

Научная статья

УДК 81"33

DOI: <https://doi.org/10.18721/JHSS.14206>



## ЗАПИСЬ И АПРОБАЦИЯ НАБОРА РЕЧЕВЫХ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ НЕГАТИВНЫХ ЭМОЦИЙ В РЕЧИ

А.А. Поволоцкая<sup>1</sup>  , В.В. Евдокимова<sup>2</sup> ,  
П.А. Скрелин<sup>2</sup> 

<sup>1</sup> Санкт-Петербургский Федеральный исследовательский центр  
Российской академии наук, Санкт-Петербург, Российская Федерация;

<sup>2</sup> Санкт-Петербургский государственный университет,  
Санкт-Петербург, Российская Федерация

✉ [anastasiia.povolotskaia@gmail.com](mailto:anastasiia.povolotskaia@gmail.com)

**Аннотация.** В статье описывается подход записи набора данных, на основе специально разработанного списка фраз. В перечень эмоций исследования вошли 7 эмоций, как базовых: отвращение, печаль, презрение, радость, раздражение, страх, удивление, также в перечень вошла эмоция нейтрального произнесения высказываний, и эмоция ехидства, которая является сложной эмоцией и может быть интерпретирована людьми по-разному. Общий список фраз состоял из 40 высказываний различной длины, общий список был разделен на 5 подсписков по 8 фраз для каждого диктора. Диктору был представлен лексический состав целевой фразы и соответствующий ей контекст, установки, т.е. эмоции были удалены для чистоты эксперимента. Данный подход к записи набора данных был выбран с целью получения более естественных реализаций эмоций в рамках студийной записи. После записи дикторов, был сформирован первичный набор данных, который прошел апробацию в рамках перцептивного эксперимента. Предварительно все аудиторы прошли тестирование на определение уровня эмоционального интеллекта по методике Н. Холла.

**Ключевые слова:** распознавание эмоций, распознавание негативных эмоций, запись набора данных, набор данных для распознавания эмоций, набор данных для распознавания негативных эмоций, интонация.

**Для цитирования:** Поволоцкая А.А., Евдокимова В.В., Скрелин П.А. Запись и апробация набора речевых данных для распознавания негативных эмоций в речи // Terra Linguistica. 2023. Т. 14. № 2. С. 59–76. DOI: 10.18721/JHSS.14206



## RECORDING AND EVALUATION OF SPEECH DATA SET FOR NEGATIVE EMOTIONS RECOGNITION IN SPEECH

A.A. Povolotskaia<sup>1</sup>  , V.V. Evdokimova<sup>2</sup> ,  
P.A. Skrelin<sup>2</sup> 

<sup>1</sup> St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russian Federation;

<sup>2</sup> St. Petersburg State University, St. Petersburg, Russian Federation

 [anastasiia.povolotskaia@gmail.com](mailto:anastasiia.povolotskaia@gmail.com)

**Abstract.** The article describes the approach of recording a database, based on a specially developed list of phrases. The list of emotions included 7 emotions, as the basic ones: disgust, sadness, contempt, joy, irritation, fear, surprise; the list also included the emotion of neutrality, and the emotion of smirk, which is a complex emotion and can be interpreted by people in different ways. The total list of phrases consisted of 40 utterances of different lengths, the total list was divided into 5 sublists of 8 phrases for each speaker. The speaker was presented with the lexical composition of the target phrase and its corresponding context, attitudes, i.e. emotions were removed for the clarity of the experiment. This approach to recording the dataset was chosen in order to obtain more natural realizations of emotion within the studio. After recording the speakers, a primary dataset was generated and validated with auditory experiment. Preliminarily to the experiment all the auditors were tested to determine the level of emotional intelligence according to the method of N. Hall.

**Keywords:** emotion recognition, negative emotion recognition, dataset recording, dataset for emotion recognition, dataset for negative emotion recognition, intonation.

**Citation:** A.A. Povolotskaia, V.V. Evdokimova, P.A. Skrelin, Recording and evaluation of speech data set for negative emotions recognition in speech, *Terra Linguistica*, 14 (2) (2023) 59–76. DOI: 10.18721/JHSS.14206

### Введение

Эмоции играют важную роль в жизни человека, например, с помощью эмоций мы воспринимаем мир и выражаем свое мнение. Тема классификации эмоций является одной из актуальных задач машинного обучения.

Активное развитие систем, подразумевающих человеко-машинное взаимодействие, повышает актуальность задачи распознавания эмоций. При анализе речи с помощью алгоритмов машинного обучения первоочередной задачей является поиск большого количества данных, которые можно использовать для обучения и тестирования моделей. Распознавание эмоций с использованием алгоритмов машинного обучения применимо в таких областях как: определение депрессии на основе анализа выражаемых человеком эмоций [40], распознавание эмоций клиентов и работников колл-центров [33], интеграция имитации эмоционального интеллекта в роботизированные устройства и голосовые ассистенты [28], выявление усталости [17].

Прежде чем решать задачу классификации данных, необходимо ответить на вопрос – какие данные использовать: студийные записи с участием профессиональных актеров или собирать материал на просторах интернета (in the wild), какие эмоции должны быть представлены в наборе данных, какую методику записи набора данных выбрать.

### Постановка проблемы

При записи наборов данных с эмоциональной речью чаще всего применяется следующая методика: приглашаются профессиональные актеры, задача которых озвучить одну или несколько



фраз в соответствии с определенными эмоциональными стимулами (нейтральность, гнев, страх, удивление и др.) [18, 22, 25, 32, 37].

Выбирают актеров, поскольку те реализации, которые они воспроизводят проще понять аудиторам, которые в дальнейшем примут участие в перцептивном эксперименте. Несмотря на то, что при записи актеров упрощается организационный момент, поскольку вокальная реализация различных эмоций входит в список их профессиональных компетенций, запись актеров имеет один существенный недостаток — неестественная реализация эмоционального состояния человека, которая свойственна ему в повседневной жизни.

В результате того, что в записи принимают участие люди с профессиональным актерским образованием, наборы данных получаются однородными. Впоследствии классификация аудиозаписей с применением алгоритмов машинного обучения показывает высокие результаты. Связано это с тем, что данные в своей реализации сходны, что ведет к высоким показателям при их классификации с помощью алгоритмов машинного обучения, но если на вход подать аудиофрагмент, который не имеет сходства с данными, представленными в наборе, то, скорее всего, данный фрагмент будет классифицирован неверно.

Сбор данных *in-the-wild* (из открытых источников, таких как, фильмы, ТВ-шоу, YouTube, и другие ресурсы) может быть необъективным. Необходимо определить не только источники, из которых будут заимствоваться эмоциональные высказывания, но и объективность их классификации по эмоциональным группам. Разметка и классификация данных, как правило, осуществляется небольшой группой аннотаторов, что может привести к необъективности при распределении данных по классам эмоций [37].

