



Методика и программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи

А. Н. Величко^а, научный сотрудник, orcid.org/0000-0002-8503-8512

А. А. Карпов^а, доктор техн. наук, профессор, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su

^аСанкт-Петербургский Федеральный исследовательский центр РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

Введение: в последнее время растет число работ, в которых разрабатываются системы паралингвистического анализа различных деструктивных явлений в речи. Однако лишь немногие работы рассматривают несколько явлений в совокупности, не смотря на то, что существуют теоретические и практические работы, в которых описываются взаимосвязи между различными паралингвистическими явлениями. **Цель:** разработать методику и программную систему интегрального анализа деструктивных паралингвистических явлений (лжи, агрессии, депрессии) в разговорной речи. **Результаты:** предложена методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи диктора. Данная методика учитывает результаты классификации каждого из трех методов определения рассматриваемых явлений (лжи, агрессии и депрессии) и на их основе вычисляет интегральную оценку с использованием ряда правил. Представленная архитектура программной системы включает в себя комплекс методов для определения рассматриваемых явлений, а также предложенную методику. На основе полученных в экспериментальных исследованиях результатов классификации деструктивных паралингвистических явлений для оценки программной системы можно вычислить интегральную среднюю F1-меру и интегральную невзвешенную среднюю полноту, которые составляют 76,8 и 75,0 % соответственно. **Практическая значимость:** использование представленной архитектуры программной системы позволяет получить интегральную оценку степени выраженности деструктивных паралингвистических явлений в речи диктора. Такая программная система может применяться для первичной оценки состояния пациентов при консультации с медицинскими специалистами в качестве одного из методов оценки психологического состояния пациентов наряду с классическими подходами (опросниками, тестами и т. д.).

Ключевые слова – анализ речи, речевые технологии, компьютерная паралингвистика, деструктивные явления, автоматическое определение лжи в речи, автоматическое определение агрессии в речи, автоматическое определение депрессии в речи.

Для цитирования: Величко А. Н., Карпов А. А. Методика и программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи. *Информационно-управляющие системы*, 2023, № 4, с. 2–11. doi:10.31799/1684-8853-2023-4-2-11, EDN: FHUUWJ

For citation: Velichko A. N., Karpov A. A. Methods and a software system for integrative analysis of destructive paralinguistic phenomena in colloquial speech. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2023, no. 4, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2023-4-2-11, EDN: FHUUWJ

Введение

Паралингвистика изучает различные невербальные аспекты речи и коммуникации человека, а компьютерная паралингвистика, в свою очередь, использует автоматизированные средства для усовершенствования систем анализа паралингвистических явлений в речи человека. В настоящее время существует множество подходов к решению задачи определения деструктивных паралингвистических явлений в речи человека по отдельным явлениям лжи, агрессии и депрессии. Эти подходы представлены как на соревнованиях по компьютерной паралингвистике (например, INTERSPEECH Computational Paralinguistics Challenge, Audio/Visual Emotion Challenge and Workshop и др.), так и вне рамок этих соревнований (например, [1–3]). Однако большинство известных подходов имеют ряд ограничений: недостаток обучающих данных ввиду сложности

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. Кроме того, существующие на данный момент программные решения по определению рассматриваемых деструктивных явлений в разговорной речи имеют следующие недостатки: 1) низкую эффективность распознавания явлений; 2) использование сложных нейросетевых архитектур, требовательных к вычислительным ресурсам; 3) большое время обучения моделей (обучение некоторых моделей может длиться нескольких суток и даже недель); 4) отсутствие программных решений, анализирующих рассматриваемые деструктивные явления в совокупности.

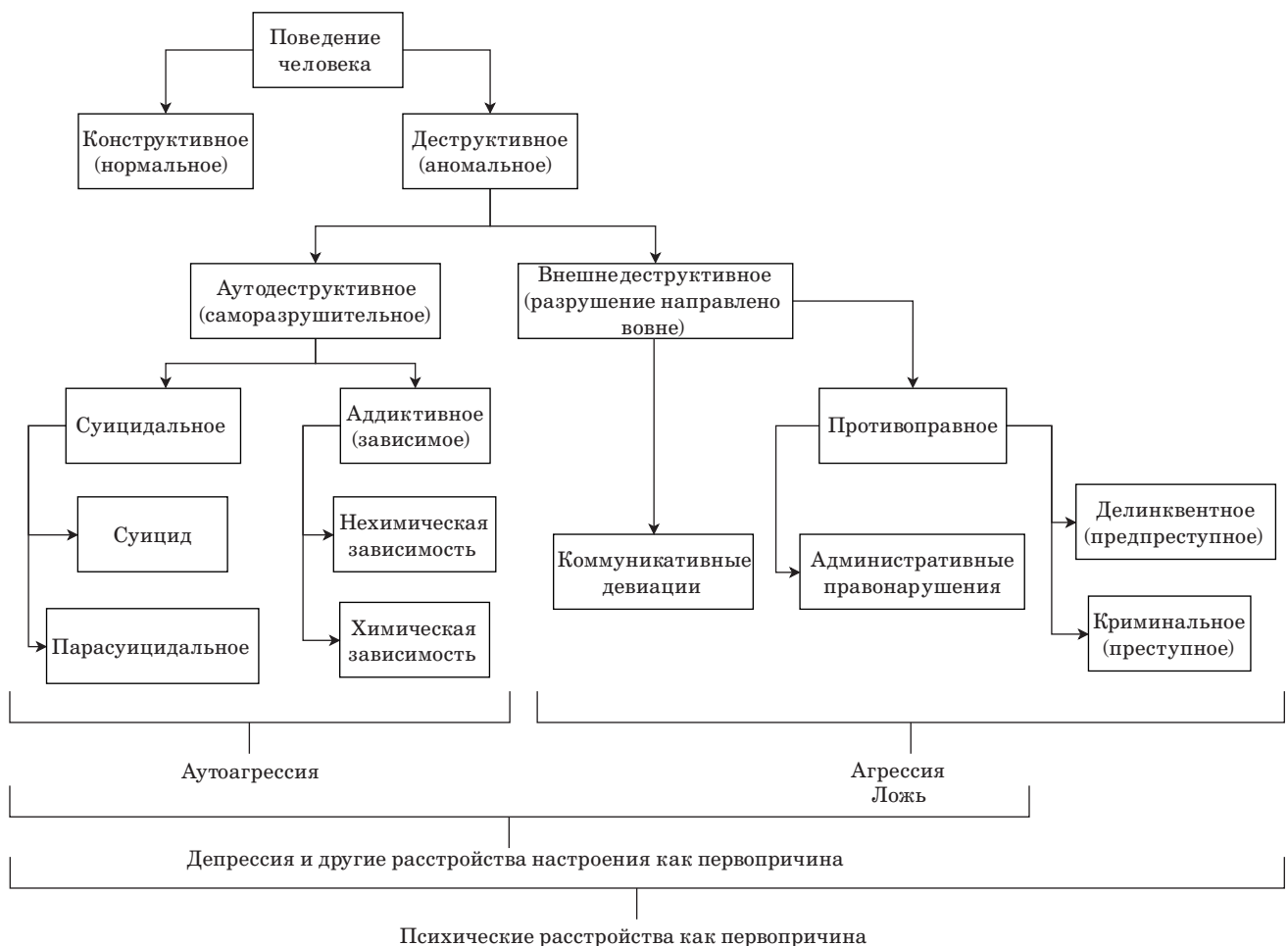
Таким образом, актуальна разработка программной системы, которая могла бы при низ-

