

На правах рукописи



ВЕЛИЧКО Алёна Николаевна

**МЕТОДЫ И ПРОГРАММНАЯ СИСТЕМА ИНТЕГРАЛЬНОГО
АНАЛИЗА ДЕСТРУКТИВНЫХ ПАРАЛИНГВИСТИЧЕСКИХ
ЯВЛЕНИЙ В РАЗГОВОРНОЙ РЕЧИ**

Специальность: 2.3.5 – Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание ученой степени кандидата технических наук

Санкт-Петербург

2023

Работа выполнена в Федеральном государственном бюджетном учреждении науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) в лаборатории речевых и многомодальных интерфейсов (ЛРМИ).

**Научный
руководитель:**

КАРПОВ Алексей Анатольевич

доктор технических наук, профессор, главный научный сотрудник, ЛРМИ Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»

**Официальные
оппоненты:**

МЕЛЬНИКОВ Сергей Юрьевич

доктор физико-математических наук, доцент кафедры прикладной информатики и теории вероятностей Федерального государственного автономного образовательного учреждения высшего образования «Российский университет дружбы народов имени Патриса Лумумбы»

КОРЕНЕВСКИЙ Максим Львович

кандидат физико-математических наук, старший научный сотрудник научно-исследовательского департамента Общества с ограниченной ответственностью «Центр речевых технологий»

**Ведущая
организация:**

Федеральное государственное бюджетное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ТУСУР)

Защита состоится «7» декабря 2023 г. в 14 часа 00 минут на заседании диссертационного совета 24.1.206.01, созданного на базе Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) по адресу: 199178, Санкт-Петербург, 14-я линия В.О., 39, каб. 401, e-mail: dc@spcras.ru. Факс: (812) 328-44-50, тел: (812) 328-33-11.

С диссертацией и авторефератом можно ознакомиться в отделе аспирантуры (каб. 402а) Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) и на сайте <https://dc.spcras.ru/>

Автореферат разослан «22» сентября 2023 г.

Ученый секретарь
диссертационного совета 24.1.206.01
кандидат технических наук



АБРАМОВ
Максим Викторович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации. Деструктивное поведение пользователей при коммуникации в сети Интернет разрушительно влияет как на самого человека, так и на других. В связи с этим актуально выявление деструктивных (девиантных, агрессивных и враждебных) действий и обеспечение психологического комфорта пользователей социальных сетей (Станкевич М., Huang Z.).

Деструктивное поведение может проявляться как вербально (словами) или невербально (поведением). Объектом деструктивного поведения чаще всего являются эмоциональное и физическое состояния субъекта, предметы материального мира, социальные связи, коммуникация между людьми, их отношения и т.д. Под девиантным поведением чаще всего подразумевается поведение личности, которое отклоняется от общепринятого, устоявшихся и общественных норм (Майсак Н.).

В диссертационной работе рассматриваются различные деструктивные явления в поведении человека: передача ложных речевых сообщений (преднамеренная ложь/обман), депрессивные состояния, проявления агрессии к другим людям.

Существующие на данный момент автоматические программные решения по определению рассматриваемых деструктивных явлений в разговорной речи имеют следующие недостатки: 1) низкая эффективность распознавания явлений, 2) использование сложных нейросетевых архитектур, требовательных к вычислительным ресурсам, 3) большое время обучения моделей (обучение некоторых моделей может достигать до нескольких суток и даже недель), 4) отсутствие программных решений, анализирующих рассматриваемые деструктивные явления в совокупности. Таким образом, актуальна разработка программной системы, которая могла бы при низких требованиях к вычислительным ресурсам и малом количестве обучающих данных эффективно определять рассматриваемые деструктивные паралингвистические явления, в том числе с учетом взаимозависимостей между ними.

Степень разработанности темы. Автоматическое определение деструктивных паралингвистических явлений в разговорной речи является относительно новой областью, но уже существуют многочисленные работы, представленные на конференциях, семинарах и соревнованиях по компьютерной паралингвистике. Такие российские ученые как Матвеев Ю.Н., Савченко А.В., Мельников С.Ю., Шуранов Е.В., Ляксо Е.Е., Потапова Р.К., Комалова Л.Р., Мещеряков Р.В., Костюченко Е.Ю. и др., а также ряд зарубежных ученых, включая Schuller B., Batliner A., Rigoll G., Eyben F., Hirschberg J., Lefter I., Kaya H., Salah A.A., Minker W., Levitan S.I. и др. занимаются анализом паралингвистических аспектов в разговорной речи, в т.ч. задачей определения деструктивных явлений, негативных эмоций и

аффективных состояний в речи. Однако известные подходы имеют ряд ограничений: недостаток обучающих данных в виду сложностей при записи речевых корпусов, содержащих рассматриваемые паралингвистические явления; дисбаланс данных для обучения и оценивания, который является естественным из-за того, что рассматриваемые паралингвистические явления не проявляются так же часто, как нейтральное состояние и т.д.

Цель диссертационной работы: повышение эффективности автоматического определения деструктивных паралингвистических явлений в разговорной речи. Цель диссертационной работы предусматривает выполнение следующих задач:

1. Разработка новых методов автоматического определения различных деструктивных паралингвистических явлений в разговорной речи.
2. Разработка методики интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи.
3. Разработка программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи и проведение экспериментальных исследований разработанных методов и программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи.

Решаемая научная задача: разработка комплекса методов и программной системы для эффективного интегрального анализа различных деструктивных паралингвистических явлений в разговорной речи.

Важность и значимость решаемой задачи обусловлены возможностью применения фундаментальных результатов исследователями деструктивных паралингвистических явлений в разговорной речи, а также специалистами в области психологии для автоматизации первичного обследования пациентов путем бесконтактного определения деструктивных паралингвистических явлений в речи и предотвращения/уменьшения негативных последствий этих явлений.

Объектом исследования являются характеристики деструктивных паралингвистических явлений в разговорной речи.

Предметом исследования являются методы, модели и системы автоматического определения деструктивных паралингвистических явлений в разговорной речи.

Научная новизна исследования заключается в том, что:

1. Предложен комплекс методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи, отличающийся использованием оригинальных наборов анализируемых акустических признаков и применением новых многоуровневых методов (для определения ложности/истинности и агрессии в разговорной речи), а также

нейросетевого классификатора для табличных данных (для определения депрессии в разговорной речи).

2. Предложена методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в речевом сигнале диктора, отличающаяся использованием информации о взаимозависимостях между ложью, агрессией и депрессией для вычисления оценки степени выраженности рассматриваемых явлений в речи диктора.