Авторами статьи было принято решение: для того, чтобы получить более естественные и вариативные реализации эмоций, необходимо составить список фраз различной длины, при этом, дикторы будут принимать самостоятельное решение об эмоции высказывания и степени ее интенсивности (т.е. будет отсутствовать эталон эмоциональной реализации фразы) с учетом контекста. Подобная методика записи набора данных была выбрана с целью исследования качества классификации эмоциональной речи методами машинного обучения на фразах различной длины и различной реализации.

### **Описание корпусов с эмоциональной речью**

Перед организацией записи был проведен сравнительно-сопоставительный анализ корпусов с эмоциональной речью, которые находятся в открытом доступе. Был составлен сводный перечень эмоций по каждому набору данных, изучена методика записи и апробации записанного материала.

При формулировании целей исследования одной из первоначальных является определение перечня эмоций, на которые будет осуществляться дальнейшая классификация. Например, для мониторинга качества обслуживания клиентов специалистами call-центров достаточно двух эмоциональных состояний: доволен или недоволен клиент качеством обслуживания. Для развития технологий диалоговых систем, алгоритмы должны распознать и классифицировать уже значительно более широкий спектр человеческих эмоций. В рамках данного исследования будут рассматриваться наборы данных с несколькими эмоциями.

Большинство из них на английском языке: RAVDESS [25], SAVEE [22], TESS [32], IEMOCAP [19], CREMA-D [20]; 2 на русском: RUSLANA и RAMAS [27, 30]; 1 на немецком, но также есть записи на английском языке: Emo-DB [18]; несколько языков представлено в наборе данных ASVP-ESD [23].

Примерный перечень эмоций для каждого набора данных включает в себя следующий ряд эмоций: спокойствие, нейтральность, счастье, грусть, злость/гнев/раздражение, испуг, удивление, отвращение и др. [18, 19, 20, 22, 23, 25, 27, 30, 32]. Сводная информация о всех эмоциях,



представленных в вышеупомянутых наборах данных, представлена в табл. 1, в скобках отмечены эмоции, которые были добавлены при аннотации корпусов.

**Таблица 1. Сводная таблица эмоций, представленных в каждом наборе данных**  
**Table 1. Summary table of emotions represented in each data set**

Эмоция	Набор данных	Emo-DB	RAVDESS	SAVEE	TESS	IEMOCAP	CREMA-D	RUSLANA	RAMAS	ASYP-ESD
Anger (раздражение, злость, гнев, ярость)		+	+	+	+	+	+	+	+	+
Anxiety (беспокойство, тревога)		+								
Boredom (скука)		+								
Calm (спокойствие)			+							
Disappointment (разочарование)						+				
Disgust (отвращение)		+	+	+	+	(+)	+		+	
Excitement (восторг)						(+)				
Fear (страх)			+	+	+	(+)	+	+	+	+
Frustration (расстройство)						+				
Happiness (радость, счастье)		+	+	+	+	+	+	+	+	+
Neutral (нейтральность)		+	+	+	+	+	+	+	+	+
Sadness (печаль)		+	+	+	+	+	+	+	+	+
Surprise (удивление)			+	+	+	(+)		+	+	+

**Emo-DB [18].** При записи набора данных Emo-DB участие приняли 10 дикторов (актеров), задача которых заключалась в том, чтобы озвучить 10 предложений, каждое с 7 различными эмоциями. Предложения являются лексически нейтральными. Записанные данные прошли валидацию путем перцептивного эксперимента с участием 20 аудиторов, которые оценивали узнаваемость эмоции и естественность звучания. Записи, которые получили оценку выше 80% и 60% по параметрам узнаваемости и естественности, соответственно, были маркированы специальными обозначениями, отмечающими качество голоса, фонационных и артикуляционных особенностей.

**RAVDESS [25].** К записи были привлечены 24 профессиональных актера - носителя английского языка, с североамериканским акцентом, которые озвучили два лексически нейтральных предложения, каждое с 8 различными эмоциями. Особенность набора данных заключается в наличии 2 эмоций, как исходных и эмоционально нейтральных – спокойствие (calm) и нейтральность (neutral). Каждая эмоция была реализована с двумя типами интенсивности: нормальная и сильная. В валидации приняли участие 319 студентов бакалавриата университета в Торонто, которые оценивали отдельно визуальную и отдельно акустическую модальности, и акустико-визуальную модальность. Задача аудиторов была оценить тип эмоции, интенсивность и естественность реализации.

**SAVEE [22].** Участие в записи данного корпуса приняли 4 профессиональных актера с британским английским. В общей сложности были записаны 480 высказываний, от каждого диктора получилось по 120 видео-аудио-фрагментов, которые затем были апробированы с участием 10 респондентов (5 носителей английского языка и остальные 5 проживали на территории Великобритании больше года). 120 видео-аудио-фрагментов от каждого диктора были разделены на 10 поднаборов данных для каждого аудитора. Таким образом, каждый аудитор соотнес 48 аудио-видео-фрагментов с 7 классами эмоций.



**TESS [32].** Во время записи набора дынных TESS, участие принимали только две женщины в возрасте 26 и 64 лет. Ключевая фраза: «Say the word \_», и 200 целевых слов. Слово необходимо было произнести с 7 различными эмоциями. Апробация с участием респондентов у данного корпуса отсутствует.

**ИЕМОСАР [19].** Перед актерами была поставлена следующая задача: в парах воспроизвести два типа сессий. Первая – сценарная, когда актеры читали сценарий, и затем воспроизводили его по памяти. Вторая – спонтанная, когда актерам необходимо было реализовать определенные эмоции в рамках импровизационной постановки. Каждый тип сессии был проигран 5 раз одной и той же парой актеров. Запись набора данных осуществлялась на аудио и видео-носители. Также, у доминирующего в диалоге актера осуществлялся 3D захват мимики, и для фиксирования жестикуляции использовались браслеты. При аннотации набора данных, каждое высказывание было оценено 3 респондентами (всего их было 6). Авторы приняли решение о дополнении исходного перечня эмоций, который использовался при записи актеров, 4 дополнительными эмоциями (отвращение, восторг, страх, удивление). Респонденты могли оценивать одно высказывание несколькими эмоциями, если их мнение не соответствовало ни одной из перечисленных эмоций, то респонденты были в праве выбрать вариант ответа «Другое» и указать свой вариант реализованной эмоции.

**CREMA-D [20].** При создании набора данных CREMA-D было привлечено большое количество аудиторов с помощью интернета. В набор данных вошли записи 91 актера (48 мужчин и 43 женщины) в возрасте от 20 до 74 лет, актеры являлись представителями различных этнических групп (афроамериканцы, представители азиатской и кавказской национальностей, испанцы, и др.). Задача актеров – озвучить 12 предложений, каждое с 6 эмоциями и различной степенью интенсивности: низкая, средняя, высокая и неопределенная (что именно подразумевается под данным термином автор не уточняет). Участие в апробации приняли 2443 аудитора, которые оценивали эмоции по видео и аудио как отдельно, так и в совместном представлении эмоции. Приблизительно 74% респондентов были представителями кавказской национальности, почти 11% – испанцы, 8% – афроамериканцы и 4.5% – представители азиатской этнической группы. Сначала респонденты соотносили эмоциональное высказывание с соответствующей ему эмоцией, затем оценивали интенсивность по 100-бальной шкале. Каждый фрагмент был оценен минимум 10 раз.