ких требованиях к вычислительным ресурсам и малом количестве обучающих данных эффективно определять рассматриваемые деструктивные паралингвистические явления, в том числе с учетом взаимозависимостей между ними. Такая программная система может быть полезна при внедрении в медицинской сфере, например при скрининге расстройств тревожно-депрессивного характера в первичных звеньях здравоохранения, на которые падает основная нагрузка при том, что именно в них наблюдается нехватка ресурсов.

Современное состояние исследований

Поведение человека может быть как конструктивным, так и деструктивным. На основе матрицы социальных девиаций [4] можно обозначить место рассматриваемых в работе деструктивных явлений: лжи, агрессии и депрессии (рис. 1). Стоит отметить, что области «Психические расстрой-

ства как первопричина» и «Депрессия как первопричина» могут как являться первопричиной деструктивных явлений, так и не являться ею. То есть деструктивные явления не обязательно могут проявляться под действием какого-либо психического заболевания или расстройства настроения (в том числе депрессии). На рисунке видно, что некоторые конечные узлы дерева относятся к нескольким деструктивным явлениям, что указывает на комплексность природы этих явлений и может означать корреляцию между ними (которая подтверждается рядом теоретических и практических работ, рассматриваемых далее). К примеру, в работах [5, 6] авторы указывают на тесную связь между гневом и депрессией, а в работе [7] эта связь изучалась с помощью межиндивидуального и внутрииндивидуального анализа. В работе [8] выявлена сильная корреляция между гневом, тревогой, депрессией и негативными эмоциями, а в работе [9] отмечена связь между посттравматическим стрессовым расстройством, депрессией,



■ **Рис. 1.** Иерархическая систематизация деструктивных явлений на основе матрицы социальных девиаций (по матрице [4])
 ■ **Fig. 1.** Hierarchical systematization of destructive phenomena based on the social deviations matrix (based on the matrix [4])

враждебностью и гневом. Связь между агрессией и ложью исследовалась в работах [10, 11], где было определено, что случайный гнев способствует неэтичному поведению и приводит к более явному проявлению имплицитных установок.

Под ложью подразумевается преднамеренный акт введения собеседника в заблуждение путем передачи неверной или вводящей в заблуждение информации [12]. Основная часть работ по автоматическому определению ложной/истинной информации в речи представлена на соревнованиях по компьютерной паралингвистике INTERSPEECH ComParE в 2016 г. Для оценки качества работы систем используется показатель невзвешенной средней полноты UAR (Unweighted Average Recall). Среди работ, представленных на этих соревнованиях, стоит упомянуть [1, 13, 14], а также работы вне соревнований [15, 16]. Для решения поставленной задачи в большинстве работ используются экспертные акустические признаки и детерминированные методы машинного обучения.

Согласно данным ВОЗ [17], депрессия является распространенным психическим расстройством и одной из основных болезней, которые приводят к ухудшению жизнедеятельности человека. Задача определения депрессии была неоднократно предложена на соревнованиях AVEC (Audio-Visual Emotion Challenge). Для оценки качества работы систем используются показатели CCC (Concordance Correlation Coefficient), RMSE (Root Mean Squared Error), F1-мера и др. Среди лучших работ за последние годы можно отметить [18–20], представленные на соревнованиях AVEC в 2019 г., и [2, 21–24], представленные вне соревнований. В них авторы используют как экспертные, так и нейросетевые акустические признаки, а для классификации и регрессии в большинстве работ используются нейросетевые методы машинного обучения либо со сверткой, либо с механизмами внимания.

Под термином агрессия в европейской культуре подразумевается деструктивное поведение, которое является мотивированным и противоречит нормам сосуществования людей. Такое поведение может быть направлено как вовне (нанесение вреда или психологического дискомфорта окружающим людям, животным, предметам), так и на себя (самоповреждение, самобичевание). Среди работ по определению агрессии за последние годы можно выделить некоторые, представленные на соревнованиях INTERSPEECH ComParE в 2021 г. Для оценки качества работы систем используется показатель UAR. Авторы лучших работ [3, 25–27] использовали как экспертные, так и нейросетевые акустические признаки, а для классификации применяли как детерминированные, так и нейросетевые методы машинного обучения.

Таким образом, можно заключить, что задача автоматического определения деструктивных

паралингвистических явлений представляется комплексной, и не существует универсального подхода к определению всех паралингвистических явлений, рассматриваемых в работе. Далее перейдем к формальной постановке задачи.

Описание разработанного подхода. Формальная постановка задачи классификации объектов (аудиозаписей)

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 аудиозаписей конечной обучающей выборки $X = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_m, y_m)\}$, 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 \mathbf{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 $A = \{A_{dec}, A_{agg}, A_{depr}\}$, состоящее из методов определения ложной/истинной информации, агрессии и депрессии в речи (A_{dec} — deception, A_{agg} — aggression, A_{depr} — depression):

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

где входные данные представлены вектором объектов-признаков множества $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l)$ 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.

Методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи

Для интегрального оценивания наличия и степени выраженности деструктивных пара-

лингвистических явлений в разговорной речи предложена методика (рис. 2), основанная на вычислении формулы

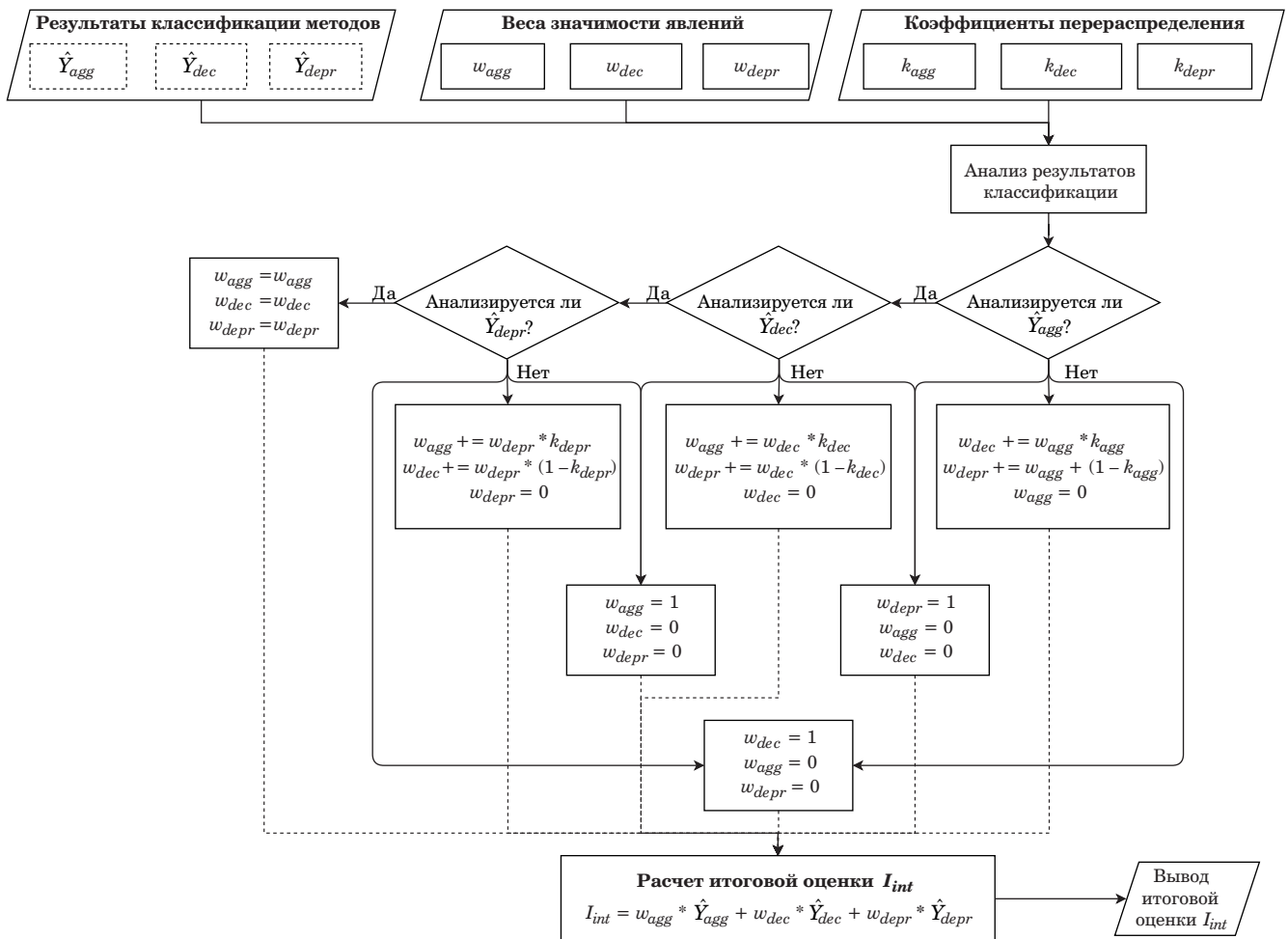
$$I_{int} = I_{agg} + I_{dec} + I_{depr} = w_{agg} \cdot \hat{y}_{agg} + w_{dec} \cdot \hat{y}_{dec} + w_{depr} \cdot \hat{y}_{depr}, \quad (2)$$

где I_{int} – интегральная оценка, $I_{int} = (I_1, \dots, I_3)$, w_i – весовые коэффициенты (веса) значимости деструктивных явлений множества весовых коэффициентов значимости $W = (w_1, \dots, w_3)$; \hat{y}_i – результаты классификации методов определения деструктивных паралингвистических явлений множества $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_3)$.

В нашем случае имеется три частных результата классификации \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 настоящей методики являются результаты классификации \hat{Y} , веса значимости явлений W , интегральная оценка степени выраженности деструктивных паралингвистических явлений I_{int} в речи диктора.

Результаты классификации методов определения ложности/истинности и депрессии могут принимать бинарные значения $\{0, 1\}$, а результаты классификации метода определения агрессии могут принимать значения из множества $\{0, 1, 2\}$. Для удобства вычислений их необходимо привести к бинарному виду. Результат со значением 0 означает отсутствие агрессии в речевом высказывании, а значения 1 и 2, означающие средний и высокий уровни агрессии соответственно, преобразуются в значение 1 – наличие агрессии в речи. Кроме того, результатами могут быть не только результаты классификации, усредненные по всему речевому сообщению, но и результаты классификации сегментов записи, а также вероятности принадлежности записи или



■ **Рис. 2.** Методика интегрального оценивания степени выраженности деструктивных паралингвистических явлений в речи диктора
 ■ **Fig. 2.** An approach for integral evaluation of destructive paralinguistic phenomena severity in speech

сегментов к классам. Значения весов значимости удовлетворяют условию $w_{dec} + w_{agg} + w_{depr} = 1$. Входные данные обрабатываются с использованием ряда формальных правил, которые основаны на экспертных оценках и теоретическом базисе корреляции между рассматриваемыми паралингвистическими явлениями. Затем обработанные данные подаются на вход блока расчета интегральной оценки I_{int} для ее вычисления (2).