3. Предложена архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи, отличающаяся возможностью одновременного комплексного определения лжи, агрессии и депрессии в разговорной речи с использованием предложенного комплекса методов и методики интегрального оценивания.

Теоретическая и практическая значимость работы. Теоретическая значимость заключается в разработке комплекса методов и новой методики для определения деструктивных паралингвистических явлений в разговорной речи. Разработанный комплекс методов предлагает новый подход к решению задачи эффективного определения деструктивного поведения человека по его речи. Он, прежде всего, ориентирован на универсальность, поэтому рассматривает несколько деструктивных паралингвистических явлений в разговорной речи, которые могут быть использованы как по отдельности (как самостоятельные средства для определения каждого рассматриваемого паралингвистического явления), так и в совокупности, в комплексном подходе, который учитывает взаимозависимости между рассматриваемыми паралингвистическими явлениями.

С практической точки зрения, разработанная программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи может быть использована как самостоятельно, так и в качестве системы комплексного анализа и распознавания многомодальной информации, полученной от человека. Такая система сможет учитывать не только аудио-, но и видеоинформацию, а также текстовые транскрипции речи, что может позволить улучшить результаты распознавания деструктивных паралингвистических явлений.

Методология и методы исследования. Для решения поставленных задач в работе используются и совершенствуются методы компьютерной паралингвистики, машинного обучения, глубокого обучения и искусственного интеллекта. В программной реализации системы использовались методы и алгоритмы, реализованные в открытых программных библиотеках Keras, TensorFlow, Scikit-learn, Catboost, XGBoost, LightGBM, TabNet, OpenSMILE и т.д.

Положения, выносимые на защиту:

1. Комплекс методов анализа речевого сигнала на основе оригинальных наборов акустических признаков, новых многоуровневых методов и нейросетевого классификатора для табличных данных.
2. Методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи диктора.
3. Архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи.

Соответствие диссертации научной специальности. Представленные результаты соответствуют специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Степень достоверности обеспечивается посредством проведения аналитического обзора современных исследований и методов паралингвистического анализа речи для определения деструктивных явлений, машинного и глубокого обучения; подтверждается согласованностью полученных результатов, успешной апробацией программной системы интегрального анализа деструктивных паралингвистических явлений в речевом сигнале, а также выступлениями с докладами на международных и российских научных конференциях, публикациями результатов исследований в ведущих рецензируемых изданиях.

Апробация результатов работы. Основные результаты работы докладывались и обсуждались на следующих конференциях:

1. Информационные технологии в управлении (ИТУ-2018), г. Санкт-Петербург, 2018.
2. 20th International Conference on Speech and Computer (SPECOM-2018), Leipzig, Germany, 2018.
3. 8-й междисциплинарный семинар «Анализ разговорной русской речи» (АРЗ-2019), г. Санкт-Петербург, 2019.
4. Intelligent Distributed Computing XIII (IDC 2019), г. Санкт-Петербург, 2019.
5. III международная конференция по инженерной и прикладной лингвистике «Пиотровские чтения 2019» (R. Piotrowski's Readings 2019), г. Санкт-Петербург, 2019.
6. International Conference on Computational Linguistics and Intellectual Technologies “Dialogue 2021”, г. Москва, 2021.
7. 23rd International Conference INTERSPEECH-2022, Incheon, Korea, 2022.

Результаты исследования были использованы в следующих проектах:

1. Разработка и исследование автоматической системы для выявления деструктивных паралингвистических явлений в разговорной речи,

РФФИ № 20-37-90144-Аспиранты (Величко А.Н.), руководитель Карпов А.А., 2020-2022 гг.

2. Разработка методов и программных средств оценки ложности передаваемых речевых сообщений, РФФИ № 16-37-60085-мол_а_дк, руководитель Будков В.Ю., 2016-2019 гг.

3. Разработка и исследование интеллектуальной системы для комплексного паралингвистического анализа речи, РФФИ № 18-11-00145, руководитель Карпов А.А., 2018-2020 гг.

4. Автоматическое бимодальное распознавание естественных эмоций в русской речи, РФФИ № 18-07-01407-а, руководитель Карпов А.А., 2018-2020 гг.

5. Интеллектуальная система многомодального распознавания аффективных состояний человека, РФФИ № 22-11-00321, руководитель Карпов А.А., 2022-2024 гг.

Публикации. По результатам выполнения диссертационного исследования опубликовано 14 печатных работ, включая 4 публикации в журналах из перечня рецензируемых научных изданий ВАК, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, 7 публикаций в изданиях, индексируемых в WoS/Scopus, 4 свидетельства о регистрации программ для ЭВМ в Роспатенте (см. приложение Б).

Структура и объем диссертационной работы. Диссертационная работа включает введение, три главы, заключение, список использованных источников (143 наименований) и три приложения. Основной текст изложен на 136 страницах машинописного текста, включая 16 рисунков и 15 таблиц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснованы важность и актуальность темы диссертационной работы, определена цель и сформулированы задачи, решение которых необходимо для ее достижения. Показаны научная новизна и практическая значимость работы. Дано краткое описание разработанных методов определения деструктивных паралингвистических явлений, методики интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи диктора и архитектуры программной системы. Представлены основные результаты их реализации в научно-исследовательских проектах и прикладных решениях.

В первой главе приводится аналитический обзор текущего состояния исследований в области компьютерной паралингвистики и рассматриваемых паралингвистических явлений, краткое описание существующих исследовательских данных, имеющиеся ограничения в разработке систем автоматического определения деструктивных паралингвистических явлений в разговорной речи, возможные пути решения и устранения этих ограничений, а также актуальные требования при разработке таких систем.

Областью паралингвистики является изучение различных невербальных аспектов в речи и коммуникации человека (например, интонации, эмоции, особенности произношения и характеристики голоса диктора, психофизиологические состояния, отражающиеся в устной речи). В этом случае информативными признаками являются просодические характеристики речи: частота основного тона (ЧОТ), форманты, темп речи, паузы, интонация и т.д. Скрыть эмоциональные изменения голоса нелегко.