**RUSLANA [27].** Данный набор данных включает в себя реализацию 6 следующих эмоций: удивление, радость, раздражение, печаль, страх и нейтральность. Задача дикторов заключалась в том, чтобы реализовать 10 лексически нейтральных предложений с 6 эмоциями, перечисленными выше. Апробация полученных записей осуществлялась в 2 этапа: 1) размещение реализованной диктором эмоции на координатной оси (ось OX – валентность, где «+» и «-» знак эмоции, позитивная или негативная; ось OY – интенсивность или активность реализованной эмоции); 2) насколько правильно была оценена реализованная эмоция по шкале от 1 до 5 (1 – очень плохо, нет сходства; 5 – полное соответствие), т.е. аннотатору представлялась запись, оценка другого аннотатора (в соответствии с пунктом 1), и эмоция, которая подразумевалась при реализации высказывания.

**RAMAS [30].** Мультимодальный набор данных, который включает в себя видео, аудио, 3D захват движения и физиологические данные. Также учитывалось участие в диалоге: лидирующая роль или доминирующая (domination) и принимающая или роль слушателя (submission). Набор данных отличается от всех остальных методикой записи: полупрофессиональным актерам были представлены 13 сценариев с описанием одной из 6 эмоций (раздражение, печаль, отвращение, радость, страх, удивление и нейтральность), реплики для обязательного воспроизведения отсутствовали. Таким образом, можно считать, что данный набор данных содержит эмоциональные вокализации, реализация которых приближена к спонтанным или повседневным эмоциям. В разметке данных приняли участие 21 аннотатор (по 5 аннотаторов на каждую запись). Задача



аннотаторов заключалась в том, чтобы с помощью специальной программы отметить начало и конец эмоции, которая кажется им наиболее естественной в своей реализации. Все аннотаторы прошли тестирование на определение уровня эмоционального интеллекта по методике J. Mayer, P. Salovey и D. Caruso, адаптированной для русскоговорящих [35]. Участие в аннотации приняли только те респонденты, которые набрали средние и высокие показатели оценки уровня эмоционального интеллекта.

**ASVP-ESD [23, 37].** При формировании данного набора данных собирались записи из открытых источников Интернета и онлайн-платформ, фильмов, записей реальной коммуникации и др. Аннотация осуществлялась 5 респондентами, путем отнесения аудио-фрагмента к одной из эмоций, выделенных в рамках исследования. В результате голосования фрагмент относился к той или иной эмоции, если количество голосов совпадало, то фрагмент случайным образом относился к одной из эмоций, набравших одинаковое количество голосов.

Информация по наборам данных представлена в сводной табл. 2. Знаком «—» обозначена информация, которая не указана в статьях авторов наборов данных.

**Таблица 2. Сводная таблица по наборам данных**  
**Table 2. Summary table by data sets**

Набор данных	Дикторы			Методика	Записи	Валидация
	Возраст	М	Ж			
Емо-DB	21 – 35	5	5	Озвучить 10 предложений с 7 эмоциями	~ 800 фрагментов	Оценка узнаваемости эмоции и естественности звучания
RAVDESS	21 – 33	12	12	Озвучить 2 предложения с 8 эмоциями	1440 фрагментов	Оценка типа эмоции, интенсивность и естественности реализации
SAVEE	27 – 31	4	0	Озвучить 10 предложений с 8 эмоциями	480 фрагментов	Соотнесение с обозначенными эмоциями
TESS	26, 64	0	2	Одно предложение и 200 целевых слов	2800 фрагментов	Отсутствует
IEMOCAP	—	5	5	Сценарные сессии и импровизационные сессии	~12 часов	Соотнесение с обозначенными эмоциями
CREMA-D	20 – 74	48	43	Озвучить 12 предложений с 6 различными эмоциями	7442 фрагментов	Соотнесение с обозначенными эмоциями и оценка интенсивности реализации
RUSLANA	16 – 28	12	49	Озвучить 10 предложений с 6 различными эмоциями	3660 фрагментов	Определение валентности и интенсивности реализованной эмоции. Оценка правильности определения валентности и интенсивности эмоции
RAMAS	18 – 28	5	5	Сценарии с описанием эмоции, которую нужно реализовать, импровизация актеров	~ 7 часов	Отметить начало и конец эмоции, которая кажется им наиболее естественной
ASVP-ESD	0 – 70	62	62	Записанные фрагменты речи из реальной коммуникации, ресурсов открытых платформ, фильмов и др	2001 фрагмент	Соотнесение с обозначенными эмоциями



На основе проведенного сравнительно-сопоставительного анализа наборов данных, можно выделить несколько следующих сходных черт, перечисленных ниже.

1. Наблюдается схожая методика записи у большинства корпусов: дикторам необходимо произнести предложение с различными эмоциями, выделенными в рамках эксперимента.
2. Все наборы данных включают эмоции: раздражения, радости, нейтральности и печали. В большинство наборов данных входят такие эмоции, как: отвращение, страх, удивление.
3. Большинство наборов данных прошли валидацию с участием респондентов.
4. Обнаруживается характерная черта для записей – отсутствие экстралингвистических явлений, возникновение которых возможно в повседневной речи. Основной фокус сконцентрирован на просодической составляющей эмоциональной реализации высказываний: интонация, интенсивность речи.
5. Отмечается отсутствие определенной методики для валидации данных и оценки качества реализованных эмоций.

В результате анализа можно прийти к следующим выводам.

1. Несмотря на то, что предложения могут иметь разную длину, так или иначе, вариативность собранного набора будет крайне низкой, в результате модель нейронной сети, обученная и протестированная на подобном наборе данных, не будет успешно классифицировать материал, который в него не входит, поскольку реализация эмоций будет сходной и любое отклонение может быть не классифицировано должным образом.
2. Нет четкого представления какие эмоции относятся к базовым, все исследователи используют различные списки эмоций.
3. Также, стоит учитывать, что реализация и восприятие эмоций зависит от культурного опыта и национальной принадлежности как диктора, так и аудитора. Поэтому, необходимо, чтобы и дикторы, и аудиторы принадлежали к одной национальности и имели приблизительно одинаковый культурный, и социальный опыт.
4. Такие невербальные явления, как вздохи, цыканья и другие, могут быть маркерами определенных эмоций, и при анализе эмоционального состояния их необходимо учитывать. Подобные вокализации часто встречаются в спонтанной речи. Таким образом, при формировании наборов данных для распознавания эмоций их необходимо учитывать, а не удалять из исходных речевых фрагментов. Изучение подобных явлений в спонтанной речи позволит производить более качественную классификацию эмоций в речи.