Предложенная методика позволяет анализировать как все три результата классификации методов определения деструктивных паралингвистических явлений, так и отсутствие результатов классификации одного или двух методов с использованием коэффициента перераспределения весов. В случае, когда не анализируются какие-либо результаты классификации, коэффициент перераспределения весов k равен $1/2$ для депрессии и $1/3$ для ложности и агрессии.

Выходными данными предложенной методики является интегральная оценка I_{int} , десятичные значения которой могут варьироваться в диапазоне $[0, 1]$. При этом значения интегральной оценки можно разделить по уровням выраженности деструктивных паралингвистических явлений: значения до 0,30 означают низкий уровень, от 0,31 до 0,6 — средний уровень, выше 0,61 — высокий уровень.

Архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи

Анализ работ на предмет наличия корреляций между рассматриваемыми явлениями показал, что существуют связи между всеми рассматриваемыми явлениями, а значит, эти связи должны учитываться при разработке программной системы.

Разработанная программная система интегрального анализа деструктивных паралингвистических явлений в разговорной речи (рис. 3) построена по модульному принципу. Ее архитектура включает в себя несколько независимых программных модулей (блоков): 1) предобработки исходных данных; 2) вычисления наборов акустических признаков из аудиоданных; 3) обработки полученного вектора акустических признаков с использованием нормализации (и аугментации) данных, а также уменьшения размерности признакового пространства; 4) получения итогового результата классификации от каждого модуля; 5) вычисления интегральной оценки анализа деструктивных явлений в речи диктора. Стоит отметить, что обучение происходит в иерархическом порядке: сначала парал-

лельно 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 работы модулей выступают в качестве входных данных в методике интегральной оценки степени выраженности деструктивных паралингвистических явлений в речи диктора.

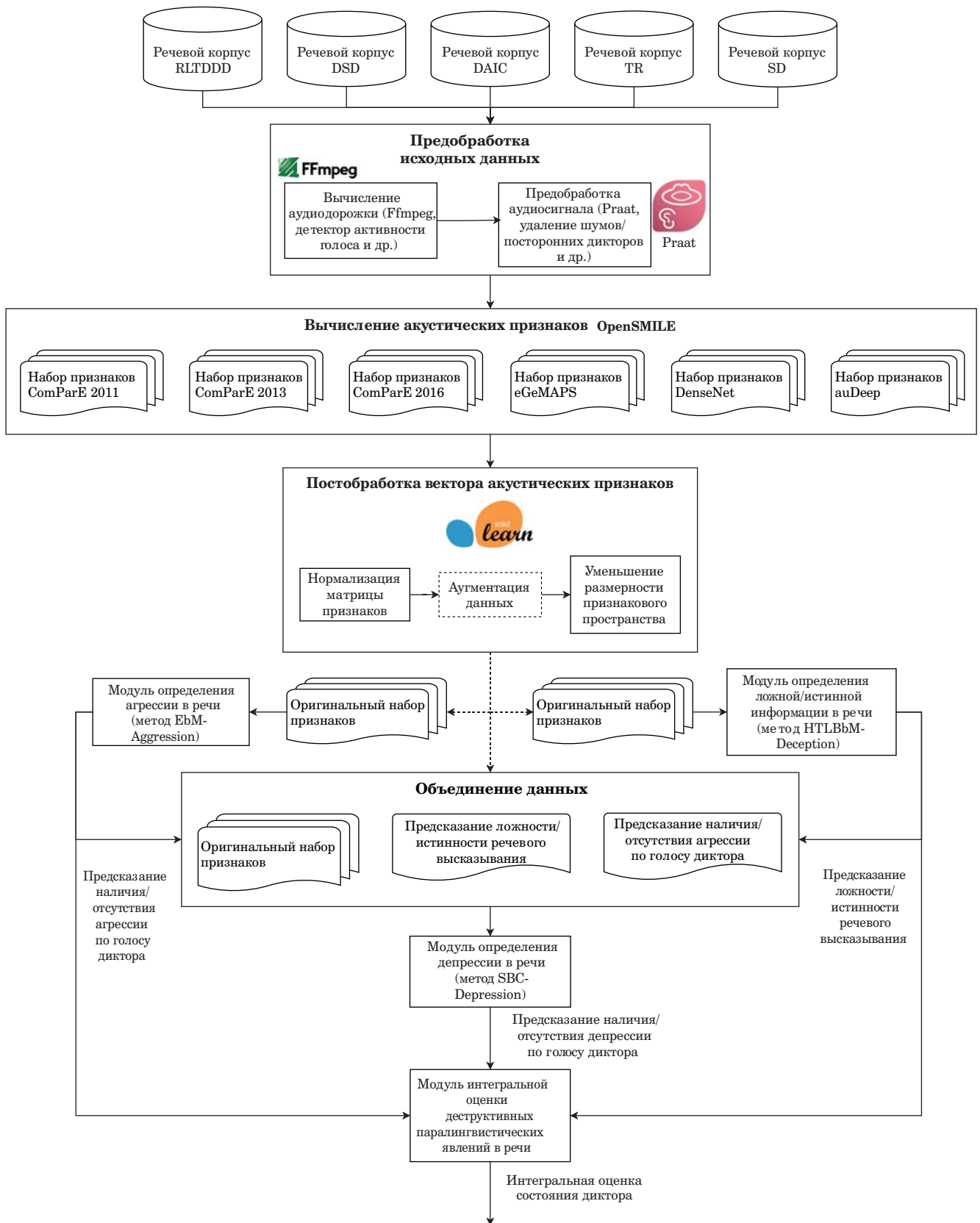
С использованием предложенной программной системы можно получить как результат классификации деструктивных паралингвистических явлений с применением отдельных методов, так и результат определения депрессии в речи с учетом корреляции акустических признаков в речевом сегменте с другими деструктивными явлениями (лжи/истинности и агрессии).

Для разработки программной системы выбран объектно ориентированный язык программирования высокого уровня Python версии 3.8 [28]. Кроме того, были использованы программные библиотеки как для обработки данных и вычисления признаков (Scikit-learn, OpenSMILE), так и для реализаций методов машинного обучения (Keras, Catboost, XGBoost, LightGBM, TabNet).