В аналитическом обзоре рассмотрен ряд работ по определению рассматриваемых деструктивных паралингвистических явлений. Многие из этих работ имеют следующие недостатки: 1) недостаточно высокая эффективность автоматического определения деструктивных паралингвистических явлений для применения на практике, что особенно касается систем определения ложной информации в речи; 2) использование сложных нейросетевых архитектур, требовательных к вычислительным ресурсам, что ограничивает возможность экспериментальных исследований; 3) большое время обучения моделей (обучение некоторых моделей может достигать до нескольких суток и даже недель), что может как вытекать из предыдущего пункта, так и идти в совокупности с ним; 4) отсутствие решений, анализирующих рассматриваемые деструктивные явления в совокупности в рамках единого программного комплекса. Кроме того, на данный момент не существует такого речевого/многомодального корпуса, который включал бы в себя все рассматриваемые явления одновременно.

Многие исследователи сталкиваются с проблемой нехватки данных, так как речевых корпусов, содержащих рассматриваемые явления, не так много, что является естественным ввиду многих факторов: 1) задачи паралингвистики являются относительно новыми, хотя все больше ученых проявляют интерес к их исследованиям; 2) процесс сбора специализированных аудиоданных является достаточно трудоемким и времязатратным, что, безусловно, сказывается на небольших объемах речевых корпусов и длительности аудиозаписей; 3) могут быть необходимы специальные сценарии, поскольку далеко не всегда возможно провести запись речевых сообщений, содержащих необходимые паралингвистические явления, в реальных условиях.

В ходе проведения обзора прослеживается рост интереса к исследованиям, связанным с определением деструктивных паралингвистических явлений. Выявлено, что не существует универсального метода определения всех рассматриваемых паралингвистических явлений, т.к. используются как определяемые экспертом акустические признаки, так и нейросетевые методы извлечения акустических признаков, а также детерминированные и нейросетевые методы машинного обучения. Эти методы исследуются во второй главе диссертации.

Во второй главе приводится описание и исследование методов вычисления акустических признаков, машинного и глубокого обучения,

которые используются при разработке систем автоматического определения деструктивных паралингвистических явлений в разговорной речи, и приводится обоснование их выбора для программной реализации, дается подробное описание предложенных методов для решения задачи определения деструктивных паралингвистических явлений в разговорной речи. Кроме того, приведено формальное описание методики интегрального оценивания степени выраженности деструктивных паралингвистических явлений в речи диктора.

Формальная постановка задачи. Пусть имеется множество аудиозаписей $S = (s_1, \dots, s_m)$ и множество меток классов $Y = (y_1, \dots, y_m)$ этих аудиозаписей. Существует неизвестная целевая зависимость – отображение $A : X \rightarrow Y$, при этом ее метки классов известны только для векторов объектов-признаков $X = \{(x_1, y_1), \dots, (x_m, y_m)\}$ аудиозаписей конечной обучающей выборки, а X получен с использованием метода вычисления акустических признаков из аудиозаписей $F : S \rightarrow X$. Тогда требуется найти метод $A : X \rightarrow \hat{Y}$, который сможет классифицировать вектор объекта-признака x множества X . Здесь множество меток классов $Y = (y_1, \dots, y_m)$ описывает истинные значения классов объектов обучения, а множество меток классов $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_m)$ описывает значения результатов классификации.

В нашем случае необходимо найти множество методов $A = \{A_{dec}, A_{agg}, A_{depr}\}$ для методов определения ложной/истинной информации (A_{dec} – deception), агрессии (A_{agg} – aggression) и депрессии (A_{depr} – depression) в разговорной речи:

$$A_{dec} : X_{dec} \rightarrow \hat{Y}_{dec}, \quad A_{agg} : X_{agg} \rightarrow \hat{Y}_{agg}, \quad A_{depr} : X_{depr} \rightarrow \hat{Y}_{depr}, \quad (1)$$

где входные данные представлены вектором объектов-признаков $X = (x_1, x_2, \dots, x_l)$ длины l , а целевые значения меток классов \hat{y} множества \hat{Y} представлены либо бинарными значениями $\{0, 1\}$, где для множества \hat{Y}_{dec} 0 обозначает истинное высказывание, а 1 – ложное; для множества \hat{Y}_{depr} 0 обозначает отсутствие депрессии, а 1 – ее наличие; либо конечным множеством $\{0, 1, 2\}$, где для множества \hat{Y}_{agg} 0 обозначает низкий уровень агрессии или ее отсутствие, 1 – средний уровень агрессии, а 2 – высокий уровень агрессии.

Акустические признаки, имеющие корреляцию с психоэмоциональным состоянием диктора, проявляются на относительно длительных временных участках речевого сигнала, которые соответствуют как минимум одному произнесенному слову (супрасегментные признаки). Системы автоматического определения паралингвистических явлений в разговорной речи используют пространства акустических признаков большого размера, например, энергетические, спектральные и просодические признаки. Одним из наиболее популярных наборов таких признаков является OpenSMILE (для соревнований ComParE). Он содержит более 6 тысяч компонент в векторах признаков, включая математические функционалы, которые можно вычислить при помощи открытого программного инструментария OpenSMILE.

Эффективным также является использование искусственных нейронных сетей для вычисления информативных признаков из аудиосигнала, например, VGG-16, AlexNet, DenseNet и т.д.

Для решения задач классификации существует множество методов, среди которых можно выделить две категории: классические (детерминированные) методы и нейросетевые методы. Программные реализации этих методов имеются в открытых программных библиотеках, например, WEKA, Scikit-learn, Keras, Tensorflow и др.

На основе базовых методов вычисления акустических признаков и методов машинного обучения, описанных в разделах 2.3-2.4 диссертации, предлагаются новые и усовершенствованные методы, входящие в комплекс методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи.

В разделе 2.2 предложен **комплекс методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи**, представленный на рисунке 1. На вход программной системы подаются речевые записи S . Из этих данных с использованием метода вычисления интегральной совокупности акустических признаков вычисляются оригинальные наборы векторов-признаков $X = \{X_{dec}, X_{agg}, X_{depr}\}$. Полученные наборы признаков подаются на вход комплекса методов определения деструктивных паралингвистических явлений в разговорной речи A , состоящего из методов определения ложной и истинной информации A_{dec} , агрессии A_{agg} и депрессии A_{depr} в речи. Данные методы применяются в иерархическом порядке: результаты классификации методов определения ложной и истинной информации и агрессии в речи $\hat{y}_{dec}, \hat{y}_{agg}$ в виде бинарных значений $\{0, 1\}$ объединяются с вектором признаков X_{depr} , поступающим на вход метода определения депрессии в речи. Далее результаты классификации $\{\hat{y}_{dec}, \hat{y}_{agg}, \hat{y}_{depr}\}$ всех трех методов являются входными данными для предложенной методики, в результате использования которой вычисляется интегральная оценка степени выраженности деструктивных паралингвистических явлений в разговорной речи I_{int} .