В результате анализа аналогов цель исследования была сформулирована следующим образом: разработать набор высказываний и соответствующие им контексты таким образом, чтобы диктор самостоятельно принял решение о том, какая эмоция заключена в высказывании, чтобы получить более естественные реализации эмоций и сформировать вариативный набор данных.

Изучив наборы данных, которые находятся в открытом доступе, было принято решение о составлении списка фраз различной длины. Чтобы получить реализации эмоций, близкие к повседневным и естественным, было принято решение о том, что дикторам не предоставляются четкие установки/эмоции, т.к. фразы и контекст будут подобраны таким образом, чтобы диктор определил эмоцию и степень ее интенсивности самостоятельно. Позднее записанный набор данных был апробирован с участием респондентов, которые приняли участие в перцептивном эксперименте.

#### **Методология и методика исследования. Обоснование перечня эмоций**

Для определения перечня эмоций, которые войдут в исследование, была проанализирована литература ряда ученых, на основе исследований которых была составлена сводная табл. 3 базовых эмоций [2, 10, 11, 12, 14, 16, 34].



**Таблица 3. Сводная таблица базовых эмоций**  
**Table 3. Summary table of basic emotions**

Эмоция	Автор	Дж. Б Уотсон [12]	Р. Плутчик [14, 34]	К. Изард [2, 14]	П. Экман [16]	В.М. Смирнов [11, 14]	П.В. Симонов [10, 14]
Страх		+	+	+	+	+	+
Раздражение (ярость/злость/гнев)		+	+	+	+	+	+
Любовь		+					
Доверие			+				
Ожидание			+				
Удивление			+	+	+		
Печаль (грусть/горе/страдание)			+	+	+		
Отвращение			+	+	+		+
Радость			+	+	+		
Интерес – волнение				+			
Презрение				+	+		
Вина				+			
Стыд				+			
Удовольствие						±	+
Дискомфорт							+

В результате, в табл. 3, где пересечений получилось больше двух, было принято решение о представлении данных эмоций в списке фраз: отвращение, печаль, презрение, радость, раздражение, страх, удивление. Также, в материал были включены нейтральные стимулы (без выраженной эмоциональности).

В качестве эксперимента была рассмотрена эмоция ехидства, которая, являясь разновидностью косвенной агрессии и формой девиантного поведения, может быть интерпретирована людьми, как одна из базовых эмоций (например, печаль, раздражение или нейтральный комментарий). Восприятие данной эмоции зависит от уровня эмоционального интеллекта человека [4].

Обоснованным представляется перечень эмоций, в который вошло 7 базовых эмоций (по классификации П. Экмана), сложная эмоция ехидства, и нейтральность.

#### **Составление перечня высказываний**

При составлении списка фраз был проведен обзор научной литературы по эмотивной вербальной русской речи [4, 8, 9, 13, 15] и интонационному выражению эмоций [1, 6, 26]. Впоследствии составлен список из 40 фраз, каждой из которых предоставлен контекст. Примеры фраз, вошедших в общий список, представлены в табл. 4. Соотношение эмоций и количества фраз в общем списке получилось следующим: ехидство – 1 фраза; удивление – 2 фразы; отвращение – 3 фразы; страх – 3 фразы; радость – 4 фразы; нейтральность – 6 фраз; печаль – 6 фраз; презрение – 6 фраз; раздражение – 9 фраз. Весь список из 40 фраз был поделен на 5 подсписков по 8 фраз для каждого диктора, таким образом, чтобы в списке были представлены несколько различных эмоций. Первая колонка, колонка с эмоциями, была удалена для чистоты эксперимента. При записи дикторам предоставлялся список из двух колонок: контекст и текст фразы.

#### **Отбор и запись дикторов**

Выборка по дикторам была осуществлена с целью ограничения спектральных и тембральных характеристик голоса. Решение не производить запись женских голосов, было принято в связи с тем, что привлекательность женского голоса и общее физическое, и эмоциональное состояние





женщины зависит от фазы менструального цикла, изменения в структуре голосовых связок связаны с наличием беременности или наступившей менопаузой, данные явления совпадают с выраженными гормональными изменениями, что в свою очередь оказывает влияние на изменения в женском голосе [21, 24, 31].

**Таблица 4. Примеры фраз, вошедших в общий список**  
**Table 4. Examples of phrases included in the general list**

Эмоция	Контекст	Фраза
Раздражение	Близкий человек / Родственник / Друг ударил Вас, при этом этот удар не был похож на «да, я же по-дружески»	Да как ты смеешь?!
Ехидство	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.
Нейтральность	Вы интересуетесь у близкого родственника как прошел его/ее день	Как прошел твой день?
Нейтральность	Запись для объявления	Уважаемые пассажиры, говорит пилот, мы входим в зону турбулентности, просим пристегнуть ремни безопасности.
Отвращение	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	Ваша идея, в которой Вы были уверены на 100%, не увенчалась успехом. О провальности данного мероприятия предупредил(а) Вас ваш(а) друг/подруга	Да, ты была права.
Презрение	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?
Радость	Вы приготовили подарок, о котором Ваш друг / девушка / жена давно мечтал(а), Вам не терпится подарить и обрадовать человека	У меня для тебя подарок.
Удивление	Коллега поделился новостью, что выиграл в лотерею. Сумма баснословная и вы не верите в реальность случившегося	Ты выиграл 500 млн рублей?
Страх	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.

Исследователи зависимости тембра и распознавания эмоций пришли к выводу, что тембр играет важную роль, по сравнению с частотой основного тона (F0), при адаптации слушателя к вокальным реализациям эмоций говорящего [29, 36, 38]. Было принято решение о записи дикторов мужского пола с одним тембром голоса. Выбор был остановлен на теноре, как одном из самых распространенных мужских голосов, который не обладает слишком низкими или высокими акустическими характеристиками, что может вызывать неприятные ощущения у слушателей. В результате, было решено проводить отбор дикторов мужского пола с одним тембром голоса – тенор.



Перед дикторами была поставлена следующая задача: ознакомиться с контекстом и соответствующей ему фразой, затем вслух прочитать целевую фразу так, как её бы произнес диктор, оказавшись в данных условиях. После озвучивания представленного списка фраз с эмоциональной окраской, диктор произносит весь список фраз с нейтральной эмоцией, т.е. без какой бы то ни было эмоциональной окраски. Также, дикторы не были ограничены в качественной реализации целевых фраз. В то время, как записи в некоторых других корпусах не содержат экстралингвистических явлений, такие как тяжелые выдохи/вздохи, «цыканье» и др., было принято решение не отказываться от подобных явлений, а оставить их в исходных речевых файлах.

После записи дикторов записанный материал был подвергнут обработке. Каждая фраза была выделена из общего аудиофайла, записанного диктором, и сохранена в отдельный файл с помощью программы Wave Assistant [39].

