

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

РАСПОЗНАВАНИЕ ЭМОЦИЙ В ЗВУКЕ

ПОПОВА АНАСТАСИЯ СЕРГЕЕВНА

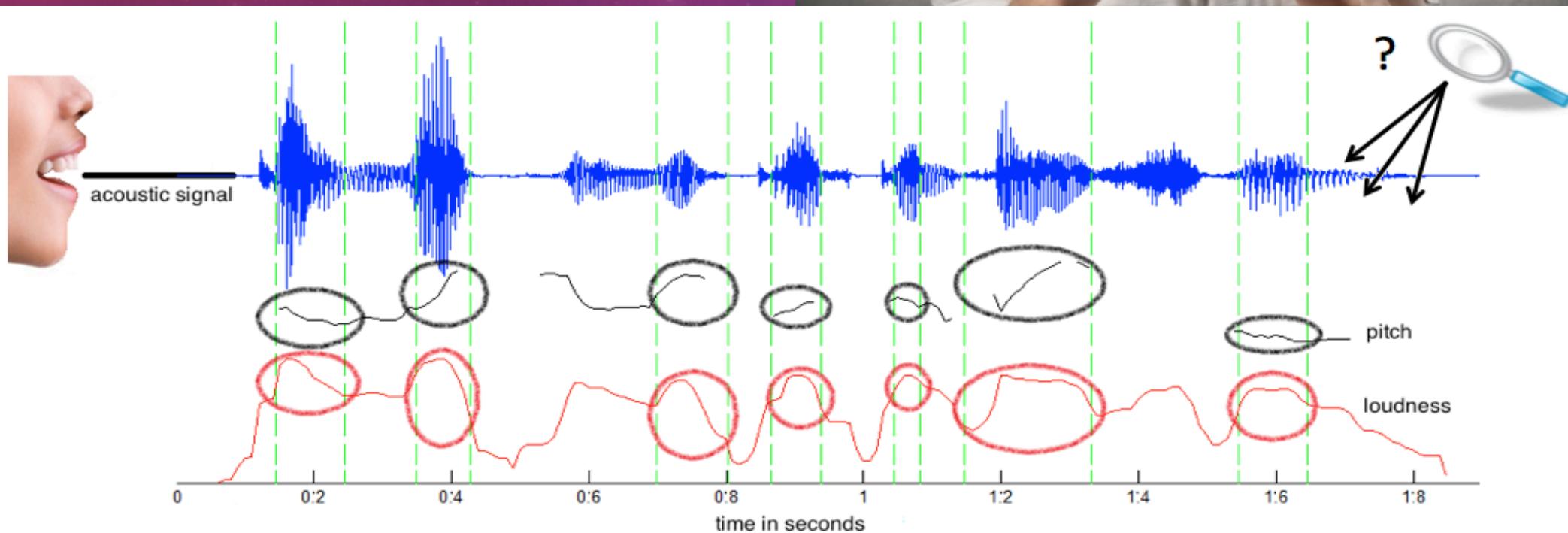
НАУЧНЫЙ РУКОВОДИТЕЛЬ: СТАРШИЙ ПРЕПОДАВАТЕЛЬ КАФЕДРЫ ПМИ
ПОНОМАРЕНКО АЛЕКСАНДР АЛЕКСАНДРОВИЧ

НИУ ВШЭ НН

2018

АКТУАЛЬНОСТЬ

Anger	0.99989
Contempt	0.00000
Disgust	0.00000
Fear	0.00008
Happiness	0.00000
Neutral	0.00002
Sadness	0.00000
Surprise	0.00001



ЦЕЛИ И ЗАДАЧИ

Улучшить точность работы алгоритма для распознавания эмоций человека по аудиозаписи его голоса, разработанный в рамках курсовой работы на 3 курсе на основе сверточных нейронных сетей.

1. Изучить другие подходы к решению поставленной задачи
2. Найти набор признаков, который позволяет улучшить модель
3. Найти наиболее подходящую модель
4. Провести серию экспериментов на открытом наборе данных. Провести анализ и сравнение полученных результатов.