Данный комплекс методов построен с использованием трех методов определения деструктивных паралингвистических явлений (лжи, агрессии, депрессии) и методики интегральной оценки деструктивных паралингвистических явлений в разговорной речи, которые описываются далее.

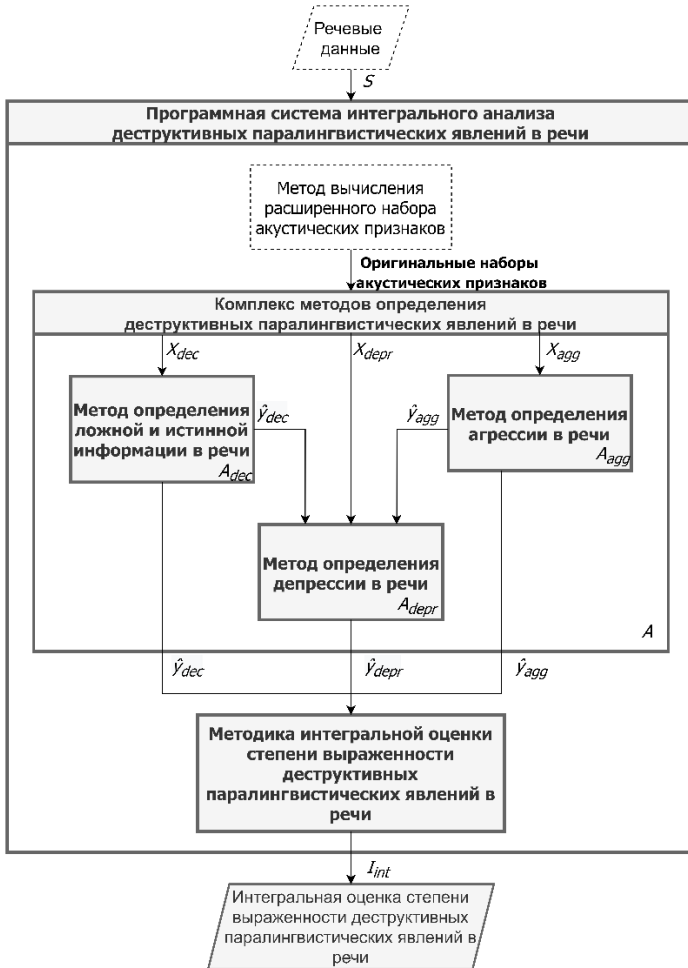


Рисунок 1 – Комплекс методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи

В разделе 2.5 предложен метод определения ложной и истинной информации в разговорной речи **HTLBbM-Deception (Hierarchical Two-level Boosting-based Method for Deception detection)**, который является иерархическим методом паралингвистического анализа речи. В общем виде схема предложенного метода HTLBbM-Deception представлена на рисунке 2.

В основе метода HTLBbM-Deception лежит двухуровневый подход на базе градиентных бустингов (Two-Level Boosting-based Method, TLBbM). Основным отличием метода HTLBbM-Deception от аналогов является

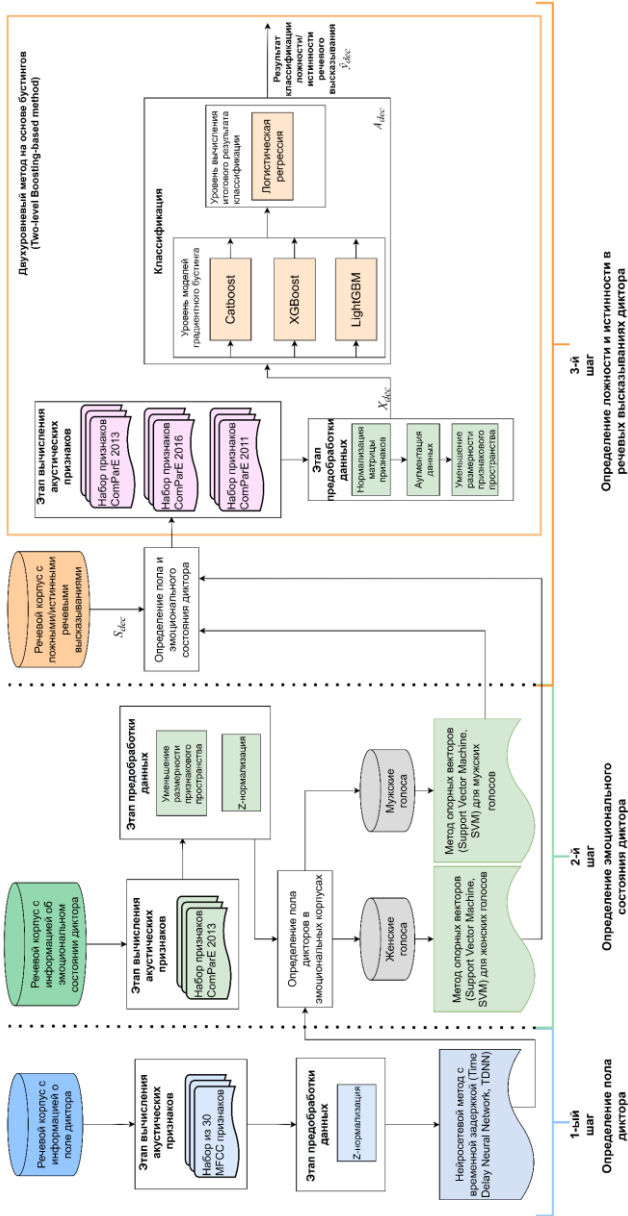


Рисунок 2 – Схема предложенного иерархического метода паралингвистического анализа речи для определения ложной и истинной информации в разговорной речи HTLbBm-Deception

иерархический подход для комплексного паралингвистического анализа речи, который включает в себя следующие шаги: 1) определение пола диктора; 2) определение эмоционального состояния диктора; 3) определение ложности/истинности речевого высказывания.

Для обучения акустической модели на шаге 3 использованы три различных реализации метода градиентного бустинга: Catboost, XGBoost, LightGBM. Для объединения результатов классификации трех методов градиентного бустинга использовался метод стекинга (stacking). Стекинг в данном методе представляет собой двухуровневый подход, где на первом уровне используются три метода градиентного бустинга, а на втором – логистическая регрессия. Логистическая регрессия объединяет результаты классификации, полученные в ходе обучения классификаторов первого уровня, и вычисляет итоговый результат классификации ложности или истинности речевого высказывания.

