Санкт-Петербургский государственный университет

ПОВОЛОЦКАЯ Анастасия Андреевна

Выпускная квалификационная работа

Распознавание негативных эмоций с использованием нейросетевых технологий

Уровень образования: магистратура

Направление 45.04.02 «Лингвистика»

Основная образовательная программа ВМ.5715. «Общая и прикладная фонетика (General and Applied Phonetics)»
Профиль «Речевые технологии»

Научный руководитель:
доцент, Кафедра фонетики и методики преподавания иностранных языков,
Евдокимова Вера Вячеславовна
Рецензент:
старший научный сотрудник,
ФГБУН «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
Кипяткова Ирина Сергеевна

Санкт-Петербург
2022

Оглавление

Введение

Данная работа посвящена вопросам создания систем распознавания эмоций по голосу с использованием нейросетевых технологий.

Системы распознавания эмоций по голосу и речи с использованием нейросетевых технологий набирают популярность, поскольку данные технологии направлены на разработку систем, совершенствующих человеко-машинное взаимодействие. При проектировании подобных систем разработчики и исследователи сталкиваются с рядом проблем: какую выбрать модель и структуру нейронной сети; какие данные подавать на вход. Главное отличие человека от машины заключается в том, что человеку информация подается по нескольким каналам: изображение, звук, текст. Также, важную роль для интерпретации эмоции играет контекст. Мультимодальные системы, получающие на вход информацию по нескольким каналам, позволяют более точно распознавать и классифицировать эмоции. На данный момент высокой точности достигла обработка визуальных данных, например, распознавание лиц и распознавание эмоций по лицу.

Задачу распознавания эмоций можно отнести к задаче классификации, которая на данный момент является важной областью применения нейронных сетей.

Объектом данного исследования является распознавание негативных эмоций по голосу с использованием нейросетевых технологий.

Предмет исследования – особенности реализации системы по распознаванию эмоций с использованием нейросетевых технологий.

Целью исследования является разработка системы распознавания негативных эмоций с использованием нейросетевых технологий.

Для достижения цели были поставлены и решены следующие задачи, перечисленные ниже.

1. Определение перечня эмоций для данной задачи на основе

научной литературы и составление списка фраз, соответствующих данным эмоциям.

2. Проведение записи дикторов.

3. Формирование обучающего набора данных.

4. Проведение перцептивного эксперимента.

5. Определение основных нейросетевых подходов, которые применяются для решения задач распознавания эмоций по речи, и выбор подходящей методики.

6. Реализация алгоритма предобработки исходных файлов, т.е. преобразования исходного аудиофайла в изображение для подачи на вход нейронной сети.

7. Подготовка обучающей и тестовой выборки.

8. Реализация, обучение и тестирование нейронной сети.

Выпускная квалификационная работа состоит из введения, трех глав, заключения, библиографии и приложений. Объем работы составляет 96 страниц, объем библиографии – 80 наименований.

В первой главе приведено описание предметной области, в рамках которой выполнена данная работа. Проводится обзор аналогичных разработок и способов решения поставленной задачи.

Во второй главе описана теоретическая составляющая: рассмотрены классификации эмоций, и обозначен перечень эмоций, включенный в исследование. Описан процесс записи набора данных и предобработка полученного материала. Представлена организация и результаты перцептивного эксперимента.

В третьей главе описаны топология нейронной сети, средства разработки, тестирование нейронной сети и полученные результаты.

В заключении представлены основные результаты выполненной работы.

Глава 1. Задача автоматического распознавания эмоций: обзор существующих подходов

1.1. Задача распознавания эмоций: подходы и решения

Задача распознавания эмоций по голосу с использованием нейросетевых технологий заключается в том, чтобы обработать, распознать и интерпретировать эмоциональное и физиологическое состояние человека по голосу в автоматическом режиме.

Сложность реализации данной задачи заключается в ряде моментов.

1. Существует множество классификаций первичных эмоциональных состояний и базовых эмоций. В различных исследованиях выделяется от нескольких единиц до нескольких десятков, которые в свою очередь формируют классы и подгруппы [5, 6, 30, 42, 77]. Как утверждает Б. И. Додонов – создание универсальной классификации эмоций невозможно, поскольку классификация эмоций, эффективная для решения одного круга задач, оказывается непригодной при решении другого круга задач [8].

2. Отсутствует универсальность методов распознавания эмоций у различных людей. Успешность в распознавании эмоций собеседника зависит от возраста, нации, особенностей состояния индивида в данный момент времени (депрессия, беременность, и др.) [12, 13, 18, 19, 51].

3. Существует высокая междикторская вариативность. Различие в реализации эмоций варьируется от человека к человеку, в зависимости от его культуры, образования, стиля и скорости речи [45, 50].

4. Речь и эмоции зависят от контекста и длительности высказываний [48, 59]. Это может выражаться в концентрации эмоции на определенной части высказываний, а не на всей фразе целиком [49]. Также, ключевым является отличие естественной речи в повседневной жизни и наборов данных, записанных в студии с полной шумоизоляцией, при участии профессиональных актёров [52, 53, 76].

Вышеперечисленный список моментов, затрудняющих решение задачи распознавания эмоций, не является исчерпывающим.

Еще одним важным пунктом является необходимость извлечения речевых признаков. Речевые признаки, которые чаще всего извлекаются из аудио-сигнала: контур частоты основного тона, среднее значение частоты основного тона, диапазон изменения частоты основного тона, спектрограмма, MFCC, кохлеаграмма, и другие [24, 44, 46, 48].

Однако, несмотря на обширное количество признаков, которое можно получить из сигнала, успешность распознавания эмоций с использованием нейросетевых технологий не превышает 70 – 75%, если система является дикторонезависимой [48, 66]. В случае с дикторозависимыми системами вероятность правильных ответов при классификации может достигать 98% [44, 75]. Однако, стоит принимать во внимание тот факт, что дикторозависимые системы имеют ограниченный спектр применения. Главная сложность заключается в том, что обработке подлежат только те, явления, которые можно наблюдать (мимика, жестикуляция, речь, и т.д.), и преобразовать их в параметры, которые могут быть позднее проанализированы, а не истинные чувства и намерения человека [55, 61, 71].

1.2. Обзор научной литературы

Speech emotion recognition using recurrent neural networks with directional self-attention [61]

В данной статье предлагается модель двунаправленной нейронной сети с долгой краткосрочной памятью с направленным самостоятельным вниманием. Нейронная сеть с долгой краткосрочной памятью может изучать долгосрочные зависимости на основе изученных локальных функций.

Информация о наборе данных IEMOCAP представлена на рисунке 1 и 2. Эксперименты проводились с 3 наборами данных: набор данных IEMOCAP со спонтанной речью; набор данных IEMOCAP со сценарной речью; полный набор данных IEMOCAP, совместивший в с себе наборы

данных со спонтанной и сценарной речью. Подобно IEMOCAP, набор данных EMO-DB также является одной из самых популярных баз данных, используемых в системах SER (Speech Emotion Recognition). Он включает записи 5 мужских и 5 женских голосов дикторов, которые записали 10 предложений (пять коротких и пять длинных) для семи различных эмоций. В данном исследовании были использованы 4 класса эмоций: радость, печаль, злость и нейтральность.

IEMOCAP Script Database					
Session	Person	Anger	Happiness	Neutral	Sadness
Session1	1F	68	53	107	24
	1 M	99	49	54	66
Session2	2F	56	28	90	50
	2 M	59	35	55	47
Session3	3F	66	46	79	53
	3 M	84	29	43	62
Session4	4F	114	20	37	36
	4 M	129	14	47	26
Session5	5F	54	23	49	56
	5 M	85	14	48	56

Рисунок 1. Количество эмоциональных предложений для каждого диктора в блоке сценарного набора данных IEMOCAP [61]

IEMOCAP Spontaneous Database					
Session	Person	Anger	Happiness	Neutral	Sadness
Session1	1F	29	24	119	43
	1 M	33	9	104	61
Session2	2F	9	21	119	33
	2 M	13	33	98	67
Session3	3F	22	22	77	94
	3 M	68	38	121	96
Session4	4F	63	5	91	38
	4 M	21	26	83	43
Session5	5F	24	49	143	82
	5 M	7	57	144	51

Рисунок 2. Количество эмоциональных предложений для каждого диктора в блоке спонтанной речи набора данных IEMOCAP [61]

Фреймворк BLSTM-DSA (полное написание, Bi-directional Long-Short Term Memory with Directional Self-Attention) модели состоит из двух частей. Первая часть – это процесс декодирования BLSTM. В этой части двунаправленный LSTM (полное написание, Long-Short Term Memory) используется для декодирования параметров речевого сигнала вперед и

назад. Вторая часть представляет собой процесс кодирования механизма направленного внутреннего внимания. В этой части к временному шагу для прямых и обратных функций добавляется механизм самоконтроля с использованием двунаправленных декодированных функций LSTM. Модель нейронной сети BLSTM-DSA представлена на рисунке 3.

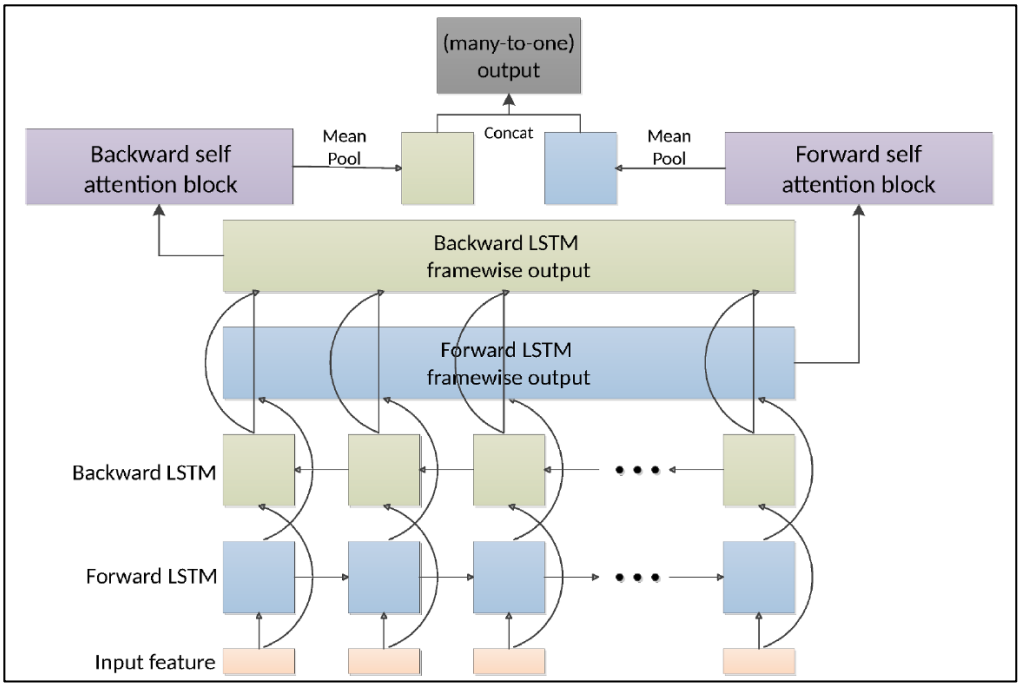

Рисунок 3. Архитектура BLSTM-DSA [61]

Для кодирования были использованы 4 числа, т.е. 4 эмоции. 0 представляет собой эмоцию злости, 1 – радость, 2 – нейтральность, 3 – печаль.

Параметры, которые были извлечены из аудио-файлов, представлены на рисунке 4, где символ △ обозначает first order difference.

Feature	Describe
MFCC(1–13) + △	Mel-Frequency Cepstral Coefficients (1–12)
Spectral roll off + △	Part of the frequency
Spectral flux + △	The rate of change of the power spectrum
Spectral centroid + △	The center of mass of the spectrum
Spectral entropy + △	The disturbance in the signal frequency spectrum
Spectral spread + △	Distribution of signals around the spectrum Center
Zero-crossing rate + △	The rate of sign-changes
Chroma(1–12) + △	Represents 12 different pitches
Pitch + △	Fundamental frequency of sound
Standard deviation of chroma + △	Deviation of 12 different pitches from central values
Energy entropy + △	Time Domain Distribution of Signals
Energy + △	The energy of sound

Рисунок 4. Выделенные параметры [61]

Результаты работы разработанной нейронной сети (BLSTM-DSA) в сравнении с классическими вариантами (CNN, LSTM, BLSTM) представлены в таблице 1 и 2.

Таблица 1. Результаты работы нейронной сети на основе набора данных IEMOCAP [61]

Название нейронной сети	WA (weighted accuracy)
IEMOCAP Spontaneous Database (набор данных со спонтанной речью)	
CNN	57,75%
LSTM	61,89%
BLSTM	62,01%
BLSTM-DSA	62,16%
IEMOCAP Script Database (набор данных со сценарной речью)	
CNN	45,70%
LSTM	47,85%
BLSTM	51,05%
BLSTM-DSA	53,09%
IEMOCAP Complete Database (набор данных сценарная речь + спонтанная речь)	
CNN	49,60±0,0080%
LSTM	57,88±0,0049%
BLSTM	60,60±0,0021%
BLSTM-DSA	61,20±0,0019%

Таблица 2. Результаты работы нейронной сети на основе набора данных EMO-DB [61]

Название нейронной сети	WA (weighted accuracy)
EMO-DB (Berlin Database of Emotional Speech)	
CNN	61,34±0,0226%
LSTM	81,03±0,0061%
BLSTM	81,12±0,0069%
BLSTM-DSA	85,95±0,0049%

Экспериментальные результаты демонстрируют, что предложенный алгоритм показывает особенно хорошие результаты на наборе данных EMO-DB.

Attention gated tensor neural network architectures for speech emotion recognition [68]

Для задачи распознавания эмоций в речи (SER) авторами были предложены тензорная факторизованная нейронная сеть (полное написание, Tensor Factorized Neural Network, далее TFNN) и тензорно-факторизованная нейронная сеть с тензорным управлением (полное написание, Attention Gated Tensor Factorized Neural Network, далее AG-TFNN).

Преимущество TFNN модели заключается в уменьшенном количестве параметров и вычислительной сложности по сравнению с архитектурой CNN + LSTM. Каждый скрытый слой в TFNN модели имеет обучаемые факторные матрицы, равные количеству мод во входном тензоре. Кроме того, созданный тензор признаков может быть сжат путем подходящего усечения матриц факторов, что дает контроль над тем, насколько мы хотим сократить количество параметров. Последовательное вычисление рекуррентных нейронных сетей ограничивает распараллеливание в обучающих примерах, что имеет решающее значение в сценариях, где длина последовательности больше. В отличие от архитектуры CNN + LSTM, где часть CNN – это фиксация локальной информации, а слой LSTM отвечает за отслеживание долгосрочных контекстных зависимостей, модель TFNN тесно интегрирует локальное изучение функций и долгосрочные глобальные зависимости благодаря уровню тензорной факторизации. Поскольку входную спектрограмму не нужно векторизовать, пространственная корреляция пикселей не нарушается, и, следовательно, долгосрочные зависимости сохраняются в TFNN слое. Таким образом, сложную архитектуру CNN + LSTM можно заменить блоком TFNN без особого ущерба для производительности, но в то же время получить преимущества меньшего количества параметров и вычислительной сложности, что дополнительно подтверждается экспериментальными результатами. Архитектура модели AG-TFNN представлена на рисунке 5.

На вход параллельной сети AG-TFNN для SER подаются мел-спектрограммы и модуляционные спектрограммы.

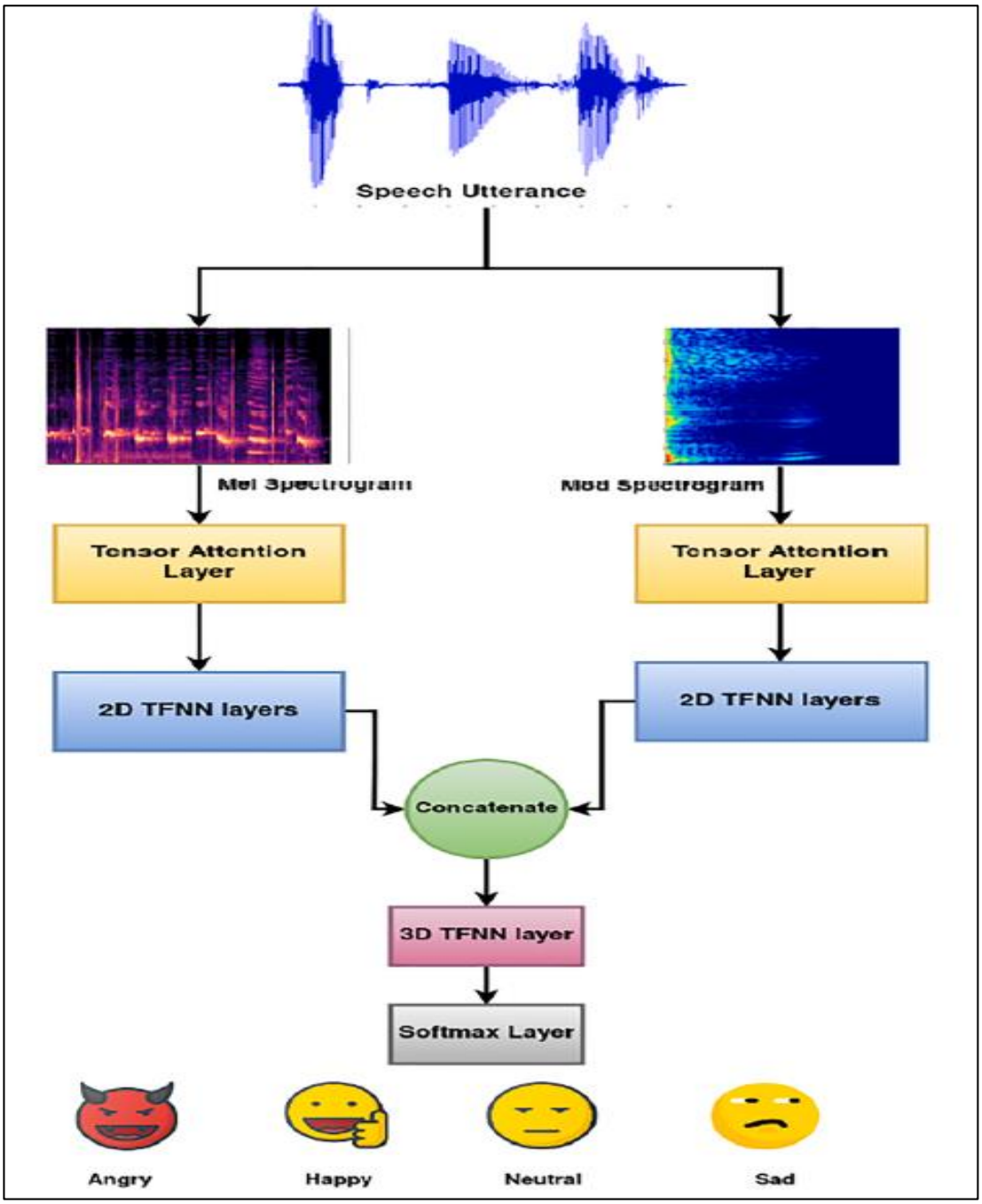

Рисунок 5. Предлагаемая параллельная сеть AG-TFNN [68]

Основные наборы данных описаны ниже.

1. Набор данных Emo-DB (Berlin Database of Emotional Speech). Набор данных Emo-DB является одной из широко используемых баз данных в исследованиях распознавания речи и эмоций. Он включает в себя семь категорий эмоций – злость, отвращение, страх, радость, печаль, удивление и нейтральность – и был записан в 2005 году. Десять немецких актеров – 5 мужчин и 5 женщин озвучили 10 предложений – 5 длинных и 5 коротких. Сигналы записывались с частотой дискретизации 48 кГц, а затем понижались

до 16 кГц. Высказывания имеют продолжительность от 1 до 4 секунд. Набор данных состоит из 535 высказываний, распределенных по семи категориям эмоций.

2. Набор данных IEMOCAP. Interactive Emotional Dyadic Motion Capture (IEMOCAP) – это набор данных с эмоциональной речью на английском языке, состоящий из данных о движении (захват мимики лица, браслеты на руках, и повязка на голове), аудиофайлов, текстовых транскрипций, видео. Набор данных содержит около 12 часов данных. Набор данных записывался по сценарию и в импровизационной форме, специально выбранной для выявления эмоциональных выражений. Эмоции, представленные в наборе данных: злость, волнение, разочарование, радость, нейтральность, печаль, удивление, также материал аннотировался метками валентности и доминирования.

Для исследования перечень эмоций был сокращен до 4: злость, радость, нейтральность и печаль. Общее количество высказываний по двум наборам данных представлено на рисунке 6.

Dataset	Emotion Classes				Total Utterances
	Anger	Happy	Neutral	Sad	
Emo-DB	128	71	79	61	339
IEMOCAP	1103	1636	1708	1084	5531

Рисунок 6. Распределение высказываний по классам эмоций. Наборы данных Emo-DB и IEMOCAP [68]

Предложенная архитектура показала результат в 53,15% по критерию Weighted Accuracy для набора данных IEMOCAP, и 85,56% для набора Emo-DB.

A novel dual attention–based BLSTM with hybrid features in speech emotion recognition [48]

Авторами статьи была предложена архитектура BLSTM (полное написание, Bidirectional Long-Short Term Memory) модели нейронной сети с двойным вниманием, представленная на рис. 7. Этой структуре на вход

подается три параметра: MFCC, дельта и дельта-дельты, где дельта – это производная по времени от MFCC, которая представляет собой скорость изменения MFCC во времени, а дельта-дельты – это производная времени от дельты.

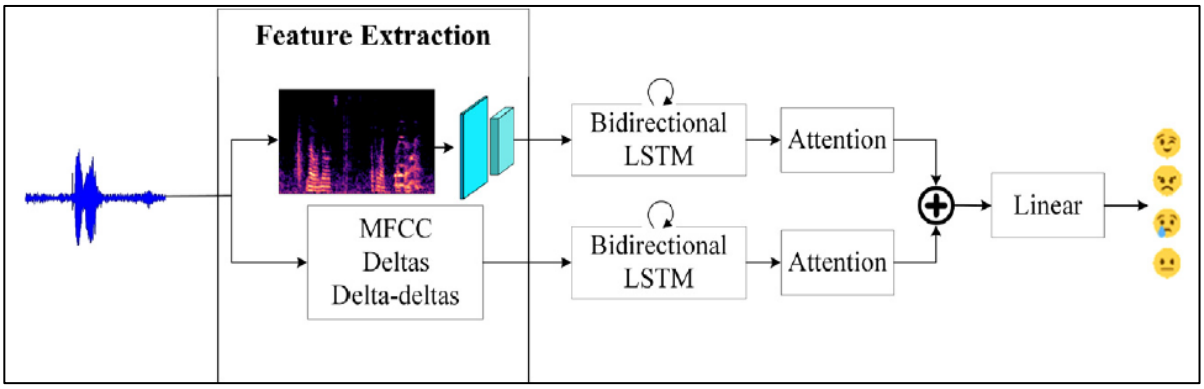

Рисунок 7. Предложенная структура BLSTM модели на основе двойного внимания [48]

Структура BLSTM модели нейронной сети с механизмом внимания (рис. 8), состоит из четырех компонентов, которые перечислены ниже.

1. Входной слой. В модель поступают значения MFCC и дельта признаки (MFCC, Delta, Delta-deltas).

2. Уровень BLSTM. BLSTM используется для извлечения высокоуровневых представлений на этапе 1. Уровень состоит из двух LSTM слоев, которые отдельно обрабатывают каждый параметр, поступивший на входной слой.

3. Слой внимания. Где генерируются веса векторов, и характеристики уровня кадра для каждого временного шага. Они умножаются на весовой вектор для формирования вектора характеристик уровня.

4. Выходной слой (Linear). Используется для вывода функций уровня генеративного высказывания.

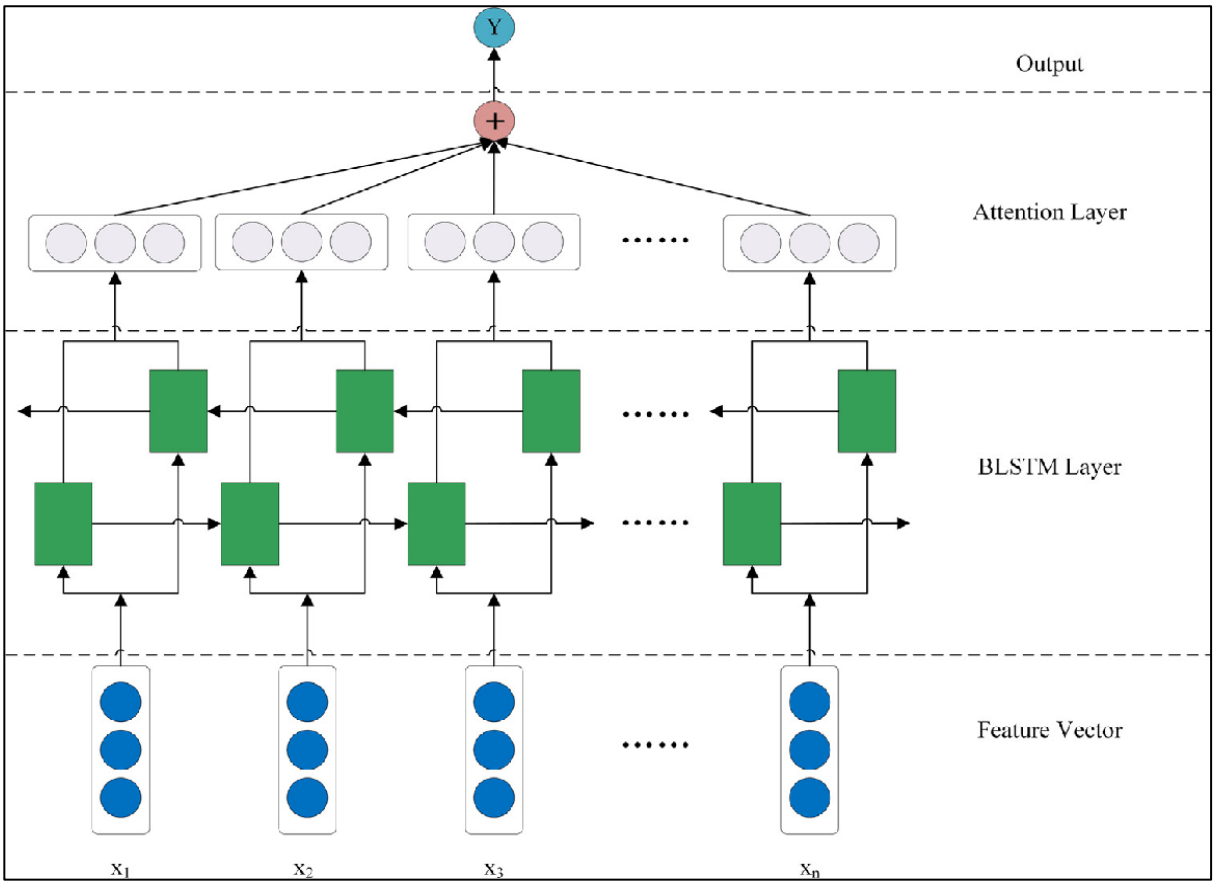

Рисунок 8. Структура сети BLSTM на основе внимания [48]

Обучение и тест модели проходил на IEMOCAP database (Interactive Emotional Dyadic Motion Capture). В набор данных входит 12 часов записи аудио-видео фрагментов, сопровождаемые текстовой транскрипцией. Средняя длина фрагмента составляет 4.5 секунды. Эмоции, которые вошли в набор данных: злость, волнение, разочарование, радость, нейтральность, печаль, удивление. Для исследования был взят 5 531 аудио-фрагмент, и перечень эмоций сокращен до 4: радость, нейтральность, печаль, злость. Сводная таблица и процентное соотношение аудио-фрагментов представлено на рисунке 9.

Used emotions	Used audio files	Contribution ratio
Happy	1636	29.58%
Neutral	1708	30.88%
Sad	1084	19.60%
Angry	1103	19.94%

Рисунок 9. Количество фрагментов и их процентное соотношение. Набор данных IEMOCAP [48]

Эксперименты на наборе данных IEMOCAP показывают преимущество предложенного подхода. Средняя точность распознавания составляет 70,29%.

Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition [43]

В данной статье для распознавания эмоций в речи авторами предлагается новый подход, основанный на трехмерной сверточной нейронной сети (CNN) с вниманием и модели долговременной краткосрочной памяти (LSTM). Предлагаемая трехмерная модель с вниманием CNN-LSTM обучается сквозным (end-to-end) образом. Для преобразования входных речевых сигналов в «изображение» речи используются методы: спектрограммы, мел-частотные кепстральные коэффициенты (MFCC), кохлеограммы и методы фрактальной размерности. Полученные изображения объединяются в четырехмерный объем данных и используются в качестве входных данных для разработанной 28-уровневой интегрированной трехмерной модели CNN-LSTM. В 3D-модели CNN-LSTM есть шесть 3D-сверточных слоев, два слоя пакетной нормализации (BN), пять слоев Rectified Linear Unit (ReLu), три слоя 3D-Max Pooling, один уровень внимания (Attention), один уровень LSTM, один слой Sequence Folding и один слой Sequence Unfolding, два полносвязных слоя (Fully Connected). Слой внимания связан со слоями свертки. Структура 3D-модели CNN-LSTM с вниманием представлена на рисунке 10.