БАЗА ДАННЫХ (RAVDESS DATABASE)

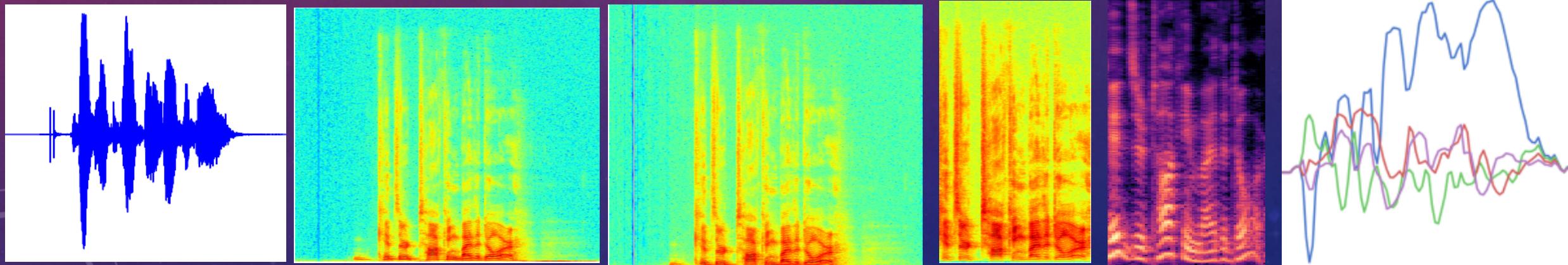
►	1neutral
►	2calm
►	3happy
►	4sad
►	5angry
►	6fearful
►	7disgust
►	8surprised

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

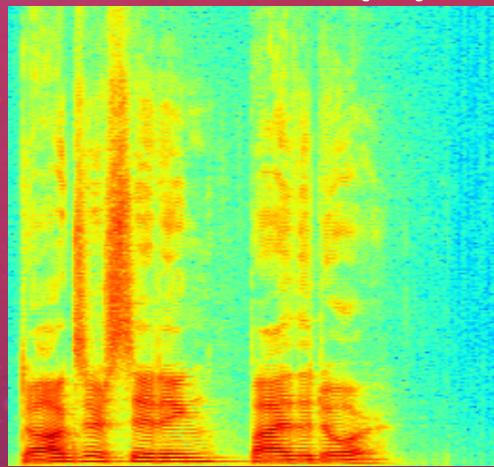
<http://neuron.arts.ryerson.ca/ravdess/?f=3>

РАЗРАБОТАННЫЙ АЛГОРИТМ

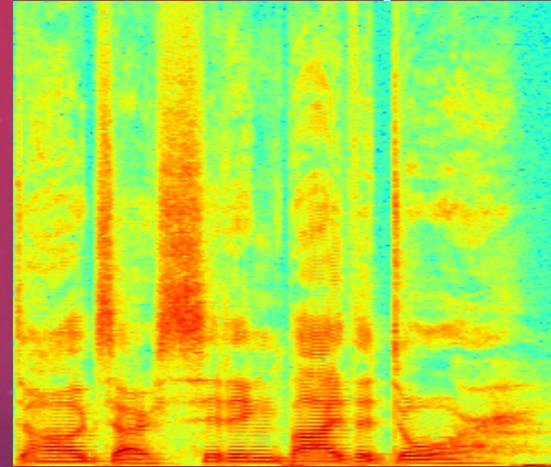
Метод LSTM	
Входной сигнал	
Предобработка	Выравнивание длины
	Выравнивание громкости
	Highpass&Lowpass фильтры
	Voice Activity Detection
	Получение мел-частотных кепстральных коэффициентов
Классификация	1-й слой LSTM (64 ячейки)
	2-й слой LSTM (64 ячейки)
	Dropout(0.5)
	Dense(8)
Принятие решения	



