

## ОПЫТ ВЫЯВЛЕНИЯ ВЕРБАЛЬНЫХ И НЕВЕРБАЛЬНЫХ МАРКЕРОВ АГРЕССИВНОГО ПОВЕДЕНИЯ ДЛЯ ЦИФРОВОГО ПРОФАЙЛИНГА

Гофман О.О.<sup>1</sup>, Лаушкина А.А.<sup>2</sup>, Андрюнин К.М.<sup>2</sup>, Самигулин Т.<sup>2</sup>,  
Гарбуз А.О.<sup>1</sup>

<sup>1</sup>Санкт-Петербургский государственный университет,  
г. Санкт-Петербург, Россия

<sup>2</sup>Санкт-Петербургский национальный исследовательский университет  
информационных технологий, механики и оптики, г. Санкт-Петербург, Россия

В статье рассматривается агрессивность как один из факторов неблагонадежности соискателя и предполагается, что выявление вербальных и невербальных характеристик личности поможет HR-специалистам при отборе. Определены закономерности проявления и предрасположенности проявления агрессивного поведения с использованием психодиагностических методик и результатов нейронной сети.

**Ключевые слова:** цифровой профайлинг, агрессивность, вербальные и невербальные маркеры.

В современном мире всё большее распространение получает искусственного интеллекта. Отдельной областью исследований, которая занимается изучением и разработкой систем и устройств, которые могут распознавать, интерпретировать, обрабатывать и имитировать человеческие аффекты является *affective computing* [1]. Однако многое до сих пор не известно. Так, например, редко при распознавании агрессии учитывают личностные особенности испытуемого – агрессивность, открытость и другие личностные черты. Распознавание происходит обычно в лабораторной ситуации, а не в ситуации близкой к реальности (например, переговорам). Проблемой таких исследований является, во-первых, рассмотрение испытуемого через призму отдельных свойств и черт, а не как целостную систему; во-вторых, используются готовые алгоритмы распознавания, без внимания к психологическим инструментам и качественному анализу данных. В нашем исследовании мы попытались частично сузить имеющиеся пробелы.

Цель работы – определить вербальные и невербальные маркеры агрессивного поведения для последующей разработки цифрового профайлинга. Цифровой профайлинг – это информационно-компьютерная технология, применяемая для верификации данных и

прогнозирование поведения человека на основе анализа наиболее информативных частных признаков, характеристик внешности, невербального и вербального поведения с целью обеспечения безопасности людей и общественных систем. Агрессивное поведение может быть фактором неблагонадежности, который требует обязательной диагностики и оценки.

Объект исследования: молодые люди, средний возраст 23 года. В первоначальном психологическом тестировании приняли участие 59 человек, из них 40 человек прошли экспериментальное исследование. После отсева по техническим причинам в выборке осталось 16 человек (от 21 до 25 лет).

Предмет исследования: признаки агрессивного поведения, которые доступны для выявления из видео и аудио записи монолога человека. С учетом технических возможностей детекции признаков агрессивного поведения, были использованы следующие маркеры:

- мимика: опущенные и сведённые брови; наличие вертикальной складки между бровями; верхние веки напряжены и приподняты, нижние веки также напряжены; губы находятся в одном из двух положений: крепко сжаты, уголки губ прямые или опущены, или раскрыты и напряжены, как при крике; ноздри могут быть расширены;
- движение глаз: глаза неподвижные, интенсивность взглядов увеличивается, закатывание глаз;
- голос: резкое увеличение силы голоса, его высоты и резкости (беглость произношения).
- движение тела: вертикальное положение головы; приподнятые плечи; руки «открыты» (руки разведены, могут вытягиваться вперёд и в стороны, активные движения руками и телом (сильные, резкие);
- признаки вербальной агрессии: 1) ласково-уничижительные и уменьшительно-ласкательные суффиксы, суффиксы – ища, – ина, – ишк – выражают недовольство, неприязнь; 2) оскорбление, оценочные высказывания по отношению к объекту агрессии; 3) риторические вопросы (риторические по мнению того, кто спрашивает, в тексте до этого могут быть маркеры того, что спрашивающий считает ответ на вопрос очевидным); 4) наличие ненормативной лексики и жаргона; 5) использование повелительного наклонения; 6) выраженное в словесной форме намерение нанести вред; 7) насмешки; 8) акцент на личных достижениях в разговоре.

В исследовании была использована классификация А. Басса, где многообразие агрессивных действий можно отобразить тремя

независимыми категориями: физическая – вербальная; активная – пассивная; прямая-косвенная.

Первичное психологическое тестирование включало: определение агрессивности (методика Басса-Дарки в апробации Хвана, 2008), личностные особенности (TIPI RU в апробации Сергеевой и др., 2016), темпераментальные особенности (Русалов, 1990).

Собственно, эксперимент состоял из нескольких серий, и в каждой серии было 2 этапа: 1) участникам предлагались к просмотру нейтральные или агрессивные видео-стимулы для индуцирования эмоций; 2) участники выполняли задание (задачи с картинками для выявления фruстрационных реакций Розенцвейга, игра на основе дилеммы заключённого [2]). Всего было 4 серии: 1) нейтральные видео + задача с картинкой; 2) агрессивное видео + задача с картинкой; 3) нейтральное видео + дилемма заключённого; 4) агрессивное видео + дилемма заключённого. Последовательности предъявления у испытуемых различались. Видео были взяты из открытой верифицированной научной базы видеороликов FilmStim (Schaefer A. et al., 2010, ст-1 [5]). После каждого ролика осуществлялась проверка корректности индуцированной эмоции: участнику необходимо было выбрать, какому схематическому выражению лица (смайлики злости, радости и нейтрального состояния) соответствуют его текущие эмоции.