В разделе 2.6 предложен **метод определения депрессии в разговорной речи SBC-Depression (Single Binary Classifier for Depression Detection)**, который основан на одном машинном классификаторе с использованием нескольких наборов акустических признаков. Схема предложенного метода определения депрессии в разговорной речи SBC-Depression представлена на рисунке 3.

В этом методе также вычисляются и объединяются несколько наборов акустических признаков: eGeMAPS (состоит из 88 признаков) и DenseNet (состоит из 1024 признаков). Далее признаки подвергаются нормализации, а затем уменьшению размерности признакового пространства. Таким образом, отличием данного метода является то, что формируется оригинальный набор акустических признаков, который затем подается на вход классификатора.

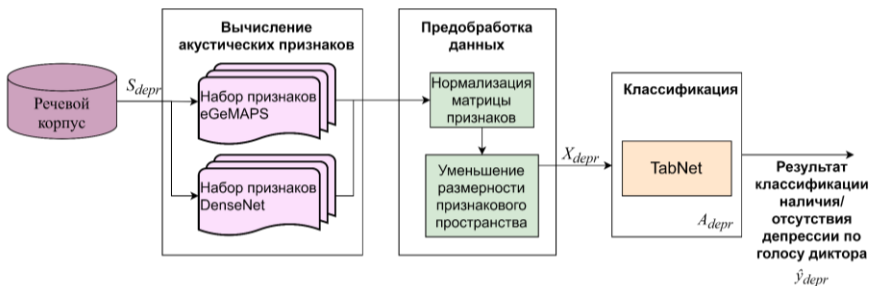


Рисунок 3 – Схема предложенного метода определения депрессии SBC-Depression

Для бинарной задачи определения депрессии используется архитектура глубокой нейронной сети TabNet. Данная архитектура была ранее разработана специально для моделирования и анализа табличных данных, которыми являются вектора признаков, подаваемые на вход классификатора. TabNet

отличает то, что метод требует небольших вычислительных ресурсов, а обучение занимает малое количество времени, что повышает ее конкурентоспособность.

В разделе 2.7 предложен метод определения агрессии в разговорной речи **EbM-Aggression (Ensemble-based Method for Aggression Detection)**, который построен на основе нескольких наборов акустических признаков и ансамбля из нескольких методов случайного леса. Он решает задачу мультиклассовой классификации состояния агрессии диктора. Схема предложенного метода для определения агрессии по речи представлена на рисунке 4. На вход метода могут подаваться обучающие данные, которые размечены согласно трем уровням агрессии: низкий, средний и высокий. Из аудиоданных вычисляются наборы акустических признаков OpenSMILE ComParE 2013, DenseNet и auDeep.

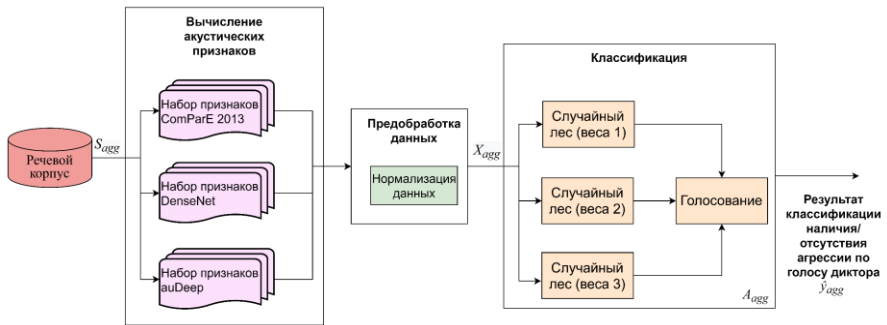


Рисунок 4 – Схема предложенного метода определения агрессии EbM-Aggression

После вычисления акустических признаков производится нормализация вектора признаков, в результате чего формируется оригинальный набор акустических признаков, что является отличием данного метода от аналогичных. Далее применяется ансамбль из методов классификации случайного леса с различными значениями весов для классов и голосованием по большинству. В качестве весов для методов случайного леса эмпирическим путем для каждого набора признаков подобраны три весовых коэффициента.

Методы случайного леса обучаются на всех трех наборах признаков, имеющих различные веса, т.е. внутри одного блока случайного леса на рисунке используются три метода для каждого набора признаков. Голосование по большинству происходит сначала между методами с одинаковыми весами (т.е. между методами, обученными на разных наборах признаков, имеющих одинаковые веса), а затем между методами с различными весами.

Кроме того, особенностью данного ансамбля является и то, что, несмотря на ансамблирование методов случайного леса (комплексирование), он не требует большого количества вычислительных и временных ресурсов.

В разделе 2.8 предложена **методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи.**

Для интегрального оценивания наличия паралингвистических деструктивных явлений в разговорной речи предложена методика, предусматривающая определение интегрального показателя:

$$I_{int} = w_{agg} \cdot \hat{y}_{agg} + w_{dec} \cdot \hat{y}_{dec} + w_{depr} \cdot \hat{y}_{depr}, \quad (2)$$

где: I_{int} – интегральная оценка, w_i – весовые коэффициенты (веса) значимости деструктивных явлений множества весовых коэффициентов значимости $W = (w_1, \dots, w_3)$, \hat{y}_i – результат классификации методов определения деструктивных явлений множества $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_3)$. Методика вычисления интегральной оценки представлена в виде схемы на рисунке 5 (пример анализа всех трех явлений).

Имеется три частных показателя \hat{y} , на основе которых вычисляется интегральная оценка: ложность/истинность высказывания (\hat{y}_{dec}), наличие агрессии в высказывании (\hat{y}_{agg}) и наличие состояния депрессии у диктора (\hat{y}_{depr}).

Входными данными методики являются результаты классификации \hat{Y} , веса значимости явлений W , удовлетворяющие условию $w_{dec} + w_{agg} + w_{depr} = 1$, и начальное значение итоговой оценки $I_{int} = 0$.

Входные данные обрабатываются с использованием ряда формальных правил, которые основаны на экспертных оценках и теоретическом базисе корреляции между рассматриваемыми паралингвистическими явлениями (подробное описание в разделе 1.1 диссертации). Затем обработанные данные подаются на вход блока расчета интегральной оценки I_{int} , где происходит вычисление интегральной оценки I_{int} с использованием формулы 2.

Выходными данными предложенной методики является интегральная оценка I_{int} , десятичные значения которой могут варьироваться в диапазоне $[0, 1]$. Предложенная методика позволяет анализировать как все три результата классификации методов определения деструктивных паралингвистических явлений, так и отсутствие результата классификации одного или двух методов с использованием коэффициента перераспределения весов (подробнее в разделе 2.8).