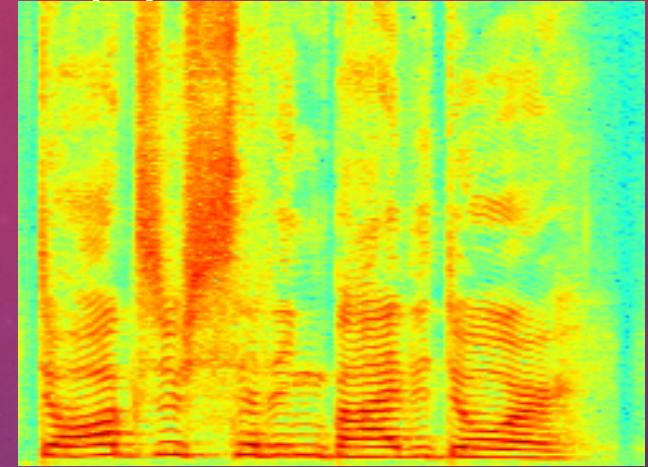
НАГЛЯДНАЯ РАЗНИЦА МЕЖДУ КЛАССАМИ (ГИПОТЕЗА)



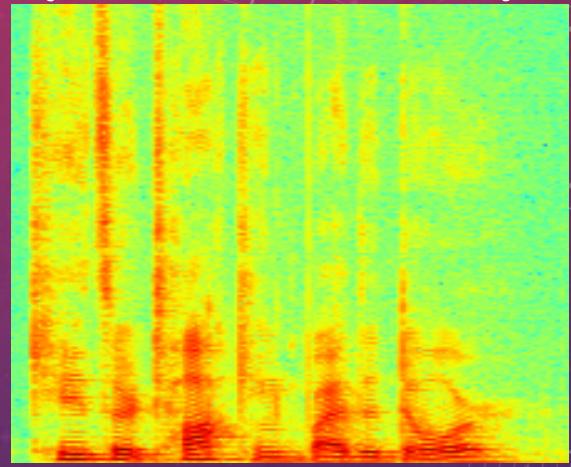
нейтральный



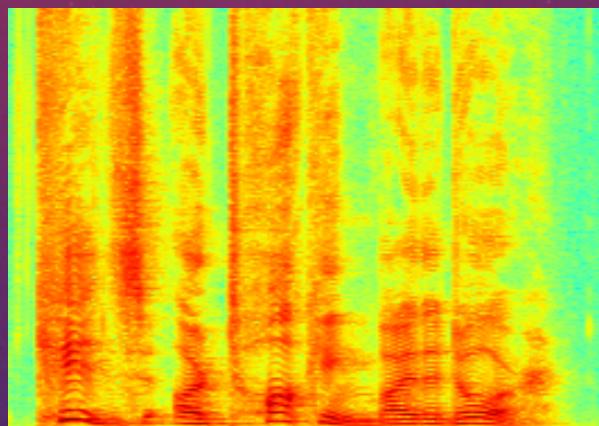
спокойный



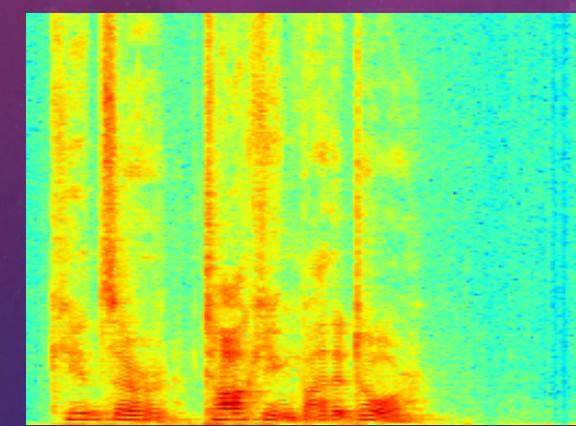
счастливый



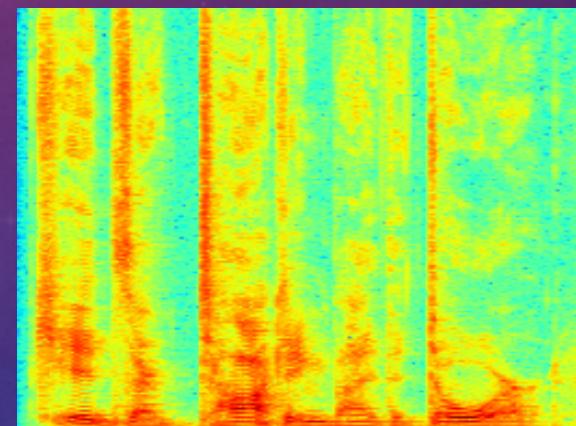