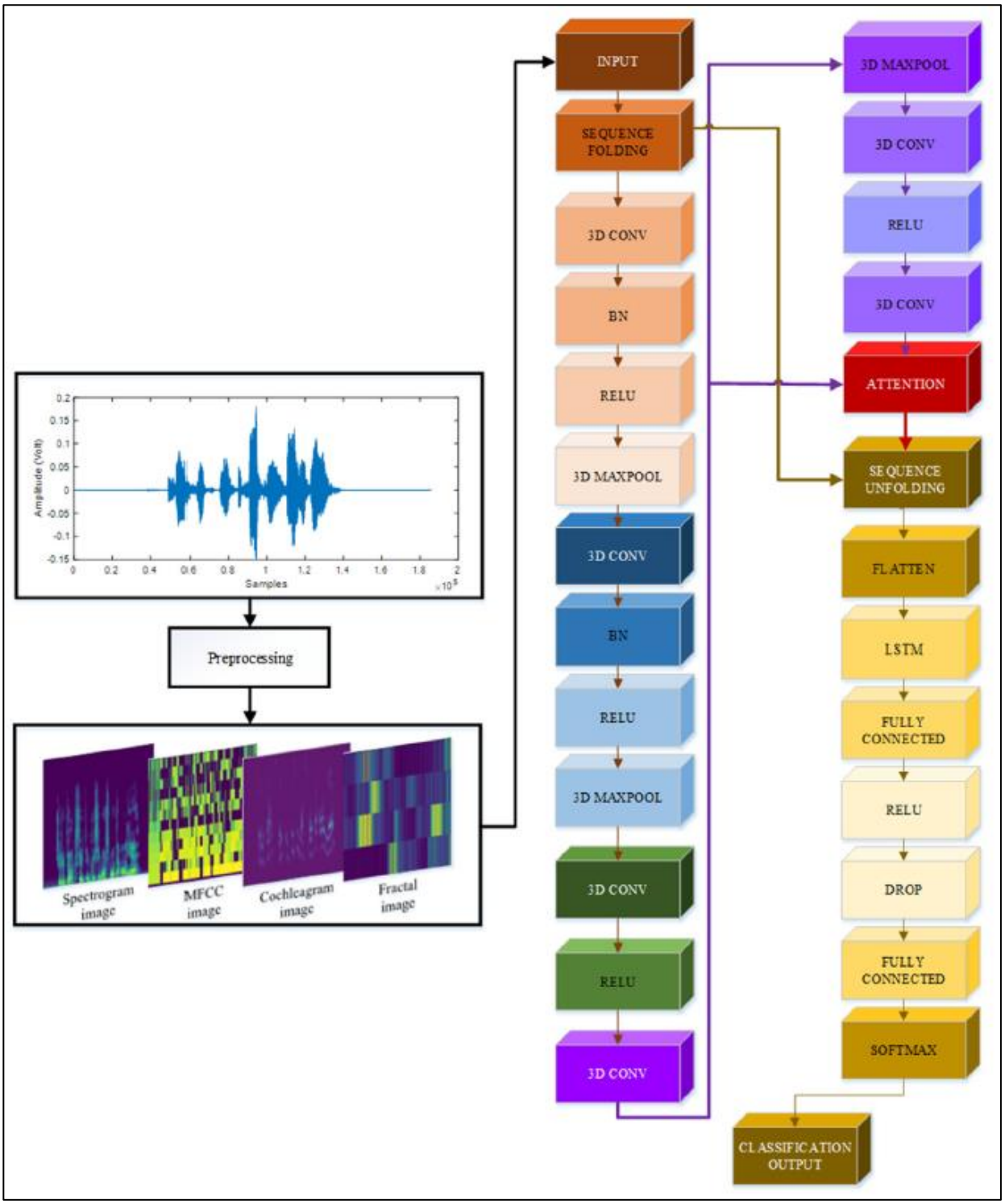

Рисунок 10. Предлагаемая 3D-модель CNN-LSTM с вниманием [43]

В исследовании было задействовано 4 набора данных: RML, SAVEE, RAVDESS, ALL.

Набор данных RML, состоит из 720 файлов: аудиовизуальное выражение эмоций. Во внимание были приняты шесть основных эмоций: злость, отвращение, страх, радость, печаль и удивление. В записи данного

набора данных приняли участие восемь дикторов, говорящих на разных языках: персидском, итальянском, английском, урду, мандаринском и панджаби. Дикторы также говорили с разными акцентами на китайском и английском языках.

В наборе данных SAVEE четыре диктора мужского пола озвучили семь основных эмоций: злость, отвращение, страх, радость, печаль, удивление, нейтральность. В набор данных были записаны 480 выражений: 60 образцов для каждой эмоции и 120 для нейтральной. Дикторы видели текстовые подсказки, относящиеся к каждой эмоции на мониторе во время записи. Набор данных SAVEE содержит фоновый шум.

Набор данных RAVDESS содержит аудиовизуальные записи 12 женщин и 12 мужчин – профессиональных актеров с североамериканским акцентом, произносящих английские предложения с восемью различными эмоциональными выражениями: злость, спокойствие, отвращение, страх, радость, грусть, удивление, нейтральность. Каждое выражение произносилось на нормальном и сильном уровнях эмоциональной интенсивности. Набор данных состоит из 1 440 образцов, по 192 образца для каждого класса эмоций, и 96 образцов для нейтральной эмоции.

Вышеупомянутые наборы данных были смешаны и названы «ВСЕ наборы данных». Набор данных «ALL» также был использован в экспериментальных целях.

Соотношение высказываний по классам эмоций из 4 наборов данных представлено на рисунке 11.

	Angry	Calm	Disgusted	Fearful	Happy	Neutral	Sad	Surprised
RAVDESS	192	192	192	192	192	96	192	192
SAVEE	60	-	60	60	60	120	60	60
RML	120	-	120	120	120	-	120	120
ALL	372	192	372	372	372	216	372	372

Рисунок 11. Распределение количества записей для каждой эмоции [43]

На рисунке 12 приведены рассчитанные средние значения точности для каждой эмоции у каждого из четырех наборов данных.

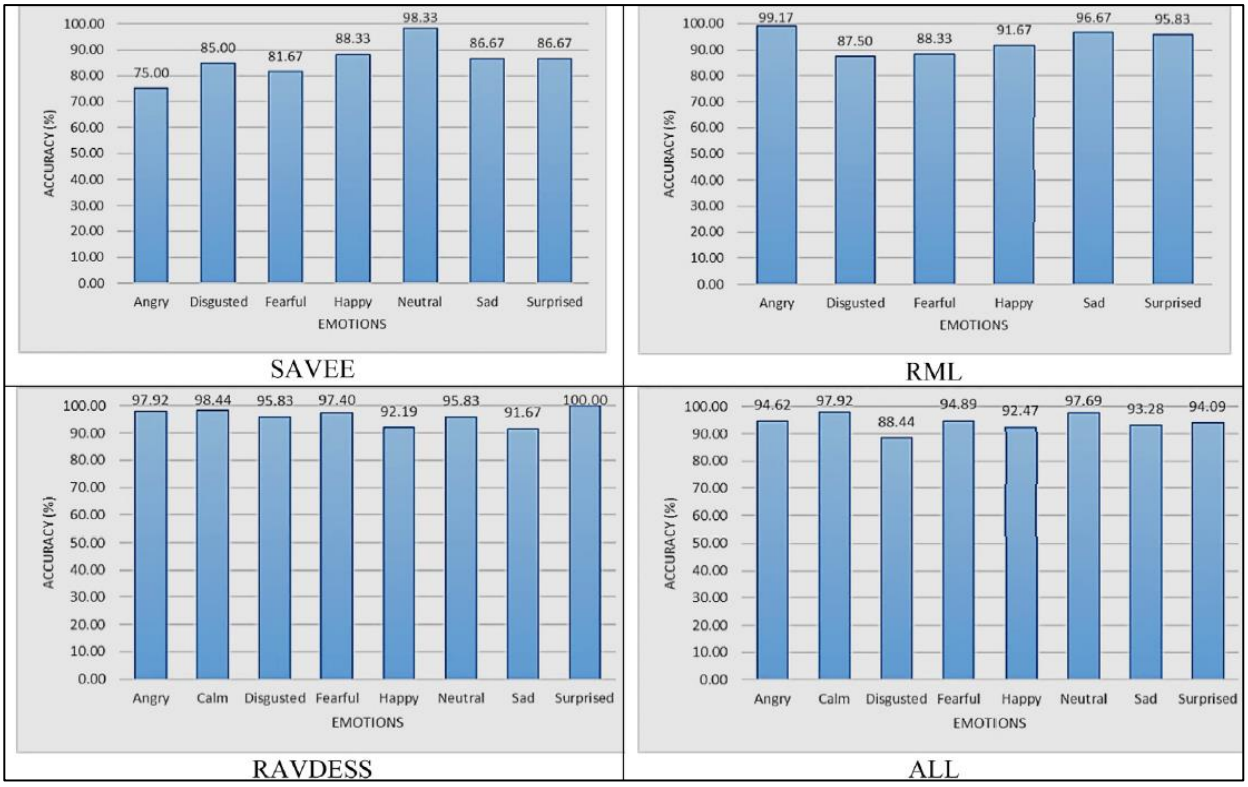

Рисунок 12. Рассчитанные средние оценки точности для каждой эмоции [43]

В данной статье тестируется 3D сеть CNN-LSTM с вниманием. Предлагаемый метод достаточно эффективен, поскольку полученные оценки точности для исследованных наборов данных улучшаются при рассмотрении соответствующих опубликованных результатов в других исследованиях: для наборов данных SAVEE, RAVDESS и RML улучшения составили 2,71%, 8,75% и 7,81% соответственно.

MLT-DNet: Speech emotion recognition using 1D dilated CNN based on multi-learning trick approach [58]

Авторы статьи предлагают сквозную (end-to-end) модель SER в режиме реального времени, основанную на одномерной расширенной сверточной нейронной сети (DCNN). Были предложены три структуры CNN для SER. Для выбора оптимальной модели было проведено множество экспериментов по подбору оптимальных параметров при одинаковых условиях для всех

предложенных методов. Подробные методы и производительность обучения для всех архитектур перечислены ниже.

Модель 1: первоначально предложенная модель для данного исследования. Общая структура CNN Модель 1 показана на рис. 13. В модели использовалась одномерная расширенная CNN с использованием трех остаточных блоков с пропуском соединения для извлечения акустических эмоциональных характеристик из необработанного фрагмента речи. Кроме того, эти изученные функции были переданы в FCN (Fully Connected Network) и переданы из слоя с функцией Softmax с метками на уровне сегмента для получения результатов.

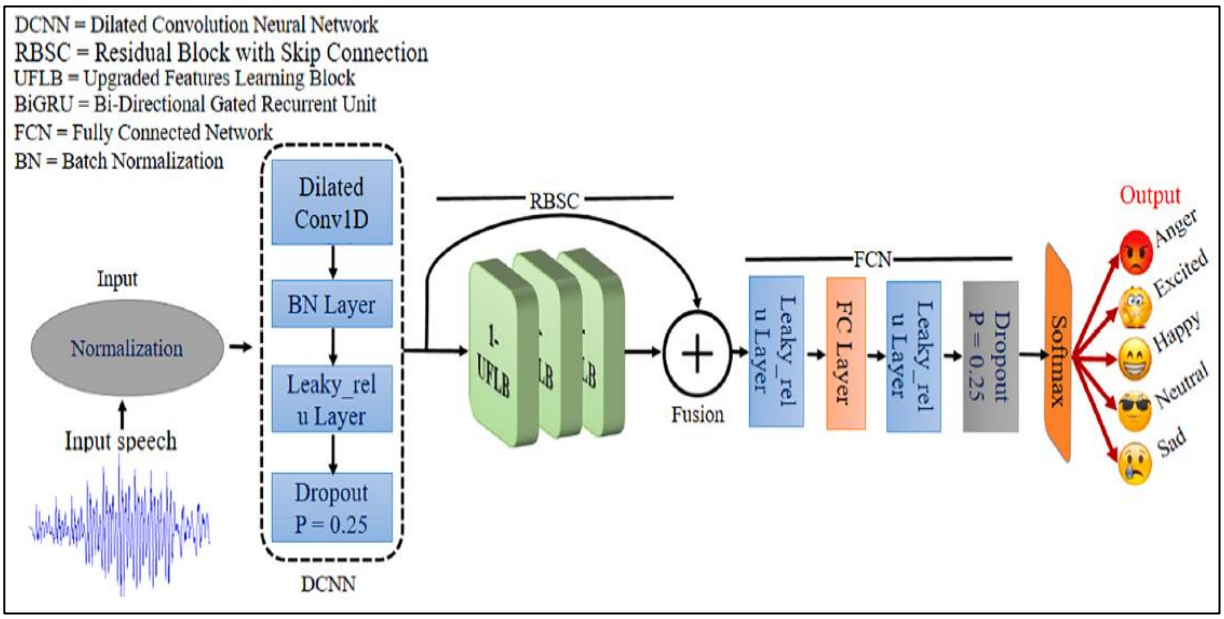

Рисунок 13. Модель 1. Простая расширенная модель CNN, использующая остаточные блоки с пропущенным соединением без приема множественного обучения (MLT) для распознавания эмоций в речи [58]

Модель 2: Модель 1 не может распознавать временные метки в речевых сигналах из-за отсутствия последовательных слоев. Чтобы решить эту проблему, было принято решение добавить последовательные уровни в Модель 1 для изучения временной информации во фрагментах речи, как показано на рис. 14. Был добавлен слой GRU, чтобы изучить долгосрочные зависимости в сигнале. За блоком FCN идет слой с функцией Softmax для получения результатов классификации.

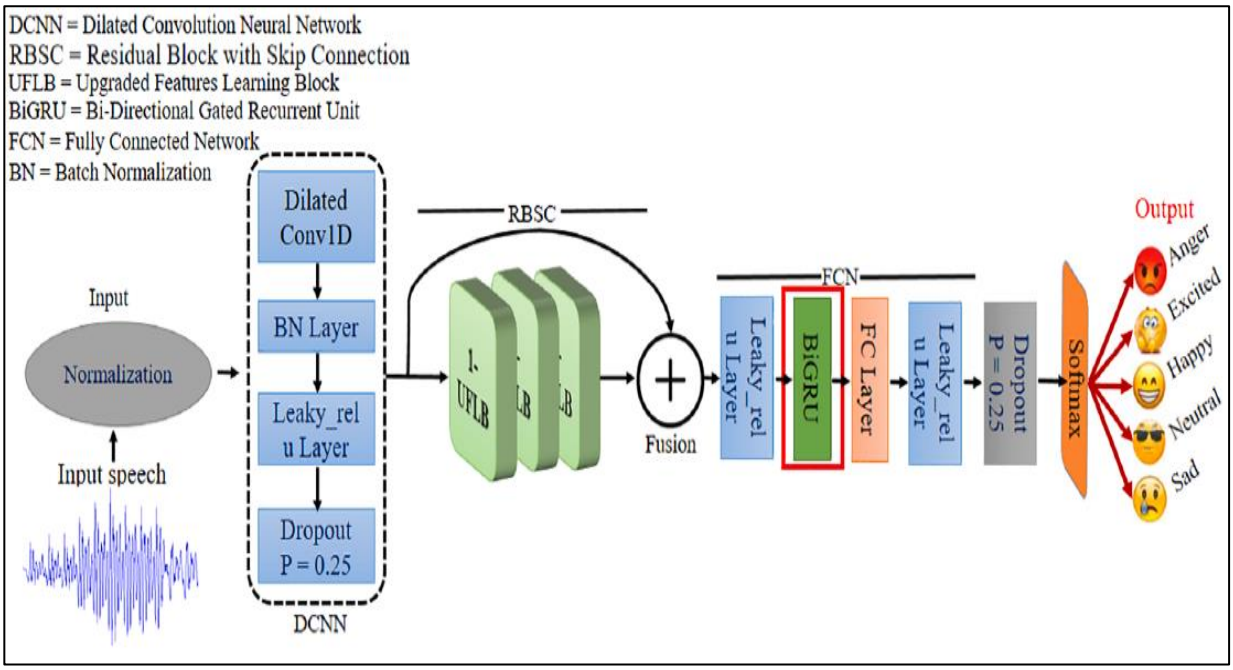

Рисунок 14. Модель 2. Расширенная модель CNN с использованием остаточных блоков с пропущенным соединением с использованием блока GRU без метода множественного обучения (MLT) для распознавания эмоций в речи [58]

Модель 3: в этой модели был использован метод множественного обучения (multi-learning trick/MLT), который показан на рис. 15. Была применена та же стратегия, но изменен метод обучения. Передача начальных акустических признаков с помощью скипового соединения из модуля RBSC (Residual Block with Skip Connection). Точно также первоначальные функции были переданы из модуля Seq_L (Sequence Learning) для изучения последовательной информации с использованием сети GRU со стеком. Наконец, объединение информации и передача ее в блок FCN, чтобы распознать глобальные особенности. Затем передача от классификатора Softmax для получения значений вероятности. Результат этой модели был лучше, чем у двух других моделей для набора данных IEMOCAP и набора данных EMO-DB.

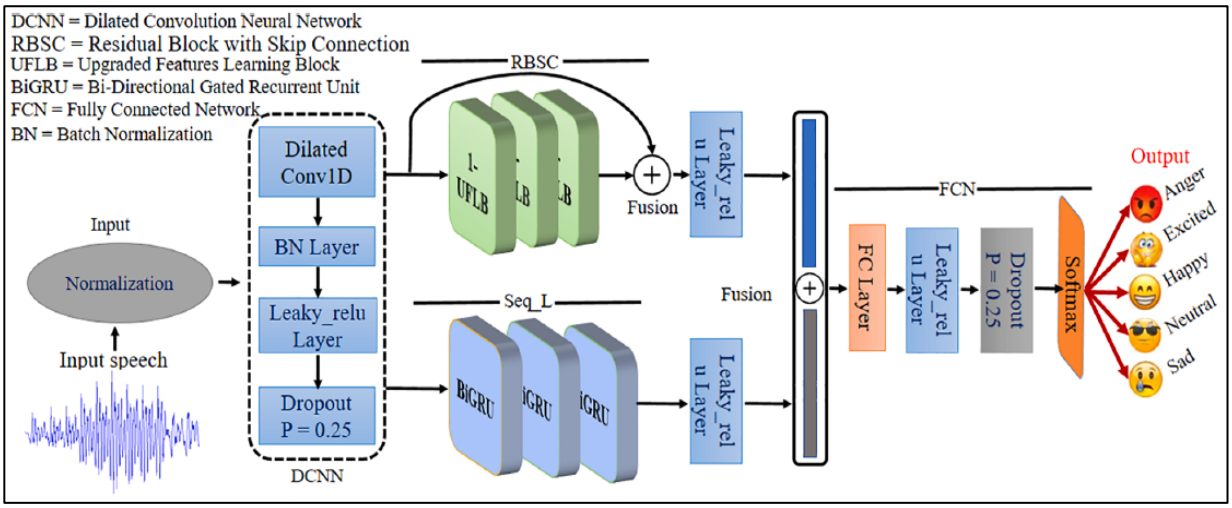

Рисунок 15. Модель 3. Предложенная расширенная модель CNN для распознавания эмоций в речи [58]

Модель нейронной сети тестировалась на двух наборах данных: IEMOCAP и EMO-DB. Эмоции, которые были выбраны для эксперимента из набора данных IEMOCAP: злость, печаль, радость, нейтральность. Из набора данных EMO-DB были выбраны высказывания с эмоциями злости, печали, радости, отвращения, страха, скуки, нейтральности. Количество высказываний и их процентное соотношение представлено на рис. 16 и 17.

Emotion/Class	Total utterances	Participation in (%)
Anger	1103	19.94
Sadness	1084	19.60
Happy	1636	29.58
Neutral	1708	30.88

Рисунок 16. Набор данных IEMOCAP с общим количеством высказываний и процентное соотношение каждого класса [58]

Emotion/Class	Total utterances	Participation in (%)
Anger	127	23.74
Sadness	62	11.59
Happy	71	13.27
Neutral	79	14.77
Disgust	46	8.60
Fear	69	12.90
Boredom	81	15.14

Рисунок 17. Набор данных EMO-DB с общим количеством высказываний и процентное соотношение каждого класса [58]

Результаты представлены на рис. 18 и 19 для каждого набора данных.

Emotion	Precision
Anger	0.85
Happiness	0.51
Neutral	0.87
Sadness	0.65
Weighted	0.75

Рисунок 18. Результаты распознавания по классам эмоций для набора данных IEMOCAP [58]

Emotion	Precision
Anger	0.93
Boredom	0.99
Disgust	1.00
Fear	0.94
Happiness	0.95
Neutral	0.79
Sadness	0.61
Weighted	0.92

Рисунок 19. Результаты распознавания по классам эмоций для набора данных EMO-DB [58]

Результаты, представленные на рис. 18 и 19, показывают, что предложенная система может применяться для распознавания эмоций в реальном времени в речевых сигналах с использованием графического процессора и высокопроизводительных вычислительных устройств.

Understanding human emotions through speech spectrograms using deep neural network [54]

В данной статье представлен сравнительный анализ базовых нейронных сетей: CNN, GRU, DNN, LSTM. Для обучения и тестирования моделей был выбран набор данных RAVDESS. Извлекаемый признак из речевых высказываний – мел-частотный кепстральный коэффициент. Классификация происходила на шесть классов эмоций: радость, печаль, нейтральность, спокойствие, отвращение, страх. Результаты представлены в таблице 3.

Таблица 3. Результаты классификации нескольких типов нейронных сетей [54]

Модель DNN	Результат (в %)
CNN	66
DNN	67
LSTM	70
GRU	68

Основываясь на результатах, которые были достигнуты при помощи простой LSTM модели, можно судить о том, что при создании гибридных моделей глубокого обучения, которые обеспечивают двунаправленные и рекуррентные более плотные архитектуры нейронных сетей можно обеспечить улучшение классификации эмоций.

1.3. Выводы по главе 1

В главе 1 были рассмотрены основные принципы, методы, и подходы к вопросу распознавания эмоций с использованием нейросетевых технологий. Описаны последние тенденции в данной области: модели нейронных сетей, которые чаще всего используются; определен ряд основных методов, которые применятся для решения данных задач; результаты, которые демонстрируют нейронные сети; наборы данных, на основе которых производится обучение и тестирование реализованных нейронных сетей.

На основе проанализированной литературы и наборов данных, находящихся в открытом доступе, было принято решение провести запись набора данных и реализовать сверточную нейронную сеть, на вход которой будут подаваться аудио-фрагменты, преобразованные в изображения.

Глава 2. Получение речевого материала для задачи автоматического распознавания эмоций

2.1. Обоснование выбора перечня эмоций

Первым этапом при решении вопроса распознавания эмоций с использованием нейросетевых технологий является определение списка эмоций для классификации.

«Эмоции (emotion – волнение, возбуждение) – субъективное состояние человека и животных, возникающие в ответ на воздействие внешних и внутренних раздражителей и проявляющиеся в форме непосредственных переживаний (удовольствие/неудовольствие, радость, гнев, и т.д.)» [7, 11].

Специалист по экспериментальной психологии П. Фресс говорит о том, что самой по себе ситуация не может быть эмоциогенной, реакция на ситуацию зависит от мотивации и возможностей человека, т.е. это не просто совокупность сложившихся обстоятельств, но и также самостоятельная оценка происходящего человеком, его отношение к ситуации в зависимости от имеющихся у него потребностей и целей [37]. Таким образом, именно человеческая оценка является первоочередным шагом на пути к созданию эмоциогенной ситуации, а обстоятельства являются предпосылкой возникновения эмоциогенной ситуации. Эмоциогенными становятся лишь те обстоятельства, которые человек считает лично для себя значимыми (т.е. удовлетворяют или не удовлетворяют его потребности, приносят удовольствие или дискомфорт и т.д.) [11].

Эмоции являются социально детерминированными, т.е. они обуславливаются нормами морали и права, свойственные общественно-экономическим формациям [7].

Эмоции характеризуются не только внутренними переживаниями, но и внешними или телесными проявлениями:

1) мимика (изменение положения губ, бровей, расширение, сужение глаз, и т.д.);

2) пантомимика (жестикуляция, поза);

3) тон голоса (вокальная мимика);

4) вегетативные проявления (частота сердечных сокращений, частота дыхания, покраснение или бледность кожи, изменение тонуса мышц, дрожь в теле, интенсивность потоотделения, и т.д.);

5) биохимические изменения (выработка надпочечниками адреналина, повышение уровня сахара в крови, и т.д.) [7].

Таким образом, изучение телесных проявлений помогает в изучении вопроса эмоций.

Важно провести четкую грань между эмоциями и чувствами. С. Л. Рубентштейн в своей работе «Основы психологии» характеризует эмоции следующим образом:

1. Эмоции выражают состояние субъекта и его отношение к объекту.

2. Эмоции имеют полярность. Эмоции взаимодействуют при формировании сложных человеческих чувств, и образуют сложные противоречивые единства.

3. Эмоции имеют характер, который связан с личностным «я», и олицетворяют человека [26].

Отличием чувств от эмоций является то, что чувства – это сложные целостные образования, связанные с социальными потребностями личности. В то время как, эмоции соотносятся с биологическими потребностями человека, т.е. характеризуют удовлетворение или неудовлетворение потребностей [26].

Советский психолог А. Н. Леонтьев пишет о том, что если рассматривать эмоциональные процессы в широком смысле, то к ним относятся аффекты, непосредственно эмоции и чувства [16].

Аффекты характеризуют сильные, кратковременные эмоциональные переживания, которые сопровождаются ярко выраженными двигательными и висцеральными проявлениями, содержание и характер этих проявлений

зависит от воспитания и самовоспитания индивида, а также может изменяться на протяжении всей жизни. Особенностью аффектов является то, что они являются ответной реакцией на уже наступившее/случившееся событие, и в этом смысле они сдвинуты во времени к концу события.

Эмоциями являются более длительные состояние, которые могут слабо проявляться во внешнем поведении. В отличие от аффектов, эмоции выражают оценочное личностное отношение к складывающимся и предстоящим событиям, и возникают на основе пережитых или воображаемых ситуаций.

Выделение чувств как особого подкласса эмоциональных процессов является более условным и менее общепринятым. Они выделяются на основе обобщения эмоций, связанных с некоторым объектом, который может быть как конкретным, так и обобщенным, либо отвлеченным (к таким относится чувство любви (любовь к родине, человеку), чувство ненависти). Чувства характеризуются устойчивым эмоциональным отношением [16].

В данной работе чувства не будут рассматриваться в связи с их сложностью, и зависимостью от пережитого опыта. Рассмотрены будут эмоции и аффекты, поскольку они подлежать внешнему выражению, пусть и порой слабому, и являются реакцией на случившееся, случающееся, или то, что случится в будущем.

Постановка проблемы базовых эмоций – является основополагающей при определении перечня эмоций для классификации материала с использованием нейронных сетей, т.к. из базовых эмоций формируются все остальные чувства и эмоции [16].

Основатель бихевиоризма, американский психолог Джон Б. Уотсон в результате наблюдения за младенцами в первые месяцы их жизни установил группу эмоциональных реакций, которые принадлежат к основной природе человека: страх, ярость, любовь [35].

В 1980 году американский профессор психологии Роберт Плутчик, выделив 8 базовых эмоций и связанные с ними сложные эмоции, представил

концепцию, которая получила название «Колесо эмоций» (Рис. 20). Восемь первичных (базовых эмоций): anger (злость); anticipation (ожидание); joy (радость); trust (доверие); fear (страх); surprise (удивление); sadness (грусть); disgust (отвращение). В центре расположены аффекты (самые интенсивные проявления эмоций); средний круг представляет базовые эмоции; внешний круг – некоторые сложные эмоции. Чем дальше от центра, тем менее интенсивными являются эмоции [70].

