



Нижегородский Государственный Технический Университет
Институт Радиоэлектроники и Информационных Технологий

Кафедра: «Вычислительные системы и технологии»

Модель и алгоритмы распознавания эмоционального состояния диктора по голосу с позиций теории активного восприятия

Выполнил:

студент группы М18-ИВТ-3

Кузнецова Полина Викторовна

Научный руководитель:

к.т.н., доцент

Гай Василий Евгеньевич

Актуальность, цели и задачи проекта

Актуальность

Задача распознавания эмоционального состояния диктора по голосу является актуальной в таких сферах как:

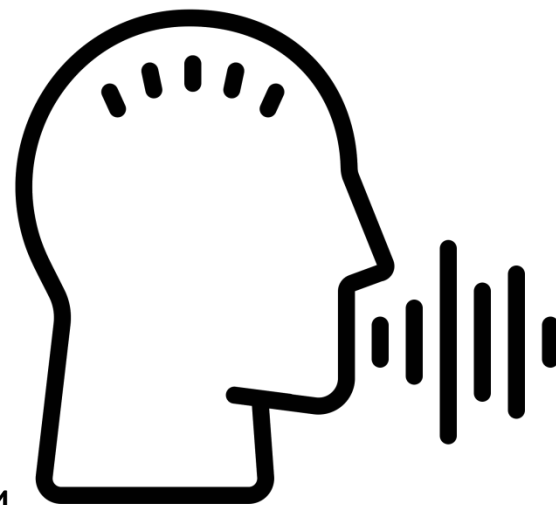
- медицина;
- образование;
- маркетинговые исследования;
- человеко-машинное взаимодействие.

Цель

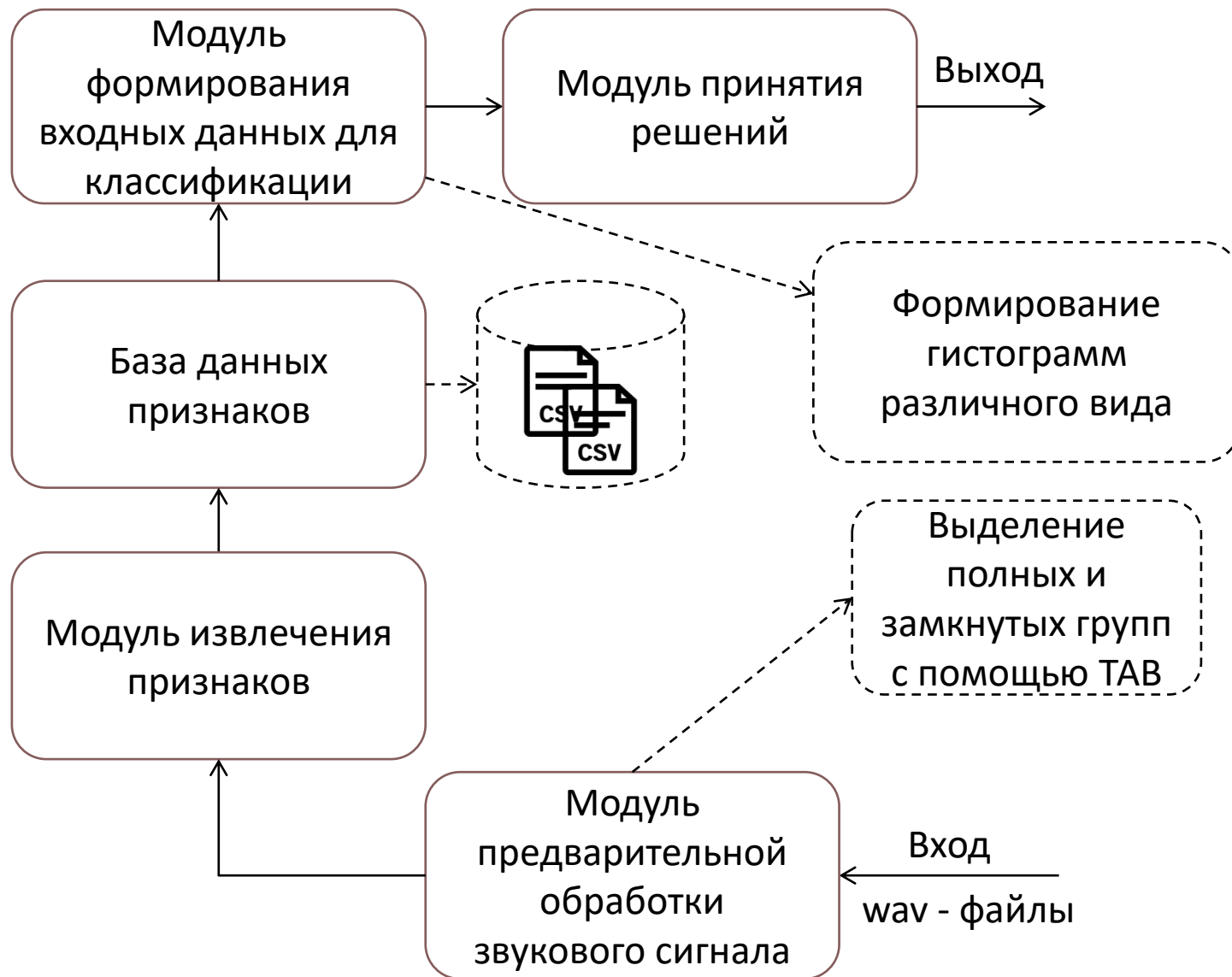
Разработка нового подхода к решению задачи распознавания эмоционального состояния диктора по голосу и проверка данного подхода с помощью вычислительного эксперимента.