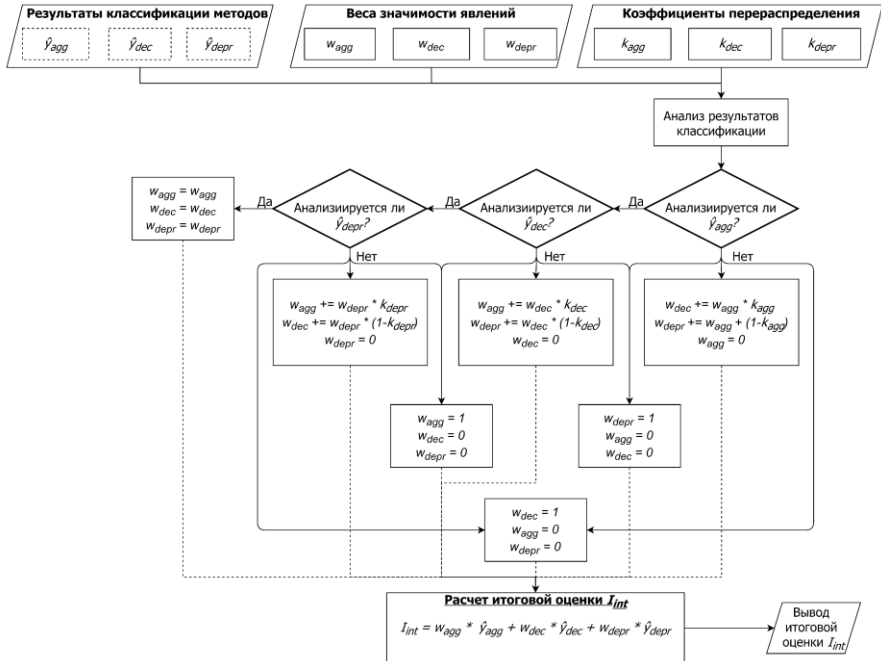


Рисунок 5 – Методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в речи диктора

В третьей главе описывается архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи DesBDet, использованные открытые программные библиотеки и графический пользовательский интерфейс. Приводится подробная информация об исследовательских данных, показатели оценивания эффективности работы предложенных методов, результаты экспериментальных исследований и сравнение предложенных методов с аналогами, известными в литературе.

Разработана программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи, получившая название DesBDet (Destructive Behaviour Detection) представлена на рисунке 6. Она построена по модульному принципу и включает в себя несколько независимых программных модулей: 1) предобработка исходных данных; 2) вычисление наборов акустических признаков из аудиоданных; 3) обработка полученного вектора акустических признаков с использованием методов нормализации (и аугментации) данных, а также уменьшения размерности признакового пространства; 4) получение результата классификации от каждого модуля; 5) вычисление интегральной оценки

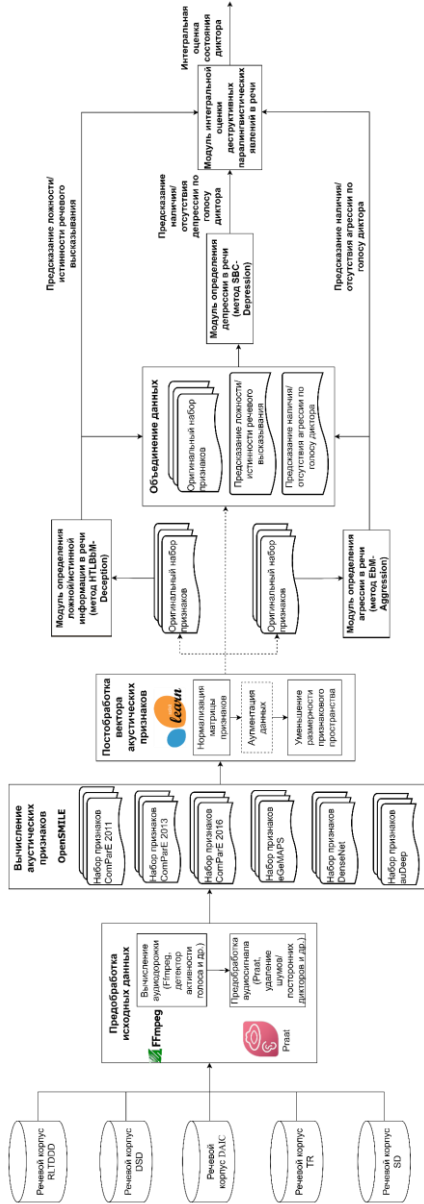


Рисунок 6 – Архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи

анализа деструктивных явлений в речи диктора.

Определение деструктивных явлений в разговорной речи происходит в иерархическом порядке: сначала параллельно работают модуль определения ложной/истинной информации и модуль определения агрессии, а затем их результаты классификации в бинарном виде $\{0, 1\}$ добавляются в качестве дополнительных признаков в признаковое пространство, которое подается на вход модуля определения депрессии. При этом гипотезы о ложности и агрессии приобретают более высокий вес по сравнению с остальными признаками. После чего все три результата классификации трех модулей выступают в качестве входных данных в методике интегральной оценки степени выраженности деструктивных паралингвистических явлений в речи диктора.

При разработке программной системы DesBDet использован объектно-ориентированный язык программирования Python версии 3.8, а также следующие открытые программные библиотеки: Scikit-learn, Catboost, XGBoost, LightGBM, Keras, Tensorflow. Графический интерфейс программной системы разработан с использованием библиотеки PyQt6.

В качестве экспериментальных данных для оценивания предложенных методов использованы несколько речевых и многомодальных корпусов, содержащих рассматриваемые деструктивные паралингвистические явления: 1) корпус DSD (1059 аудиозаписей) и корпус RLTDDD (121 запись) для метода определения ложной и истинной информации; 2) корпус aGender (65364 записей) для метода определения половой принадлежности, корпуса EmoDB (535 записей) и Ruslana (3661 записей) для определения эмоционального состояния диктора; 3) корпус DAIC (219 записей) для метода определения депрессии; 4) корпуса SD и TR (893 записи) для метода определения агрессии.

Все экспериментальные исследования были проведены с использованием 5-кратной перекрестной валидации. Для количественного оценивания эффективности работы предложенных методов использовались показатели: F1-мера и невзвешенная средняя полнота (UAR).