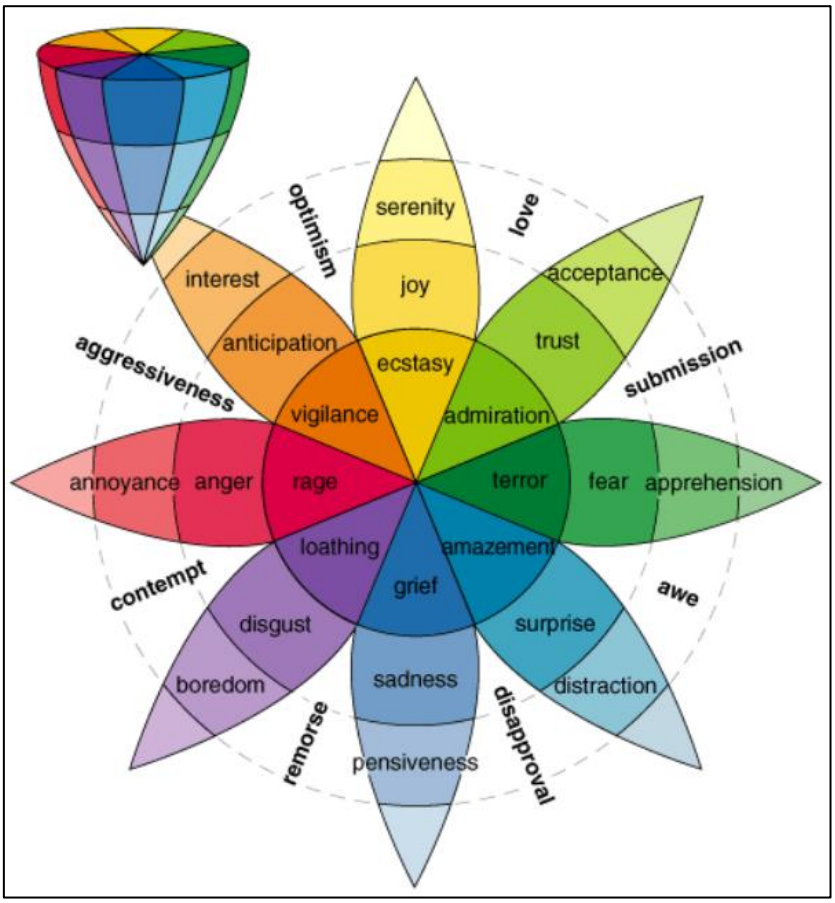

Рисунок 20. «Колесо эмоций» Р. Плутчика [70]

Американский психолог, специалист в области психологии эмоций Кэррол Изард, выделил 10 фундаментальных эмоций: интерес – волнение; радость; удивление; горе – страдание; гнев; отвращение; презрение; страх; стыд; вина [10].

Пол Экман считает неправильным деление эмоций на позитивные и негативные, т.к. удивление, страх, презрение, могут иметь и положительную, и отрицательную полярность. При этом психолог полагает, что есть семь базовых эмоций: гнев (злость); печаль (грусть); презрение; отвращение;

страх; удивление; радость. В жизни эти эмоции встречаются как в чистом, так и смешанном виде, считает психолог [40].

Исследования в области нейропсихологии и нейрофизиологии позволили подойти к решению вопроса классификации эмоций с помощью методов разрушения и стимуляции определенных отделов мозга. Также, были получены данные, свидетельствующие о том, что в мозгу человека присутствуют отдельные структуры, стимулирование или разрушение, которых приводит к определенным эмоциональным состояниям (например, удовлетворение – неудовлетворение) или эмоциям, либо полностью лишает человека способности испытывать те или иные эмоции [38].

Согласно В. М. Смирнову, электрическая стимуляция миндалины головного мозга провоцирует у пациентов эмоции страха, гнева, ярости и редко удовольствия [32].

Как отмечает П. В. Симонов, при прямом раздражении мозга провоцируются только пять эмоциональных состояний: гнев; страх; удовольствие; и противоположные удовольствию – отвращение и дискомфорт. Вполне возможно, что именно эти пять эмоций и являются базисным фондом эмоций [28].

В результате изучения научных работ, в общий перечень базовых эмоций были внесены: страх; ярость; злость; гнев; любовь; доверие; ожидание; радость; удивление; печаль; грусть; отвращение; интерес – волнение; горе – страдание; презрение; стыд; вина; удовольствие; отвращение; дискомфорт.

Было принято решение объединить в группы следующие эмоции. Раздражение: ярость, злость, гнев. Печаль: горе, страдание, грусть. Это связано с тем, что они являются сходными по своей природе, но различаются по степени интенсивности. Полученная сводная таблица базовых эмоций представлена в таблице 4.

Таблица 4. Сводная таблица базовых эмоций

Автор / Эмоция	Дж. Б Уотсон	Р. Плутчик	К. Изард	П. Экман	В. М. Смирнов	П. В. Симонов
Страх	+	+	+	+	+	+
Раздражение (ярость/ злость/гнев)	+	+	+	+	+	+
Любовь	+					
Доверие		+				
Ожидание		+				
Удивление		+	+	+		
Печаль (Грусть/горе/ страдание)		+	+	+		
Отвращение		+	+	+		+
Радость		+	+	+		
Интерес – волнение			+			
Презрение			+	+		
Вина			+			
Стыд			+			
Удовольствие					+/–	+
Дискомфорт						+

В сводной таблице 4 можно обнаружить пересечения, на основании этого, там, где пересечений два и более, было принято решение включить их в перечень базовых эмоций данной работы.

Таким образом, базовый перечень эмоций включает в себя страх, раздражение, удивление, печаль, отвращение, радость, презрение.

Также, было принято решение о включении в перечень более сложной, составной эмоции – ехидства. Ехидство является формой проявления агрессии. Агрессия относится к форме девиантного поведения, которая проявляется в физической либо вербальной форме, целью которого является нанесение вреда кому-либо. Агрессия может быть прямой или косвенной. Прямая агрессия проявляется в виде действий: физическое насилие, нападение, и т.д. Косвенная агрессия: неприязненное отношение, ироничные высказывания в адрес человека, ехидство, сарказм. Косвенная агрессия

оказывает психологическое давление и влияние на жертву [34]. Доминирующими характеристиками является антипатия, раздраженность и недовольство поведением окружающих. Однако, трактоваться она может, как печаль, раздражение или вовсе не восприниматься собеседником [20]. Именно ехидство, как форма косвенной агрессии и проявления раздражения была выбрана в качестве сложной эмоции, которая вошла в список эмоций данного исследования.

В повседневной жизни каждый человек сталкивается с нейтральными высказываниями, это могут быть вопросы, объявления, констатация фактов, и т.д. Поэтому в список эмоций была также включена эмоция нейтральности, которая не несет в себе никакого эмотивного содержания.

В результате, в общий перечень эмоций данного исследования вошли: страх, раздражение, удивление, печаль, отвращение, радость, презрение, и нейтральность, как базовые. Эмоция ехидства, которая является сложной составной эмоцией, была также включена в общий перечень эмоций в экспериментальных целях, для того, чтобы проследить, куда данная эмоция будет отнесена респондентами: как самостоятельная эмоция ехидства или ее составляющие, т.е. раздражение, печаль, и т.д.

2.2. Обоснование списка фраз

Для записи корпусов эмоциональной речи обычно приглашают профессиональных актеров [3,47, 52, 73, 76].

Методика, разработанная П. В. Симоновым и его коллегами для изучения восприятия эмоционального состояния по выражению лица и голосу, содержит проективные элементы и направлена на обнаружение доминирующей эмоции [14]. Именно по этой методике формируется большинство корпусов для подобных исследований. Материал для предоставления стимула на слух респонденту записывается с участием профессиональных актеров. В случае эксперимента П. В. Симонова, использовался метод К. С. Станиславского, т.е. актёр мысленно представляет

себе ситуацию в соответствии с эмоциональным состоянием, которое необходимо было реализовать [27]. Отбор материала осуществлялся на основе принципа наличия эмоционального напряжения, т.е. соответствие мимики и интонации, частоты сердечных сокращений, дыхание и т.д. В итоге были получены последовательности из 10 высказываний, которые были произнесены одним диктором с разной эмоцией: радость, горе, страх, гнев, нейтральность [14, 27, 29]. Суть эксперимента заключалась в том, что было два набора данных: первый – высказывание было одно и то же, задача актера была произнести одну фразу с различными эмоциональными состояниями, второе – фразы, которые получили название конфликтные пробы, т.е. фразы, содержание которых противоречит ее эмоциональному выражению. Задача конфликтной пробы состояла в том, чтобы выявить слово, которое провоцирует помеху для распознавания эмоции [14, 27]. Респонденты должны были:

1) при первом прослушивании каждого высказывания определить эмоциональное состояние говорящего;

2) при втором прослушивании все записи с одним и тем же выражением эмоции повторялись подряд;

3) при третьем прослушивании сопоставить фотографии лица человека, выражающего различные эмоции, с интонацией в голосе [14, 21, 22, 27].

В своих экспериментах В. П. Морозов прибегнул к такому же методу, что и П. В. Симонов. В состав эксперимента вошла речь профессиональных актеров и вокальная речь музыкальных исполнителей. Одна и та же фраза была воспроизведена дикторами с различными эмоциями. Записанный материал проходил отбор и классификацию по характеру естественности и степени выраженности той или иной эмоции методом экспертных оценок с участием специалистов по сценической речи [21, 22].

Подобными экспериментами доказывается тот факт, что лексическое содержание речи независимо от его эмоционального выражения речи, по

этой причине, лексический состав высказывания и его эмотивную реализацию можно считать самостоятельными.

В целях научных исследований существует несколько способов получения эмоционально окрашенной речи:

1) естественное эмоциональное состояние человека;

2) внушение эмоционального состояния под гипнозом;

3) болезненные состояния (депрессия, эйфория, шизофрения и т.д.) [22].

В рамках данной работы будет рассмотрен и применен метод естественного эмоционального состояния человека.

Стоит обратить внимание на то, что при записи любого из вышеперечисленных наборов данных, участие принимали профессиональные актёры, которые получают специальное образование в области актерского мастерства и сценической речи. Подобные реализации эмоций, какие мы встречаем в театре или в кино, могут быть неестественными в условиях реальной жизни.

Для респондентов, которые будут принимать участие в перцептивном эксперименте, будет важна естественность речи и реализации эмоций, т.к. гиперболизированная реализация высказываний может сбить респондентов с толку, а в повседневной жизни, люди не склонны реализовывать свои мысли так, как это делают люди со специальным образованием. В связи с этим, необходимо учитывать такие факторы, как эмоциональный слух и эмоциональный интеллект.

Понятие «эмоциональный слух» (ЭС) было впервые введено В. П. Морозовым и представлено в статье 1985 года «Эмоциональный слух человека». В своем труде В. П. Морозов пишет о том, что есть люди эмоционально тугоухие, т.е. они плохо слышат эмоциональную интонацию, и люди с утонченным эмоциональным слухом [29].

Морозов В. П. характеризует ЭС как способность воспринимать и понимать язык эмоций. Теоретически ЭС, является частью невербальной

системы коммуникации. Человек способен адекватно оценивать эмоциональную информацию, которая представлена в звуковой форме (речь, пение). ЭС можно протестировать и оценить. В большей степени, ЭС является природным качеством человека, которое также можно развить с помощью специальных упражнений [21].

В исследовании эмоционального слуха выделяют следующие направления исследований перечисленные ниже.

1. Выделение спектральных, временных и динамических характеристик акустического сигнала голоса человека, посредством которых может передаваться информация об эмотивности речи [1, 24, 57].

2. Разработка систем по автоматическому распознаванию эмоций [56, 63].

3. Выделение признаков, при контроле которых удастся фиксировать эмоциональное и физическое состояние людей, деятельность которых подразумевает высокие риски и жертвы (авиадиспетчеры, водители общественных и рейсовых маршрутов, т.д.) [62, 65, 78, 80].

4. Изучение мимико-интонационных, физиологических, вегетативных выражений эмоций [23, 41, 69].

Понятие «эмоциональный интеллект» (ЭИ) впервые появилось в зарубежной психологии в 1990 году благодаря Джону Мэйеру и Питеру Сэловея [74].

Позднее их модель эмоционального интеллекта была модернизирована и включила критерии, которые составляют структуру эмоционального интеллекта, они перечислены ниже.

1. Идентификация эмоций. Восприятие эмоций, т.е. факт наличия/отсутствия эмоции и их идентификация; адекватное выражение своих эмоций; различение подлинных и искусственных эмоций.

2. Использование эмоций для повышения эффективности мышления и деятельности. Способность использовать собственные эмоции

для концентрации внимания на важном событии, вызывать эмоции, которые способствуют решению поставленных задач и т.д.

3. Понимание эмоций. Способность понимать комплексы эмоций; связи между эмоциями; переходы от одной эмоции к другой; причины эмоций; информацию, которая выражается об эмоциях вербально.

4. Управление эмоциями. Контроль эмоций и их интенсивности [11].

Понятия ЭС и ЭИ являются схожими друг с другом. Уровень эмоционального интеллекта и слуха варьируется от человека к человеку в связи с чем, при оценке материала результаты опросов могут быть неоднородными.

В связи с тем, что тест на уровень эмоционального слуха В. П. Морозова найти не удалось, было принято решение провести тестирование респондентов, которые примут участие в перцептивном эксперименте, на уровень эмоционального интеллекта по методике Н. Холла [Приложение В]. Методика Холла направлена на выявление способности человека: понимать отношения личности, передающиеся посредством эмоций, и управлять эмоциональной сферой. Тест Н. Холла состоит из 30 утверждений, и 6 вариантов ответа. Н. Холл рассматривает ЭИ по нескольким критериям, которые перечислены ниже.

1. Эмоциональная осведомленность. Понимание и осознание своих эмоций, непрерывное пополнение собственного словаря эмоций. Чем выше данный показатель, тем выше уровень осведомленности человека о его эмоциональном состоянии.

2. Управление своими эмоциями. Эмоциональная гибкость, способность управлять своими эмоциями.

3. Самомотивация. Управление своими эмоциями.

4. Эмпатия. Понимание эмоций других людей, способность сопереживать эмоциональному состоянию другого человека, готовность оказать поддержку.

5. Распознавание эмоций других людей. Способность понимать и воздействовать на эмоциональное состояние других людей [11].

Важность, в рамках данной работы, представляют такие критерии, как эмоциональная осведомленность, эмпатия и распознавание эмоций других людей, т.к. чем выше значения этих критериев, тем выше вероятность того, что человек сможет правильно понимать и трактовать эмоции другого человека (см. 2.4.1. Результаты перцептивного эксперимента).

Для получения фраз, которые будут естественными по своей эмоциональной составляющей, было высказано предположение – можно составить эмотивные высказывания и предоставить соответствующий к ним контекст, таким образом, что при реализации высказываний дикторы без подсказки экспериментаторов и просьбы воспроизвести конкретную эмоцию, смогут самостоятельно догадаться об эмоции, которая заключена в высказывании и сопутствующем ей контексте, и воспроизвести одни и те же высказывания с одинаковой эмоцией, которая будет отличаться лишь в степени своей интенсивности.

Таким образом, в данной работе было решено, что для того, чтобы итоговый корпус соответствовал более естественным реализациям эмоций и был вариативен – необходимо составить перечень фраз и соответствующие им контексты. Воспользовавшись методом К. С. Станиславского, дикторам обозначили контексты, погрузившись в которые, они должны будут воспроизвести эмотивные фразы [27].

Был проанализирован перечень литературы об эмотивной вербальной русской речи [2, 17, 25, 36, 39]. Составлен список эмотивных фраз и соответствующие им контексты (см. Приложение А).

2.3. Описание изначального набора данных

При подготовке материала для обучения и тестирования нейронной сети использовался собственный набор данных. Набор данных (Context-

dependent emotional speech dataset) представляет из себя корпус с эмоциональными высказываниями.

Этапы создания набора данных представлены ниже.

1. Был составлен список из 40 фраз, в который входили фразы со следующими эмоциями: раздражение, печаль, презрение, страх, отвращение, радость, удивление, и нейтральность. Также, в список входит одна фраза с эмоцией ехидства, список фраз представлен в Приложении А.

2. Каждой фразе был предоставлен контекст, в котором фраза должна прозвучать.

3. Были найдены и приглашены дикторы, которых просили озвучить сформированные списки фраз, при этом эмоции были удалены из списка для чистоты эксперимента. Дикторам было представлено две колонки: контекст и текст фразы.

4. Задача была поставлена следующим образом: про себя прочитать контекст фразы, мысленно погрузиться в него и озвучить соответствующую контексту фразу так, как озвучил бы её диктор, оказавшись в данной ситуации.

5. Каждому диктору был представлен один из 5 списков, в который входило 8 фраз (см. Приложение Б) из общего списка, после прочтения списка фраз с добавлением эмоциональной окраски, дикторов просили озвучить тот же самый список фраз, но уже в нейтральном прочтении.

По итогам проделанной работы получились следующие результаты:

1. Было записано 72 диктора мужского пола в возрасте от 20 до 60 лет.

2. В общий корпус записанных фраз в результате вошли 1 442 аудио-фрагмента. Набор данных, был подвергнут перцептивному эксперименту.

3. Тип файлов – WAVE.

4. Частота дискретизации – 22 050 Гц.

5. Общий временной объем файлов составил 1 час 17 минут.

Результаты проделанной работы представлены в таблице 5. В связи с тем, что возникла разбалансировка данных (неравномерное соотношение материалов по классам), было принято решение о сокращение количества единиц материала в классе «Нейтральность», данное значение является медианой других значений эмоций представленных в колонке. Набор данных без разбалансировки также представлен в таблице 5.

Таблица 5. Начальные наборы данных

Эмоция	Количество записей	
	Начальный набор данных с разбалансировкой	Начальный набор данных без разбалансировки
Ехидство	15	15
Нейтральность	781	86
Отвращение	57	57
Печаль	115	115
Презрение	105	105
Радость	86	86
Раздражение	173	173
Страх	56	56
Удивление	54	54
Итого:	1442	747
Время:	1 час 17 минут	39 минут 15 секунд

Для бинарной классификации начальный набор данных представлен в таблице 6.

Таблица 6. Начальный набор данных для бинарной классификации

Эмоция	Количество записей
Негативная	521
Нейтральная	781
Итого:	1302
Время:	1 час 9 минут

2.4. Перцептивный эксперимент

Для того, чтобы проверить записанный материал: действительно ли у дикторов получилось реализовать предполагаемые эмоции и получилось ли

это сделать таким образом, что воспроизводимые ими фразы будут восприниматься одинаково в большинстве случаев; было решено организовать перцептивный эксперимент.

Перцептивный эксперимент – (от лат. *perceptio* «восприятие») – исследование лингвистических реакций носителя языка на речевые сигналы, которое заключается в способности опознавать и отличать звуковые единицы языка, слоги, слова в разных экспериментальных условиях, оценивать и фонологически интерпретировать изменение акустических характеристик речевого сигнала и др [15].

Перцептивный эксперимент был организован на платформе Google-Forms. Было необходимо проверить 1 442 аудио-файла.

Было составлено 14 опросов, включающие в себя от 85 до 117 вопросов. Ссылки на опросы представлены ниже.

1. Опрос №1. URL: https://forms.gle/CqQX6whGT2VFoRmTA.
2. Опрос №2. URL: https://forms.gle/Cu7WVwXq5VGTqNrx6.
3. Опрос №3. URL: https://forms.gle/dCP2G9Y7bFuGczyk9.
4. Опрос №4. URL: https://forms.gle/RYFa2WneWCGVPd2G8.
5. Опрос №5. URL: https://forms.gle/RvJ66mTtNBuAov627.
6. Опрос №6. URL:https://forms.gle/eAgQw4K8bdrvLHZR9.
7. Опрос №7. URL: https://forms.gle/i1bBxrCSnPvHL5er5.
8. Опрос №8. URL: https://forms.gle/ucmqMbfwrgUJja7C9.
9. Опрос №9. URL: https://forms.gle/upd3zWsK56nEbFAJ8.
10. Опрос №10. URL: https://forms.gle/NKE648MQeoL4TUuV8.
11. Опрос №11. URL: https://forms.gle/ihwkT9ndWm8zRyz89.
12. Опрос №12. URL: https://forms.gle/xDXTbrfgdbfDupxd9.
13. Опрос №13. URL: https://forms.gle/4r56iCtCiGq3cpx69.
14. Опрос №14. URL: https://forms.gle/jnqJ7WWk1UuYANmq8.

Вопрос к каждому аудио-фрагменту был сформулирован следующий: «Какую эмоцию реализует в данном высказывании диктор?». Указана ссылка на файл, на Google-диске, и к вопросу предложено 9 вариантов ответа:

раздражение (ярость, гнев, раздражение, злость, ненависть), печаль, презрение, страх, отвращение, радость, удивление, нейтральность, ехидство. Также, десятым вариантом была оставлена свободная форма для заполнения: «Другое…», чтобы респонденты могли указать свой вариант.

Каждый опрос начинается с просьбы к респонденту указать свое имя и возраст.

Пример опроса представлен на рис. 21 и 22.

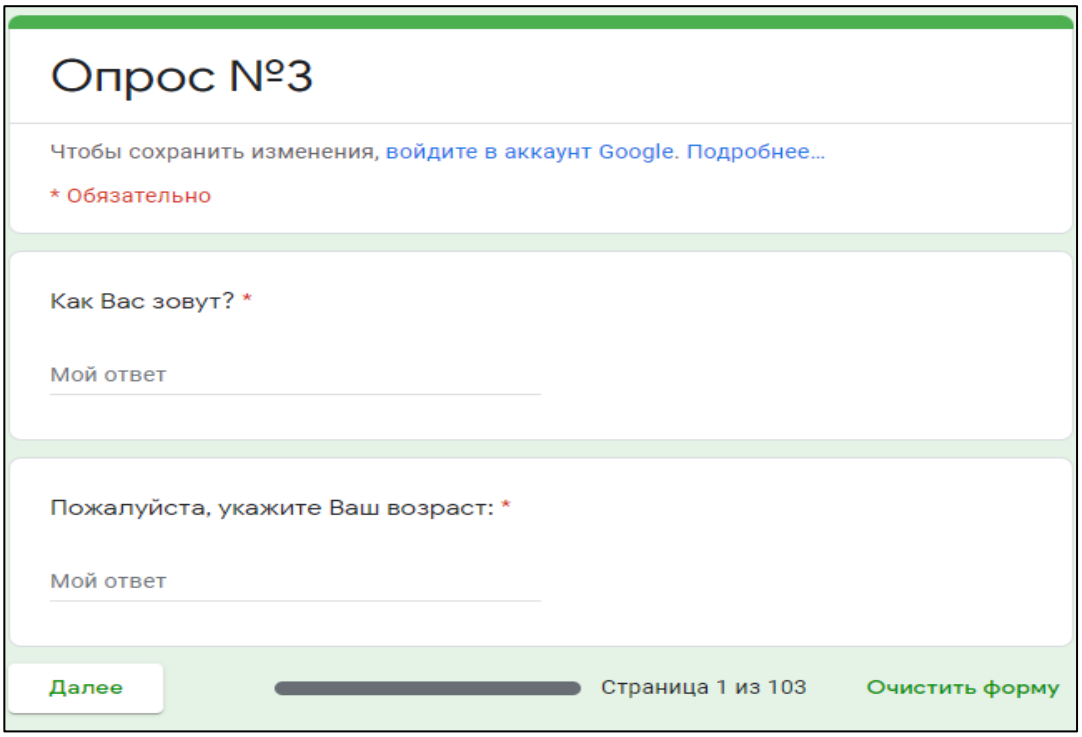

Рисунок 21. Начальная страница опроса №3

Рисунок 22. Страница опроса №2

По результатам перцептивного эксперимента был сформирован набор данных на основе следующих критериев.

Записи распределялись по папкам с эмоциями, если получалось:

1) 100–процентное согласие респондентов об эмоции высказывания;

2) если большинство респондентов (70–99%) сошлись во мнении об эмоции высказывания;

Результаты сортировки записей на основе перцептивного эксперимента представлены в разделе 2.6 Описание набора данных.

2.4.1. Результаты перцептивного эксперимента

В общей сложности, в перцептивном эксперименте участие приняли 14 респондентов. В возрасте от 23 до 74 лет. 3 мужчины и 11 женщин. 11

респондентов имеют высшее образование, 2 – среднее технические и 1 – полное среднее образование. Сводные таблицы опросов и респондентов, принявших в них участие, представлены ниже.

Таблица 7. Респонденты опросов №1, №2, №3, №4

№ Опроса / № ПП	№1		№2		№3		№4	
	Имя (И)	Воз – раст (В)	И	В	И	В	И	В
1	Валерия	23	Валерия	23	Валерия	23	Валерия	23
2	Екатерина	23	Елена	24	Елена	24	Елена	24
3	Елена	24	Алексей	40	Алексей	40	Алексей	40
4	Алексей	40	Ирина	45	Ирина	45	Ирина	45
5	Наталья	45	Наталья	45	Наталья	45	Наталья	45
6	Ирина	45	Андрей	48	Андрей	48	Андрей	48
7	Андрей	48	Ирина	54	Ирина	54	Ирина	54
8	Наталья	51	Тамара	64	Тамара	64	Тамара	64
9	Ирина	54	Нина	69	Нина	69	Нина	69
10	Тамара	64	Валентина	74	Валентина	74	Валентина	74
11	Нина	69	–		–		–	
12	Валентина	74	–		–		–	

Таблица 8. Респонденты опросов №5, №6, №7, №8

№ Опроса / № ПП	№5		№6		№7		№8	
	Имя (И)	Воз – раст (В)	И	В	И	В	И	В
1	Валерия	23	Валерия	23	Валерия	23	Валерия	23
2	Елена	24	Елена	24	Елена	24	Елена	24
3	Алексей	41	Алексей	40	Алексей	41	Ксения	34
4	Ирина	45	Ирина	45	Ирина	45	Алексей	41
5	Наталья	45	Наталья	45	Наталья	45	Ирина	45
6	Андрей	48	Андрей	48	Андрей	48	Андрей	48
7	Ирина	54	Ирина	54	Ирина	54	Ирина	54
8	Тамара	64	Тамара	64	Тамара	64	Тамара	64
9	Нина	69	Нина	69	Нина	69	Нина	69
10	Валентина	74	Валентина	74	Валентина	74	Валентина	74

Таблица 9. Респонденты опросов №9, №10, №11, №12

№ Опроса / № ПП	№9		№10		№11		№12	
	Имя (И)	Воз – раст (В)	И	В	И	В	И	В
1	Валерия	23	Валерия	23	Валерия	23	Валерия	23
2	Елена	24	Елена	24	Елена	24	Елена	24
3	Ксения	34	Ксения	34	Ксения	34	Ксения	34
4	Алексей	41	Алексей	41	Виталий	31	Виталий	31
5	Ирина	45	Ирина	45	Ирина	45	Ирина	45
6	Андрей	48	Андрей	48	Андрей	48	Андрей	48
7	Ирина	54	Ирина	54	Ирина	54	Ирина	54
8	Тамара	64	Тамара	64	Тамара	64	Тамара	64
9	Нина	69	Нина	69	Нина	69	Нина	69
10	Валентина	74	Валентина	74	Валентина	74	Валентина	74

Таблица 10. Респонденты опросов №13, №14

№ Опроса / № ПП	№13		№14	
	Имя (И)	Возраст (В)	И	В
1	Валерия	23	Валерия	23
2	Елена	24	Елена	24
3	Ксения	34	Ксения	34
4	Виталий	31	Виталий	31
5	Ирина	45	Ирина	45
6	Андрей	48	Андрей	48
7	Ирина	54	Ирина	54
8	Тамара	64	Тамара	64
9	Нина	69	Нина	69
10	Валентина	74	Валентина	74

Каждому респонденту был предоставлен для выполнения тест Н. Холла (см. Приложение В). В рамках способности распознавать и понимать эмоции другого человека во внимание принимаются такие критерии, как: эмоциональная осведомленность, эмпатия, распознавание эмоций других людей. Результаты тестирования представлены в Таблице 11.

Таблица 11. Результаты тестирования респондентов по методике Холла

№ П П	Имя (возраст)	Уровень эмоциональной осведомленности	Уровень эмпатии	Распознавание эмоций других людей
1	Валерия (23)	Средний уровень (15 баллов)	Средний уровень (12 балла)	Средний уровень (14 баллов)
2	Екатерина (23)	Средний уровень (15 балл)	Средний уровень (14 баллов)	Низкий уровень (10 баллов)
3	Елена (24)	Средний уровень (15 баллов)	Средний уровень (15 баллов)	Средний уровень (15 баллов)
4	Виталий (31)	Средний уровень (13 баллов)	Средний уровень (12 баллов)	Средний уровень (13 баллов)
5	Ксения (34)	Средний уровень (15 баллов)	Высокий уровень (17 баллов)	Высокий уровень (17 баллов)
6	Алексей (40)	Низкий уровень (11 балла)	Низкий уровень (10 балла)	Низкий уровень (10 балла)
7	Наталья (45)	Высокий уровень (16 баллов)	Средний уровень (14 баллов)	Средний уровень (14 баллов)
8	Ирина (45)	Средний уровень (13 баллов)	Средний уровень (13 баллов)	Средний уровень (12 баллов)
9	Андрей (48)	Высокий уровень (17 баллов)	Средний уровень (14 баллов)	Средний уровень (15 баллов)
10	Наталья (51)	Низкий уровень (9 балла)	Низкий уровень (10 балла)	Низкий уровень (9 балла)
11	Ирина (54)	Высокий уровень (16 баллов)	Средний уровень (13 баллов)	Средний уровень (13 баллов)
12	Тамара (64)	Высокий уровень (17 баллов)	Средний уровень (14 баллов)	Средний уровень (14 баллов)
13	Нина (69)	Средний уровень (15 баллов)	Низкий уровень (11 баллов)	Средний уровень (15 баллов)
14	Валентина (74)	Средний уровень (14 баллов)	Средний уровень (15 баллов)	Низкий уровень (11 баллов)

Таким образом, большинство респондентов имеют средние и высокие показатели тестирования по важным для исследования критериям. Тем

самым, можно предположить, что удастся создать корпус, естественно реализованных эмоций, апробированный группой респондентов. Результаты сортировки записей на основе перцептивного эксперимента представлены в разделе 2.6. Описание набора данных.