Задачи

- Выбор средств разработки;
- Разработка модели системы;
- Разработка алгоритма распознавания эмоционального состояния диктора по голосу;
- Выявление новых информативных признаков голосового сигнала, отражающих эмоциональное состояние диктора;
- Программная реализация системы;
- Тестирование реализованной системы.

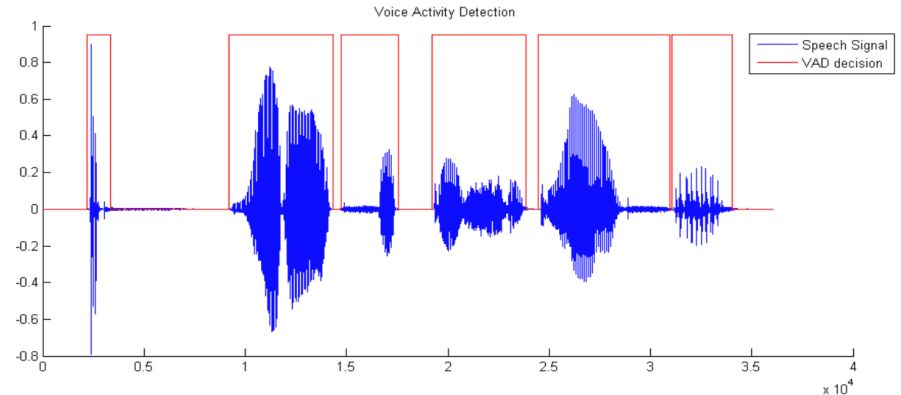


Модель системы распознавания эмоционального состояния диктора по голосу



Предварительная обработка

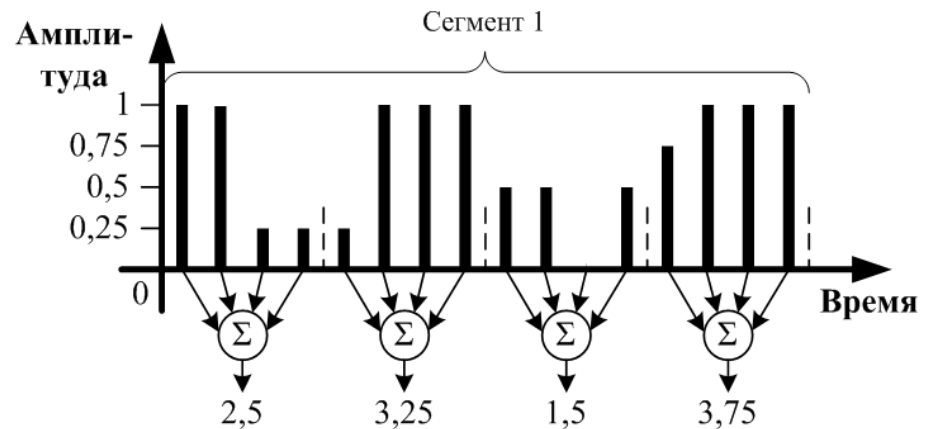
- Voice Activity Detection (VAD);
- Деление исходного сигнала на сегменты;
- Q-преобразование