F1-мера ($F1$) основывается на показателях матрицы спутывания и является гармоническим средним между полнотой Recall и точностью Precision классификации:

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall}, \quad (3)$$

$$UAR = \frac{1}{k} \sum_{i=1}^k \frac{N_c^{(i)}}{N_0^{(i)}}, \quad (4)$$

где $N_c^{(i)}$ - число верно распознанных элементов i -го класса, $N_0^{(i)}$ - общее число объектов в i -ом классе, N - общее число объектов, а k - число классов.

Результаты экспериментальных исследований и сравнение с основными аналогами приведены в таблице 1. Подробнее экспериментальные исследования представлены в разделах 3.5-3.7 диссертации. Среди известных

работ по определению ложной и истинной информации можно выделить работы [Mendels G. и др., 2017] и [Montacié C. и др., 2016], достигшие значений $F1$ -меры = 63,9% и $UAR = 74,9\%$, соответственно. Среди методов определения депрессии работы [Xezonaki D. и др., 2020] и [Huang Zh. и др., 2020] достигли значений $F1$ -меры = 70,3%, $UAR = 70,3\%$, и $UAR = 88,0\%$, соответственно. Среди методов определения агрессии можно выделить работу [Schuller B. и др., 2021] с результатом $UAR = 72,2\%$ и работу [Egas-López J.V. и др., 2021] с $UAR = 77,8\%$.

Таблица 1 – Результаты экспериментальных исследований предложенных методов

Метод	Результаты классификации, %
Иерархический метод определения ложной/истинной информации в речи HTLBbM-Deception	$F1 = 88,4 \pm 1,5$ $UAR = 88,5 \pm 1,3$
Метод определения депрессии в разговорной речи SBC-Depression	$F1 = 64,0 \pm 0,7$ $UAR = 60,0 \pm 0,5$
Метод определения агрессии в разговорной речи EbM-Aggression	$F1 = 78,0 \pm 1,5$ $UAR = 76,5 \pm 1,3$

На основе проведенных экспериментальных исследований можно сделать вывод, что некоторые предложенные методы определения деструктивных паралингвистических явлений в разговорной речи в ряде случаев превосходят известные аналоги по эффективности. Также выявлено, что увеличение количества данных для обучения и тестирования как естественным (использование нескольких исследовательских баз данных), так и искусственным путем (аугментация данных), оказывает положительное влияние на итоговый результат распознавания. Можно также отметить, что применение ансамблевого подхода и многоуровневых методов классификации оказывается успешным благодаря повышенной устойчивости таких подходов к переобучению и их способности к более эффективному выявлению зависимостей в признаковом пространстве.

На основе полученных результатов классификации можно вычислить интегральную среднюю $F1$ -меру ($F1_{int}$) и интегральную невзвешенную среднюю полноту (UAR_{int}) для оценки предложенной программной системы:

$$F1_{int} = \frac{1}{3} \sum_{i=1}^3 F1_i = 76,8\%, \quad (5)$$

где $F1_i$ – $F1$ -мера i -го метода из 3 методов,

$$UAR_{int} = \frac{1}{3} \sum_{i=1}^3 UAR_i = 75,0\%, \quad (6)$$

где UAR_i – UAR i -го метода из 3 методов.

В заключении диссертации приведены основные научно-практические результаты, полученные в ходе диссертационного исследования, и их основные характеристики.

ЗАКЛЮЧЕНИЕ

В диссертационной работе сформулирована и решена новая научно-техническая задача повышения эффективности автоматического определения деструктивных паралингвистических явлений в разговорной речи. Решенная задача имеет важное значение для совершенствования методов и программных решений, используемых при паралингвистическом анализе разговорной речи в условиях недостаточного количества обучающих данных и их дисбалансе, а также при ограничениях в вычислительных ресурсах.

В процессе выполнения диссертационного исследования получены новые научные результаты, составляющие **итоги** исследования:

1. Предложен и исследован комплекс методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи, включающий в себя: 1) метод определения ложности/истинности в разговорной речи HTLBbM-Deception; 2) метод определения депрессии в разговорной речи SBC-Depression; 3) метод определения агрессии в разговорной речи EbM-Aggression.

2. Разработана методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи диктора, которая учитывает результаты классификации каждого из трех методов и на их основе вычисляет интегральную оценку на s использованием ряда правил.

3. Разработана архитектура и прототип программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи DesBDet, проведены экспериментальные исследования программных реализаций методов определения деструктивных паралингвистических явлений и получены следующие количественные результаты: 1) метод определения ложности/истинности в речевых высказываниях достигает F1-меры = 88,4%; 2) метод определения депрессивного состояния в речевых высказываниях достигает F1-меры = 64,0%; 3) метод агрессии в речевых высказываниях достигает UAR = 76,5%; 4) интегральные F1-мера и невзвешенная средняя полнота для комплекса методов: $F1_{int} = 76,8\%$, $UAR_{int} = 75,0\%$.

Рекомендации. Предложенные методы и разработанная программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи могут быть использованы как самостоятельно, так и в составе сложной системы комплексного анализа и распознавания речи человека. Такая система сможет учитывать не только аудио-, но также и видеоинформацию, и текстовые транскрипции, что, согласно исследованиям аналогичных многомодальных систем, может улучшить результаты распознавания деструктивных паралингвистических явлений. Ограничения предложенного комплекса методов связаны с относительно небольшим

количеством исследовательских данных и дисбалансом классов в них, а также ограниченными доступными вычислительными ресурсами.

Основным эффектом от использования предлагаемого комплекса методов анализа речевого сигнала для определения деструктивных паралингвистических явлений в разговорной речи является психологический комфорт пользователей при взаимодействии в сети Интернет. Кроме того, разработанная программная система на основе предложенного комплекса методов может применяться для первичной оценки состояния пациентов при консультации с медицинскими специалистами в качестве одного из методов оценки психологического состояния пациента наряду с классическими подходами первичной оценки пациента (опросники, тесты и т.д.). Результаты диссертационного исследования были успешно апробированы в СПИИРАН, входящем в СПб ФИЦ РАН, и ООО «Первый психотерапевтический» (см. раздел 3.8 текста диссертации).

Перспективы дальнейшей разработки темы состоят в развитии предложенного комплекса методов и программной реализации в нескольких направлениях, в частности, увеличение количества модальностей, а именно анализ не только акустических характеристик речи, но и смысловой ее составляющей (текста высказываний), а также анализ визуальных образов (видеоданных лица человека). Развитие программной реализации возможно в направлении использования новейших программных библиотек, которые смогут позволить повысить как скорость исполнения программного кода, так и улучшить дизайн и повысить удобство использования графического пользовательского интерфейса.