На базе кафедры фонетики и методики преподавания иностранных языков СПбГУ, для проведения экспериментального исследования было привлечено 72 диктора мужского пола с тембром тенор. После обработки записей получилось 1442 аудио-фрагмента, каждый из которых был сохранен в отдельный файл формата WAV, с частотой дискретизации 22 050 Гц, общий временной объем файлов составил 1 час 17 минут. Все дикторы являются либо гражданами РФ, либо продолжительное время проживающими на ее территории.

#### **Перцептивный эксперимент**

С целью получения объективного представления о классификации записей по эмоциям было принято решение о проведении перцептивного эксперимента с предварительным тестированием auditors на определение уровня эмоционального интеллекта. Для проведения перцептивного эксперимента к участию были привлечены 14 auditors (носители русского языка, в возрасте от 23 до 74 лет, 3 мужчины и 11 женщин, 11 auditors имеют высшее образование, 2 – среднее техническое, и 1 – полное среднее образование. Все респонденты являются гражданами РФ), при этом предварительно оценивался уровень их эмоционального интеллекта.

В каждом из 14 опросов участие приняли 10 респондентов. Все респонденты были протестированы по методике Н. Холла для определения уровня эмоционального интеллекта [3, с. 645]. В рамках определения уровня эмоционального интеллекта при тестировании учитывались три ключевых критерия: распознавание эмоций других людей, уровень эмпатии и уровень эмоциональной осведомленности. В результате тестирования, 9 из 14 респондентов показали средний и высокий показатели по трем ключевым критериям (табл. 5). Шкала каждого критерия уровня развития эмоционального интеллекта разделена на три уровня: низкий, средний и высокий уровень развития. Максимальный балл по каждому критерию составляет 18 баллов. 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.

Было принято решение о том, чтобы 7 auditors со средними и высокими показателями уровня эмоционального интеллекта прошли все 14 опросов, прослушав каждый аудио-фрагмент. Также, были учтены ответы респондентов с низкими показателями (по 3 в каждом опросе). Перцептивный эксперимент был организован таким образом, чтобы сформировать более репрезентативную выборку, поскольку в реальной жизни не все люди являются обладателями высокого эмоционального интеллекта.

Перцептивный эксперимент был организован на платформе Google-Forms. Весь материал, 1442 аудио-фрагмента, был представлен в форме анкеты на сайте вышеуказанной платформы, где были составлены 14 опросов. Начальная страница опроса и страница опроса представлены на рис. 1 и 2.

Задача для auditors была сформулирована следующим образом: прослушайте аудио-фрагмент и ответьте на вопрос: «Какую эмоцию реализует в данном высказывании диктор?». На выбор



предложены девять вариантов ответов, соответствующие 9 вышеупомянутым эмоциям, и десятый вариант – свободная форма для заполнения, чтобы аудиторы могли указать свой вариант ответа. Все вопросы представлялись для каждого диктора в случайном порядке. Также был проведен предварительный инструктаж с каждым аудитором:

- 1) выполнять опросы в комфортной обстановке и тихом помещении без посторонних шумов, желательно наличие наушников;
- 2) пребывать в спокойном эмоциональном состоянии, не рекомендуется выполнять опрос в конце дня перед сном или после рабочего дня;
- 3) делать перерыв на несколько минут после каждых двадцати вопросов для того, чтобы отдохнуть и не спровоцировать перенапряжение, и возможное «накладывание» впечатлений о ранее прослушанных записях.

**Таблица 5. Результаты тестирования по методике Н. Холла на определение уровня эмоционального интеллекта**  
**Table 5. The results of testing according to the method of N. Hall to determine the level of emotional intelligence**

№ ПП	Имя (возраст)	Уровень эмоциональной осведомленности (кол-во баллов)	Уровень эмпатии (кол-во баллов)	Распознавание эмоций других людей (кол-во баллов)
1	Валерия (23)	Средний уровень (15 баллов)	Средний уровень (12 баллов)	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 балла)	Средний уровень (15 баллов)
14	Валентина (74)	Средний уровень (14 баллов)	Средний уровень (15 баллов)	Низкий уровень (11 баллов)



**Опрос №2**

Чтобы сохранить изменения, [войдите в аккаунт Google](#). [Подробнее...](#)

**\* Обязательно**

Пожалуйста, укажите Ваш возраст: \*

Мой ответ \_\_\_\_\_

Как Вас зовут? \*

Мой ответ \_\_\_\_\_

[Далее](#)  Страница 1 из 107 [Очистить форму](#)

Никогда не используйте формы Google для передачи паролей.  
Компания Google не имеет никакого отношения к этому контенту. [Сообщение о нарушении](#) - [Условия использования](#) - [Политика конфиденциальности](#)

Google Формы

Рис. 1. Начальная страница опроса № 2  
Fig. 1. The initial page of the survey No. 2

**Опрос №2**

Чтобы сохранить изменения, [войдите в аккаунт Google](#). [Подробнее...](#)

**\* Обязательно**

**Какую эмоцию реализует в данном высказывании диктор?**

ПОЯСНЕНИЕ: Перейдите по ссылке и прослушайте аудио-фрагмент.

<https://is.gd/us9EAd> \*

Раздражение (т.е. ярость, гнев, раздражение, злость, ненависть)

Печаль

Презрение

Страх

Отвращение

Радость

Удивление

Нейтральность

Ехидство

Другое: \_\_\_\_\_

[Назад](#) [Далее](#)  Страница 2 из 107 [Очистить форму](#)

Никогда не используйте формы Google для передачи паролей.  
Компания Google не имеет никакого отношения к этому контенту. [Сообщение о нарушении](#) - [Условия использования](#) - [Политика конфиденциальности](#)

Google Формы

Рис. 2. Страница опроса № 2  
Fig. 2. Survey page No. 2



## Результаты исследования

В случае, если 70 и более процентов респондентов сходились во мнении относительно эмоции аудио-фрагмента, то данный аудио-фрагмент относился к соответствующему классу эмоций.

Результаты записи дикторов («Начальный набор данных») и результаты проведенного перцептивного эксперимента («Набор данных после перцептивного эксперимента») представлены в табл. 6.

**Таблица 6. Результаты записи и проведенного перцептивного эксперимента**  
**Table 6. Results of recording and conducted perceptual experiment**

Эмоция	Количество записей	
	Начальный набор данных	Набор данных после перцептивного эксперимента
Ехидство	15	6
Нейтральность	781	397
Отвращение	57	0
Печаль	115	40
Презрение	105	2
Радость	86	63
Раздражение	173	110
Страх	56	2
Удивление	54	26
Итого:	1442	646
Время:	1 час 17 минут	30 минут 43 секунды

В дальнейшем собранный набор данных был разбит на поднаборы. Была построена модель сверточной нейронной сети (CNN), на вход которой подавались: спектрограммы, мел-спектрограммы и изображения контуров частоты основного тона [7].

Собранный набор данных получил название «CoCSE» – Corpus of Context Stimulated Emotions.