Эксперименты

Результаты приведенных в таблице экспериментальных исследований подробнее описаны в работах [29–31]. В качестве данных для обучения и оценивания методов определения деструктивных паралингвистических явлений в речи были использованы следующие речевые и модальные корпуса: 1) корпус DSD (1059 аудиозаписей) [32] и корпус RLTD (121 запись) [33] для метода определения ложной/истинной информации; 2) корпус DAIC (219 записей) [34] для метода определения депрессии; 3) корпуса SD и TR (893 записи) [35] для метода определения агрессии. Все экспериментальные исследования были проведены с пятикратной перекрестной валидацией. Для количественного оценивания эффективности работы предложенных методов выбраны показатели F1-мера и UAR.

Проведенные экспериментальные исследования позволяют сделать вывод, что некоторые методы определения деструктивных паралингвистических явлений в разговорной речи, входящие в состав предложенной архитектуры программной системы, в ряде случаев превосходят



■ **Рис. 3.** Архитектура программной системы интегрального анализа деструктивных паралингвистических явлений в разговорной речи
 ■ **Fig. 3.** An architecture of the software system for integral analysis of destructive paralinguistic phenomena in colloquial speech

■ Результаты экспериментальных исследований методов определения деструктивных паралингвистических явлений в речи

■ Results of the experimental studies on the methods for destructive paralinguistic detection in speech

Метод определения явления	Результаты классификации, %
Ложной/истинной информации в разговорной речи	$F1 = 88,4 \pm 1,5$ $UAR = 88,5 \pm 1,3$
Депрессии в разговорной речи	$F1 = 64,0 \pm 0,7$ $UAR = 60,0 \pm 0,5$
Агрессии в разговорной речи	$F1 = 78,0 \pm 1,5$ $UAR = 76,5 \pm 1,3$

известные аналоги по эффективности. 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:

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

$$UAR_{int} = \frac{1}{3} \sum_{i=1}^3 UAR_i = 75,0\%, \quad (4)$$

где $F1_i$, UAR_i — показатели i -го метода из трех методов.

Заключение

Предложенные в работе методика и архитектура программной системы интегрального оценивания степени выраженности деструктивных паралингвистических явлений в разговорной речи рассматривают в совокупности несколько деструктивных паралингвистических явлений (ложь, агрессию и депрессию), а также учитывают корреляцию между ними. Полученные результаты экспериментальных исследований свидетельствуют о практической значимости предложенных методики и архитектуры программной системы и позволяют сделать вывод, что в задаче определения деструктивных паралингвистических явлений необходимо учитывать тесные взаимосвязи между ними. Это поможет улучшить результаты определения рассматриваемых явлений как по отдельности, так и в совокупности.

При проведении экспериментальных исследований был преодолен ряд ограничений технического характера, а именно: 1) относительно небольшое количество обучающих данных ввиду

того, что сбор специфичных данных является трудоемким и сложным процессом; 2) дисбаланс в классах, поскольку в реальной жизни рассматриваемые явления встречаются реже, чем нормальное состояние человека. Для преодоления этих ограничений были использованы аугментация данных, выбор информативных признаков, а также показатели точности работы системы, учитывающие дисбаланс в данных.

В дальнейшем планируется разработка пользовательского интерфейса программной системы, после чего будет возможна апробация в реальных условиях под контролем специалистов в области психологии. Вероятными ограничениями при практическом применении можно назвать условия записи речи (различные характеристики устройств записи, шумы при записи и т. д.), которые могут оказать влияние на качество работы моделей определения деструктивных явлений. Способами преодоления этих ограничений являются: 1) использование качественного оборудования и (или) методов цифровой обработки данных (удаление шумов из записи); 2) сбор нового речевого корпуса, содержащего все рассматриваемые деструктивные явления (и последующее дообучение существующих моделей с использованием собранных данных). Также одним из направлений продолжения исследований является совершенствование моделей и повышение качества определения деструктивных явлений. Помимо ограничений технического характера при практическом внедрении, стоит также упомянуть ограничения теоретического характера, которые включают в себя возможные упущения авторов при разработке предложенной методики ввиду отсутствия профессиональных знаний из областей психологии и психиатрии, что решается тесной работой со специалистами в этих областях.

Финансовая поддержка

Работа выполнена при финансовой поддержке фонда РФ (проект № 22-11-00321).

Литература

1. Montacié C., Caraty M.-J. Prosodic cues and answer type detection for the deception sub-challenge *Proc. of INTERSPEECH-2016*, 2016, pp. 2016–2020. doi:10.21293/1818-0442-2016-19-2-56-60
2. Huang Zh., Epps J., Joachim D., Joachim D., Stasak B., Williamson J. R., Quatieri T. F. Domain adaptation for enhancing speech based depression detection in natural environmental conditions using di-