2.5. Предобработка

2.5.1. Алгоритм преобразования аудио-файла в спектрограмму

Весь аудио-материал был преобразован в изображения, а именно в спектрограммы, на основе которых проводилось обучение и тестирование реализованной сверточной нейронной сети.

Код для преобразования аудио-фрагмента в спектрограмму представлен на рис. 23. Пример спектрограммы, который получается после обработки, представлен на рис. 24. Формат изображения png.

```
from pathlib import Path
import matplotlib.pyplot as plot
from scipy.io import wavfile


path = Path('E:/Context-dependent emotional speech dataset').glob('**/*.wav')
wavs = [str(wavf) for wavf in path if wavf.is_file()]
wavs.sort()
number_of_files = len(wavs)
spk_ID = [wavs[i].split('/')[-1].lower() for i in range(number_of_files)]
for i in range(number_of_files):
    samplingfrequency, signaldata = wavfile.read(wavs[i])
    pxx, freq, bins, im = plot.specgram(x = signaldata, Fs = samplingfrequency,
noverlap = 384, NFFT = 512, cmap = 'viridis')
    plot.axis('off')
    plot.savefig("{}.png".format(spk_ID[i]), bbox_inches = 'tight', dpi = 800)
```

Рисунок 23. Преобразование аудио-фрагментов в спектрограммы

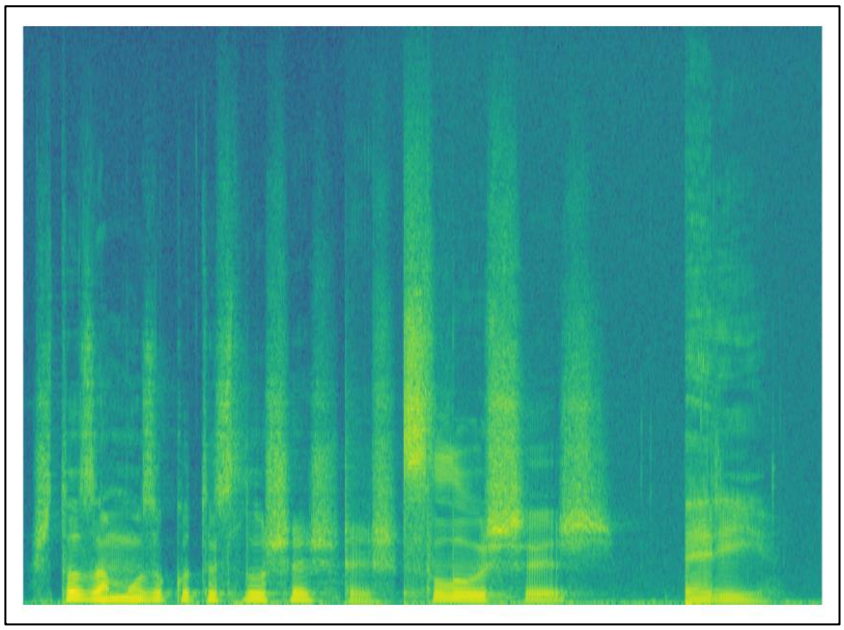

Рисунок 24. Пример спектрограммы

2.5.2. Алгоритм преобразования аудио-файла в мел-спектрограмму

Все аудио-фрагменты были преобразованы в мел-спектрограммы, на основе которых проводилось обучение и тестирование реализованной сверточной нейронной сети.

Код для преобразования аудио-фрагмента в мел-спектрограмму представлен на рис. 25. Пример мел-спектрограммы, который получается после обработки аудио-фрагмента представлен на рисунке 26. Формат изображения png.

```
import librosa
import librosa.display
import numpy as np
import matplotlib.pyplot as plt


n_fft = 2048
hop_length = 512
n_mels = 128
filename = "…" # название файла
y, sr = librosa.load(filename)
signal, _ = librosa.effects.trim(y)
D = np.abs(librosa.stft(signal[:n_fft], n_fft = n_fft, hop_length = n_fft + 1))
D = np.abs(librosa.stft(signal, n_fft = n_fft, hop_length = hop_length))
DB = librosa.amplitude_to_db(D, ref = np.max)
mel = librosa.filters.mel(sr = sr, n_fft = n_fft, n_mels = n_mels)
S = librosa.feature.melspectrogram(signal, sr = sr, n_fft = n_fft,
hop_length = hop_length, n_mels = n_mels)
S_DB = librosa.power_to_db(S, ref = np.max)
img = librosa.display.specshow(S_DB, sr = sr, hop_length = hop_length,
x_axis = 'off', y_axis = 'off')
img.figure.savefig("{}{}.png".format(filename, "_melspec"), dpi = 800)
```

Рисунок 25. Преобразование аудио-фрагментов в мел-спектрограммы

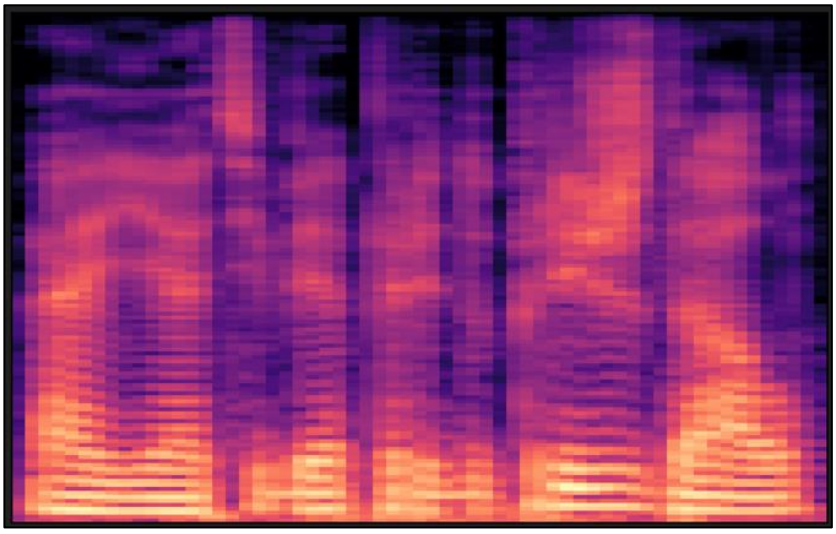

Рисунок 26. Пример мел-спектрограммы

2.5.3. Предобработка аудио-файла. Графики Основного тона

Для обработки файлов использовалась программа Wave Assistant[1]. Данная программа является специализированным звуковым редактором, в которой реализованы различные режимы обработки сигнала; возможность разметки сигнала (глубиной до 12 уровней), включая фонетическую разметку и текстовку.

Стартовая страница программы Wave Assistant представлена на рис. 27 – 29.

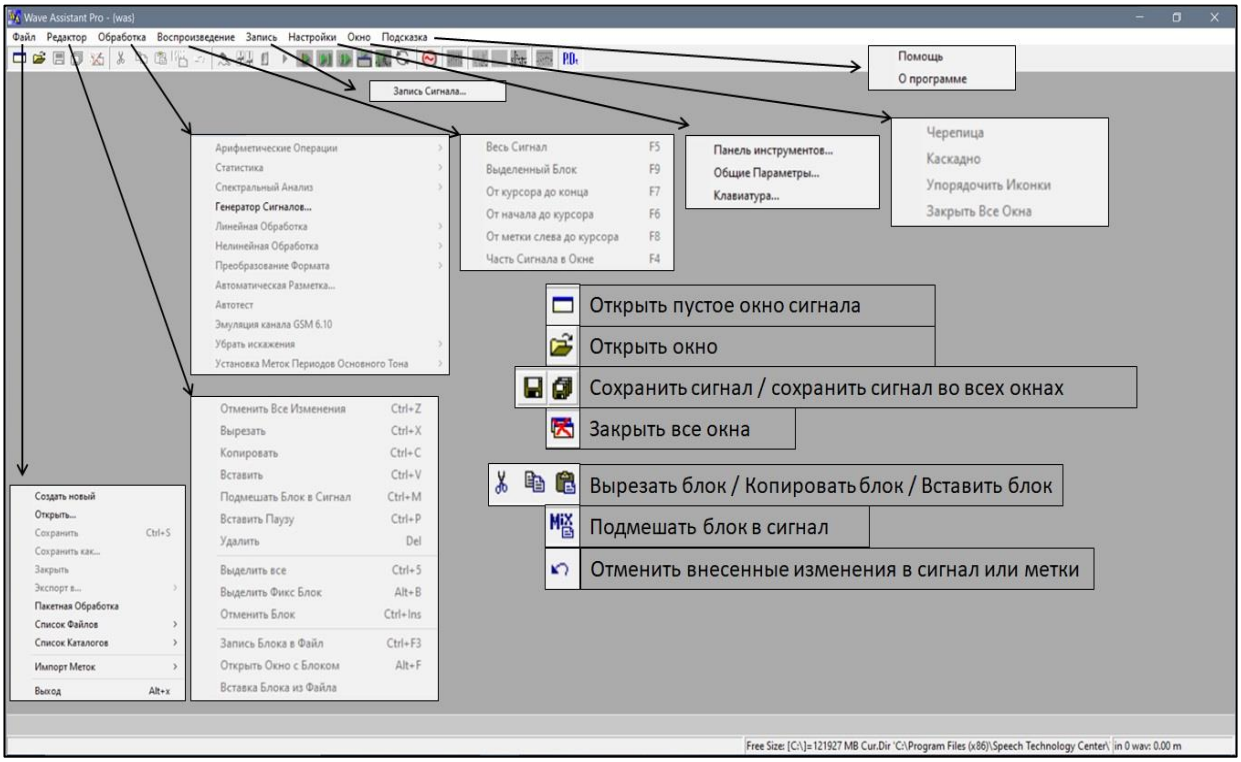

Рисунок 27. Стартовая страница программы Wave Assistant

[1] URL: https://vk.com/wave_assistant

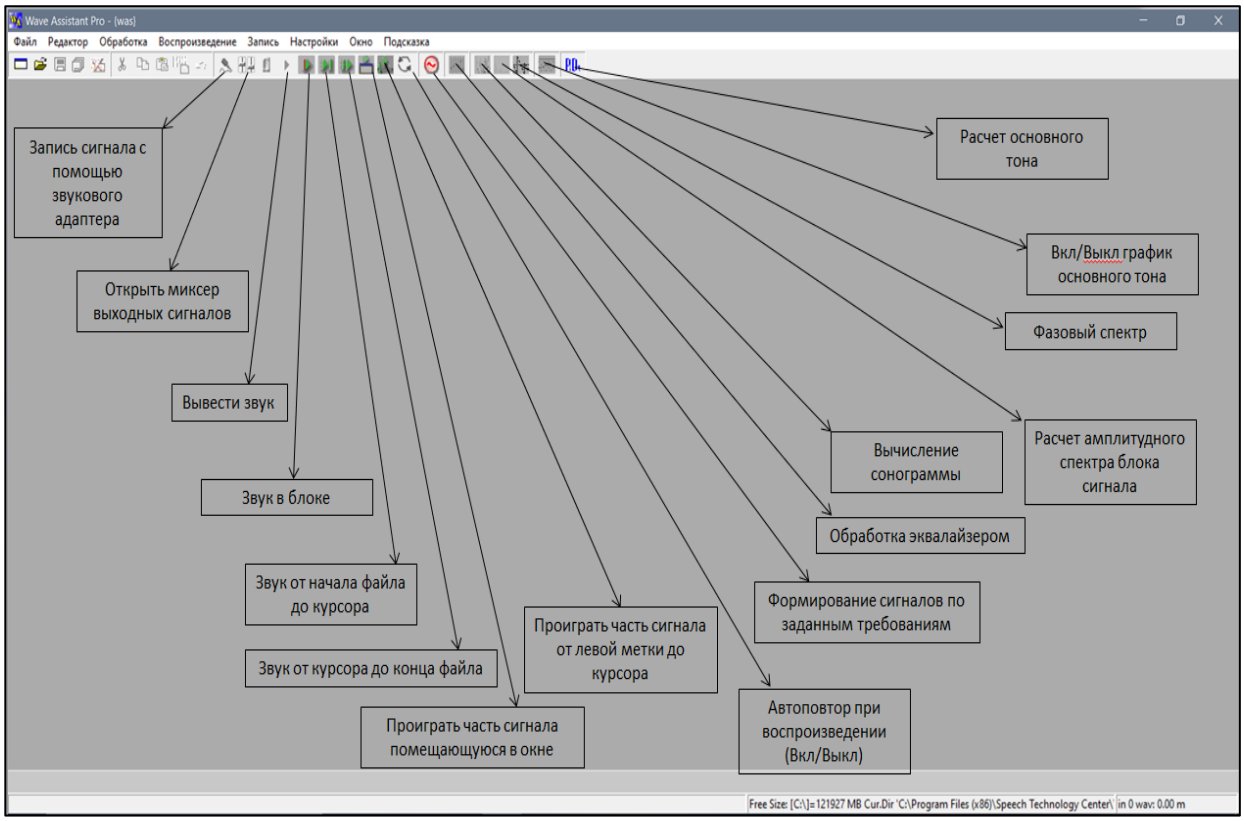

Рисунок 28. Стартовая страница программы Wave Assistant

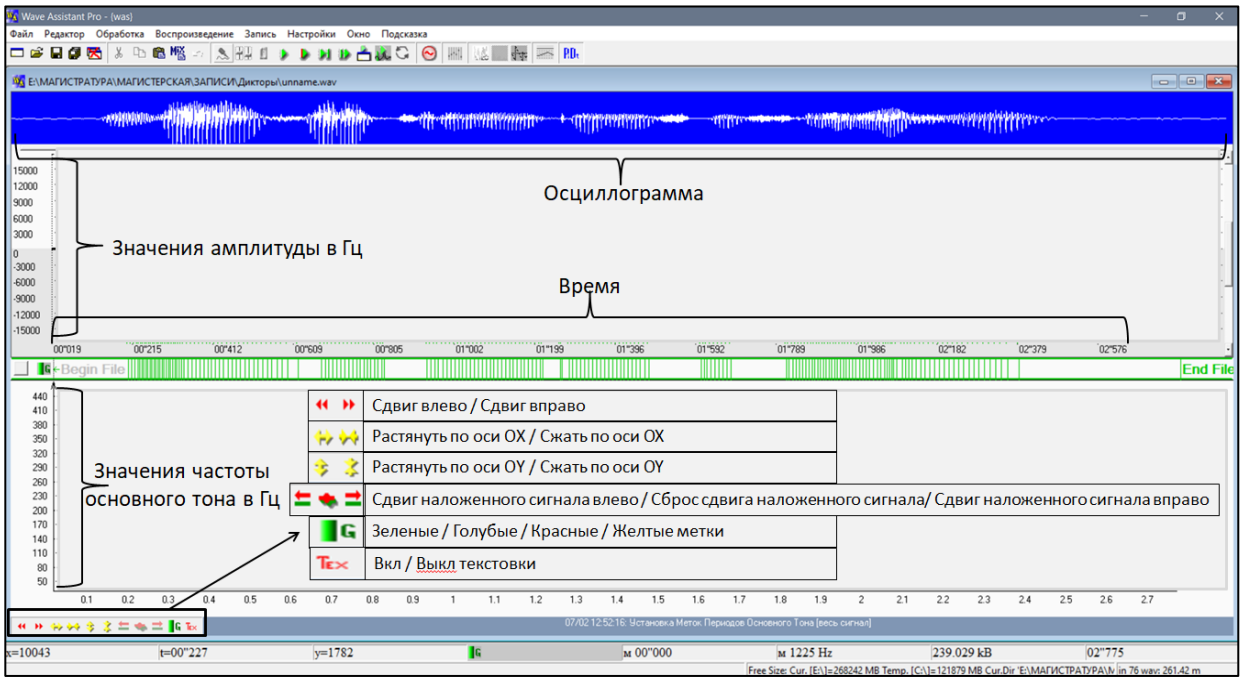

Рисунок 29. Страница работы с аудио-файлом программы Wave Assistant

Этапы обработки аудио-файла описаны ниже.

1. Открытие записи в программе Wave Assistant (рис. 30).

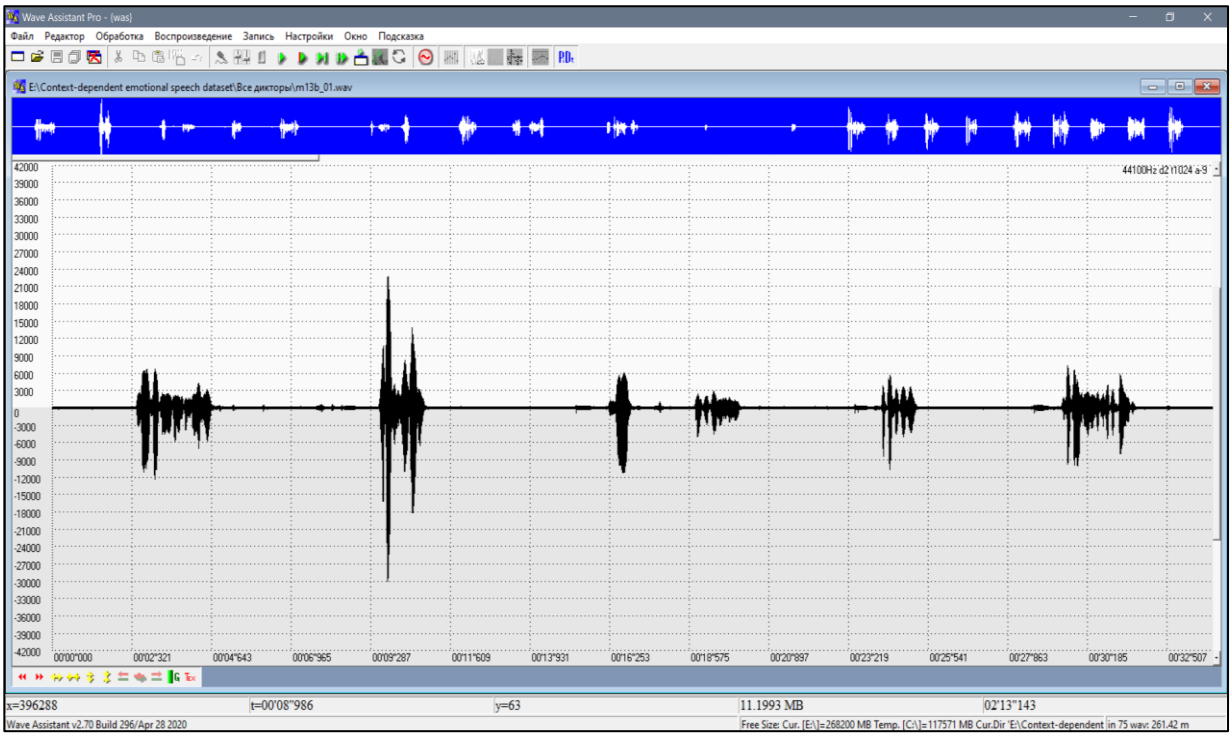

Рисунок 30. Изображение открытого аудио-файла

2. Поскольку, одни фразы дикторы произносили тихо, а другие громко, для их выравнивания по громкости перед делением общего файла на фрагменты производилась линейная обработка файла, амплитуда приводилась к максимальному значению (рис. 31).

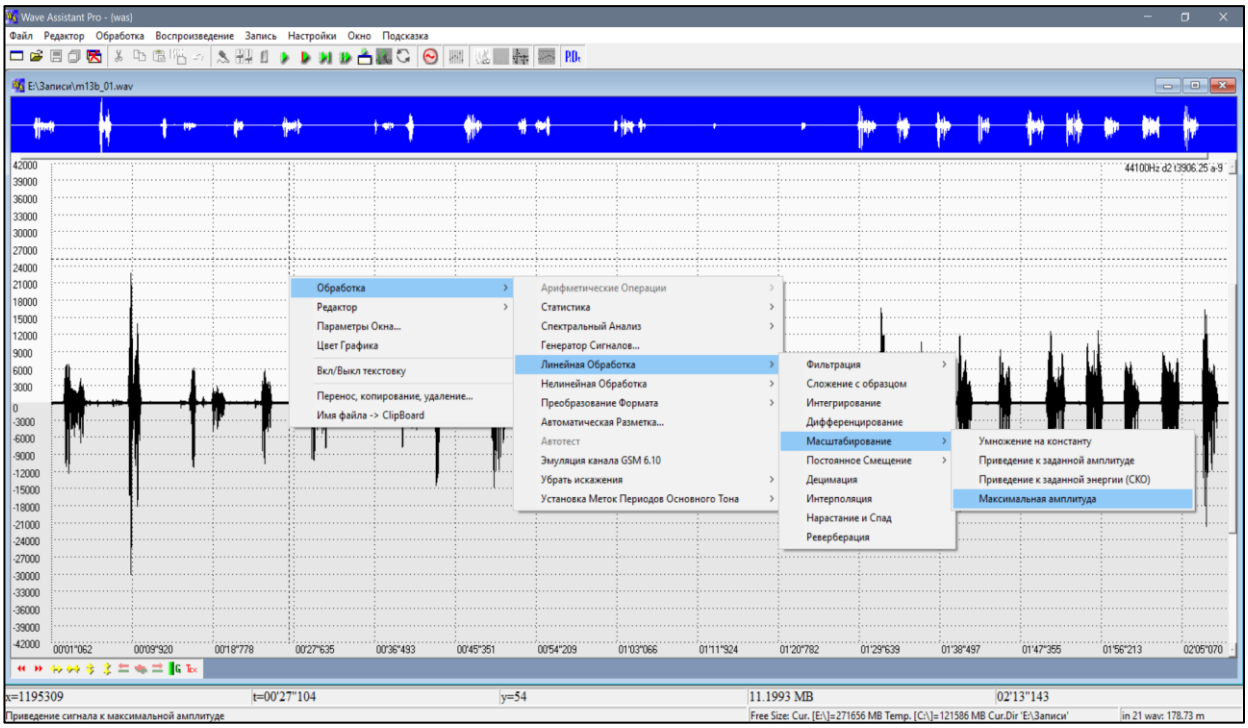

Рисунок 31. Линейная обработка. Приведение амплитуды к максимальному значению

3. Левой кнопкой мыши отмечается начало, правой кнопкой мыши – конец сигнала, который необходимо скопировать. Затем производится копирование фрагмента файла (фразы), (рис. 32).

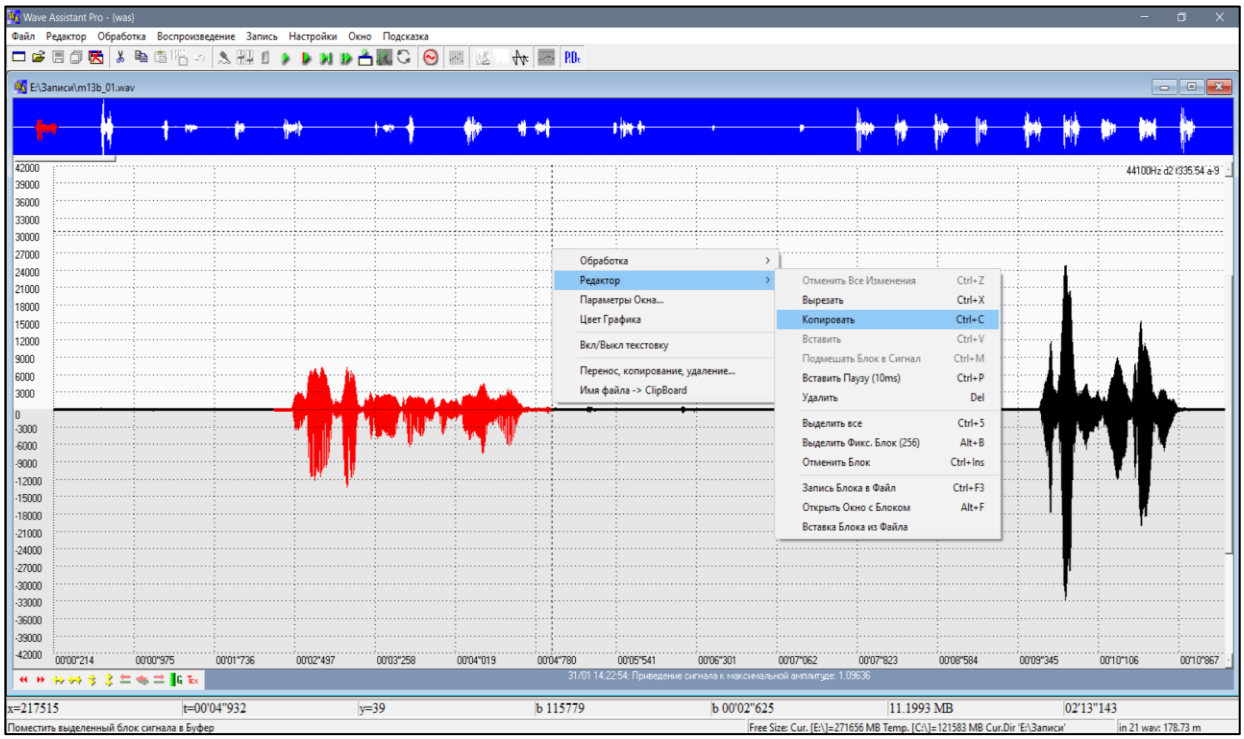

Рисунок 32. Копирование фрагмента файла (фразы) из записи

4. Для переноса аудио-фрагмента открывается новое пустое окно сигнала (рис. 33).

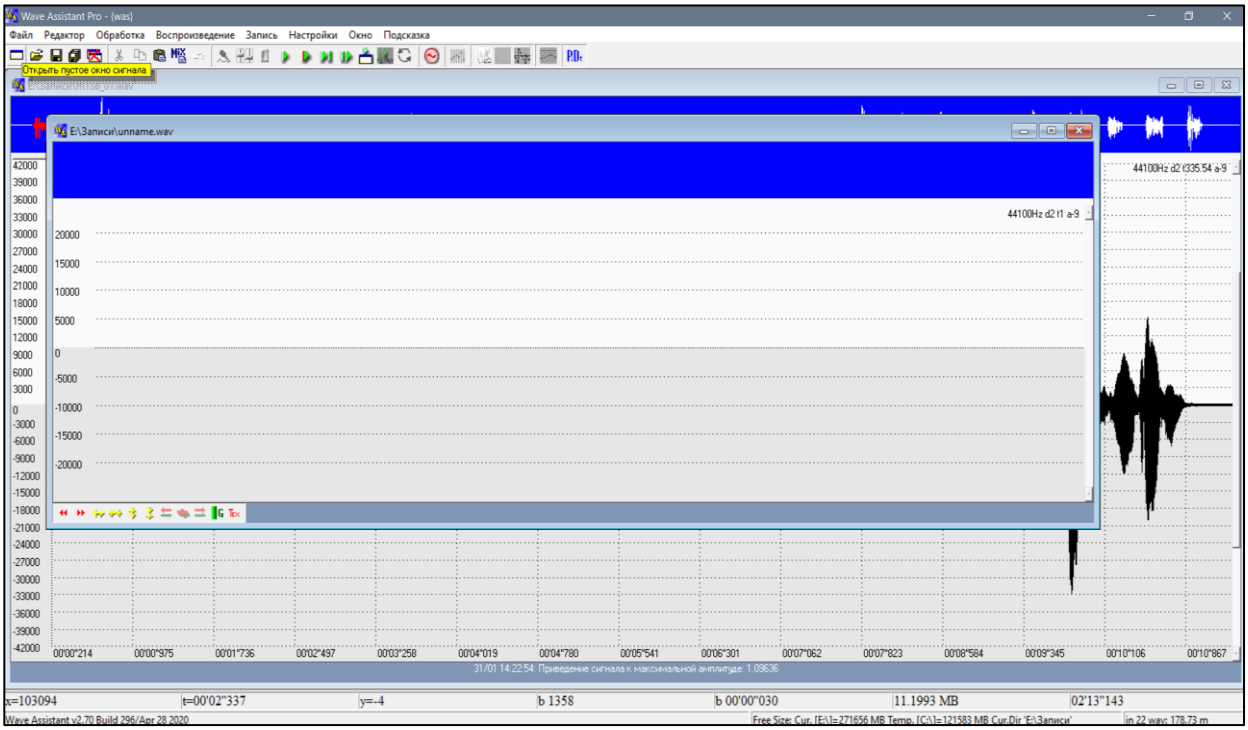

Рисунок 33. Открытие пустого окна для вставки сигнала

5. Вставка скопированного фрагмента записи в новое пустое окно (рис. 34).