## Заключение

Эмоции субъективны, их сложно фиксировать, и все люди воспринимают и интерпретируют их по-разному, в результате сложно сформировать объективные наборы данных. Более того, сложно получить естественные реализации эмоций в ситуациях, созданных искусственно, например, при записи в студии.

На основе проведенного анализа литературы и сравнительно-сопоставительного анализа наборов данных с эмоциональной речью было высказано следующее предположение: возможно составить перечень фраз и соответствующих им контекстов без указания эмоций (конкретных стимулов), для получения более естественных и не наигранных реализаций эмоций при записи в студии.

В результате проведенной работы был записан уникальный набор данных, основанный на списке фраз различной длины.

В данном исследовании при записи набора данных отбирались дикторы мужского пола и только с одним тембром голоса. Данное решение обосновывается тем, что аудиторы адаптируются к распознаванию эмоций в рамках одного тембра и наличие в записях голосов различных тембров может привести к дизадаптации аудиторов, и впоследствии некорректному классифицированию записанного аудиоматериала.

Перед проведением перцептивного эксперимента было организовано тестирование на определение уровня эмоционального интеллекта аудиторов разной возрастной категории. Данное



тестирование было организовано для того, чтобы сформировать репрезентативную выборку для каждого из 14 опросов и получить вариативность ответов с учетом людей как способных с легкостью распознавать человеческие эмоции, так и людей, испытывающих трудности.

В результате перцептивного эксперимента количество записей по каждому классу эмоций значительно сократилось. Что может свидетельствовать о том, что проведение перцептивных экспериментов при формировании наборов данных с эмоциональной речью может позволить учитывать мнение респондентов с различным уровнем эмоционального интеллекта и исключать записи, которые не соответствуют характеристикам эмоционального высказывания по аудиальным параметрам.

### СПИСОК ИСТОЧНИКОВ

1. **Галунов В.И.** О возможности определения эмоционального состояния по речи // Речевые технологии. 2008. № 1. С. 60–66.
2. **Изард К.** Эмоции человека. М.: Директ-Медиа. 2008. 954 с.
3. **Ильин Е.П.** Эмоции и чувства. Изд. 2-е. СПб.: Питер. 2011. 783 с.
4. **Маслечкина С.В.** Выражение эмоций в языке и речи // Вестник Брянского государственного университета. 2015. № 3. С. 231–236.
5. **Менделевич В.Д.** Психология девиантного поведения. Учебное пособие. – СПб.: Речь. 2005. 445 с.
6. **Первезенцева О.А.** Роль просодических маркеров в регуляции устно-речевого дискурса // Актуальные вопросы лингвистики и лингводидактики: традиции и инновации: Материалы международной научно-практической конференции, посвященной 70-летию Института иностранных языков, Москва, 22–24 ноября 2018 года / Под ред. Е.А. Никулиной, Е.Е. Беляевой. Том Часть 1. Москва: Московский педагогический государственный университет. 2018. С. 153–159.
7. **Поволоцкая А.А.** Распознавание негативных эмоций с использованием нейросетевых технологий // Фонетический лицей. СПб: ООО «Скифия-принт». 2022. С. 83–91.
8. **Романенко В.О.** Эмоциональные характеристики речи и их связь с акустическими параметрами // Общество. Среда. Развитие (Terra Humana). 2010. № 4. С. 119–123.
9. **Романов Д.А.** Языковая репрезентация эмоций: уровни, функционирования и системы исследований (на материале русского языка): автореф. дис. ...док. филол. наук: Белгород. 2004. 52 с.
10. **Симонов П.В.** Высшая нервная деятельность человека. Мотивационно-эмоциональные аспекты. М.: Изд-во: Ленанд. 2021. 176 с.
11. **Смирнов В.М., Резникова Т.Н., Губачев Ю.М., Дорничев В.М.** Мозговые механизмы психофизиологических состояний. Л.: Наука. 1989. 148 с.
12. **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.
13. **Уфимцева А.А.** Слово в лексико–семантической системе языка. М.: УРСС. 1968. 286 с.
14. **Шадриков В.Д.** Введение в психологию: эмоции и чувства. М.: Логос. 2002. 156 с.
15. **Шаховский В.И.** Категоризация эмоций в лексико–семантической системе языка: монография. Изд. 3-е. М.: УРСС. 2009. 296 с.
16. **Экман П.** Психология эмоций. СПб.: Изд-во: Питер, 2019. 448 с.
17. **Bayerl S.P., Wagner D., Baumann I., Riedhammer K., Bocklet T.** Detecting Vocal Fatigue with Neural Embeddings // Journal of Voice. 2023. DOI: 10.1016/j.jvoice.2023.01.012
18. **Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B.** A database of German emotional speech // 9<sup>th</sup> European Conference on Speech Communication and Technology. 2005. № 5. pp. 1517–1520. DOI: 10.21437/Interspeech.2005–446
19. **Busso C., Bulut M., Lee C.C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S., Narayanan S.S.** IEMOCAP: Interactive emotional dyadic motion capture database // Journal of Language Resources and Evaluation. 2008. No. 4. Pp. 335–359.