$$m(\overline{T}) = \sum_{n=0}^{N-1} (s_{dig}[n]),$$

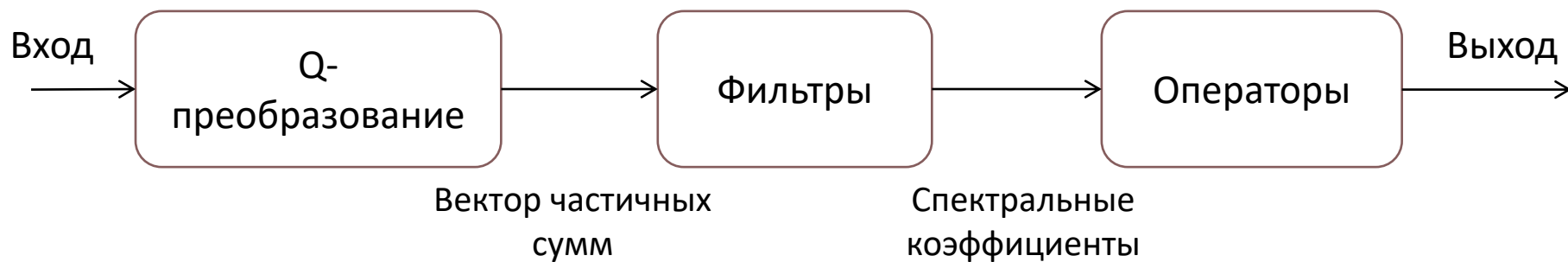
где N – число отсчётов сигнала s_{dig} ,