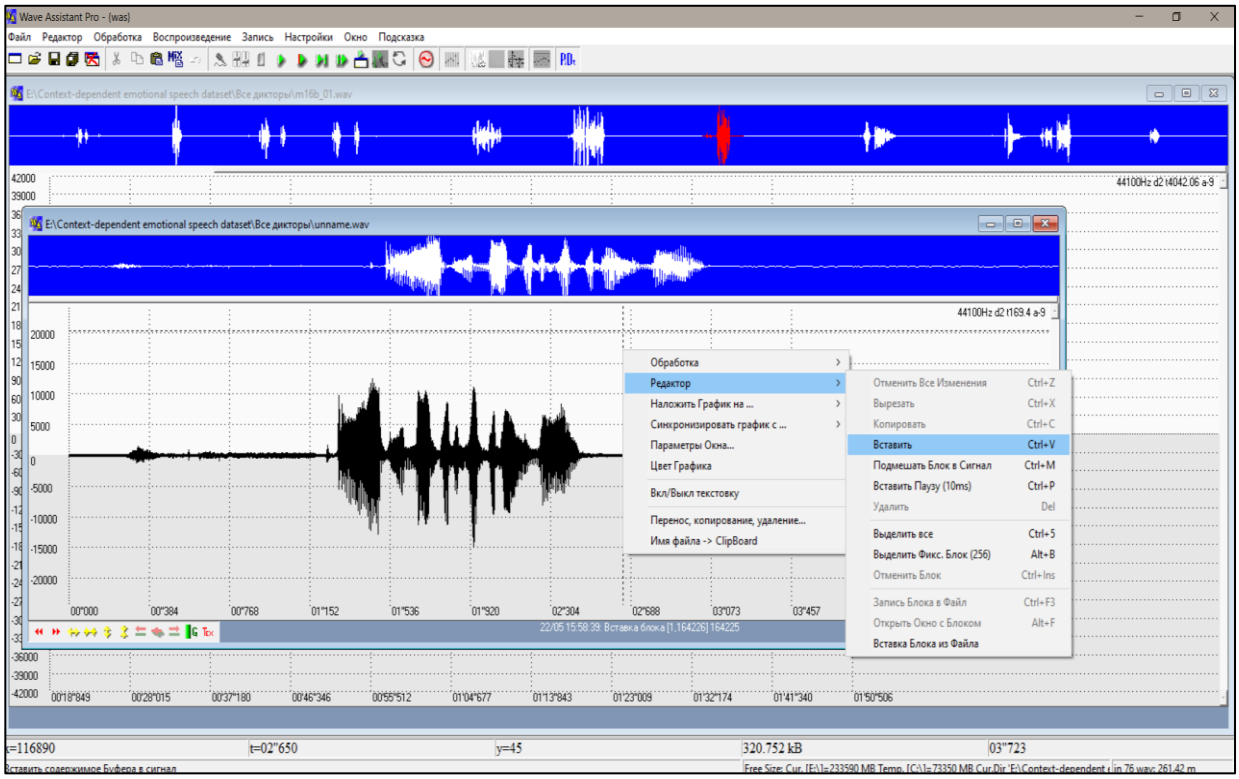

Рисунок 34. Вставка фрагмента файла (фразы) из записи в новое (отдельное) окно

6. Производится автоматическая установка меток основного тона (ОТ) (рис. 35). Метки ОТ устанавливаются на зеленом уровне. Производится визуализация графика ОТ (рис. 36 и 37) [4].

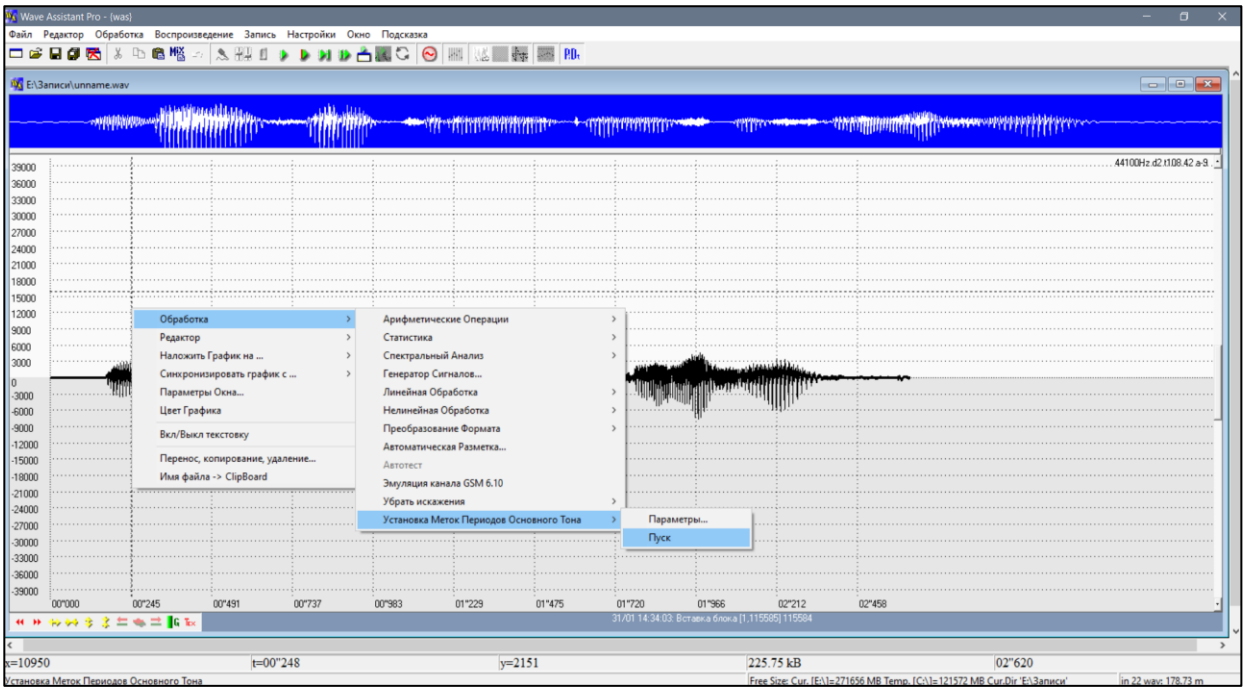

Рисунок 35. Автоматическая установка меток периодов основного тона (зеленый уровень)

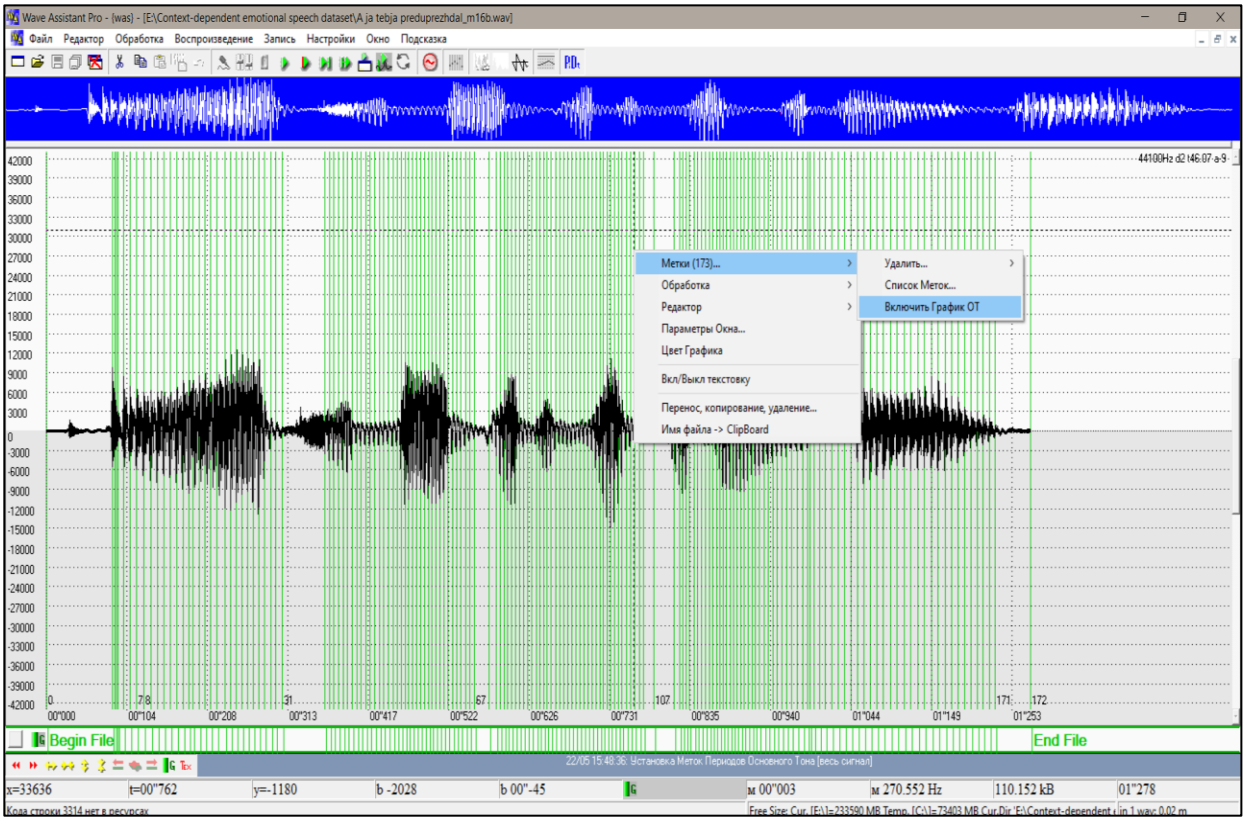

Рисунок 36. Визуализация графика основного тона

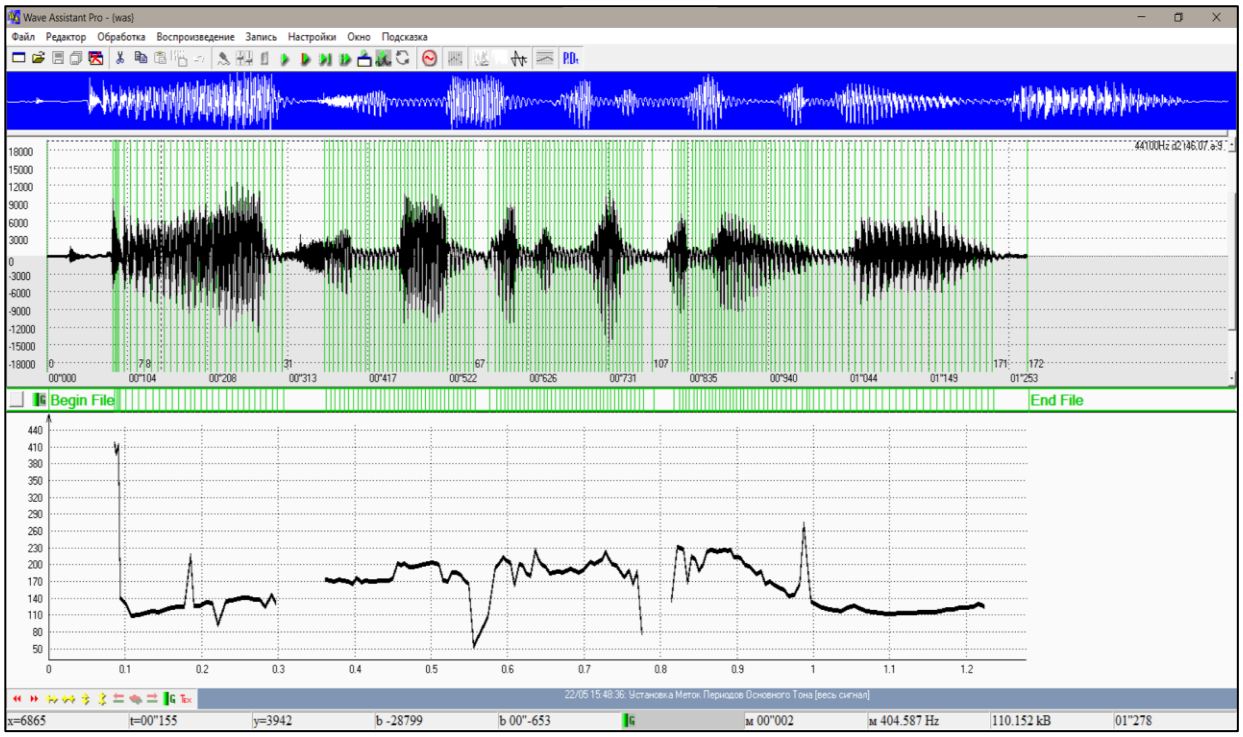

Рисунок 37. Визуализированный график основного тона

7. Добавление второго уровня разметки (синий уровень) – аудио-фрагмент подписывается соответствующей ему фразе (рис. 38).

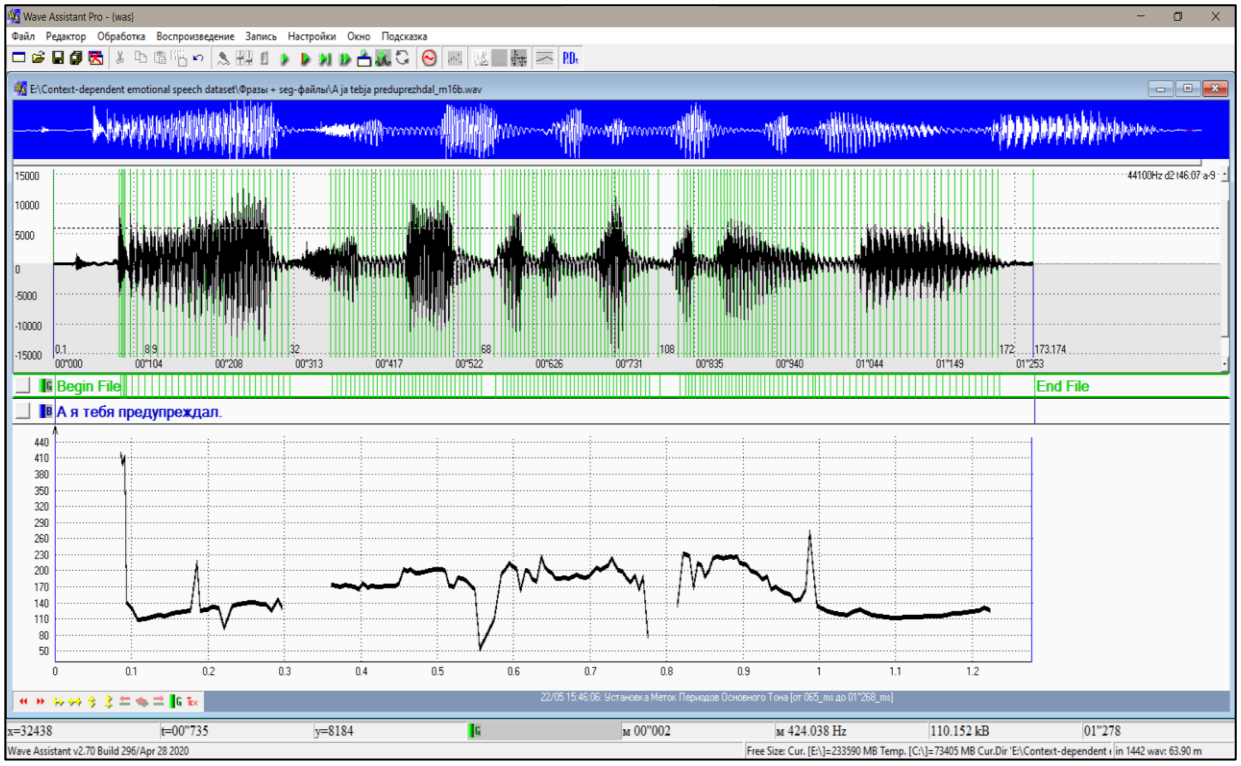

Рисунок 38. Добавление второго уровня разметки (синий уровень) – текст фразы

8. При автоматической установке меток ОТ (зеленый уровень) возникают сбои в виде смещения меток. В результате чего, график ОТ искажается и возникает необходимость редактирования меток основного тона в ручную, которая заключается в следующем:

1) удаление меток расположенных не на линии пересечения с нулем [31];

2) удаление меток искажающих график ОТ;

3) выставление дополнительных меток, если это необходимо (рис. 39).

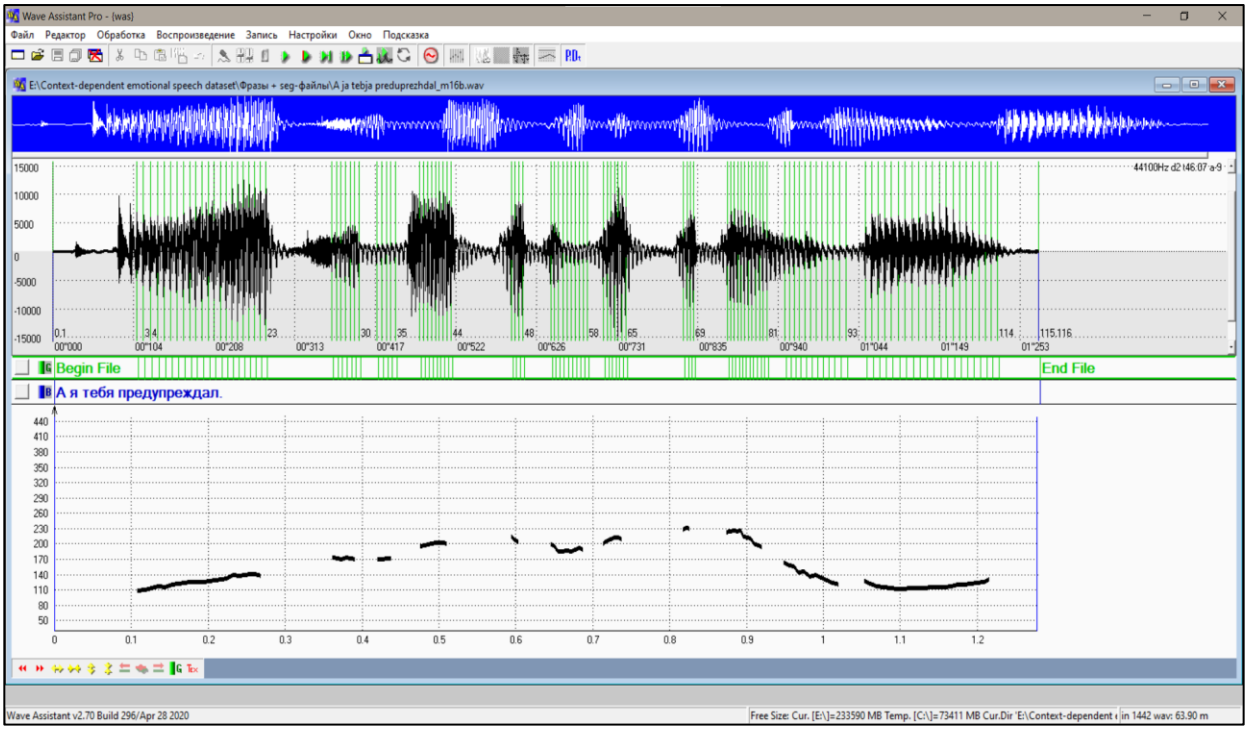

Рисунок 39. Итоговый результат обработки фразы

9. Сохранение итогового результата: фраза_диктор (рис. 40). В отдельные файлы формата seg сохраняются зеленый уровень (SEG_G1) и синий уровень (SEG_B1). Как выглядят seg-файлы – представлено на рис. 41 и 42.

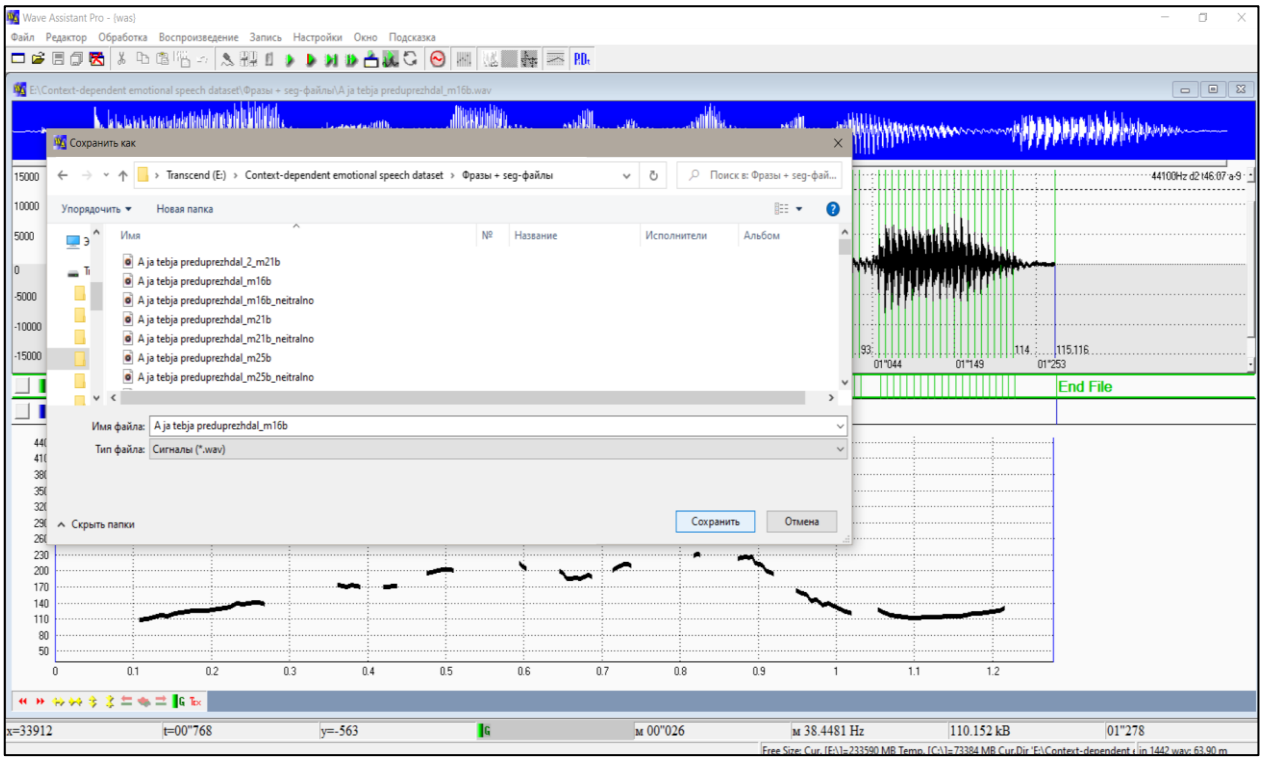

Рисунок 40. Сохранение итогового результата обработки в отдельный файл, название(латиницей)_номер диктора

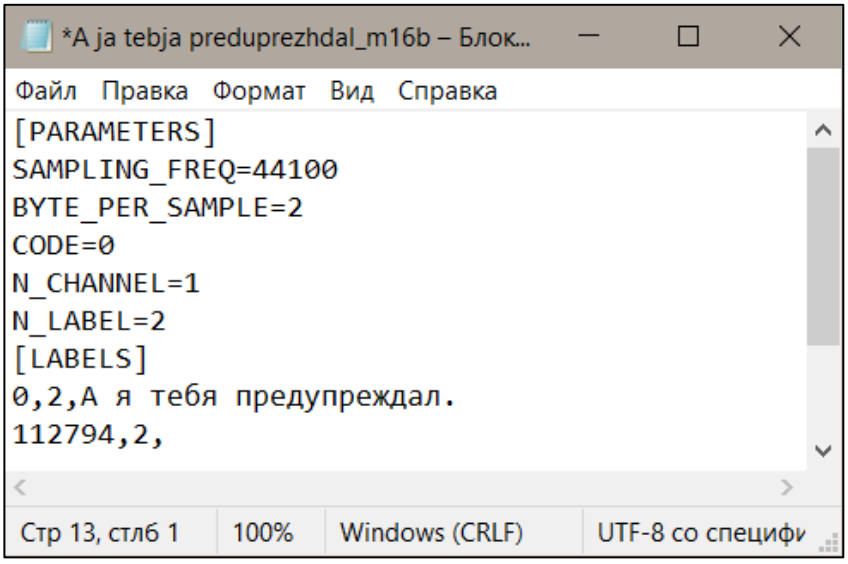

Рисунок 41. Seg-файл уровня высказывания (синий уровень)

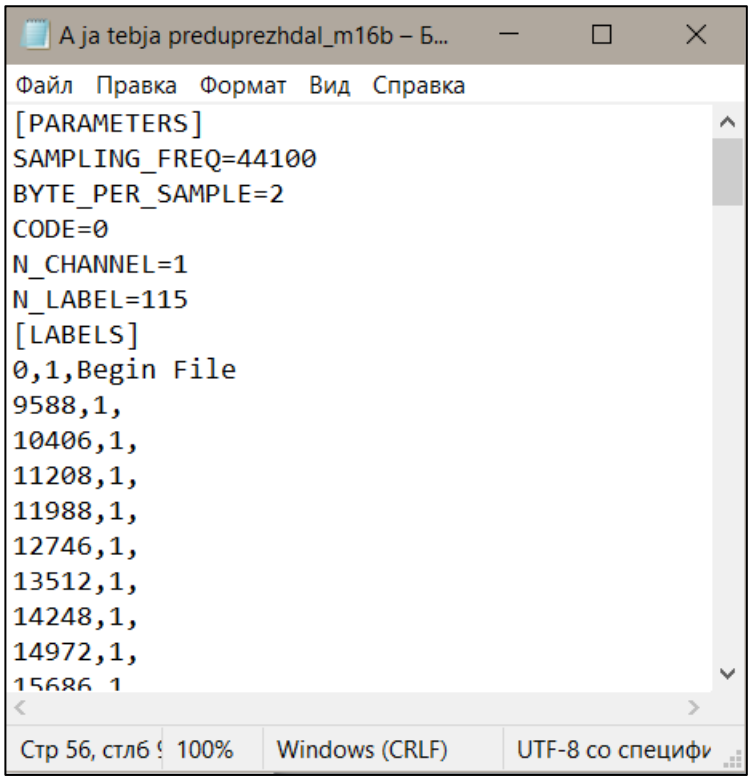

Рисунок 42. Seg-файл уровня ОТ (зеленый уровень)

2.5.3.1. Предобработка изображения

После получения результата обработки аудио-фрагмента в программе Wave Assistant, использовалась программа «Ножницы», с помощью которой «вырезался» фрагмент, на котором изображен график основного тона (рис. 43).

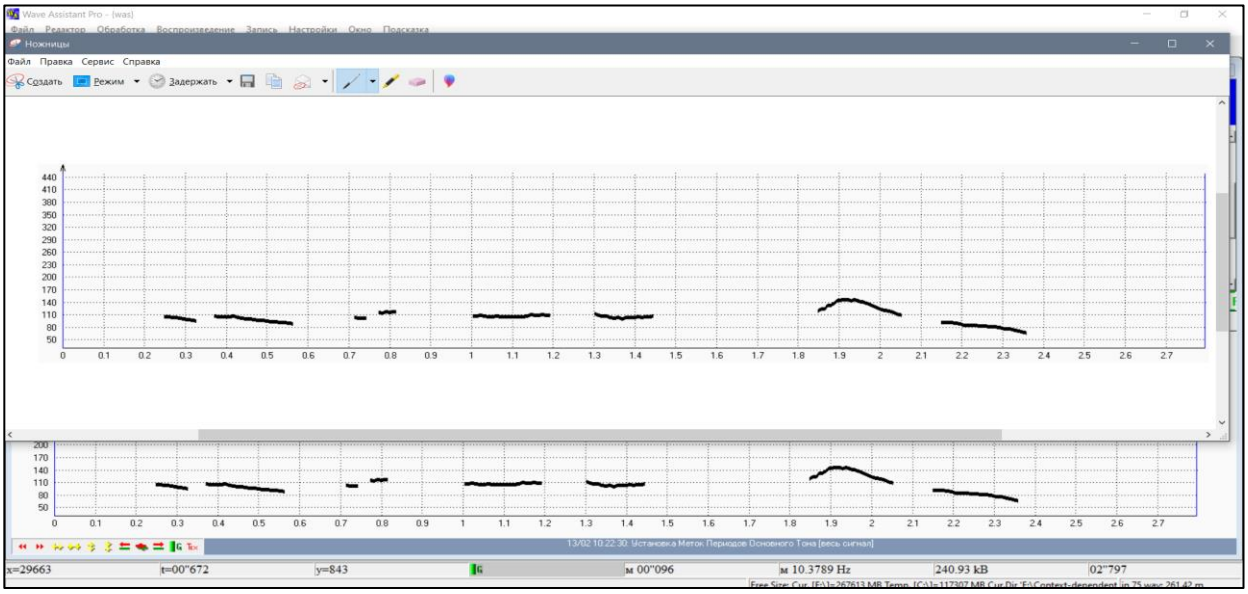

Рисунок 43. Создание изображения графика основного тона высказывания

Изображение сохранялось под таким же названием, что и аудио-фрагмент (рис. 44).