- lated CNNs. *Proc. of INTERSPEECH-2020*, 2020, pp. 4561–4565. doi:10.21437/Interspeech.2020-3135
3. **Egas-López J. V., Vetráb M., Tóth L., Gosztolya G.** Identifying conflict escalation and primates by using ensemble X-vectors and Fisher vector features. *Proc. of INTERSPEECH-2021*, 2021, pp. 476–480. doi:10.21437/Interspeech.2021-1173
 4. **Майсак Н. В.** Матрица социальных девиаций: классификация типов и видов девиантного поведения. *Современные проблемы науки и образования*, 2010, № 4, с. 78–86. <https://science-education.ru/ru/article/view?id=4505> (дата обращения: 10.03.2023).
 5. **Balsamo M.** Anger and depression: Evidence of a possible mediating role for rumination. *Psychological Reports*, 2010, no. 106, pp. 3–12. doi:10.2466/PRO.106.1.3-12
 6. **Luutonen S.** Anger and depression – theoretical and clinical considerations. *Nordic Journal of Psychiatry*, 2007, no. 61, pp. 246–251. doi:10.1080/08039480701414890
 7. **Robbins P., Tanck R.** Anger and depressed affect: Interindividual and intraindividual perspectives. *The Journal of Psychology*, 1997, no. 131, pp. 489–500. doi:10.1080/00223989709603537
 8. **Ng T., Sorensen K., Zhang Y., Yim F. H. K.** Anger, anxiety, depression, and negative affect: Convergent or divergent? *Journal of Vocational Behavior*, 2018, no. 110, pp. 186–202. doi:10.1016/j.jvb.2018.11.014
 9. **Bhardwaj V., Angkaw A., Franceschetti M., Rao R., Baker D. G.** Direct and indirect relationships among posttraumatic stress disorder, depression, hostility, anger, and verbal and physical aggression in returning veterans. *Aggressive Behavior*, 2019, no. 45(4), pp. 417–426. doi:10.1002/ab.21827
 10. **Huntsinger J. R.** Anger enhances correspondence between implicit and explicit attitudes. *Emotion*, 2013, no. 13(2), pp. 350–357. doi:10.1037/a0029974
 11. **Yip J. A., Schweitzer M. E.** Mad and misleading: Incidental anger promotes deception. *Organizational Behavior and Human Decision Processes*, 2016, no. 137, pp. 207–217. doi:10.1016/j.obhdp.2016.09.006
 12. **Горшков Ю. Г., Дорофеев А. В.** Речевые детекторы лжи коммерческого применения. *Информационный мост (ИНФОРМОСТ). Радиоэлектроника и Телекоммуникация*, 2003, № 6, с. 13–15.
 13. **Kaya H., Karpov A.** Fusing acoustic feature representations for computational paralinguistics tasks. *Proc. of INTERSPEECH-2016*, 2016, pp. 2046–2050. doi:10.21437/Interspeech.2016-995
 14. **Pan X., Zhao H., Zhou Y.** The application of fractional Mel cepstral coefficient in deceptive speech detection. *PeerJ*, 2015. doi:10.7717/peerj.1194
 15. **Levitan S. I., Levitan Y., An G., Levine M., Levitan R., Rosenberg A., Hirschberg J.** Identifying individual differences in gender, ethnicity, and personality from dialogue for deception detection. *Proc. NAACL Workshop on Computational Approaches to Deception Detection*, 2016, pp. 40–44. doi:10.18653/v1/W16-0806
 16. **Zhang J., Levitan S. I., Hirschberg J.** Multimodal deception detection using automatically extracted acoustic, visual, and lexical features. *Proc. of INTERSPEECH-2020*, 2020, pp. 359–363. doi:10.21437/Interspeech.2020-2320
 17. *Depression and Other Common Mental Disorders: Global Health Estimates*. Technical Report. World Health Organization, 2017. Licence: CC BY-NC-SA 3.0 IGO. <https://apps.who.int/iris/handle/10665/254610> (дата обращения: 10.03.2023).
 18. **Kaya H., Fedotov D., Dresvyanskiy D., Doyran M., Mamontov D., Markitantov M., Salah A. A. A., Kavcar E., Karpov A., Salah A. A.** Predicting depression and emotions in the cross roads of cultures, paralinguistics, and non-linguistics. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 27–35. doi:10.1145/3347320.3357691
 19. **Makiuchi M. R., Warnita T., Uto K., Shinoda K.** Multimodal fusion of BERT-CNN and gated CNN representations for depression detection. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 55–63. doi:10.1145/3347320.3357694
 20. **Fan W., He Z., Xing X., Cai B., Lu W.** Multi-modality depression detection via multi-scale temporal dilated CNNs. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 73–80. doi:10.1145/3347320.3357695
 21. **Haque A., Guo M., Miner A. S., Fei-Fei L.** Measuring depression symptom severity from spoken language and 3D facial expressions. *Machine Learning for Health (ML4H) Workshop at NeurIPS 2018*, 2018. <http://arxiv.org/abs/1811.0859> (дата обращения: 10.03.2023).
 22. **Niu M., Tao J., Liu B., Fan C.** Automatic depression level detection via lp-norm pooling. *Proc. of INTERSPEECH-2019*, 2019, pp. 4559–4563. doi:10.21437/Interspeech.2019-1617
 23. **Rohanian M., Hough J., Purver M.** Detecting depression with word-level multimodal fusion. *Proc. of INTERSPEECH-2019*, 2019, pp. 1443–1447. doi:10.21437/Interspeech.2019-2283
 24. **Zhao Z., Li Q., Cummins N., Liu B., Wang H., Tao J., Schuller B. W.** Hybrid network feature extraction for depression assessment from speech. *Proc. of INTERSPEECH-2020*, 2020, pp. 4956–4960. doi:10.21437/Interspeech.2020-2396
 25. **Lefter I., Jonker C. M.** Aggression recognition using overlapping speech. *Seventh Intern. Confe. on Affective Computing and Intelligent Interaction (ACII)*, 2017, pp. 299–304. doi:10.1109/ACII.2017.8273616
 26. **Sahoo S., Routray A.** Detecting aggression in voice using inverse filtered speech features. *IEEE Transactions on Affective Computing*, 2018, no. 9(2), pp. 217–226. doi:10.1109/TAFFC.2016.2615607
 27. **Zhou Z., Xu Y., Li M.** Detecting escalation level from speech with transfer learning and acoustic-lexical in-

- formation fusion. *arXiv Preprints*, 2021. doi:10.48550/arXiv.2104.06004
28. *About Python*. Python.org, 2022. <https://www.python.org/about/> (дата обращения: 10.03.2023).
29. 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. doi:10.17308/sait/1995-5499/2022/4/180-188
30. Velichko A., Markitantov M., Kaya H., Karpov A. Complex paralinguistic analysis of speech: Predicting gender, emotions and deception in a hierarchical framework. *Proc. of INTERSPEECH-2022*, 2022, pp. 4735–4739. doi:10.21437/Interspeech.2022-11294
31. 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: Proc. of the Intern. Conf. "Dialogue 2021"*, 2021, no. 20(27), pp. 698–704. doi:10.28995/2075-7182-2021-20-698-704
32. Schuller B., Steidl S., Batliner A., Hirschberg J., Burgoon J. K., Baird A., Elkins A., Zhang Y., Coutinho E., Evanini K. The INTERSPEECH 2016 computational paralinguistic challenge: Deception, sincerity & native language. *Proc. of INTERSPEECH-2016*, 2016, pp. 2001–2005. doi:10.21437/Interspeech.2016-129
33. Pérez-Rosas V., Abouelenien M., Mihalcea R., Burzo M. Deception detection using real-life trial data. *Proc. of the 2015 ACM Intern. Conf. on Multimodal Interaction*, 2015, pp. 59–66. doi:10.1145/2818346.2820758
34. Gratch J., Artstein R., Lucas G., Stratou G., Scherer S., Nazarian A., Wood R., Boberg J., DeVault D., Marsella S., Traum D., Rizzo S., Morency L.-P. The distress analysis interview corpus of human and computer interviews. *Proc. of the Ninth Intern. Conf. on Language Resources and Evaluation (LREC'14)*, 2014, pp. 3123–3128.
35. Ringeval F., Schuller B., Valstar M., Cummins N., Cowie R., Tavabi L., Schmitt M., Alisamir S., Amiriparian S., Messner E. -M., Song S., Liu S., Zhao Z., Mallol-Ragolta A., Ren Z., Soleymani M., Pantic M. AVEC 2019 workshop and challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 3–12. doi:10.1145/3347320.3357688