$m(\overline{T})$ – сумма отсчётов сегмента

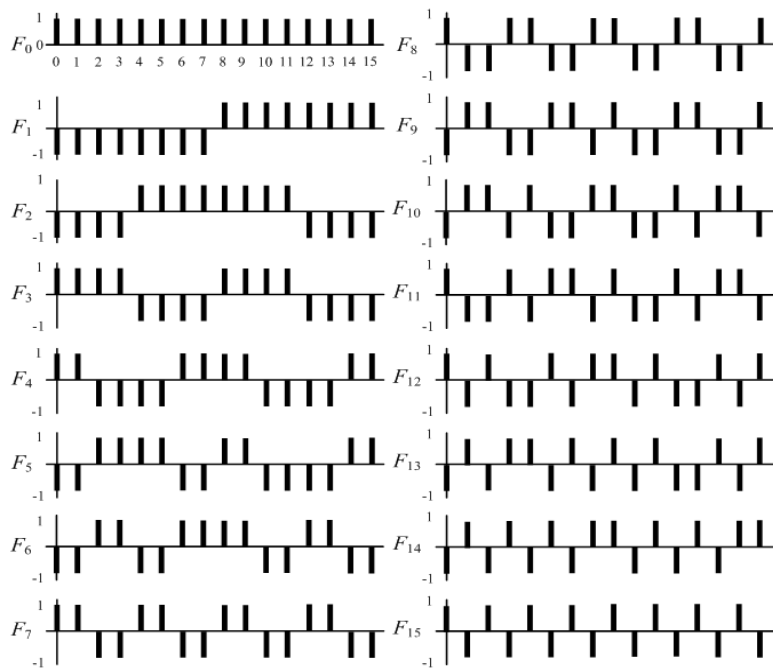


Выделение признаков

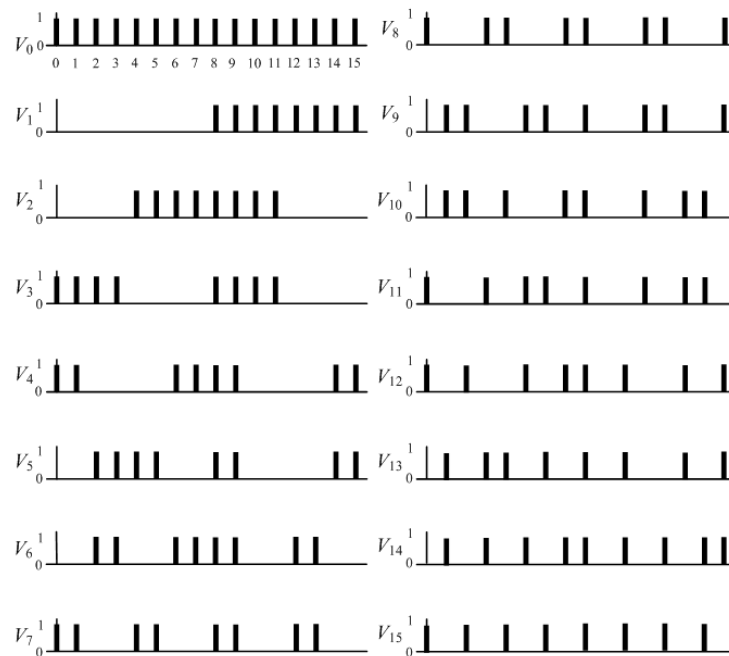
U-преобразование



Фильтры

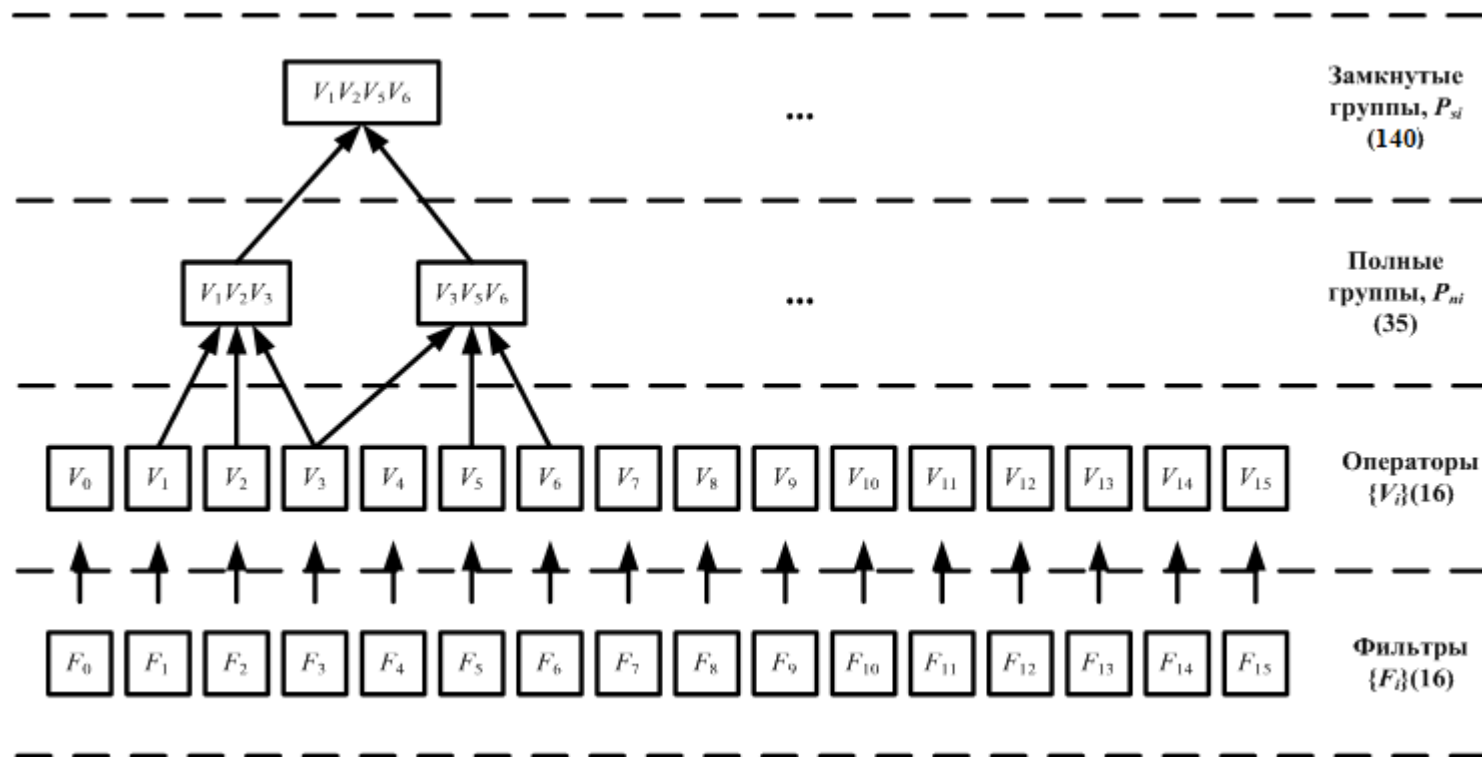


Операторы



Выделение признаков

Полные и замкнутые группы



Признаковое описание

Структура хранения признаков:

№гр.	масса	№гр.	масса	№гр.	масса	№гр.	масса
7	123.25	34	124.79	127	125.67	131	129.48