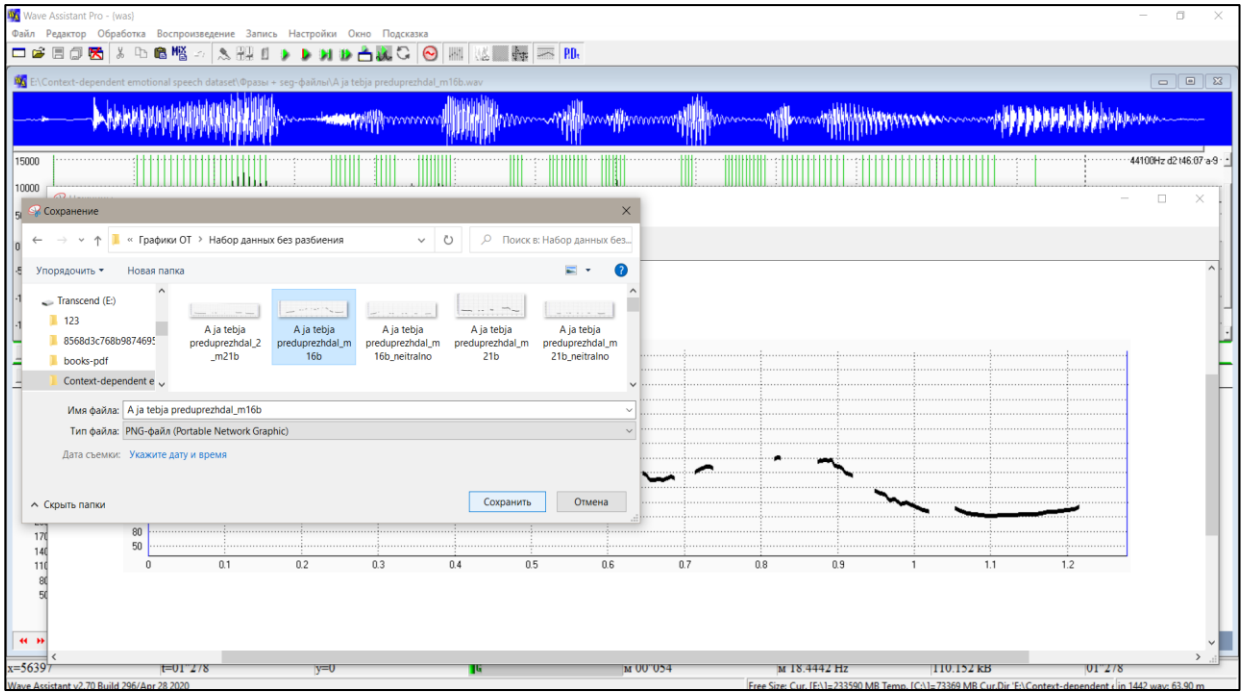

Рисунок 44. Сохранение изображения

2.5.4. Организация и хранение фалов

Аудио-файлы, изображения и seg-файлы к файлам хранятся в общей папке с названием «Context-dependent emotional speech dataset» (рис. 45).

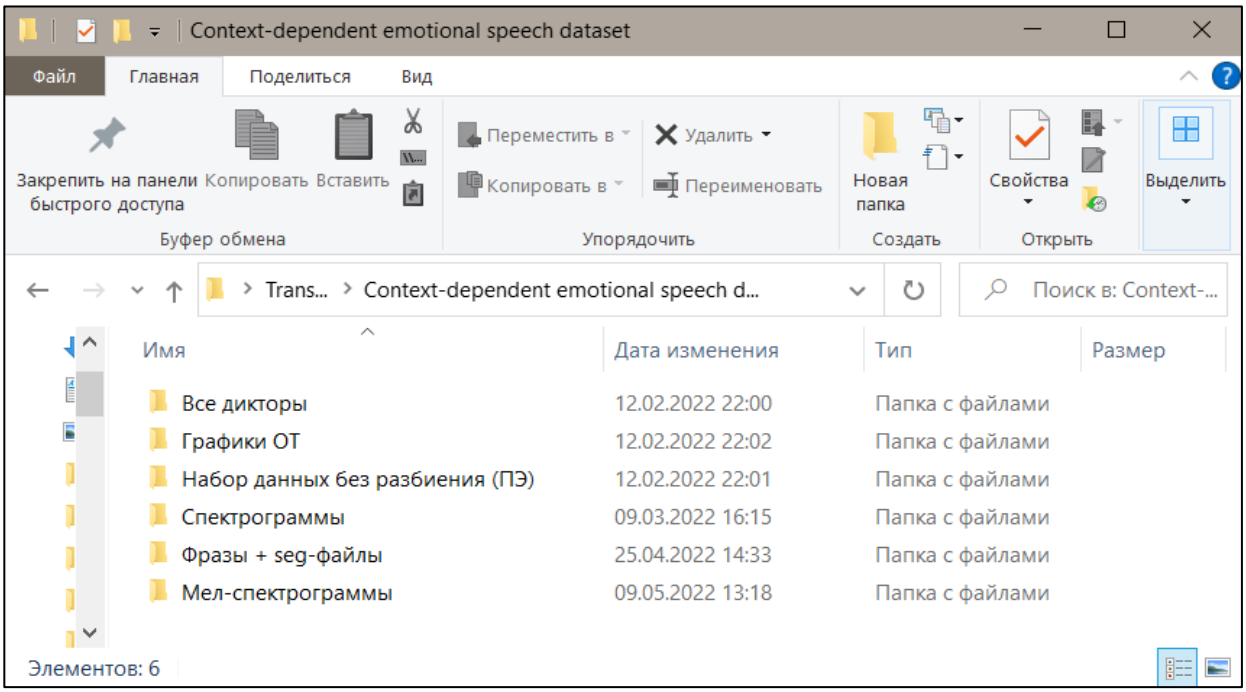

Рисунок 45. Хранение всех файлов

В папке «Все дикторы» содержатся записи всех дикторов (рис. 46).

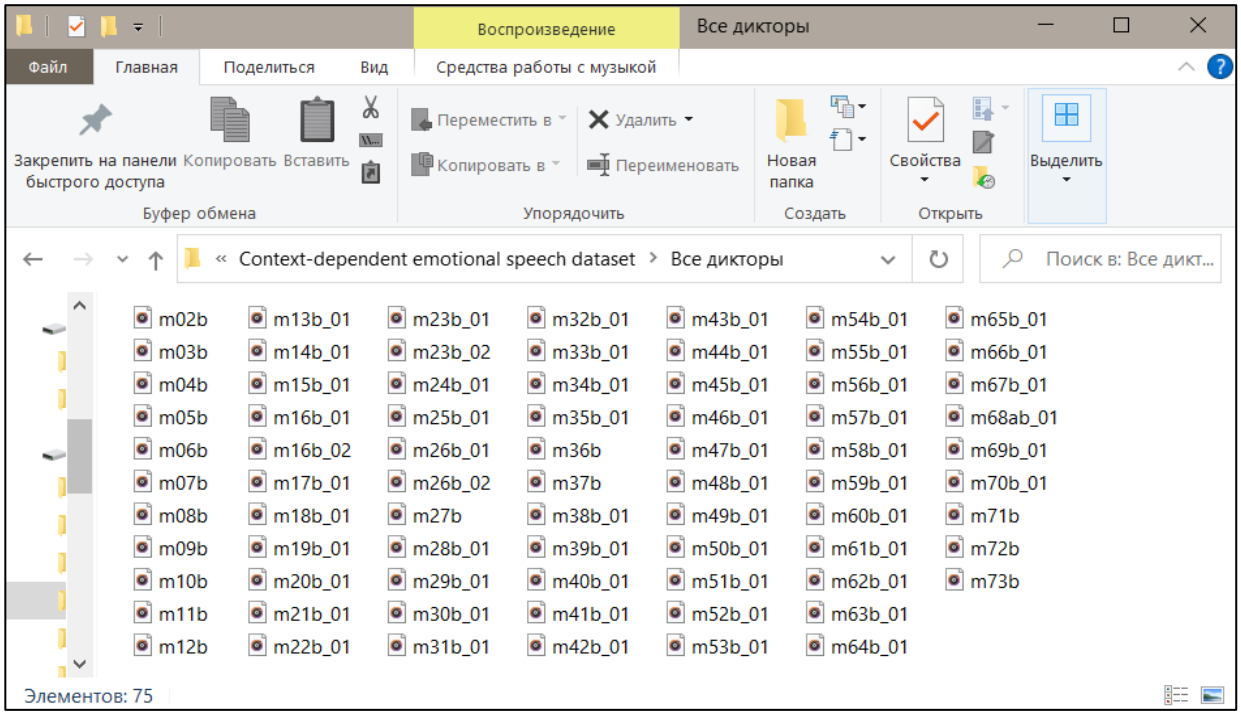

Рисунок 46. Содержимое папки «Все дикторы»

Папка «Фразы + seg-файлы» содержит в себе wav- и seg-файлы, обработанных аудио-фрагментов (рис. 47).

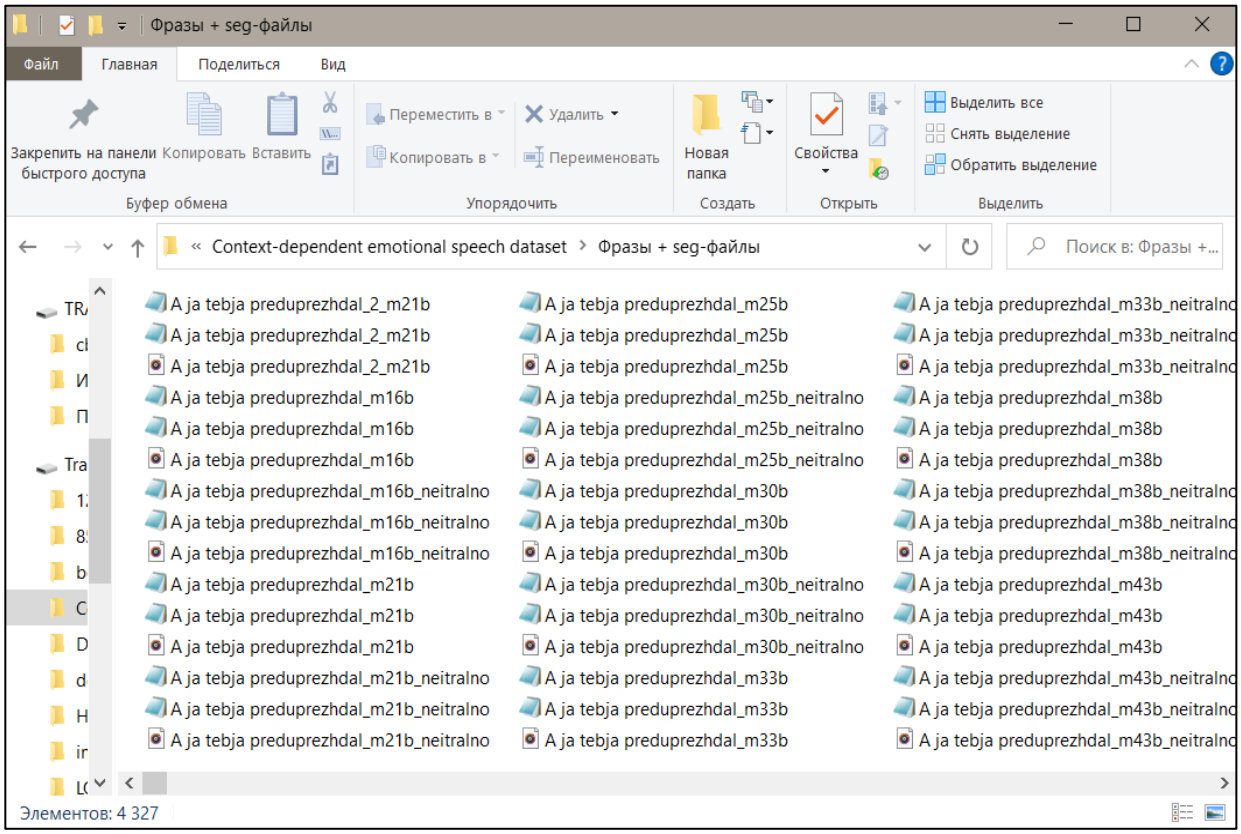

Рисунок. 47. Содержимое папки «Фразы + seg-файлы»

В папке «Графики ОТ» содержатся папки с наборами данных изображений контуров основного тона к аудио-фрагментам (рис. 48).

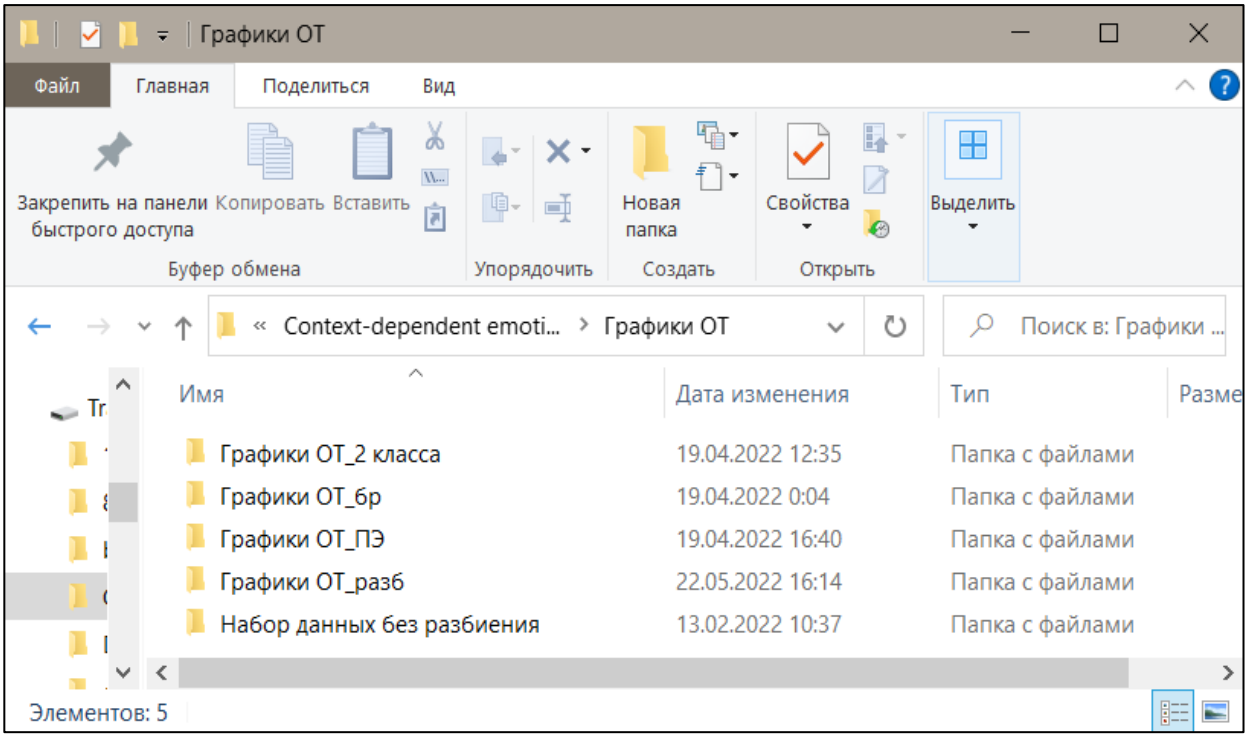

Рисунок 48. Содержимое папки «Графики ОТ»

«Графики ОТ_разб» включает обучающую и тестовую выборки начального набора данных с разбалансировкой. Папка «Графики ОТ_бр» включает обучающую и тестовую выборки начального набора данных без разбалансировки (рис. 49).

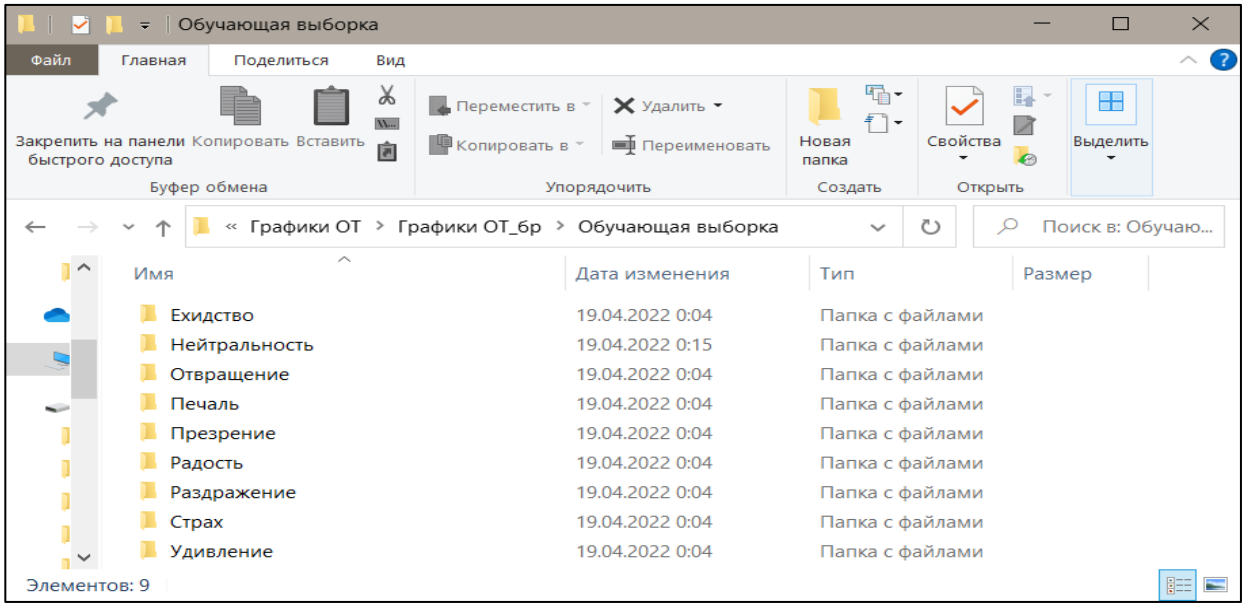

Рисунок 49. Содержимое папки «Графики ОТ_бр. Обучающая выборка»

«Графики ОТ_2 класса» содержит обучающую и тестовую выборки начального набора данных для бинарной классификации. «Графики ОТ_ПЭ» включает обучающую и тестовую выборки набора данных прошедшего перцептивный эксперимент, бинарная классификация (рис. 50).

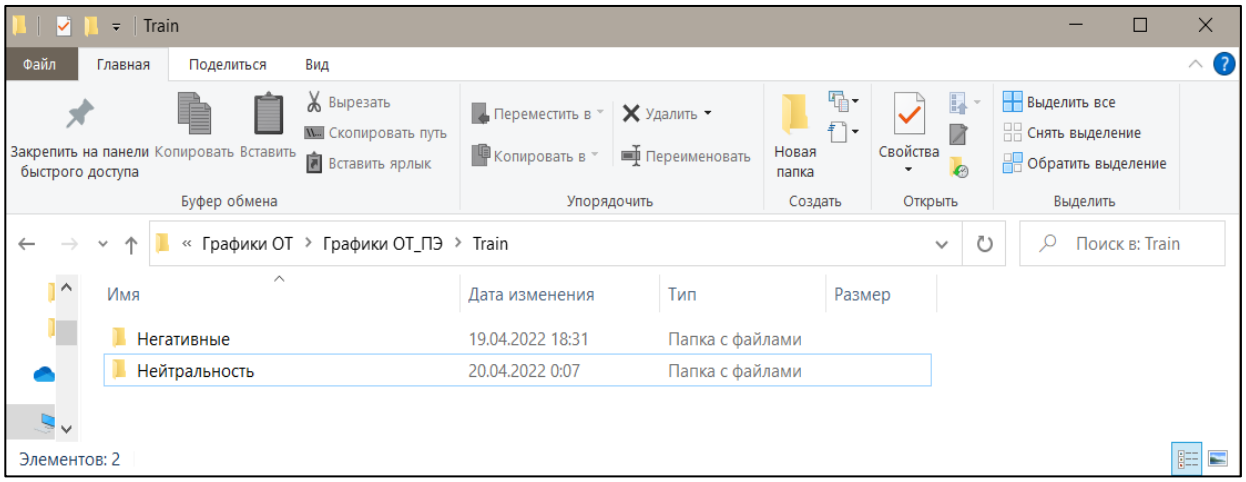

Рисунок 50. Содержимое папки «Графики ОТ_ПЭ. Тестовая выборка»

Папка «Набор данных без разбиения» содержит 1 442 изображения контура основного тона для каждого аудио-фрагмента (рис. 51). Формат изображения png.

Точно такую же структуру имеют папки «Спектрограммы» и «Мел-спектрограммы».

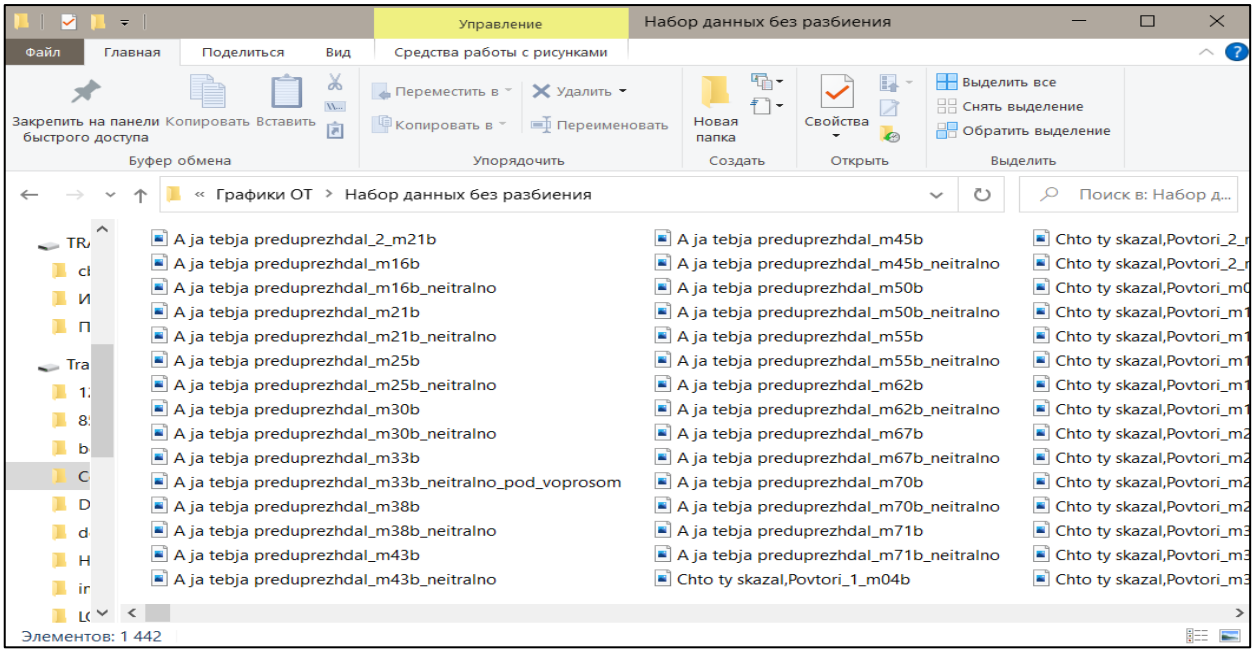

Рисунок 51. Содержимое папки «Набор данных без разбиения»

Папка «Набор данных без разбиения (ПЭ)» (рис. 52) включает в себя результаты перцептивного эксперимента. При условии если 70 и более процентов респондентов сходились во мнении относительно эмоции высказывания, то аудио-фрагмент перемещается в соответствующую папку. Все аудио-фрагменты набравшие 69 и меньше процентов перемещались в папку «Меньше 69%».

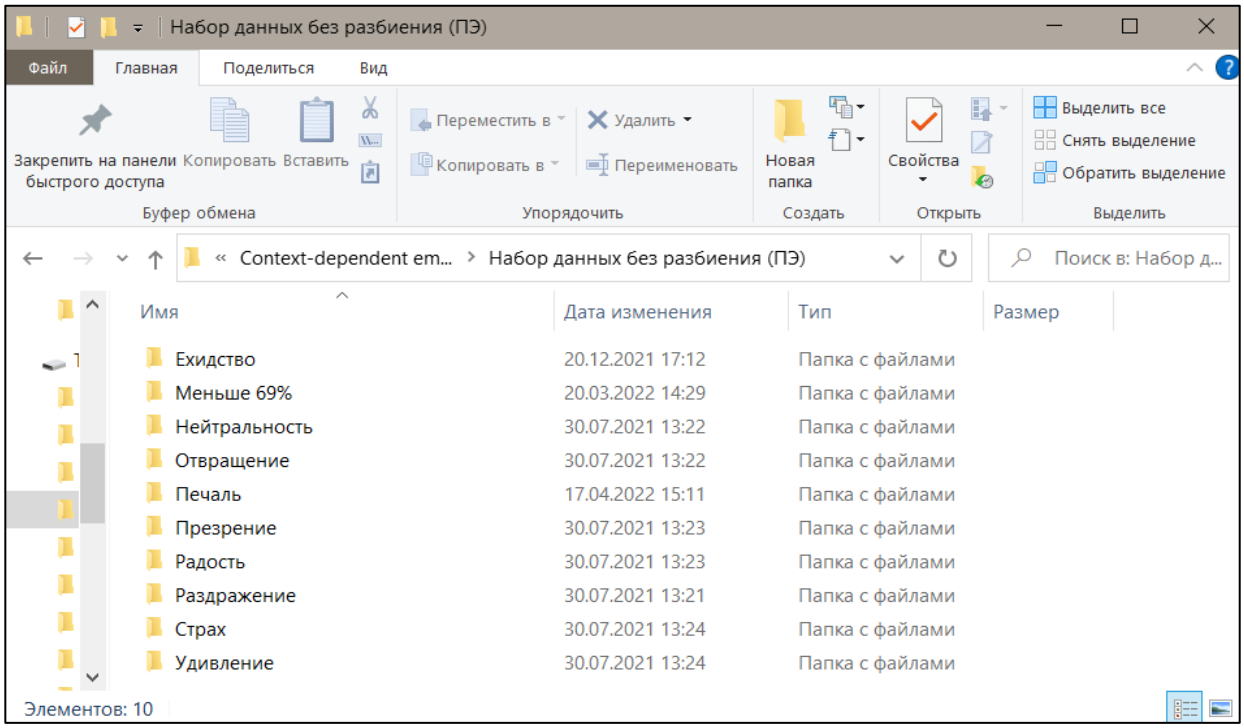

Рисунок 52. Папка «Набор данных без разбиения (ПЭ)»

2.6. Описание набора данных

В результате проведенного перцептивного эксперимента были получены ответы 14 респондентов о 1 442 файлах, которые были записаны в начале эксперимента. Результаты сортировки записей на основе перцептивного эксперимента представлены в Таблице 12.

Таблица 12. Результаты сортировки записей на основе перцептивного эксперимента

Эмоция	Количество записей
Ехидство	6
Нейтральность	397

Продолжение таблицы 12. Результаты сортировки записей на основе перцептивного эксперимента

Отвращение	0
Печаль	40
Презрение	2
Радость	63
Раздражение	110
Страх	2
Удивление	26
Итого:	646
Время:	30 минут 43 секунды

Таким образом, общее число записей, которые прошли апробацию респондентами, составило 646 аудио-фрагментов, что составило 30 минут 43 секунды. Число высказываний с отрицательными эмоциями составило 160 записей. Число высказываний с нейтральной реализацией составило 397 записей. Набор данных после перцептивного эксперимента для бинарной классификации представлен в таблице 13.

Таблица 13. Набор данных после перцептивного эксперимента для бинарной классификации

Эмоция	Количество записей
Негативная	160
Нейтральная	397
Итого:	557
Время:	26 минут 21 секунд

2.7. Выводы по главе 2

Во второй главе был обоснован перечень эмоций, включенных в исследование и список фраз, на основе которого проводилась запись материала для набора данных. Описан процесс записи набора данных и проведения перцептивного эксперимента, приведены результаты. Представлены реализованные алгоритмы для предобработки исходных файлов, которые будут подвергнуты тестированию реализованной сверточной нейронной сетью.

В результате проделанной работы были сформированы 4 набора данных: начальный набор данных с разбалансировкой; начальный набор данных без разбалансировки; начальный набор данных для бинарной классификации и набор данных после перцептивного эксперимента для бинарной классификации.

Глава 3. Реализация сверточной нейронной сети

3.1. Теория и топология сверточной нейронной сети

Появление сверточных нейронных сетей (Convolutional Neural Network – CNN) связано с изучением зрительной коры головного мозга и применяется для распознавания изображения с 1980–х годов [9, 33].

Сверточные сети, преимущественно используемые для анализа изображений, основаны на принципах, по которым работает человеческое зрение, т.е. изображение представлено в трех измерениях: ширина, высота, глубина [9, 33].

Сверточный слой (the kernel). Самым важным блоком в структуре CNN является сверточный слой, представляющий собой фильтр, который сканирует изображение. В результате этого сканирования создается признаковое изображение (feature map[2]), представленное в виде длинного массива чисел. Архитектура CNN позволяет нейронной сети сосредоточиться на низкоуровневых признаках в первом сверточном слое, затем скомпонировать их в признаки более высокого уровня в следующем сверточном слое и т.д. Причиной, почему сверточные сети настолько хорошо работают при распознавании изображений – их иерархическая структура, которая распространена в реальных изображениях [9, 33].

Пулинговый слой. Данный слой уменьшает размер изображения. С помощью него снижается вычислительная мощность, и извлекаются доминирующие признаки. Существует два типа пулинга: Max Pooling (максимальный пулинг) и Average Pooling (средний пулинг). Задача максимального пулинга возвращать максимальное значение из части изображения, которое покрывает сверточный слой, а средний пулинг возвращает среднее всех значений [9, 33].

Страйд. Количество пискселей, на которое смещается окно фильтра/кернела. Фильтр двигается с определенным значением шага/страйда

[2] Feature map или признаковое отображение –функция, которая берет векторы признаков в одном пространстве и преобразует их в векторы признаков в другом.

до тех пор, пока не пройдет всю ширину изображения. Затем, переходя к началу (слева) изображения, повторяет процесс до тех пор, пока изображение не будет «изучено» целиком [9, 33].

3.2. Средства реализации и окружение

Для реализации системы распознавания негативных эмоций с использованием нейросетевых технологий были использованы следующие средства разработки, перечисленные ниже.

1. Среда разработки: Jupyter Notebook 6.4.5.
2. Язык программирования: Python 3.9.7.