грустный



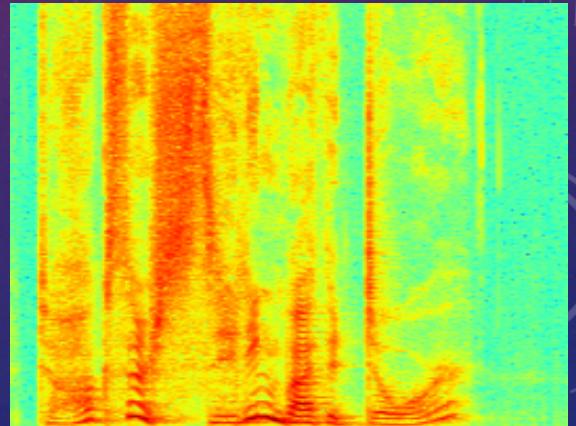
злой



напуганный

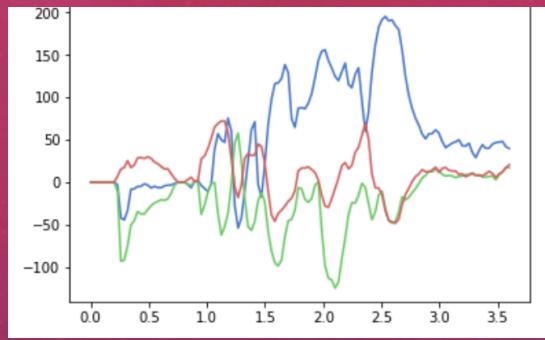


неприязнь

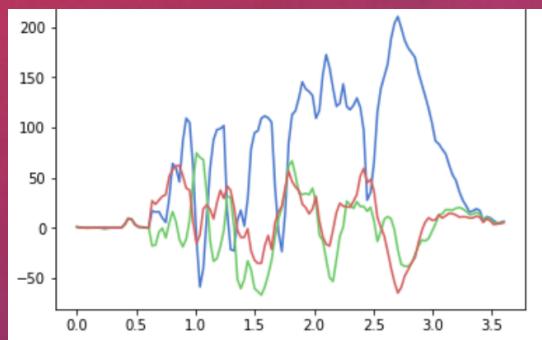


удивление

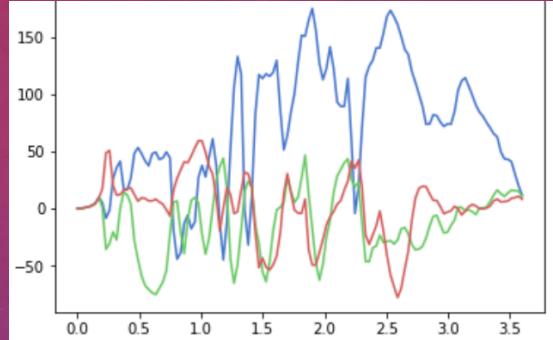
НАГЛЯДНАЯ РАЗНИЦА МЕЖДУ КЛАССАМИ (ГИПОТЕЗА)