Полученные результаты соответствуют специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в ведущих рецензируемых научных журналах и изданиях из перечня рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук:

1. Величко А.Н. Метод анализа речевого сигнала для автоматического определения агрессии в разговорной речи // Вестник ВГУ. Системный анализ и информационные технологии. – 2022. – № 4. – С. 180-188.
2. Величко А.Н., Карпов А.А. Аналитический обзор систем автоматического определения депрессии по речи // Информатика и автоматизация. – 2021. – № 3 (20). – С. 497-529.
3. Величко А.Н., Карпов А.А., Будков В.Ю. Исследование методов классификации для автоматического определения истинной или ложной информации в речевых сообщениях // Научный вестник НГТУ. – 2018. – № 3. – С. 21-32.
4. Двойникова А.А., Маркитантов М.В., Рюмина Е.В., Уздяев М.Ю., Величко А.Н., Рюмин Д.А., Ляко Е.Е., Карпов А.А. Анализ информационного и математического обеспечения для распознавания аффективных состояний человека // Информатика и автоматизация. – 2022. – № 6 (21). – С. 1097-1144.

Публикации в ведущих российских и иностранных научных изданиях, входящих в перечни WoS/Scopus:

1. Velichko A., Markitantov M., Kaya H., Karpov A. Complex Paralinguistic Analysis of Speech: Predicting Gender, Emotions and Deception in a Hierarchical Framework // Proceedings of the International Conference INTERSPEECH-2022. 2022. P. 4735-4739.
2. Velichko A.N., Karpov A.A. Automatic Detection of Deceptive and Truthful Paralinguistic Information in Speech using Two-Level Machine Learning Model // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2021". 20 (27). 2021. P. 698-704.
3. Verkholyak O., Dresvyanskiy D., Dvoynikova A., Kotov D., Ryumina E., Velichko A., Mamontov D., Minker W., Karpov A. Ensemble-Within-Ensemble Classification for Escalation Prediction from Speech // Proceedings of the International Conference INTERSPEECH-2021. 2021. P. 481-485.
4. Velichko A., Karpov A. A Study of Data Scarcity Problem for Automatic Detection of Deceptive Speech Utterances // Proceedings of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019) Saint Petersburg, Russia, November 27, 2019. Published on CEUR-WS. Vol. 2552. 2020. P. 38-46.
5. Velichko A., Budkov V., Kagirov I., Karpov A. Applying Ensemble Learning Techniques and Neural Networks to Deceptive and Truthful Information Detection Task in the Flow of Speech // Intelligent Distributed Computing XIII. IDC 2019. Studies in Computational Intelligence, Springer, Cham. Vol. 868. 2019. P. 477-482.
6. Levonevskii D., Shumskaya O., Velichko A., Uzdiaev M., Malov D. Methods for Determination of Psychophysiological Condition of User within Smart Environment Based on Complex Analysis of Heterogeneous Data // Proceedings of 14th International Conference on Electromechanics and Robotics "Zavalishin's Readings". Smart Innovation, Systems and Technologies. Springer, Cham. Vol. 154. 2019. P. 511-523.

7. Velichko A.N., Budkov V.Y., Kagirow I. A., Karpov A.A. Comparative Analysis of Classification Methods for Automatic Deception Detection in Speech // Proceedings of the 20th International Conference on Speech and Computer SPECOM-2018. Springer, LNAI. Vol. 11096. 2018. P. 737-746.

Публикации в других изданиях:

1. Величко А.Н., Будков В.Ю. Разработка прототипа системы автоматического определения ложной и истинной информации в речи // Материалы семинара «Анализ разговорной русской речи» (АРЗ-2019), Санкт-Петербург, 2019, С. 17-20.

2. Величко А.Н., Карпов А. А., Будков В. Ю. Аналитический обзор речевых корпусов для систем определения ложных речевых сообщений // Материалы конференции «Информационные технологии в управлении» ИТУ-2018 в рамках МКПУ-2018, Санкт-Петербург, 2018, С. 638-642.

3. Dvoynikova A., Markitantov M., Ryumina E., Uzdiaev M., Velichko A., Kagirow I., Kipyatkova I., Lyakso E., Karpov A. An analysis of automatic techniques for recognizing human's affective states by speech and multimodal data. Proceedings of the 24th International Congress on Acoustics ICA-2022. 2022. P. 22-33.

Свидетельства о государственной регистрации программ для ЭВМ:

1. Величко А.Н. Программное обеспечение для определения депрессивного состояния по речи человека. Свидетельство № 2021680548. Зарегистрировано в Реестре программ для ЭВМ 13.12.2021.

2. Величко А.Н., Верхоляк О.В., Карпов А.А. Программная система для распознавания эмоций в речи (ProSpER – Program for Speech Emotion Recognition). Свидетельство № 2020664234. Зарегистрировано в Реестре программ для ЭВМ 10.11.2020.

3. Верхоляк О.В., Маркитантов М.В., Величко А.Н., Кипяткова И.С., Карпов А.А. Программная система комплексного анализа паралингвистических явлений в речи (ComPAS - Complex Paralinguistic Analysis of Speech). Свидетельство № 2020664233. Зарегистрировано в Реестре программ для ЭВМ 10.11.2020.

4. Величко А.Н., Будков В.Ю., Карпов А.А. Программная система для автоматического определения ложной и истинной информации в речи. Свидетельство № 2018662956. Зарегистрировано в Реестре программ для ЭВМ 17.10.2018.

Автореферат диссертации

Величко Алёна Николаевна

**МЕТОДЫ И ПРОГРАММНАЯ СИСТЕМА ИНТЕГРАЛЬНОГО
АНАЛИЗА ДЕСТРУКТИВНЫХ ПАРАЛИНГВИСТИЧЕСКИХ
ЯВЛЕНИЙ В РАЗГОВОРНОЙ РЕЧИ**

Текст автореферата размещен на сайтах:

Высшей аттестационной комиссии при Министерстве науки и высшего
образования Российской Федерации
<https://vak.minobrnauki.gov.ru>

Федерального государственного бюджетного учреждения науки
Санкт-Петербургского Федерального исследовательского центра
Российской академии наук (СПб ФИЦ РАН)
<https://dc.spcras.ru/>

Подписано в печать «5» сентября 2023 г.
Формат 60x84 1/16. Бумага офсетная. Печать офсетная.
Усл.печ.л. 1,0. Тираж 100 экз.
Заказ № __