В процессе разработки были использованы следующие программные продукты и библиотеки:

1. Matplotlib. Библиотека двумерной графики для Python, с помощью которой можно создавать высококачественные рисунки различных форматов [64].
2. NumPy. Библиотека для Python, добавляющая поддержку многомерных матриц и множество функций для совершения операций над ними [67]. Использовалась для работы с массивами данных.
3. PyTorch. Библиотека для Python, является фреймворком машинного обучения. Используется для задач компьютерного зрения, обработки естественного языка [72].

3.3. Реализация и обучение нейронной сети

После тестирования различных топологий была выбрана модель сверточной нейронной сети, состоящая из четырех сверточных слоев и двух пулинговых слоев. Схема нейронной сети представлена на рис. 53.

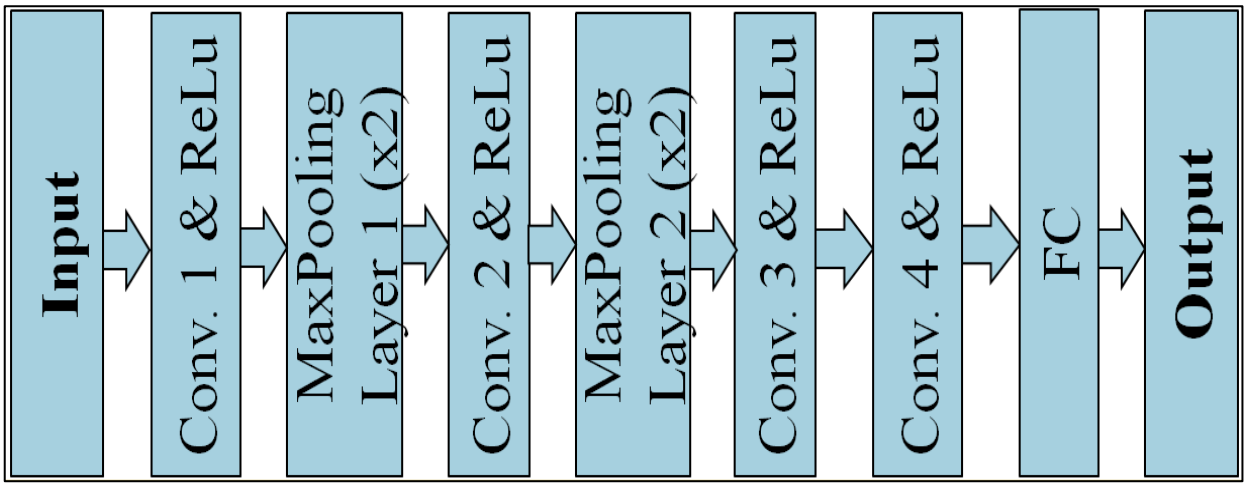

Рисунок 53. Схема реализованной сверточной нейронной сети

На вход нейронной сети подается изображение размером 256x256 пикселей. В результате применения двух пулинговых слоев исходное изображение уменьшается до размерности 64x64 пикселя. В зависимости от поставленной задачи (бинарная классификация, многоклассовая классификация), указывается количество классов: 2 или 9 (по количеству эмоций заявленных в начале работы). Реализация построения модели на языке Python 3.9.7 с использованием библиотеки PyTorch приведена на рис. 54.

```
transformer=transforms.Compose([
    transforms.Resize((256,256)),
    transforms.RandomHorizontalFlip(),
    transforms.ToTensor(),
    transforms.Normalize([0.5,0.5,0.5],
                         [0.5,0.5,0.5])
])
```

```
train_path = 'Binary classification-perc.exp/Train'
test_path = 'Binary classification-perc.exp/Test'

train_loader=DataLoader(
    torchvision.datasets.ImageFolder(train_path, transform=transformer),
    batch_size=4, shuffle=True
)
test_loader=DataLoader(
    torchvision.datasets.ImageFolder(test_path, transform=transformer),
    batch_size=4, shuffle=True
)
```

Рисунок 54. Реализация построения модели

В качестве функции активации для каждого сверточного слоя используется функция ReLu, завершает модель полносвязный слой.

Параметры реализованной сверточной нейронной сети представлены ниже.

1. optimizer: оптимизатор, алгоритм который изменяет веса и смещения во время обучения. Используем Adam, с параметрами lr = 0.001, weight_decay = 0.0001.

2. loss_function: функция потерь, так как перед нами стоит задача классификации, в качестве параметра указываем CrossEntropyLoss, которая используется для решения задач подобного типа.

3. batch_size: обучение и тестирование осуществляется пакетами по 4 изображения.

4. epochs: число проходов по всем тренировочным данным. Путем тестирования было выбрано оптимальное количество эпох – 12 эпох.

Соотношение обучающей и тестовой выборки было выбрано 80% и 20%, соответственно. Данные в обучающей и тестовой выборках не пересекаются.

Для обучения и тестирования нейронной сети использовались предобработанные ранее данные.

3.4. Реализация системы классификации эмоций

В результате обучения и тестирования реализованная сверточная нейронная сеть показала результаты, представленные в таблице 14 – 16.

На рис. 55, 57, 59 представлена графическая визуализация лучшего результата по классу (графики ОТ[3], спектрограммы, мел-спектрограммы). Данная визуализация выполнена с помощью библиотеки Visdom [79].

На рис. 56, 58, 60 представлены значения, полученные во время тестирования и обучения модели, соответствуют графикам, упомянутым выше.

[3] Графики ОТ – Графики основного тона

Таблица 14. Результаты обучения и тестирования модели. Графики ОТ

Набор данных / Результаты	Начальный набор данных (разбаланс.)	Начальный набор данных (без разбаланс.)	Начальный набор данных (бин. класс.)	Набор данных после ПЭ[4] (бин. класс.)
Training accuracy (12 эпох)	0.949565	0.948844	0.925890	0.954954
Test accuracy (12 эпох)	0.548951	0.304635	0.641732	0.725663

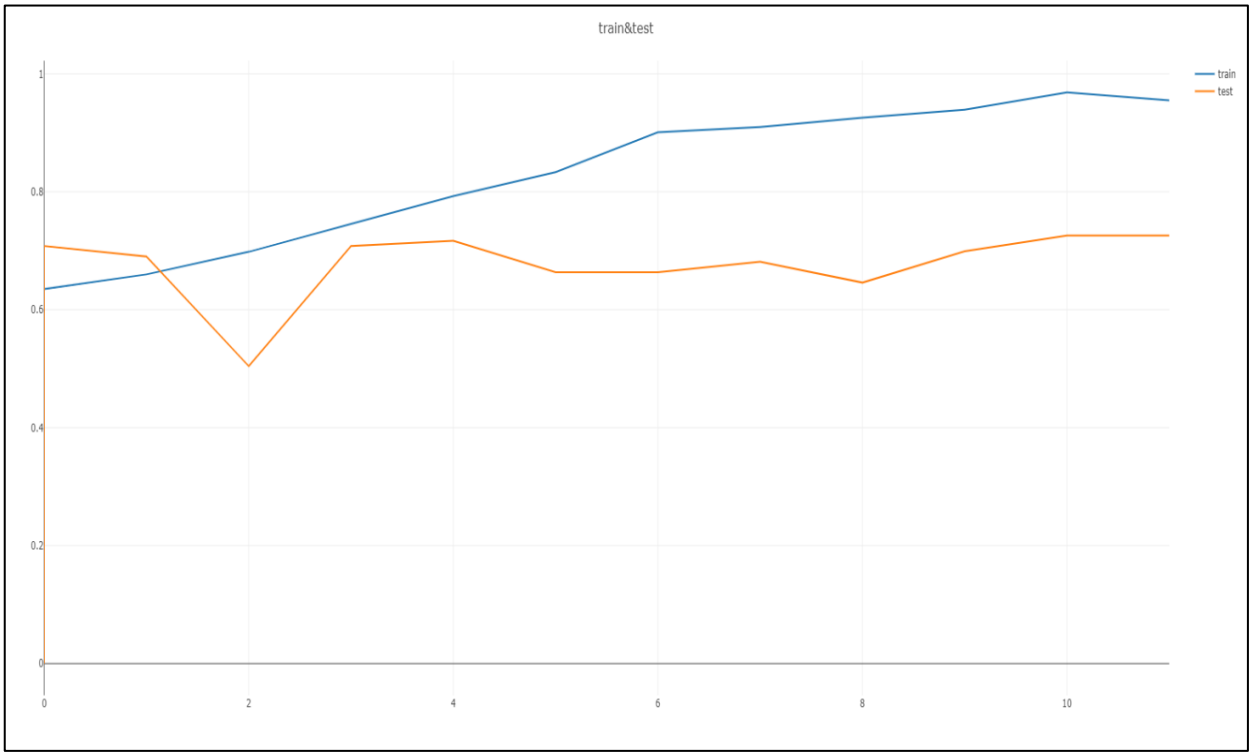

Рисунок 55. Визуализация тестирования модели. Графики основного тона. Набор данных после ПЭ (бин. класс.)

```
Epoch: 0 Train Loss: tensor(14.6484) Train Accuracy: 0.6351351351351351 Test Accuracy: 0.7079646017699115
Epoch: 1 Train Loss: tensor(6.1682) Train Accuracy: 0.6599099099099099 Test Accuracy: 0.6902654867256637
Epoch: 2 Train Loss: tensor(2.3201) Train Accuracy: 0.6981981981981982 Test Accuracy: 0.504424778761062
Epoch: 3 Train Loss: tensor(1.0035) Train Accuracy: 0.7454954954954955 Test Accuracy: 0.7079646017699115
Epoch: 4 Train Loss: tensor(0.4908) Train Accuracy: 0.7927927927927928 Test Accuracy: 0.7168141592920354
Epoch: 5 Train Loss: tensor(0.3956) Train Accuracy: 0.8333333333333334 Test Accuracy: 0.6637168141592921
Epoch: 6 Train Loss: tensor(0.2963) Train Accuracy: 0.9009009009009009 Test Accuracy: 0.6637168141592921
Epoch: 7 Train Loss: tensor(0.2391) Train Accuracy: 0.9099099099099099 Test Accuracy: 0.6814159292035398
Epoch: 8 Train Loss: tensor(0.1711) Train Accuracy: 0.9256756756756757 Test Accuracy: 0.6460176991150443
Epoch: 9 Train Loss: tensor(0.1746) Train Accuracy: 0.9391891891891891 Test Accuracy: 0.6991150442477876
Epoch: 10 Train Loss: tensor(0.1067) Train Accuracy: 0.9684684684684685 Test Accuracy: 0.7256637168141593
Epoch: 11 Train Loss: tensor(0.1130) Train Accuracy: 0.954954954954955 Test Accuracy: 0.7256637168141593
```

Рисунок 56. Результаты тестирования модели. Графики основного тона. Набор данных после ПЭ (бин. класс.)

[4] ПЭ – Перцептивный Эксперимент

Таблица 15. Результаты обучения и тестирования модели. Спектрограммы

Набор данных / Результаты	Начальный набор данных (разбаланс.)	Начальный набор данных (без разбаланс.)	Начальный набор данных (бин. класс.)	Набор данных после ПЭ (бин. класс.)
Training accuracy (12 эпох)	0.830582	0.993377	0.936538	0.977528
Test accuracy (12 эпох)	0.563573	0.757961	0.746987	0.964601

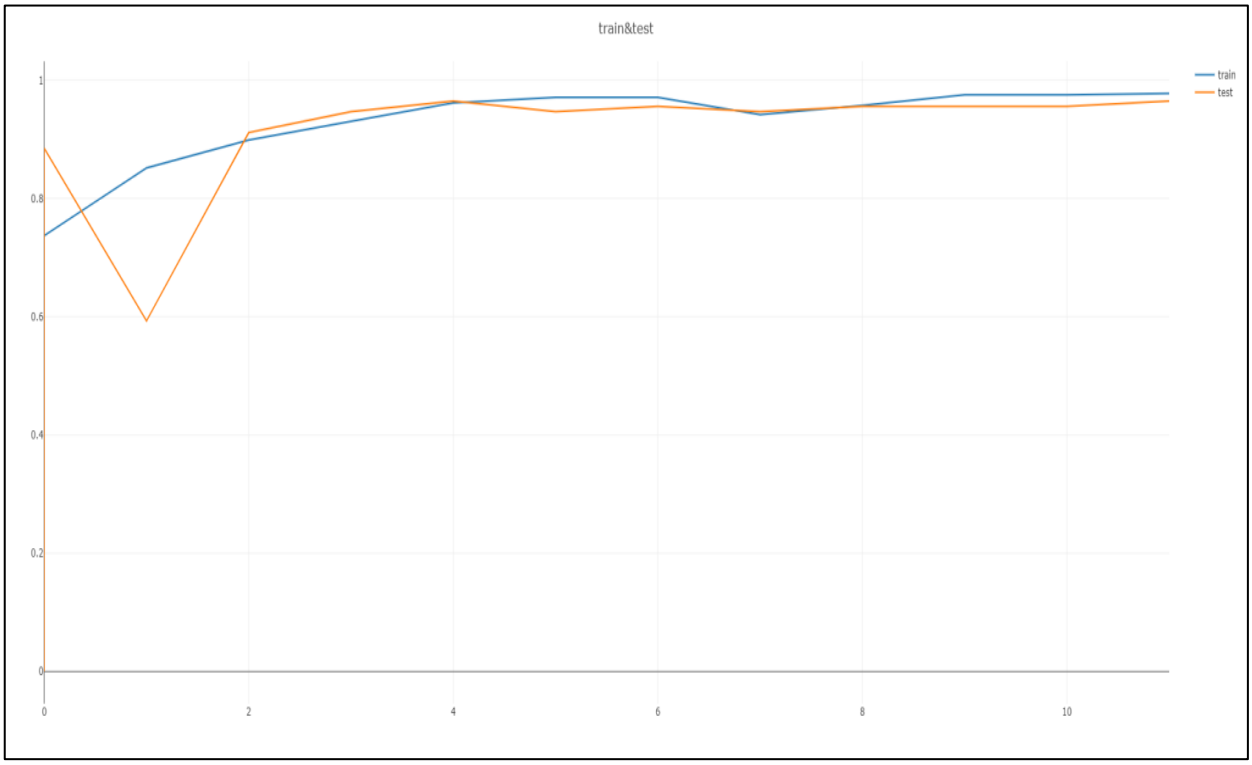

Рисунок 57. Визуализация тестирования модели. Спектрограммы. Набор данных после ПЭ (бин. класс.)

```
Epoch: 0 Train Loss: tensor(11.3674) Train Accuracy: 0.7370786516853932 Test Accuracy: 0.8849557522123894
Epoch: 1 Train Loss: tensor(3.9236) Train Accuracy: 0.851685393258427 Test Accuracy: 0.5929203539823009
Epoch: 2 Train Loss: tensor(1.7546) Train Accuracy: 0.898876404494382 Test Accuracy: 0.911504424778761
Epoch: 3 Train Loss: tensor(1.0115) Train Accuracy: 0.9303370786516854 Test Accuracy: 0.9469026548672567
Epoch: 4 Train Loss: tensor(0.4149) Train Accuracy: 0.9617977528089887 Test Accuracy: 0.9646017699115044
Epoch: 5 Train Loss: tensor(0.1708) Train Accuracy: 0.9707865168539326 Test Accuracy: 0.9469026548672567
Epoch: 6 Train Loss: tensor(0.1489) Train Accuracy: 0.9707865168539326 Test Accuracy: 0.9557522123893806
Epoch: 7 Train Loss: tensor(0.6478) Train Accuracy: 0.9415730337078652 Test Accuracy: 0.9469026548672567
Epoch: 8 Train Loss: tensor(0.2691) Train Accuracy: 0.9573033707865168 Test Accuracy: 0.9557522123893806
Epoch: 9 Train Loss: tensor(0.1532) Train Accuracy: 0.9752808988764045 Test Accuracy: 0.9557522123893806
Epoch: 10 Train Loss: tensor(0.2794) Train Accuracy: 0.9752808988764045 Test Accuracy: 0.9557522123893806
Epoch: 11 Train Loss: tensor(0.1378) Train Accuracy: 0.9775280898876404 Test Accuracy: 0.9646017699115044
```

Рисунок 58. Результаты тестирования модели. Спектрограммы. Набор данных после ПЭ (бин. класс.)

Таблица 16. Результаты обучения и тестирования модели.

Мел-спектрограммы

Набор данных / Результаты	Начальный набор данных (разбаланс.)	Начальный набор данных (без разбаланс.)	Начальный набор данных (бин. класс.)	Набор данных после ПЭ (бин. класс.)
Training accuracy (12 эпох)	0.922943	0.954173	0.9875	0.959550
Test accuracy (12 эпох)	0.658536	0.543352	0.773076	0.821428

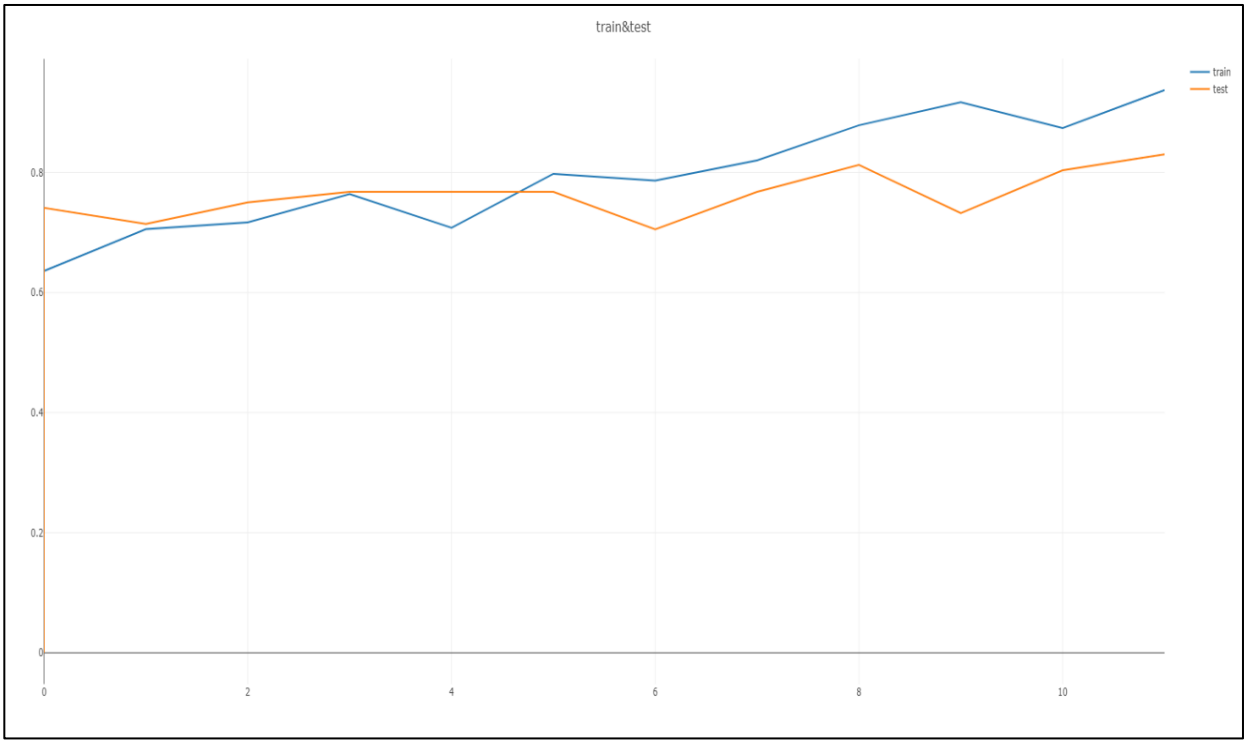

Рисунок 59. Визуализация тестирования модели. Мел-спектрограммы. Набор данных после ПЭ (бин. класс.)

```
Epoch: 0 Train Loss: tensor(13.3857) Train Accuracy: 0.6359550561797753 Test Accuracy: 0.7410714285714286
Epoch: 1 Train Loss: tensor(4.3034) Train Accuracy: 0.7056179775280899 Test Accuracy: 0.7142857142857143
Epoch: 2 Train Loss: tensor(3.4401) Train Accuracy: 0.7168539325842697 Test Accuracy: 0.75
Epoch: 3 Train Loss: tensor(2.1959) Train Accuracy: 0.7640449438202247 Test Accuracy: 0.7678571428571429
Epoch: 4 Train Loss: tensor(2.7814) Train Accuracy: 0.7078651685393258 Test Accuracy: 0.7678571428571429
Epoch: 5 Train Loss: tensor(1.0528) Train Accuracy: 0.797752808988764 Test Accuracy: 0.7678571428571429
Epoch: 6 Train Loss: tensor(0.8771) Train Accuracy: 0.7865168539325843 Test Accuracy: 0.7053571428571429
Epoch: 7 Train Loss: tensor(0.6064) Train Accuracy: 0.8202247191011236 Test Accuracy: 0.7678571428571429
Epoch: 8 Train Loss: tensor(0.2961) Train Accuracy: 0.8786516853932584 Test Accuracy: 0.8125
Epoch: 9 Train Loss: tensor(0.2208) Train Accuracy: 0.9168539325842696 Test Accuracy: 0.7321428571428571
Epoch: 10 Train Loss: tensor(0.3347) Train Accuracy: 0.8741573033707866 Test Accuracy: 0.8035714285714286
Epoch: 11 Train Loss: tensor(0.1479) Train Accuracy: 0.9370786516853933 Test Accuracy: 0.8303571428571429
```

Рисунок 60. Результаты тестирования модели. Мел-спектрограммы. Набор данных после ПЭ (бин. класс.)

Реализованная сверточная нейронная сеть показала средние показатели на следующих наборах данных.

1. Графики основного тона: набор данных после перцептивного эксперимента для бинарной классификации – 0.725663.
2. Спектрограммы: начальный набор данных без разбалансировки – 0.757961; начальный набор данных для бинарной классификации – 0.746987.
3. Мел-спектрограммы: начальный набор данных для бинарной классификации – 0.773076.

Результаты выше среднего были зафиксированы на наборе после перцептивного эксперимента для бинарной классификации у спектрограмм и мел-спектрограмм, 0.964601 и 0.821428, соответственно.

Так как, эксперименты были проведены на наборах данных, прошедших проверку перцептивным экспериментом, следовательно, результаты в большей степени согласовываются с тем, как воспринимает и интерпретирует эмоции человек.

Подобные результаты могут свидетельствовать о том, что проведение перцептивных экспериментов и апробация материала с участием респондентов может повысить успешность и качество классификации эмоций с использованием нейросетевых технологий.

3.5. Выводы по главе 3

В главе 3 приведено описание реализованной сверточной нейронной сети и результаты экспериментов.

Реализованная сверточная нейронная сеть показала результаты выше среднего на наборах данных прошедший перцептивный эксперимент у спектрограмм и мел-спектрограмм, что может свидетельствовать о том, что применение подобной методики может усовершенствовать технологию распознавания эмоций с использованием нейросетевых технологий.

Заключение

В рамках данной работы была разработана система, которая позволяет классифицировать негативные эмоции с использованием нейросетевых технологий. При этом были решены следующие задачи:

1. Записан уникальный набор речевых данных. Были записаны 72 диктора мужского пола, в возрасте от 20 до 60 лет. После обработки записанного материала в общий корпус вошли 1 442 аудио-фрагмента общей продолжительностью 1 час 17 минут. В общий перечень эмоций данного исследования вошли: страх, раздражение, удивление, печаль, отвращение, радость, презрение, нейтральность, ехидство.

2. Проведен перцептивный эксперимент. Участие приняли 14 респондентов в возрасте от 23 до 74 лет. В результате перцептивного эксперимента начальный набор данных сократился до 646 аудио-фрагментов, общей продолжительностью 30 минут 43 секунды.

3. На основе исследования научной литературы выбрана методика обработки речевых данных. В связи с тем, что было принято решение о реализации сверточной нейронной сети, на вход которой подаются изображения, необходимо было преобразовать аудио-фрагменты в изображения. Было принято решение преобразовать каждый аудио-фрагмент в три типа изображений: изображение графика основного тона, спектрограмму и мел-спектрограмму.

4. Разработаны и реализованы алгоритмы предобработки данных. В результате аудио-фрагменты были преобразованы в спектрограммы и мел-спектрограммы. Изображения графиков основного тона извлекались вручную.

5. Сформированы обучающие и тестовые выборки. В общей сложности на данный момент сформированы 4 набора данных. Временная продолжительность каждого представлена ниже:

1) начальный набор данных с разбалансировкой – 1 час 17 минут;
2) начальный набор данных без разбалансировки – 39 минут

15 секунд;

3) начальный набор данных для бинарной классификации – 1 час 9 минут;

4) набор данных после ПЭ для бинарной классификации – 26 минут 21 секунда.

6. Реализована, обучена и протестирована нейронная сеть.

Максимальные показатели, которых удалось достичь, представлены ниже.

Графики основного тона: 0.725663 – набор данных после перцептивного эксперимента для бинарной классификации.

Спектрограммы: 0.964601 – набор данных после перцептивного эксперимента для бинарной классификации.

Мел-спектрограммы: 0.821428 – набор данных после перцептивного эксперимента для бинарной классификации.

Дальнейшая работа может быть направлена на улучшение точности классификации. Для этого в первую очередь необходимо существенно расширить размер наборов данных, рассмотреть другие форматы предобработки аудио-материала и, возможно, изменить топологию нейронной сети.

Список литературы

1. Алдошина И. Связь акустических параметров с эмоциональной выразительностью речи и пения // Звукорежиссёр. 2003. № 2. С. 17 – 25.
2. Бабенко Л.Г. Лексические обозначения эмоций в русском языке. Свердловск: Из–во Урал. Ун–та. 1989. 184 с. ISBN 5–7525–0061–3.
3. Беляев А. (2019) Мультимодальное распознавание эмоций [видеозапись презентации мультимодального корпуса для распознавания эмоций на конференции Moscow Data Science Major (31.08.2019), секция Fail/Success story] // YouTube. 7.10.2019 (https://www.youtube.com/watch?v=UJKqls7RsuY).
4. Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики: Учеб. пособие для студ. филол. и лингв. фак. высш. учеб. заведений. – 4–е издание., испр. – СПб.: Филологический факультет СПбГУ; М.: Издательский центр “Академия”, 2004. – 160с. ISBN 5–8465–0177–X (Филол.фак. СПбГУ), ISBN 5–7695–1658–5 (Изд.центр “Академия”).
5. Бэн А. Психология. М., 1906. Т. 2 (Кн. 3 – 4).
6. Вундт В. Основы физиологической психологии: Чувства и аффекты. СПб., 1880. Вып. 55 (Т. 3, гл. XVI). 216 с.
7. Гиноян Р.В., Хомутов А.Е. Физиология эмоций. Учебно–методическое пособие. Изд–во Нижегородского госуниверситета. 2010. 66 с.
8. Додонов Б.И. Эмоция как ценность. – М.: Политиздат, 1978. – 272 с.
9. Жерон О. Прикладное машинное обучение с помощью Scikit–Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. – Пер. с англ. – СпБ.: ООО "Альфа–книга', 2018. – 688 с. ISBN 978–5–9500296–2–2.
10. Изард К. Эмоции человека. – Изд–во Питер, 2002, 464 с.
11. Ильин Е**.** П**.** Эмоции и чувства. 2–е изд. — СПб.: Питер, 2011. – 783 с. ISBN 978–5–4237–0059–1.
12. Карабущенко Н.Б., Сунгурова Н.Л., Чхиквадзе Т.В., Пилишвили Т.С. Особенности распознавания эмоций студентами из России и стран Азии (интеллектуальные основания) // Вестник ТвГУ. Серия "Педагогика и

психология". – 2020. – №1 (50). – С. 104–113. DOI: 10.26456/vtpsyped/2020.1.104

13. Карелина И.О. Развитие понимания эмоций в период дошкольного детства: психологический ракурс : монография. – Прага : Vědecko vydavatelské centrum «Sociosféra–CZ», 2017. – 178 с. ISBN 978–80–7526–228–8.

14. Кислова О.О., Русалова М.Н. Восприятие эмоций в речи. Обзор исследований в психологии и физиологии // Успехи физиологических наук. – 2013. – том 44, № 2. – С. 41 – 61.

15. Кодзасов С.В., Кривнова О.Ф. Общая фонетика: Учебник. М.:Рос. гос. гуманит. ун–т, 2001. 592 с. ISBN 5–7281–0347–2.

16. Леонтьев А.Н. Потребности, мотивы и эмоции: Конспект лекций. – М., 1971.

17. Маслечкина С.В. Выражение эмоций в языке и речи // Вестник БГУ. 2015. №3. С. 231 – 236.

18. Мелёхин А.И., Сергиенко Е.А. Когнитивные смещения при распознавании эмоций по лицу в пожилом возрасте [Электронный ресурс] // Клиническая и специальная психология. 2019. Том 8. № 2. С. 53 – 79. doi: 10.17759/psyclin.2019080204.

19. Мельников М.Е., Безматерных Д.Д., Козлова Л.И., Натарова К.А., Штарк М.Б. Стиль привязанности и распознавание эмоциональной мимики при депрессии. Бюллетень сибирской медицины. 2021; № 20(1), С. 90 – 97. https://doi.org/10.20538/1682–0363–2021–1–90–97.