UDC 004.934.2

doi:10.31799/1684-8853-2023-4-2-11

EDN: FHUWJ

Methods and a software system for integrative analysis of destructive paralinguistic phenomena in colloquial speech

A. N. Velichko^a, Research Fellow, orcid.org/0000-0002-8503-8512

A. A. Karpov^a, Dr. Sc., Tech., Professor, orcid.org/0000-0003-3424-652X, karpov@iias.spb.su

^aSt. Petersburg Federal Research Center of the RAS, 39, 14th Line, 199178, Saint-Petersburg, Russian Federation

Introduction: There has been a growth in the number of studies devoted to the systems for paralinguistic analysis of various destructive phenomena in speech. However, only a few of them consider several phenomena altogether, despite the fact that there are also some theoretical and practical studies that describe the relationships between different paralinguistic phenomena. **Purpose:** To develop an approach, methods and a software system for integrative analysis of destructive paralinguistic phenomena (deception, aggression, depression) in colloquial speech. **Results:** We present an approach for the integrated evaluation of destructive paralinguistic phenomena in speech. This approach takes into account classification results from each of the three methods for detecting the abovementioned paralinguistic phenomena (deception, aggression and depression) and uses them to calculate the integrated score according to the set of rules. The proposed architecture of a software system includes a complex of methods for the detection of paralinguistic phenomena and the proposed approach. On the basis of the experimental classification results of the detection of destructive paralinguistic phenomena, we can calculate the integral average F1-measure and integral unweighted average recall which result in 76.8% and 75.0% accordingly. **Practical relevance:** The use of the proposed architecture of the software system makes it possible to make integrated evaluation of destructive paralinguistic phenomena in colloquial speech. Such a software system can be applied by medical professionals during an initial assessment of patients in consultation as one of the methods of psychological evaluation as well as other classical approaches (questionnaires, tests, etc.).

Keywords – speech analysis, speech technologies, computational paralinguistics, destructive phenomena, automatic deception detection in speech, automatic aggression detection in speech, automatic depression detection in speech.

For citation: Velichko A. N., Karpov A. A. Methods and a software system for integrative analysis of destructive paralinguistic phenomena in colloquial speech. *Informatsionno-upravliayushchie sistemy* [Information and Control Systems], 2023, no. 4, pp. 2–11 (In Russian). doi:10.31799/1684-8853-2023-4-2-11, EDN: FHUWJ

Financial support

This work was supported financially by the Russian Science Foundation (project No. 22-11-00321).