20. **Cao H., Cooper D.G., Keutmann M.K., Gur R.C., Nenkova A., Verma R.** CREMA–D: Crowdsourced Emotional Multimodal Actors Dataset // *IEEE Transactions on Affective Computing*. 2014. № 4. pp. 377–390. DOI: 10.1109/TAFFC.2014.2336244
21. **Ghaemi H., Dehqan A., Mahmoodi-Bakhtiari B., Scherer R.S.** Voice Changes During Pregnancy Trimesters in Iranian Pregnant Women // *Journal of Voice*. 2020. № 34. pp. 358–363. DOI: 10.1016/j.jvoice.2018.09.016
22. **Jackson P.** Surrey audio-visual expressed emotion (SAVEE) database. University of Surrey: Guildford, UK. 2014. URL: <http://kahlan.eps.surrey.ac.uk/savee/> (дата обращения: 14.03.2023).
23. **Landry D.T.T., He Q., Yan H., Li Y.** ASVP-ESD: A dataset and its benchmark for emotion recognition using both speech and non-speech utterances // *Global Scientific Journals*. 2020. No. 8. Pp. 1793–1798.
24. **Li D., Zhang L., Wang X.** The Effect of Menstrual Cycle Phases on Approach–Avoidance Behaviors in Women: Evidence from Conscious and Unconscious Processes // *Brain Science*. 2022. No. 12. Pp. 1417. DOI: 10.3390/brainsci12101417
25. **Livingstone S.R., Russo F.A.** The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // *PLoS ONE*. 2018. No. 13. DOI: 10.1371/journal.pone.0196391
26. **Luengo I., Navas E., Hernández I.** Automatic emotion recognition using prosodic parameters // *INTERSPEECH-2005*. 2005. Pp. 493–496.
27. **Makarova V., Petrushin V.A.** RUSLANA: a database of Russian emotional utterances // *Proc. 7<sup>th</sup> International Conference on Spoken Language Processing*. 2002. Pp. 2041–2044. DOI: 10.21437/ICSLP.2002-560
28. **Marcos-Pablos S., García-Peñalvo F.J.** Emotional Intelligence in Robotics: A Scoping Review // *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*. 2021. No. 1410. Pp. 66–75. DOI: 10.1007/978-3-030-87687-6\_7
29. **Nussbaum C., von Eiff C.I., Skuk V.G., Schweinberger S.R.** Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency // *Cognition*. 2022. No. 219. Pp. 104967. DOI: 10.1016/j.cognition.2021.104967
30. **Perepelkin O., Kazimirova E., Konstantinova M.** RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing // *Speech and Computer*. 2018. No. 11096. Pp. 501–510. DOI: 10.1007/978-3-319-99579-3\_52
31. **Pi Lā F.M.B., Ardura D.** What Voice-Related Metrics Change with Menopause? A Systematic Review and Meta-Analysis Study // *Journal of Voice*. 2022. № 36. pp. 438.e1-438.
32. **Pichora-Fuller M.K., Dupuis K.** Toronto emotional speech set (TESS). 2010. URL: <https://tspace.library.utoronto.ca/handle/1807/24487>. DOI: <https://doi.org/10.5683/SP2/E8H2MF>
33. **Plaza M., Kazala R., Koruba Z., Kozłowski M., Lucińska M., Sitek K., Spyrka J.** Emotion Recognition Method for Call/Contact Centre Systems. // *Applied Sciences*. 2022. No. 12. Pp. 10951. DOI: 10.3390/app122110951
34. **Plutchik R.A.** general psychoevolutionary theory of emotion // *Theories of Emotion*. N.–Y.: Academic Press. 1980. Pp. 3–33. DOI: 10.1016/B978-0-12-558701-3.50007-7
35. **Sergienko G., Vetrova I.I., Volochkov A.A., Popov A.Y.** Adaptation of J. Mayer P. Salovey and D. Caruso emotional intelligence test on russian-speaking sample. 2010. No. 31. Pp. 55–73.
36. **Skuk V.G., Dammann L.M., Schweinberger S.R.** Role of timbre and fundamental frequency in voice gender adaptation // *The Journal of the Acoustical Society of America*. 2015. No. 138. Pp. 1180–1193. DOI: 10.1121/1.4927696
37. **Tientcheu T.L.D., Qianhua H., Wei X.** Audio, Speech and Vision Processing Lab Emotional Sound database (ASVP-ESD). 2021. DOI: 10.5281/zenodo.4782712
38. **Tursunov A., Kwon S., Pang H-S.** Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features. *Applied Sciences*. 2019. No. 9. Pp. 2470. DOI: 10.3390/app9122470
39. Wave Assistant – специализированный звуковой редактор с поддержкой пакетной обработки. URL: [https://vk.com/wave\\_assistant](https://vk.com/wave_assistant) (дата обращения: 14.03.2023).
40. **Wu W., Wu M., Yu K.** Inspecting Depression Detection via Emotion Recognition. // *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2022. Pp. 6262–6266. DOI: 10.1109/ICASSP43922.2022.9746634



## REFERENCES

- [1] **V.I. Galunov**, O vozmozhnosti opredeleniya emotsionalnogo sostoyaniya po rechi [About the possibility of determining the emotional state by speech], *Rechevyye tekhnologii* [Speech technologies]. 1 (2008) 60–66.
- [2] **K. Izard**, *Emotsii cheloveka* [Human emotions]. M.: Direkt-Media. 2008.
- [3] **Ye.P. Ilin**, *Emotsii i chuvstva* [Emotions and feelings]. Izd. 2. SPb.: Piter. 2011.
- [4] **S.V. Maslechkina**, Vyrazheniye emotsiy v yazyke i rechi [Expression of emotions in language and speech], *Bulletin of the Bryansk State University*. 3 (2015) 231–236.
- [5] **V.D. Mendeleviya**, *Psikhologiya deviantnogo povedeniya* [Psychology of deviant behavior]. SPb.: Rech. 2005.
- [6] **O.A. Pervezentseva**, The role of prosodic markers in the regulation of oral-speech discourse, *Topical issues of linguistics and linguodidactics: Traditions and Innovations: Materials of the International scientific and Practical conference dedicated to the 70<sup>th</sup> anniversary of the Institute of Foreign Languages, Moscow, November 22–24, 2018*, edited by E.A. Nikulina, E.E. Belyaeva. Volume Part 1. Moscow: Moscow Pedagogical State University. 2018. Pp. 153–159.
- [7] **A.A. Povolotskaya**, Raspoznaniye negativnykh emotsiy s ispolzovaniyem neyrosetevykh tekhnologiy [Recognition of negative emotions using neural network technologies], *Foneticheskiy litsey*. SPb.: OOO “Skifiya-print”. 2022. Pp. 83–91.
- [8] **V.O. Romanenko**, Emotsionalnyye kharakteristiki rechi i ikh svyaz s akusticheskimi parametrami [Emotional characteristics of speech and their connection with acoustic parameters], *Society. Wednesday. Development (Terra Humana)*. 4 (2010) 119–123.
- [9] **D.A. Romanov**, Linguistic representation of emotions: levels, functioning and research systems (based on the material of the Russian language): autoref. dis. ...doc. Philol. sciences: Belgorod. 2004.
- [10] **P.V. Simonov**, Vysshaya nervnaya deyatelnost cheloveka. Motivatsionno–emotsionalnyye aspekty [Higher nervous activity of a person. Motivational and emotional aspects]. M.: Izd-vo: Lenand. 2021.
- [11] **V.M. Smirnov, T.N. Reznikova, Yu.M. Gubachev, V.M. Dornichev**, *Mozgovyye mekhanizmy psikhofiziologicheskikh sostoyaniy* [Brain mechanisms of psychophysiological states]. L.: Nauka. 1989.
- [12] **D.B. Uotson**, Osnovnyye napravleniya psikhologii v klassicheskikh trudakh. Bikhevizm. Printsipy obucheniya, osnovannyye na psikhologii. Psikhologiya kak nauka o povedenii [The main directions of psychology in classical works. Behaviorism. Principles of teaching based on psychology. Psychology as a science of behavior]. M.: OOO “Publishing House AST–LTD”. 1998.
- [13] **A.A. Ufimtseva**, Slovo v leksiko–semanticheskoy sisteme yazyka [The word in the lexico–semantic system of the language]. M.: URSS. 1968.
- [14] **V.D. Shadrikov**, Vvedeniye v psikhologiyu: emotsii i chuvstva [Introduction to psychology: emotions and feelings]. M.: Logos. 2002.
- [15] **V.I. Shakhovskiy**, Kategorizatsiya emotsiy v leksiko–semanticheskoy sisteme yazyka: monografiya [Categorization of emotions in the lexico–semantic system of language: monograph]. Izd. 3. M.: URSS. 2009.
- [16] **P. Ekman**, *Psikhologiya emotsiy* [Psychology of emotions]. SPb.: Piter, 2019.
- [17] **S.P. Bayerl, D. Wagner, I. Baumann, K. Riedhammer, T. Bocklet**, Detecting Vocal Fatigue with Neural Embeddings, *Journal of Voice*. 2023. DOI: 10.1016/j.jvoice.2023.01.012
- [18] **F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, B. Weiss**, A database of German emotional speech, 9<sup>th</sup> European Conference on Speech Communication and Technology. 5 (2005) 1517–1520. DOI: 10.21437/Interspeech.2005–446
- [19] **C. Busso, M. Bulut, C.C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J.N. Chang, S. Lee, S.S. Narayanan**, IEMOCAP: Interactive emotional dyadic motion capture database, *Journal of Language Resources and Evaluation*. 4 (2008) 335–359.
- [20] **H. Cao, D.G. Cooper, M.K. Keutmann, R.C. Gur, A. Nenkova, R. Verma**, CREMA–D: Crowdsourced Emotional Multimodal Actors Dataset, *IEEE Transactions on Affective Computing*. 4 (2014) 377–390. DOI: 10.1109/TAFFC.2014.2336244
- [21] **H. Ghaemi, A. Dehqan, B. Mahmoodi-Bakhtiari, R.S. Scherer**, Voice Changes During Pregnancy Trimesters in Iranian Pregnant Women, *Journal of Voice*. 34 (2020) 358–363. DOI: 10.1016/j.jvoice.2018.09.016
- [22] **P. Jackson**, Surrey audio-visual expressed emotion (SAVEE) database. University of Surrey: Guildford, UK. 2014. URL: <http://kahlan.eps.surrey.ac.uk/savee/> (accessed: 14.03.2023).