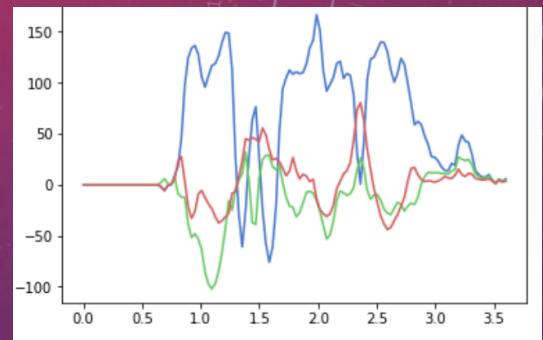
нейтральный



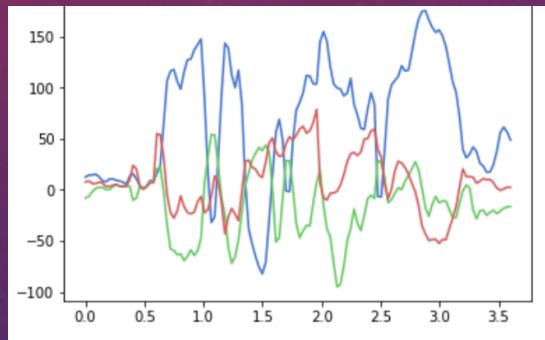
спокойный



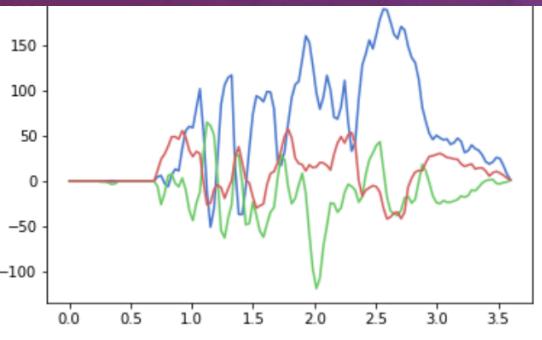
счастливый



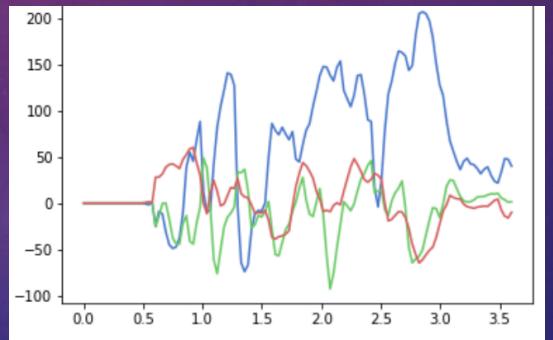
грустный



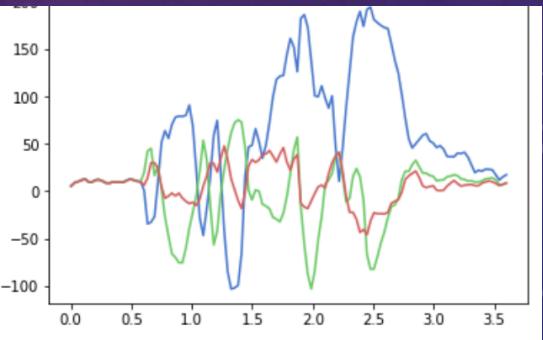
злой



напуганный



неприязнь



удивление

РЕЗУЛЬТАТЫ

Модель	Точность
Генератор случайных чисел	12.5%
KNN + обработанный сигнал	24%
Random forest + обработанный сигнал	29%
SVM + обработанный сигнал	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%

	neutral	calm	happy	sad	angry	fearfull	disgust	surprised
neutral	21	8	0	0	0	0	0	0
calm	7	46	1	1	0	1	2	0
happy	0	1	26	7	6	9	5	4
sad	0	2	10	31	2	3	9	1
angry	0	1	1	0	43	2	5	6
fearfull	0	1	3	2	6	34	6	6
disgust	0	0	0	3	2	3	49	1
surprised	0	0	2	1	1	8	12	12