References

- Montacié C., Caraty M.-J. Prosodic cues and answer type detection for the deception sub-challenge. *Proc. of INTERSPEECH-2016*, 2016, pp. 2016–2020. doi:10.21293/1818-0442-2016-19-2-56-60
- Huang Zh., Epps J., Joachim D., Joachim D., Stasak B., Williamson J. R., Quatieri T. F. Domain adaptation for enhancing speech based depression detection in natural environmental conditions using dilated CNNs. *Proc. of INTERSPEECH-2020*, 2020, pp. 4561–4565. doi:10.21437/Interspeech.2020-3135
- Egas-López J. V., Vetráb M., Tóth L., Gosztolya G. Identifying conflict escalation and primates by using ensemble X-vectors and Fisher vector features. *Proc. of INTERSPEECH-2021*, 2021, pp. 476–480. doi:10.21437/Interspeech.2021-1173
- Maysak N. V. The matrix of social deviations: classification of types and sorts of deviant behaviour. *Modern Problems of Science and Education*, 2010, no. 4, pp. 78–86. Available at: <https://science-education.ru/ru/article/view?id=4505> (accessed 10 March 2023) (In Russian).
- Balsamo M. Anger and depression: Evidence of a possible mediating role for rumination. *Psychological Reports*, 2010, no. 106, pp. 3–12. doi:10.2466/PRO.106.1.3-12mel
- Luutonen S. Anger and depression – Theoretical and clinical considerations. *Nordic Journal of Psychiatry*, 2007, no. 61, pp. 246–251. doi:10.1080/08039480701414890
- Robbins P., Tanck R. Anger and depressed affect: Interindividual and intraindividual perspectives. *The Journal of Psychology*, 1997, no. 131, pp. 489–500. doi:10.1080/00223989709603537
- Ng T., Sorensen K., Zhang Y., Yim F. H. K. Anger, anxiety, depression, and negative affect: Convergent or divergent? *Journal of Vocational Behavior*, 2018, no. 110, pp. 186–202. doi:10.1016/j.jvb.2018.11.014
- Bhardwaj V., Angkaw A., Franceschetti M., Rao R., Baker D. G. Direct and indirect relationships among posttraumatic stress disorder, depression, hostility, anger, and verbal and physical aggression in returning veterans. *Aggressive Behavior*, 2019, no. 45(4), pp. 417–426. doi:10.1002/ab.21827
- Huntsinger J. R. Anger enhances correspondence between implicit and explicit attitudes. *Emotion*, 2013, no. 13(2), pp. 350–357. doi:10.1037/a0029974
- Yip J. A., Schweitzer M. E. Mad and misleading: Incidental anger promotes deception. *Organizational Behavior and Human Decision Processes*, 2016, no. 137, pp. 207–217. doi:10.1016/j.obhdp.2016.09.006
- Gorshkov Yu. G., Dorofeev A. V. Rechevye detektory lzhi kommercheskogo primeneniya. *Informacionnyj most (INFORMOST). Radioelektronika i Telekommunikaciya*, 2003, no. 6, pp. 13–15 (In Russian).
- Kaya H., Karpov A. Fusing acoustic feature representations for computational paralinguistics tasks. *Proc. of INTERSPEECH-2016*, 2016, pp. 2046–2050. doi:10.21437/Interspeech.2016-995
- Pan X., Zhao H., Zhou Y. The application of fractional Mel cepstral coefficient in deceptive speech detection. *PeerJ*, 2015. doi: 10.7717/peerj.1194
- Levitan S. I., Levitan Y., An G., Levine M., Levitan R., Rosenberg A., Hirschberg J. Identifying individual differences in gender, ethnicity, and personality from dialogue for deception detection. *Proc. NAACL Workshop on Computational Approaches to Deception Detection*, 2016, pp. 40–44. doi:10.18653/v1/W16-0806
- Zhang J., Levitan S. I., Hirschberg J. Multimodal deception detection using automatically extracted acoustic, visual, and lexical features. *Proc. of INTERSPEECH-2020*, 2020, pp. 359–363. doi:10.21437/Interspeech.2020-2320
- Depression and Other Common Mental Disorders: Global Health Estimates*. Technical Report. World Health Organization, 2017. Licence: CC BY-NC-SA 3.0 IGO. Available at: <https://apps.who.int/iris/handle/10665/254610> (accessed 10 March 2023).
- Kaya H., Fedotov D., Dresvyanskiy D., Doyran M., Mamonov D., Markitantov M., Salah A. A., Kavcar E., Karpov A., Salah A. A. Predicting depression and emotions in the crossroads of cultures, paralinguistics, and non-linguistics. *Proc. of the 9th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 27–35. doi:10.1145/3347320.3357691
- Makiuchi M. R., Warnita T., Uto K., Shinoda K. Multimodal Fusion of BERT-CNN and Gated CNN Representations for Depression Detection. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 55–63. doi:10.1145/3347320.3357694
- Fan W., He Z., Xing X., Cai B., Lu W. Multi-modality depression detection via multi-scale temporal dilated CNNs. *Proc. of the 9th ACM International Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 73–80. doi:10.1145/3347320.3357695
- Haque A., Guo M., Miner A. S., Fei-Fei L. Measuring depression symptom severity from spoken language and 3D facial expressions. *Machine Learning for Health (ML4H) Workshop at NeurIPS 2018*, 2018. Available at: <http://arxiv.org/abs/1811.0859> (accessed 10 March 2023).
- Niu M., Tao J., Liu B., Fan C. Automatic depression level detection via lp-norm pooling. *Proc. of INTERSPEECH-2019*, 2019, pp. 4559–4563. doi:10.21437/Interspeech.2019-1617
- Rohanian M., Hough J., Purver M. Detecting depression with word-level multimodal fusion. *Proc. of INTERSPEECH-2019*, 2019, pp. 1443–1447. doi:10.21437/Interspeech.2019-2283
- Zhao Z., Li Q., Cummins N., Liu B., Wang H., Tao J., Schuller B. W. Hybrid network feature extraction for depression assessment from speech. *Proc. of INTERSPEECH-2020*, 2020, pp. 4956–4960. doi:10.21437/Interspeech.2020-2396
- Lefter I., Jonker C. M. Aggression recognition using overlapping speech. *Seventh Interna. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2017, pp. 299–304. doi:10.1109/ACII.2017.8273616
- Sahoo S., Routray A. Detecting aggression in voice using inverse filtered speech features. *IEEE Transactions on Affective Computing*, 2018, no. 9(2), pp. 217–226. doi:10.1109/TAFFC.2016.2615607
- Zhou Z., Xu Y., Li M. Detecting escalation level from speech with transfer learning and acoustic-lexical information fusion. *arXiv Preprints*, 2021. doi:10.48550/arXiv.2104.06004
- About Python*. Python.org., 2022. Available at: <https://www.python.org/about/> (accessed 10 March 2023).
- Velichko A. N. A speech signal analysis method for automatic aggression detection in colloquial speech. *Proc. of Voronezh State University. Ser.: Systems Analysis and Information Technologies*, 2022, no. 4, pp. 180–188 (In Russian). doi:10.17308/sait/1995-5499/2022/4/180-188
- Velichko A., Markitantov M., Kaya H., Karpov A. Complex paralinguistic analysis of speech: Predicting gender, emotions and deception in a hierarchical framework. *Proc. of INTERSPEECH-2022*, 2022, pp. 4735–4739. doi:10.21437/Interspeech.2022-11294
- 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: Proc. of the Intern. Conf. "Dialogue 2021"*, 2021, no. 20(27), pp. 698–704. doi:10.28995/2075-7182-2021-20-698-704
- Schuller B., Steidl S., Batliner A., Hirschberg J., Burgoon J. K., Baird A., Elkins A., Zhang Y., Coutinho E., Evanini K. The INTERSPEECH 2016 computational paralinguistic challenge: Deception, sincerity & native language. *Proc. of INTERSPEECH-2016*, 2016, pp. 2001–2005. doi:10.21437/Interspeech.2016-129
- Pérez-Rosas V., Abouelenien M., Mihalcea R., Burzo M. Deception detection using real-life trial data. *Proc. of the 2015 ACM Intern. Conf. on Multimodal Interaction*, 2015, pp. 59–66. doi:10.1145/2818346.2820758
- Gratch J., Artstein R., Lucas G., Stratou G., Scherer S., Nazarian A., Wood R., Boberg J., DeVault D., Marsella S., Traum D., Rizzo S., Morency L.-P. The distress analysis interview corpus of human and computer interviews. *Proc. of the Ninth Intern. Conf. on Language Resources and Evaluation (LREC'14)*, 2014, pp. 3123–3128.
- Ringeval F., Schuller B., Valstar M., Cummins N., Cowie R., Tavabi L., Schmitt M., Alisamir S., Amiriparian S., Messner E.-M., Song S., Liu S., Zhao Z., Mallol-Ragolta A., Ren Z., Soleymani M., Pantic M. AVEC 2019 workshop and challenge: State-of-mind, detecting depression with AI, and cross-cultural affect recognition. *Proc. of the 9th ACM Intern. Workshop on Audio/Visual Emotion Challenge (AVEC'19)*, 2019, pp. 3–12. doi:10.1145/3347320.3357688