Одномерные гистограммы

1	2	3	4	5	6	7	...	139	140
1	0	2	3	0	1	2	...	0	0

Двумерные гистограммы

	1	2	3	...	140
1	2	0	1	...	1
2	1	0	2	...	2
3	1	1	3	...	4
...
140	2	0	1	...	3

Двумерные матрицы масс
полных групп

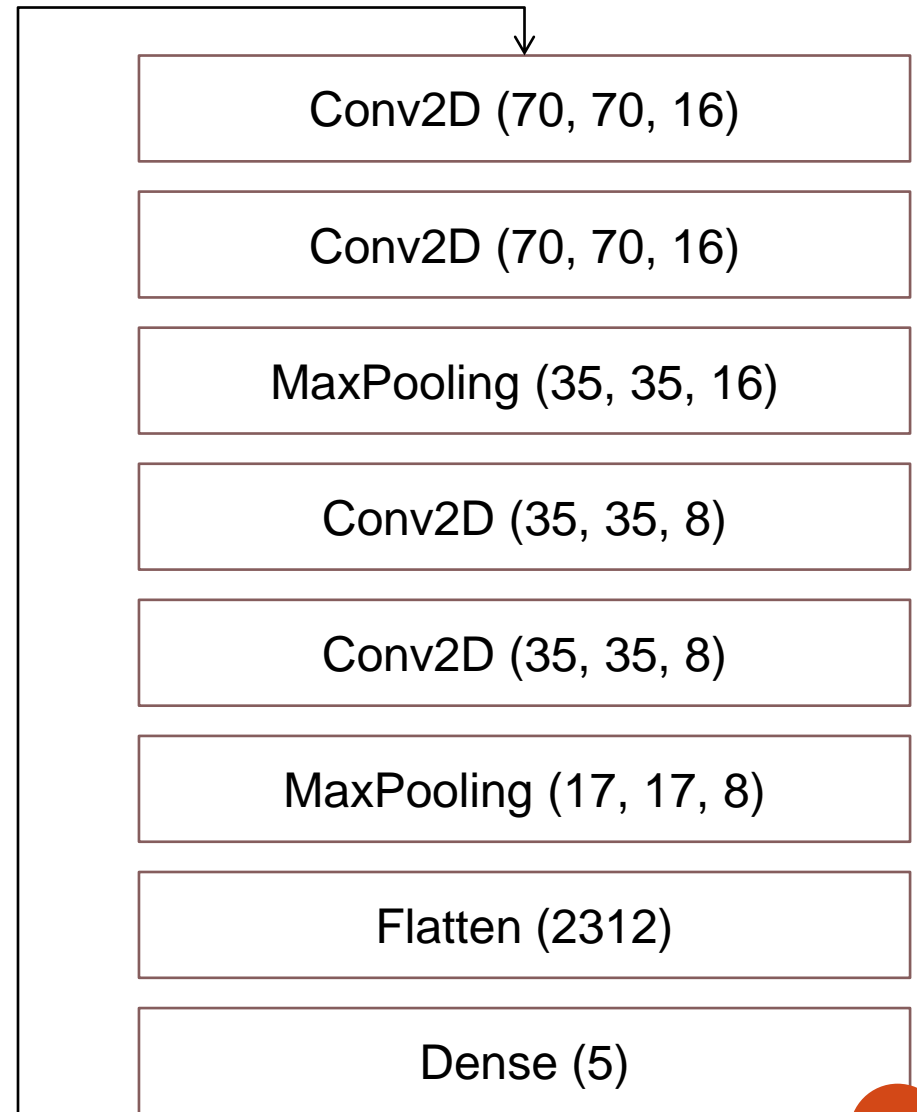
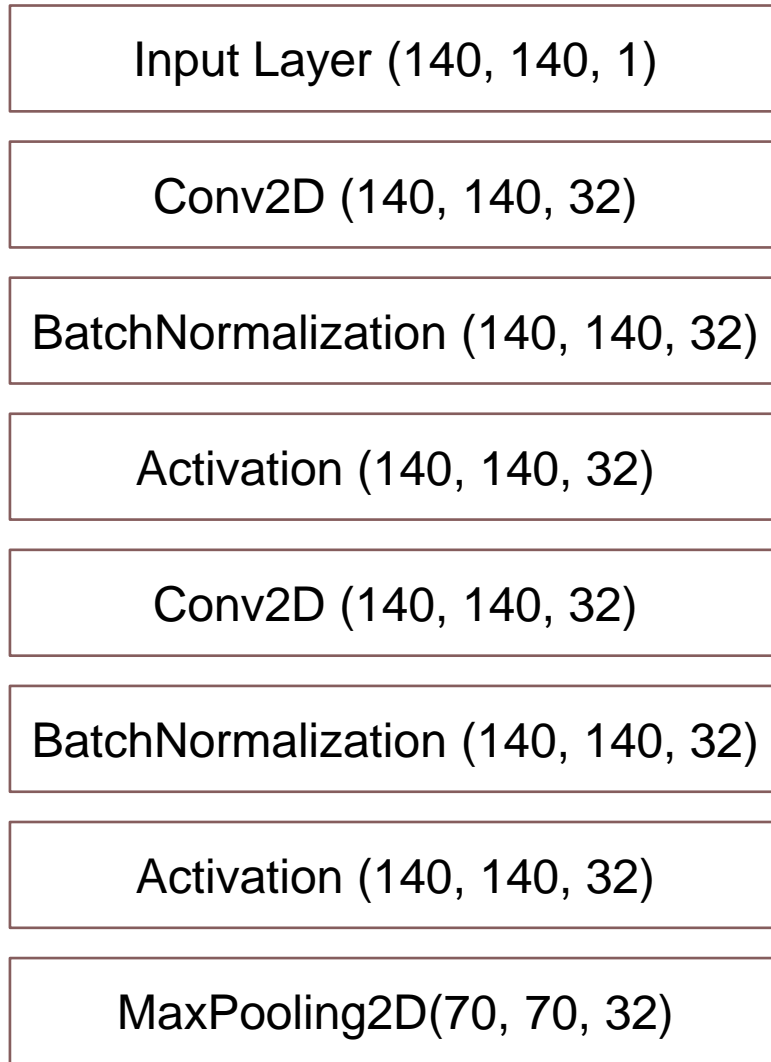
	1	2	3	...	140
1	122.7	127.5	129.1	...	130.4
2	123.9	124.7	128.6	...	123.8
3	124.4	127.2	129.0	...	125.5
...
140	126.6	125.4	133.9	...	124.1

Двумерные двоичные
матрицы

	1	2	3	...	140
1	0	0	1	...	0
2	1	0	0	...	0
3	0	1	0	...	0
...
140	0	0	1	...	0

Метод классификации

CNN



Условия эксперимента









Тестовые данные:

Ravdess - 24 актёра (12 мужчин и 12 женщин), 23 мин. чистой речи + 23 мин. исходные данные, искажённые нормальным шумом, отношение сигнал / шум - 10 дБ;

Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) - 91 актёр (48 мужчин и 43 женщины), 204 мин. чистой речи + 204 мин. исходные данные, искажённые нормальным шумом, отношение сигнал / шум - 10 дБ;

Исследуемые эмоции:


Счастье, злость, грусть, нейтральное состояние, страх

- >  Actor_01
- >  Actor_02
- ▼  Actor_03
 -  03-01-01-01-01-01-03.wav
 -  03-01-01-01-01-02-03.wav
 -  03-01-01-01-02-01-03.wav
 -  03-01-01-01-02-02-03.wav
 -  03-01-02-01-01-01-03.wav