20. Менделевия В. Д. Психология девиантного поведения. Учебное пособие. – СПб.: Речь, 2005. – 445 с. ISBN 5–9268–0387–X.

21. Морозов В.П. Эмоциональный слух человека // Эволюц. биохимии и физиологии. 1985.Т. 21. № 6. С. 569 – 577.

22. Морозов В.П., Дмитриева Е.С., Зайцева К.А. и др. Возрастные особенности восприятия человеком эмоций в речи и пении // Эволюц. Биохимии и физиологии. 1983. Т. 19. № 3. С. 289 – 292.

23. Пашина А.Х. К проблеме распознавания эмоционального контекста звуковой речи // Вопросы психологии. 1991. № 1. С. 88 – 95.

24. Романенко В.О. Эмоциональные характеристики речи и их связь с акустическими параметрами // Общество, среда, развитие. – 2010. – №4. – С. 119 – 123.

25. Романов Д.А. Языковая репрезентация эмоций: уровни, функционирования и системы исследований (на материале русского языка): автореф. дис. …док. филолог. наук: 10.02.01; 10.02.19/Романов Д.А.; Тульский государственный педагогический университет имени Л. Н. – Белгород, 2004. – 30 с.

26. Рубинштейн С.Л. Основы психологии. – Изд–во: Питер, 2018. 714 с.

27. Сидорова О.А., Симонов П.В., Цветкова Л.С. Методика изучения восприятия признаков эмоционального состояния у человека // Журн. высш. нерв. деятельности. 1978. Т. 18. Вып. 2. С. 415 – 419.

28. Симонов П.В. Высшая нервная деятельность человека. Мотивационно–эмоциональные аспекты. – Изд–во: Ленанд, 2021. – 176 с. ISBN 978–5–9710–8623–9.

29. Симонов П.В. Метод К. С. Станиславского и физиология эмоций. – М.: Книга по требованию, 2012. – 86 с. ISBN 978–5–458–31589–0.

30. Симонов П. В. Исследование эмоциональных реакций животных и человека в научных учреждениях США // Журнал высшей нервной деятельности. 1968. Вып. 5. С. 836 – 849.

31. Скрелин П.А. Сегментация и трансрипция. – СПб.: Из–во С.–Петербургского ун–та, 1999. – 108 с. ISBN 5–288–02352–2.

32. Смирнов В.М., Резникова Т.Н., Губачев Ю.М., Дорничев В.М. Мозговые механизмы психофизиологических состояний. – Л.: Наука, 1989. – 148 с.

33. Таулли Т. Основы искусственного интеллекта: нетехническое введение. – СПб.: БХВ–Петербург, 2021. – 288 с. ISBN 978–5–9775–6717–6.

34. Узеиров А.А. Девиантные формы поведения личности: учебно–

методическое пособие. – Ростов Н/Д: Изд–во РостГМУ, 2017. – 30 с.

35. Уотсон Д.Б. Основные направления психологии в классических трудах. Бихевиоризм. Принципы обучения, основанные на психологии. Психология как наука о поведении. – М.: ООО "Издательство АСТ–ЛТД", 1998. – 704 с. ISBN 5–15–000894–X (АСТ).

36. Уфимцева А.А. Слово в лексико–семантической системе языка. – М.: УРСС, 1968. – 286 с.

37. Фресс П. Эмоции // Экспериментальная психология. М., 1975. Вып. V. С. 111 – 195.

38. Шадриков В.Д. Введение в психологию: эмоции и чувства. – М.: Логос, 2002. – 156 с. ISBN 5–94010–159–3.

39. Шаховский В.И. Категоризация эмоций в лексико–семантической системе языка. М.: ЛКИ, 2008. 208 с.

40. Экман П. Психология эмоций. Издательство: Питер, 2019. – 448 с.– ISBN 978–5–4461–1304–0.

41. Adolphs R., Tranel D., Damasio A.R. Dissociable Neural Systems for Recognizing Emotions // Brain and Cognition. 2003. V. 52. № 1. P. 61–69.

42. Arnold M. B. Emotion and Personality. v. 1. Psychological aspects. v. 2. Neurological and physiological aspects. N.–Y., Columbia University Press, 1960.

43. Atila O., Sengür A. Attention guided 3D CNN–LSTM model for accurate speech based emotion recognition // Applied Acoustics. – 2021. – №182. – P. 1–11.

44. Bakhshi A.,Harimi A., Chalup S. CyTex: Transforming speech to textured images for speech emotion recognition // Speech Communication. – 2022. – №139. – P. 62 – 75.

45. Beier E.G., Zautra A.J. Identification of vocal communication of emotions across cultures // Journal of Consulting and Clinical Psychology. Vol. 39, Issue 1. – 1972, August.

46. Buermann M., van Meer T.A.J.P. Speech recognition using very deep neural networks: Spectrograms vs Cochleagrams // URL: DOI:

10.13140/RG.2.2.19111.09121 (дата обращения: 25.03.2022).

47. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W. , Weiss B. A Database of German Emotional Speech // INTERSPEECH 2005 – Eurospeech, 9th European Conference on Speech Communication and Technology,. – Lisbon, Portugal, September 4–8, 2005: September 2005. – P. 1 – 4.

48. Chen Q., Huang G. A novel dual attention–based BLSTM with hybrid features in speech emotion recognition // Engineering Applications of Artificial Intelligence. – 2021. – №102. – P. 1 – 11.

49. Chen, M., He, X., Yang, J., Zhang, H., 2018. 3–d convolutional recurrent neural networks with attention model for speech emotion recognition. IEEE Signal Process. Lett. 25 (10), P. 1440 – 1444.

50. El Ayadi, M., Kamel, M.S., Karray, F., 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recognit. №44(3), P. 572 – 587.

51. Elfenbein H.A., Ambady N. Universals and Cultural Differences in Recognizing Emotions // Current Directions in Psychological Science. – 2003. – Vol. 12, No. 5. – P. 159–164.

52. EmoDB Dataset, Emotional Speech database for classification problem. // kaggle URL: https://www.kaggle.com/datasets/piyushagni5/berlin–database–of–emotional–speech–emodb (дата обращения: 02.05.2022).

53. Fahada Md. S., Ranjan A., Yadav J., Deepak A. A survey of speech emotion recognition in natural environment // Digital Signal Processing. – 2021. – №110. – С. 1 – 28.

54. Gupta V., Juyal S., Hu Y–C. Understanding human emotions through speech spectrograms using deep neural network // The Journal of Supercomputing. – 2022. –Том 78, Выпуск 5. – P. 6944 – 6973.

55. Hoegen R., Gratch J., ParkinsonB., Shore D. Signals of Emotion Regulation in a Social Dilemma: Detection from Face and Context // 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). – Cambridge, UK: IEEE, 09 December 2019. – DOI: 10.1109/ACII.2019.8925478.

56. Jianhua T. Emotion recognition for human–computer interaction // Virtual Reality & Intelligent Hardware. – 2021. – Volume 3, Issue 1. – P. iii – iv.
57. Kotlyar G.M., Morozov V.P. Acoustic Correlates of the Emotional Content of Vocalized Speech // Sov.Physics. Acoust. 1976. № 22. P. 370 – 376.
58. Kwon M.S. MLT–DNet: Speech emotion recognition using 1D dilated CNN based on multi–learning trick approach // Expert Systems With Applications. – 2021. – №167. – P. 1 – 12.
59. Lee J., Kim S., Kim S., Park J., Sohn K. Context–Aware Emotion Recognition Networks // IEEE International Conference on Computer Vision (ICCV). – Oct. 2019.
60. Lei S., Gratch J. Smiles Signal Surprise in a Social Dilemma // 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). – Cambridge, UK: IEEE, 09 December 2019. – DOI: 10.1109/ACII.2019.8925494.
61. Li D., Liu J., Yang Z., Sun L., Wang Z. Speech emotion recognition using recurrent neural networks with directional self–attention // Expert Systems With Applications. – 2021. – №173. – P. 1 – 13.
62. Liping P., Liang G., Zhang J., Xiaoru W., Hongquan Q., Xin W., Subject–specific mental workload classification using EEG and stochastic configuration network (SCN) // Biomedical Signal Processing and Control. – 2021. – Volume 68. – P. 1 – 15.
63. Maithri M., Raghavendra U., Gudigar A., Samanth J., Barua P.D., Murugappan M., Chakole Y., Acharya U.R. Automated emotion recognition: Current trends and future perspectives // Computer Methods and Programs in Biomedicine. – 2022. – Volume 215. – P. 1 – 30.
64. Matplotlib [Электронный ресурс]. URL: https://matplotlib.org/ (дата обращения: 05.04.2022).
65. Nassif A.B., Shahin I., Elnagar A., Velayudhan D., Alhudhaif A., Polat K. Emotional speaker identification using a novel capsule nets model // Expert Systems With Applications. – 2022. – №193. – P. 1–11.

66. Ngai W. K., Xie H., Zou D., Chou K–L. Emotion recognition based on convolutional neural networks and heterogeneous bio–signal data sources // Information Fusion. – 2022. – №77. – P. 107 – 117.

67. NumPy [Электронный ресурс]. URL: https://www.numpy.org/ (дата обращения: 05.04.2022).

68. Pandey S.K., Shekhawat H.S. Prasanna S.R.M. Attention gated tensor neural network architectures for speech emotion recognition // Biomedical Signal Processing and Control. – 2022. – №71. – P. 1 – 16.

69. Pell M.D. Cerebral Mechanisms for Understanding Emotional Prosody In Speech // Brain and Language. 2006. V. 96. Issue 2. P. 221 – 234.

70. Plutchik R. A general psychoevolutionary theory of emotion // R. Plutchik, H. Kellerman (eds.). Emotion: Theory, research and experience, v. 1: Theories of emotion.N.–Y., Acad. Press, 1980a, p. 3 – 31.

71. Preto S., Emotion–reading algorithms cannot predict intentions via facial expressions., september 4, 2019 // USC News URL: https://news.usc.edu/160360/algorithms–emotions–facial–expressions–predict–intentions/ (дата обращения: 25.03.2022).

72. PyTorch [Электронный ресурс]. URL: https://pytorch.org/ (дата обращения: 05.04.2022).

73. Ryerson. Audio–Visual Database of Emotional Speech and Song (RAVDESS) // kaggle.com URL: https://www.kaggle.com/uwrfkaggler/ravdess–emotional–speech–audio (дата обращения: 22.02.2022).

74. Salovey P., Mayer J.D. Emotional intelligence meets traditional standards for an intelligence // Intelligence. – 1999. – Volume 27, Issue 4. – P. Pages 267 – 298.

75. Senthilkumar N., Karpakam S., Gayathri Devi M., Balakumaresan R., Dhilipkumar P. Speech emotion recognition based on Bi–directional LSTM architecture and deep belief networks // Materials Today: Proceedings. – Available online 28 December 2021. – Aticle in press. – https://doi.org/10.1016/j.matpr.2021.12.246.

76. The Interactive Emotional Dyadic Motion Capture (IEMOCAP) Database // University of Southern California URL: https://sail.usc.edu/iemocap/ (дата обращения: 02.05.2022).
77. Titchener E. B. A primer of psychology. N.–Y., 1899.
78. Tseng Li–Ping, Chuang Mao–Te, Liu Yung–Ching Effects of noise and music on situation awareness, anxiety, and the mental workload of nurses during operations // Applied Ergonomics. – 2022. – Volume 99. – P. 1 – 7.
79. Visdom // GitHub [Электронный ресурс]. URL: https://github.com/fossasia/visdom (дата обращения: 25.04.2022).
80. Waard D., Nes N. International Encyclopedia of Transportation // Driver State and Mental Workload. – 2021. – P. 216 – 220.

Приложения

Приложение А. Список фраз

Эмоция	Контекст	Текст
Раздра-жение (9 фраз)	Ваш коллега уже в который раз, который день подходит и спрашивает о том как нужно работать с программой/делать расчеты/как включить принтер.	Сколько раз мне нужно тебе повторить?
	Близкий человек/Родственник/Друг ударил Вас, при этом этот удар не был похож на «да, я же по-дружески».	Да как ты смеешь?!
	Ваш брат или сестра шантажирует Вас тем, что если Вы с ним не поделитесь конфетой/игрушкой/и т.д., он(а) расскажет Ваш секрет, который Вы ему/ей доверили.	Только попробуй.
	В Ваш адрес сказали обидную, неприятную фразу.	Что ты сказал? Повтори.
	Вам уже несколько дней говорят о том, чтобы Вы не забыли выполнить поручение.	Я же сказал, что я помню, не нужно мне повторять.
	Уже который день вы добиваетесь от своего сотрудника отчета, но все никак не можете его получить.	Почему я до сих пор не получил Ваш отчет?
	Вы одолжили своему другу достаточно крупную сумму денег. Он обещал вернуть долг к определенному сроку. Срок истек 4 месяца назад.	Когда долг вернешь?
	Вы пришли на фильм, который очень долго ждали. Зайдя в зал, вы увидели, что Ваше место занято.	Вы заняли мое место.
	В Вашем подъезде нет воды уже вторую неделю, вы пытаетесь дозвониться в ЖЭК/УК, но везде стоит автоответчик. В какой–то момент трубку поднимает оператор, и равнодушным голосом отвечает, что воды не будет еще месяц, и кладет трубку. Вам звонит Ваша девушка жена и интересуется не дозвонились ли Вы узнать по поводу отсутствия воды.	Я уже 150 раз звонил в этот ЖЭК, все, что мне сказал оператор, что воды не будет месяц!

Продолжение приложения А. Список фраз

Ехидство (1 фраза)	Вы предполагали исход ситуации и говорили об этом другу, но он Вас упорно не хотел слушать и верил в свой успех. В итоге, ваши предположения оказались явью.	А я тебя предупреждал.
Нейтраль-ность (6 фраз)	Вы интересуетесь у близкого родственника как прошел его/ее день.	Как прошел твой день?
	Вы заходите в кабинет к коллеге с просьбой помочь.	Надеюсь, я тебя не отвлекаю.
	Вы – учитель, спрашиваете у класса все ли понятно.	Еще есть вопросы?
	Вы успокаиваете своего собеседника.	Если что, я рядом.
	Запись для объявления.	Уважаемые пассажиры, говорит пилот, мы входим в зону турбулентности, просим пристегнуть ремни безопасности.
	Запись для объявления.	Пожарная тревога. Всем покинуть здание согласно плану эвакуации.
Отвраще-ние (3 фразы)	Вас предал человек, которому Вы доверяли.	Как можно быть таким негодяем?
	Ваш партнер сильно потолстел, но его/ее это не волнует. Его/ее состояние устраивает, вы же беспокоитесь за его/ее здоровье и то, что человек себя «запустил».	Тебе нужно начать ходить в спортзал.
	В комнату зашел Ваш знакомый/друг/подруга и принес с собой неприятный запах. Вы интересуетесь, не является ли этот человек источником этого неприятного аромата.	Ты когда последний раз мылся?
Печаль (6 фраз)	Человек, который Вам привлекателен, обладает неприятным темпераментом, но несмотря на это, Вы хотите продолжить общение с этим человеком.	Да, характер у него, конечно, сложный.

Продолжение приложения А. Список фраз

	Производится запись для электронного приложения банка. Произнесите фразу не как «бесчувственная» машина, а как понимающий сотрудник банка.	К сожалению, мы вынуждены сообщить, что наш банк не может предоставить Вам кредит.
	Ваш близкий человек подвел Вас и в момент, когда он был нужен Вам, не пришел на помощь.	Такого от тебя я не ожидал.
	Вы – ребенок и объясняете маме, чем отличается Капитан Америка от Железного Человека, в чем их суперсила и почему они крутые. Но мама не разделяет Вашего увлечения, и в итоге вы сдаетесь.	Ты ничего не понимаешь.
	Вы собираетесь сообщить собеседнику неприятную новость.	Нам нужно срочно поговорить.
	Ваша идея, в которой Вы были уверены на 100%, не увенчалась успехом. О провальности данного мероприятия предупреждал(а) Вас ваш(а) друг/подруга.	Да, ты была права.
Презрение (6 фраз)	Вы зашли в магазин, о котором слышали много положительных и восторженных отзывов. Но по итогу, ассортимент магазина оказался скудным, сотрудники – хамами, а расположение магазина дальше, чем вы предполагали.	Магазин своеобразный.
	Вы узнали, что ваш знакомый слушает музыку, которая Вам не нравится, и вы не понимаете, как подобное можно не просто назвать музыкой, но и тем более слушать.	Что за музыку ты слушаешь?
	Ваш друг/знакомый решил продемонстрировать свой талант (танец/рисунок/пение), но, по вашему мнению, хвалиться нечем.	Это не твое, может, попробуешь что–то другое?

Продолжение приложения А. Список фраз

	Вы слушаете своего собеседника, который живет в прекрасном мире, о том, что все можно изменить, не бывает безвыходных положений. Но вы так не считаете.	Смирись, ты все равно ничего не сможешь изменить…
	Вы – секретарь, клиенты/сотрудники каждую минуту заглядывают в Ваш кабинет и спрашивают можно ли пройти к начальнику.	Ждите, я вас приглашу.
	В который раз вы делаете замечание своему знакомому относительно его грамотности.	Вообще–то, правильно говорить звонИт, а не звОнит.
Радость (4 фразы)	Ваша мечта купить квартиру/машину исполнилась и Вы хотите поделиться этой новостью с другом.	У меня для тебя есть новости.
	Вы приготовили подарок, о котором Ваш друг/девушка /жена давно мечтал(а), Вам не терпится подарить и обрадовать человека.	У меня для тебя подарок.
	Ваш друг/подруга подстригся(лась) и вы в восторге от его/ее нового образа.	Тебе так идет эта прическа.
	Вы очень гордитесь достижениями своего близкого человек, и хотите его подбодрить.	Ты большая молодец, у тебя все получится.
Страх (3 фразы)	Вы разбили любимую машину мужа/жены/парня/девушки. И очень боитесь, что за содеянным, последует наказание, в какой форме – вы не знаете.	Я разбил твою машину.
	Вам сообщили диагноз, вы звоните близкому родственнику, чтобы сообщить новость.	Пришли результаты анализов, они неутешительные, рак в 4ой стадии.
	Вам пришло уведомление из банка о проведенной операции, которую Вы не совершали. Сумма приличная.	С моего счета списали деньги!
Удивление (2 фразы)	Коллега поделился новостью, что выйграл в лотерею. Сумма баснословная и вы не верите в реальность случившегося.	Ты выиграл 500 млн рублей?

Продолжение приложения А. Список фраз

	Вы – специалист кабинета ультразвукового исследования. На первичный осмотр к Вам пришла беременная женщина, в результате исследования, вы определила 3 плода. На последующем обследовании, неожиданно для себя вы увидели еще 2. И сообщаете об этом пациентке	Поздравляем, УЗИ показало, что у вас будет 3 мальчика и 2 девочки.

Приложение Б. Списки фраз для дикторов

Список 1

Контекст	Текст
Человек, который Вам привлекателен, обладает неприятным темпераментом, но несмотря на это, Вы хотите продолжить общение с этим человеком.	Да, характер у него, конечно, сложный.
Производится запись для электронного приложения банка. Произнесите фразу не как «бесчувственная» машина, а как понимающий сотрудник банка.	К сожалению, мы вынуждены сообщить, что наш банк не может предоставить Вам кредит.
Вы зашли в магазин, о котором слышали много положительных и восторженных отзывов. Но по итогу, ассортимент магазина оказался скудным, сотрудники – хамами, а расположение магазина дальше, чем вы предполагали.	Магазин своеобразный.
Вы разбили любимую машину мужа/жены/парня/девушки. И очень боитесь, что за содеянным, последует наказание, в какой форме – вы не знаете.	Я разбил твою машину.
Ваш коллега уже в который раз, который день подходит и спрашивает о том как нужно работать с программой/делать расчеты/как включить принтер.	Сколько раз мне нужно тебе повторить?
Вы интересуетесь у близкого родственника как прошел его/ее день.	Как прошел твой день?
Вас предал человек, которому Вы доверяли.	Как можно быть таким негодяем?
Вы узнали, что ваш знакомый слушает музыку, которая Вам не нравится, и вы не понимаете, как подобное можно не просто назвать музыкой, но и тем более слушать.	Что за музыку ты слушаешь?

Список 2

Контекст	Текст
Вы заходите в кабинет к коллеге с просьбой помочь.	Надеюсь, я тебя не отвлекаю.
Близкий человек/Родственник/Друг ударил Вас, при этом этот удар не был похож на «да, я же по-дружески».	Да как ты смеешь?!
Ваш близкий человек подвел Вас и в момент, когда он был нужен Вам, не пришел на помощь.	Такого от тебя я не ожидал.
Ваша мечта купить квартиру/машину исполнилась и Вы хотите поделиться этой новостью с другом.	У меня для тебя есть новости.
Вы приготовили подарок, о котором Ваш друг/девушка /жена давно мечтал(а), Вам не терпится подарить и обрадовать человека.	У меня для тебя подарок.
Вы – учитель, спрашиваете у класса все ли понятно.	Еще есть вопросы?
Вы – ребенок и объясняете маме, чем отличается Капитан Америка от Железного Человека, в чем их суперсила и почему они крутые. Но мама не разделяет Вашего увлечения, и в итоге вы сдаетесь.	Ты ничего не понимаешь.
Ваш друг/подруга подстригся(лась) и вы в восторге от его/ее нового образа.	Тебе так идет эта прическа.

Список 3

Контекст	Текст
Ваш брат или сестра шантажирует Вас тем, что если Вы с ним не поделитесь конфетой/игрушкой/и т.д., он(а) расскажет Ваш секрет, который Вы ему/ей доверили.	Только попробуй.
В Ваш адрес сказали обидную, неприятную фразу.	Что ты сказал? Повтори.

Продолжение списка 3

Ваш партнер сильно потолстел, но его/ее это не волнует. Его/ее состояние устраивает, вы же беспокоитесь за его/ее здоровье и то, что человек себя «запустил».	Тебе нужно начать ходить в спортзал.
Вам уже несколько дней говорят о том, чтобы Вы не забыли выполнить поручение.	Я же сказал, что я помню, не нужно мне повторять.
Вы предполагали исход ситуации и говорили об этом другу, но он Вас упорно не хотел слушать и верил в свой успех. В итоге, ваши предположения оказались явью.	А я тебя предупреждал.
Вы слушаете своего собеседника, который живет в прекрасном мире, о том, что все можно изменить, не бывает безвыходных положений. Но вы так не считаете.	Смирись, ты все равно ничего не сможешь изменить…
Ваш друг/знакомый решил продемонстрировать свой талант (танец/рисунок/пение), но, по вашему мнению, хвалиться нечем.	Это не твое, может, попробуешь что–то другое?
Вы успокаиваете своего собеседника.	Если что, я рядом.

Список 4

Контекст	Текст
Вы очень гордитесь достижениями своего близкого человек, и хотите его подбодрить.	Ты большая молодец, у тебя все получится.
Вам сообщили диагноз, вы звоните близкому родственнику, чтобы сообщить новость.	Пришли результаты анализов, они неутешительные, рак в 4ой стадии.
Запись для объявления.	Уважаемые пассажиры, говорит пилот, мы входим в зону турбулентности, просим пристегнуть ремни безопасности.
Вам пришло уведомление из банка о проведенной операции, которую Вы не совершали. Сумма приличная.	С моего счета списали деньги!

Продолжение списка 4

Вы – секретарь, клиенты/сотрудники каждую минуту заглядывают в Ваш кабинет и спрашивают можно ли пройти к начальнику.	Ждите, я вас приглашу.
Вы собираетесь сообщить собеседнику неприятную новость.	Нам нужно срочно поговорить.
Коллега поделился новостью, что выйграл в лотерею. Сумма баснословная и вы не верите в реальность случившегося.	Ты выиграл 500 млн рублей?
Вы – специалист кабинета ультразвукового исследования. На первичный осмотр к Вам пришла беременная женщина, в результате исследования, вы определила 3 плода. На последующем обследовании, неожиданно для себя вы увидели еще 2. И сообщаете об этом пациентке.	Поздравляем, УЗИ показало, что у вас будет 3 мальчика и 2 девочки.

Список 5

Контекст	Текст
Запись для объявления.	Пожарная тревога. Всем покинуть здание согласно плану эвакуации.
Ваша идея, в которой Вы были уверены на 100%, не увенчалась успехом. О провальности данного мероприятия предупреждал(а) Вас ваш(а) друг/подруга.	Да, ты была права.
Уже который день вы добиваетесь от своего сотрудника отчета, но все никак не можете его получить.	Почему я до сих пор не получил Ваш отчет?
Вы одолжили своему другу достаточно крупную сумму денег. Он обещал вернуть долг к определенному сроку. Срок истек 4 месяца назад.	Когда долг вернешь?
В который раз вы делаете замечание своему знакомому относительно его грамотности.	Вообще–то, правильно говорить звонИт, а не звОнит.

Продолжение списка 5

Вы пришли на фильм, который очень долго ждали. Зайдя в зал, вы увидели, что Ваше место занято.	Вы заняли мое место.
В комнату зашел Ваш знакомый/друг/подруга и принес с собой неприятный запах. Вы интересуетесь, не является ли этот человек источником этого неприятного аромата.	Ты когда последний раз мылся?
В Вашем подъезде нет воды уже вторую неделю, вы пытаетесь дозвониться в ЖЭК/УК, но везде стоит автоответчик. В какой–то момент трубку поднимает оператор, и равнодушным голосом отвечает, что воды не будет еще месяц, и кладет трубку. Вам звонит Ваша девушка жена и интересуется не дозвонились ли Вы узнать по поводу отсутствия воды.	Я уже 150 раз звонил в этот ЖЭК, все, что мне сказал оператор, что воды не будет месяц!

Приложение В. Тест по методике Н. Холла

Инструкция: Далее вам будут предложены высказывания, которые, так или иначе, отражают различные стороны вашей жизни. Пожалуйста, поставьте плюс в одном из полей «Варианты ответа» к каждому из высказываний.

№ ПП	Варианты ответа / Высказывание	Полностью не согласен (–3 балла)	В основном не согласен (–2 балла)	Отчасти не согласен (–1 балл)	Отчасти согласен (+1 балл)	В основном согласен (+2 балла)	Полностью согласен (+3 балла)
1	Для меня как отрицательные, так и положительные эмоции служат источником знания, как поступать в жизни.						
2	Отрицательные эмоции помогают мне понять, что я должен изменить в моей жизни.						
3	Я спокоен, когда испытываю давление со стороны.						
4	Я способен наблюдать изменение своих чувств.						
5	Когда необходимо, я могу быть спокойным и сосредоточенным, чтобы действовать в соответствии с запросами жизни.						
6	Когда необходимо, я могу вызвать у себя широкий спектр положительных эмоций, таких как веселье, радость, внутренний подъем и юмор.						
7	Я слежу за тем, как я себя чувствую.						
8	После того как что–то расстроило меня, я могу легко совладать со своими чувствами.						

9	Я способен выслушивать проблемы других людей.						
10	Я не зацикливаюсь на отрицательных эмоциях.						
11	Я чувствителен к эмоциональным потребностям других.						
12	Я могу действовать успокаивающе на других людей.						
13	Я могу заставить себя снова и снова встать перед лицом препятствия.						
14	Я стараюсь подходить творчески к жизненным проблемам.						
15	Я адекватно реагирую на настроения, побуждения и желания других людей.						
16	Я могу легко входить в состояние спокойствия, готовности и сосредоточенности.						
17	Когда позволяет время, я обращаюсь к своим негативным чувствам и разбираюсь, в чем проблема.						
18	Я способен быстро успокоиться после неожиданного огорчения.						
19	Знание моих истинных чувств важно для поддержания «хорошей формы».						

20	Я хорошо понимаю эмоции других людей, даже если они не выражены открыто.						
21	Я хорошо могу распознавать эмоции по выражению лица.						
22	Я могу легко отбросить негативные чувства, когда необходимо действовать.						
23	Я хорошо улавливаю знаки в общении, которые указывают на то, в чем другие нуждаются.						
24	Люди считают меня хорошим знатоком переживаний других людей.						
25	Люди, осознающие свои истинные чувства, лучше управляют своей жизнью.						
26	Я способен улучшить настроение других людей.						
27	Со мной можно посоветоваться по вопросам отношений между людьми.						
28	Я хорошо настраиваюсь на эмоции других людей.						
29	Я помогаю другим использовать их побуждения для достижения личных целей.						
30	Я могу легко отключиться от переживания неприятностей.						

Интерпретация результатов

Шкала	Пункты
Эмоциональная осведомленность	1, 2, 4, 17, 19, 25
Управление своими эмоциями	3, 7, 8, 10, 18, 30
Самомотивация	5, 6, 13, 14, 16, 22
Эмпатия	9, 11, 20, 21, 23, 28
Распознавание эмоций других людей	12, 15, 24, 26, 27, 29

Максимальное количество баллов, которое может быть получено респондентом – 18. Шкала каждого параметра разделена на три уровня: низкий, средний и высокий уровень развития. Максимальный балл по каждому критерию – 18 баллов. Полученный результат Диапазон низкого уровня составляет от 0 до 11 баллов. Средний уровень – 12 (70% от общего числа) – 15 баллов. Высокий уровень – от 16 (90% от общего числа) до 18 баллов.