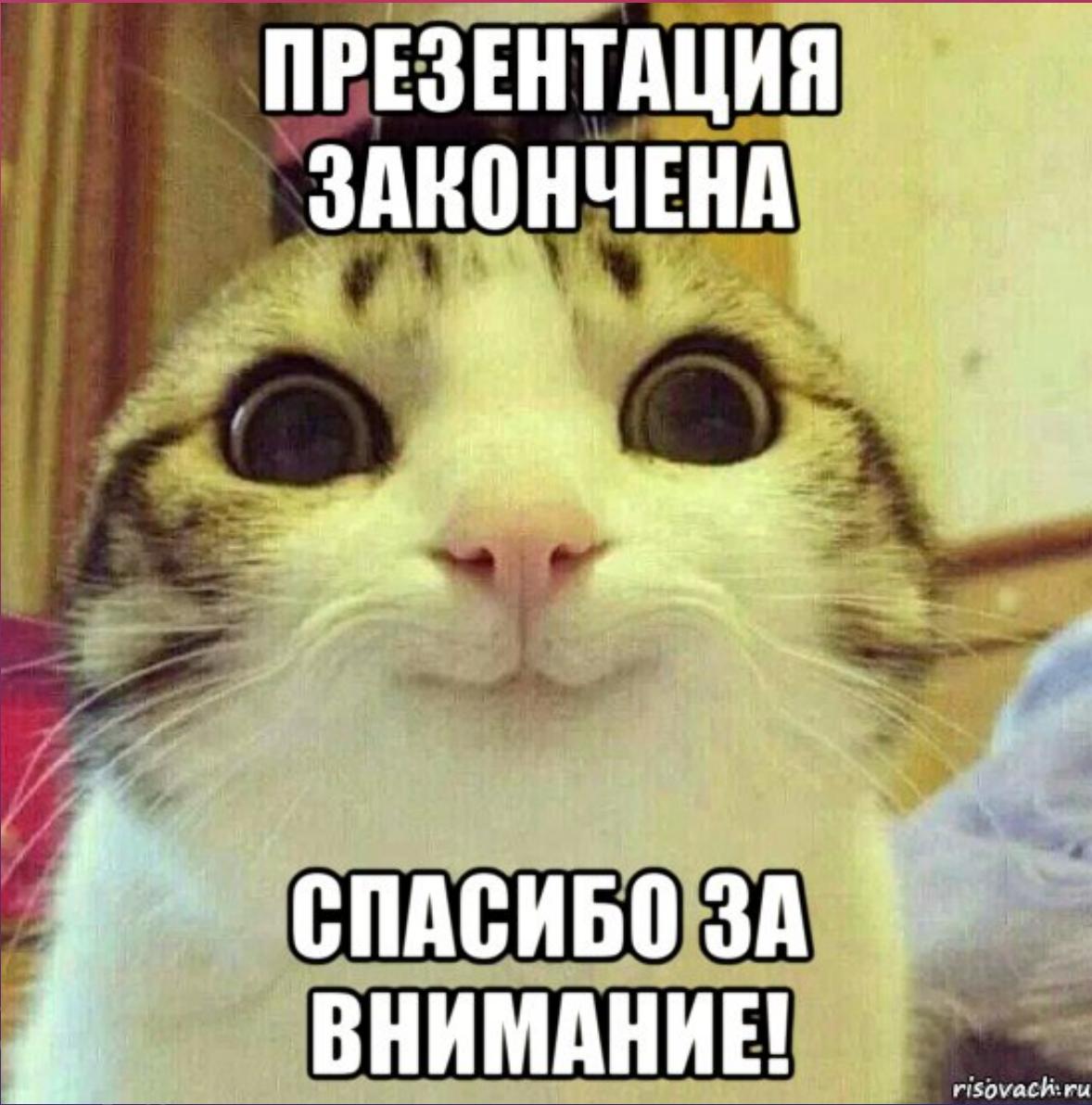
# РЕЗУЛЬТАТЫ

Model	Accuracy
Random choise	12.5%
Knn + preprocessed voice	24%
Random forest + preprocessed voice	29%
SVM + preprocessed voice	31%
Vgg11 + spectrogram	64%
Vgg16 + melspectrogram	71%

	2 mfcc	2,3 mfcc	2,3,4 mfcc
1 lstm layer	91.72%	92.19%	91.06%
2 lstm layer	97.88%	94.50%	97.27%
3 lstm layer	98.45%	99.86%	99.31%

# РЕЗУЛЬТАТЫ





**ПРЕЗЕНТАЦИЯ  
ЗАКОНЧЕНА**

**СПАСИБО ЗА  
ВНИМАНИЕ!**

*risovach.ru*

**ВОПРОСЫ?**

**stasysp.96@gmail.com**