 1001_ITS_FEA_XX.wav

 1001_ITS_HAP_XX.wav

 1001_ITS_NEU_XX.wav

 1001_ITS_SAD_XX.wav

 1001_IWL_ANG_XX.wav

 1001_IWL_FEA_XX.wav

 1001_IWL_HAP_XX.wav

 1001_IWL_NEU_XX.wav

 1001_IWL_SAD_XX.wav

 1001_IWW_ANG_XX.wav

Результаты работы программной системы распознавания диктора по голосу

Одномерные гистограммы полных групп

	ANG	HAP	SAD	FEA	NEU
ANG	56%	22%	0%	22%	0%
HAP	0%	94%	0%	6%	0%
SAD	0%	0%	100%	0%	0%
FEA	0%	0%	0%	100%	0%
NEU	27%	27%	0%	20%	27%

Количество эпох: 150

Средняя точность распознавания: 74%

Потери: 0.085

Одномерные гистограммы замкнутых групп

	ANG	HAP	SAD	FEA	NEU
ANG	100%	0%	0%	0%	0%
HAP	40%	0%	20%	0%	40%
SAD	0%	0%	100%	0%	0%
FEA	60%	0%	0%	40%	0%
NEU	0%	0%	0%	100%	0%

Количество эпох: 150

Средняя точность распознавания: 40%

Потери: 0.331

Результаты работы программной системы распознавания диктора по голосу

Двумерные гистограммы полных групп

	ANG	HAP	SAD	FEA	NEU
ANG	39%	11%	28%	16%	6%
HAP	42%	50%	0%	8%	0%
SAD	11%	0%	78%	11%	0%
FEA	0%	50%	0%	38%	12%
NEU	69%	15%	0%	16%	0%

Количество эпох: 50

Средняя точность распознавания: 38%

Потери: 0.256

Двумерные матрицы масс полных групп

	ANG	HAP	SAD	FEA	NEU
ANG	5%	3%	21%	26%	45%
HAP	5%	79%	0%	5%	11%
SAD	7%	0%	73%	13%	7%
FEA	5%	0%	32%	26%	37%
NEU	16%	16%	5%	16%	47%

Количество эпох: 50

Средняя точность распознавания: 38.2%

Потери: 0.102

Результаты работы программной системы распознавания диктора по голосу

Двумерные двоичные матрицы полных групп

	ANG	HAP	SAD	FEA	NEU
ANG	27%	12%	3%	27%	31%
HAP	17%	83%	0%	0%	0%
SAD	8%	4%	84%	4%	0%
FEA	8%	0%	0%	92%	0%
NEU	10%	7%	5%	34%	44%

Количество эпох: 50

Средняя точность распознавания: 56.4%

Потери: 0.001

Двумерные матрицы с максимальным значением массы для полных групп

	ANG	HAP	SAD	FEA	NEU
ANG	38%	23%	24%	5%	10%
HAP	0%	43%	43%	14%	0%
SAD	2%	7%	85%	3%	3%
FEA	29%	16%	18%	8%	29%
NEU	91%	0%	0%	9%	0%

Количество эпох: 50

Средняя точность распознавания: 40.9%

Потери: 0.359

Средняя точность распознавания по всем экспериментам

	одномерные гистограммы полных групп	одномерные гистограммы замкнутых групп	двумерные гистограммы полных групп	двумерные матрицы масс полных групп	двумерные двоичные матрицы полных групп	двумерные матрицы с максимальным значением массы полных групп
точность распознавания	74.0%	40.0%	38.0%	38.2%	56.4%	40.9%

Итоги

- ✓ Изучены и проанализированы существующие методы решения задачи распознавания эмоционального состояния диктора по голосу
- ✓ Разработан новый метод решения поставленной задачи
- ✓ Разработан программный продукт для проведения исследования
- ✓ Проведен ряд экспериментов, подтверждающих работоспособность предложенного метода

Публикации

П.В. Кузнецова, В.Е. Гай «Модель и алгоритмы распознавания эмоционального состояния диктора по голосу с позиций теории активного восприятия» Материалы XXVI международной научно-технической конференции «Информационные системы и технологии - 2020», ИСТ-2020, Россия, Н. Новгород, 2020г.

Спасибо за внимание!