- [23] **D.T.T. Landry, Q. He, H. Yan, Y. Li**, ASVP-ESD: A dataset and its benchmark for emotion recognition using both speech and non-speech utterances, *Global Scientific Journals*. 8 (2020) 1793–1798.
- [24] **D. Li, L. Zhang, X. Wang**, The Effect of Menstrual Cycle Phases on Approach–Avoidance Behaviors in Women: Evidence from Conscious and Unconscious Processes, *Brain Science*. 12 (2022) 1417. DOI: 10.3390/brainsci12101417
- [25] **S.R. Livingstone, F.A. Russo**, The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLoS ONE*. 13 (2018). DOI: 10.1371/journal.pone.0196391
- [26] **I. Luengo, E. Navas, I. Hernáez**, Automatic emotion recognition using prosodic parameters, *INTERSPEECH-2005*. 2005. Pp. 493–496.
- [27] **V. Makarova, V.A. Petrushin**, RUSLANA: a database of Russian emotional utterances, *Proc. 7th International Conference on Spoken Language Processing*. 2002. Pp. 2041–2044. DOI: 10.21437/ICSLP.2002-560
- [28] **S. Marcos-Pablos, F.J. García-Peñalvo**, Emotional Intelligence in Robotics: A Scoping Review // *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*. 1410 (2021) 66–75. DOI: 10.1007/978-3-030-87687-6\_7
- [29] **C. Nussbaum, C.I. von Eiff, V.G. Skuk, S.R. Schweinberger**, Vocal emotion adaptation aftereffects within and across speaker genders: Roles of timbre and fundamental frequency, *Cognition*. 219 (2022) 104967. DOI: 10.1016/j.cognition.2021.104967
- [30] **O. Perepelkin, E. Kazimirova, M. Konstantinova**, RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing, *Speech and Computer*. 11096 (2018) 501–510. DOI:10.1007/978-3-319-99579-3\_52
- [31] **F.M.B. Pi Lã, D. Ardura**, What Voice-Related Metrics Change with Menopause? A Systematic Review and Meta-Analysis Study, *Journal of Voice*. 36 (2022) 438.e1–438.
- [32] **M.K. Pichora-Fuller, K. Dupuis**, Toronto emotional speech set (TESS). 2010. URL: <https://tspace.library.utoronto.ca/handle/1807/24487>. DOI: <https://doi.org/10.5683/SP2/E8H2MF>
- [33] **M. Plaza, R. Kazala, Z. Koruba, M. Kozłowski, M. Lucińska, K. Sitek, J. Spyрка**, Emotion Recognition Method for Call/Contact Centre Systems, *Applied Sciences*. 12 (2022) 10951. DOI: 10.3390/app122110951
- [34] **R.A. Plutchik**, General psychoevolutionary theory of emotion, *Theories of Emotion*. N.–Y.: Academic Press. 1980. Pp. 3–33. DOI: 10.1016/B978-0-12-558701-3.50007-7
- [35] **G. Sergienko, I.I. Vetrova, A.A. Volochkov, A.Y. Popov**, Adaptation of J. Mayer P. Salovey and D. Caruso emotional intelligence test on russian-speaking sample. 31 (2010) 55–73.
- [36] **V.G. Skuk, L.M. Dammann, S.R. Schweinberger**, Role of timbre and fundamental frequency in voice gender adaptation, *The Journal of the Acoustical Society of America*. 138 (2015) 1180–1193. DOI: 10.1121/1.4927696
- [37] **T.L.D. Tientcheu, H. Qianhua, X. Wei**, Audio, Speech and Vision Processing Lab Emotional Sound database (ASVP-ESD). 2021. DOI: 10.5281/zenodo.4782712
- [38] **A. Tursunov, S. Kwon, H-S. Pang**, Discriminating Emotions in the Valence Dimension from Speech Using Timbre Features. *Applied Sciences*. 9 (2019) 2470. DOI: 10.3390/app9122470
- [39] Wave Assistant – specialized sound editor with batch processing support. URL: [https://vk.com/wave\\_assistant](https://vk.com/wave_assistant) (accessed: 14.03.2023).
- [40] **W. Wu, M. Wu, K. Yu**, Inspecting Depression Detection via Emotion Recognition, *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2022. Pp. 6262–6266. DOI: 10.1109/ICASSP43922.2022.9746634

## СВЕДЕНИЯ ОБ АВТОРАХ / INFORMATION ABOUT AUTHORS

Поволоцкая Анастасия Андреевна

Anastasiia A. Povolotskaia

E-mail: [anastasiia.povolotskaia@gmail.com](mailto:anastasiia.povolotskaia@gmail.com)

ORCID: <https://orcid.org/0000-0003-4957-8578>



**Евдокимова Вера Вячеславовна**

**Vera V. Evdokimova**

E-mail: ver.evdokimova@gmail.com

ORCID: <https://orcid.org/0000-0001-9742-5299>

**Скрелин Павел Анатольевич**

**Pavel A. Skrelin**

E-mail: p.skrelin@spbu.ru

ORCID: <https://orcid.org/0000-0002-8355-7378>

*Поступила: 15.03.2023; Одобрена: 27.06.2023; Принята: 29.06.2023.*

*Submitted: 15.03.2023; Approved: 27.06.2023; Accepted: 29.06.2023.*