МАТРИЦА ОШИБОК

ВЫВОДЫ

1. Предложен и апробирован подход к классификации эмоций человека в звуковом фрагменте
2. Проведён численный эксперимент
3. Сравнены результаты работы различных классификаторов, сверточных сетей (VGG-11 и VGG-16) и LSTM сетей с различными наборами признаков
4. Найден простой алгоритм обработки записи речи человека и получения хорошего набора признаков (ускорение за счет уменьшения размерности пространства признаков)
5. Выявлено, что основная информация о эмоциях содержится во втором мел-частотном кепстральном коэффициенте
6. Улучшена точность с 71% до 98%
7. В планах сбор своего большого и разнообразного набора данных для дальнейшей разработки алгоритма (Telegram: @Emotion_database_bot)

ПУБЛИКАЦИИ

1. А. С. Попова, А. Г. Рассадин, А. А. Пономаренко "Детектирование эмоций в мультимедиа контенте" //Международная научно-техническая конференция «Информационные системы и технологии» ИСТ-2017 , СЕКЦИЯ 5.3 Техническая кибернетика (информационное моделирование когнитивных процессов) – 2017. – С. 852-857.
2. Anastasiya S. Popova, Alexandr G. Rassadin, Alexander A. Ponomarenko "Emotion recognition in sound" // SPRINGER Selected Papers of the XIX International Conference on Neuroinformatics, October 2-6, 2017, Moscow, Russia
3. А. С. Попова, А. Г. Рассадин, А. А. Пономаренко "Детектирование эмоций в речи с использованием долгой краткосрочной памяти" //Международная научно-техническая конференция «Информационные системы и технологии» ИСТ-2018 , СЕКЦИЯ 5.3 Техническая кибернетика (информационное моделирование когнитивных процессов) – 2018. - С.1083-1089

ВЫСТУПЛЕНИЯ НА КОНФЕРЕНЦИЯХ

1. А. С. Попова, А. Г. Рассадин, А. А. Пономаренко “Детектирование эмоций в мультимедиа контенте” //Международная научно-техническая конференция «Информационные системы и технологии» ИСТ-2017 , СЕКЦИЯ 5.3 Техническая кибернетика (информационное моделирование когнитивных процессов) – 2017. – С. 852-857.
2. Anastasiya S. Popova, Alexandr G. Rassadin, Alexander A. Ponomarenko "Emotion recognition in sound" // The XIX International Conference on Neuroinformatics, October 2-6, 2017, Moscow, Russia
3. Anastasiya S. Popova «Emotion recognition in sound» // 2nd Winter School in Data Analytics, Nizhny Novgorod, November 03-04, 207
4. А. С. Попова, А. Г. Рассадин, А. А. Пономаренко «Детектирование эмоций в речи с использованием долгой краткосрочной памяти» //Международная научно-техническая конференция «Информационные системы и технологии» ИСТ-2018 , СЕКЦИЯ 5.3 Техническая кибернетика (информационное моделирование когнитивных процессов) – 2018. - С.1083-1089

ГРАНТЫ

1. Научно-исследовательская работа: Научно-исследовательская работа: Научно-учебная группа «Анализ мультимедийных данных»

МЕДИА

1. The Times (<https://www.thetimes.co.uk/article/ai-struggles-with-human-happiness-s98bt22gm>)
2. Российская газета (<https://rg.ru/2017/11/09/reg-pfo/rossijskie-uchenye-nauchili-kompiuter-raspoznavat-emocii.html>)
3. Cnews (http://www.cnews.ru/news/top/2017-11-08_rossiyane_nauchili_kompyuter_raspoznavat_emotsii)
4. Rambler (<https://news.rambler.ru/other/38363203-rossiyskie-uchenye-nauchili-kompyuter-raspoznavat-emotsii/>)
5. IQ HSE (<https://iq.hse.ru/news/211613179.html>)
6. Европа+
7. Столица Нижний Медиа

СПАСИБО ЗА ВНИМАНИЕ!

stasysp.96@gmail.com