Видео- и аудиоматериалы эксперимента обрабатывались при помощи нейронной сети и дополнительно оценивались экспертами-психологами по картам наблюдения. Психологическая обработка включала качественные методы (оценка карт наблюдения экспертами – психологами) и количественные методы (методы математической статистики).

Полученные данные оценивались по трем каналам: видео, аудио, текст. Далее описана работа с каждым каналом информации.

1. Видео. Для извлечения невербальных характеристик из видеозаписей использовалась сверточная нейронная сеть BlazeFace для получения вектора мимики лица человека. По 468 ориентирам рассчитывается вектор мимики лица человека внутри каждого исследуемого временного окна. Изменения вектора мимики в текущем окне относительно среднего положения для всего видео интерпретируются как движения: покусывание губ – для нижней зоны

лица; повороты головы и закрывание лица – для всей совокупности точек, а также количество и частота морганий человека.

2. Аудио. К аудиозаписям применялась предварительная обработка данных, для удаления шумов и нормализации громкости [3, 4]. Для вычисления резкого изменения громкости использовался эвристический метод, основанный на определении угла наклона между соседними значениями амплитудной огибающей и расстояния до максимального значения. Если значение амплитуды было близко к максимальному по всему речевому сигналу и происходило быстрее, то изменение громкости считалось резким. Под темпом речи понимается количество слов в минуту. Для определения данного показателя выполняется вычисление спектрального потока сигнала, затем подсчитывается количество пиковых значений, которые принимаются за начало слова.

3. Текст. Для транскрибации речи был выбран набор инструментов Vosk. Для определения маты и жаргона был собран словарь с возможными модификациями (около 140 тысяч слов). Определялось количество подобных слов. Для определения повелительного наклонения проводился морфологический разбор (библиотека ru morphy2) входящего предложения и выявлялись слова повелительного наклонения в двух вариантах: говорящий включён или говорящий не включён в действие. Для определения уменьшительно-ласкательных (пренебрежительных) выражений, слова обрабатывались с помощью стеммера Портера. А затем проверялось наличие уменьшительно-ласкательных суффиксов у слова. Для определения оскорблений и угроз были обучены две модели на основе архитектуры нейронной сети BERT для русского языка (RuBertTiny2).

Для определения эмоциональной окраски была обучена модель сверточной нейронной сети с полносвязным классификатором. Для обучения модели были объединены несколько наборов данных: CaFE, Emo-DB, RAVDESS и SAVEE. Все данные отнесены к одному из семи классов эмоций: злость, скука, тревога, радость, печаль, отвращение и нейтральное состояние.

В результате математической обработки и качественного анализа было выявлено три группы факторов, направляющих агрессивное поведение:

1. Индивидуальные особенности (ИО) испытуемых. Испытуемые склонные к аутоагgression чаще использовали повелительное наклонение. ИО носят независимый характер и не обусловлены

контекстом ситуации. Вне зависимости от контекста, который был заложен в эксперименте, ИО ярко проявились у части испытуемых.

2. Содержание. Данный фактор проявляется в том, что испытуемые после просмотра агрессивных видео-стимулов в межличностной коммуникации (дилемме заключённого) чаще используют повелительное наклонение, угрозы, нецензурную лексику и оскорблений.

3. Форма (проявлялось только в голосе). Испытуемые говорили громче во время монолога (ответы с картинки). Не проявлялось в межличностной коммуникации (игре).

Выявлено, что агрессивное поведение имеет разные предпосылки: 1) связано с индивидуальными особенностями испытуемых (например, экстраверсией или интроверсией), агрессивное поведение проявляется вне зависимости от контекста, 2) ситуационным контекстом (так называемые, ситуационно-зависимые испытуемые, которые реагировали на контекст «здесь и сейчас»).

Дополнительно обнаружено: 1) высокие показатели агрессивности соответствуют неверbalным маркерам поведения при просмотре видео. Так, при просмотре агрессивных видео-стимулов увеличивается лицевая экспрессия (движение глаз, бровей, лба); 2) выявлена специфичные маркеры в разной лицевой экспрессии: активность «верхней» части лица (брови – лоб), «нижней» (губы) или лица в целом; 3) высокие показатели эмоциональности (свойства темперамента) положительно связаны с лицевой экспрессией; 4) внешние признаки атоагgressии и гетероагgressии могут отличаться.

При сверке результатов работы нейронной сети и качественного анализа экспертов-психологов было обнаружено, что нейронная сеть корректно распознавала агрессию в половине случаев. Это говорит о том, что технологии алгоритмов нейронной сети требуется доработка с учётом полученных психологических результатов.

Проведя пилотажное исследование, мы выявили психологические характеристики испытуемых, группы факторов и их соотношения, особенности верbalных и неверbalных проявлений на агрессивные видео-стимулы и рекомендации по проведению исследования и техническому совершенствованию алгоритмов нейронной сети. Это позволит нам при проведении полномасштабного исследования избежать уже допущенных ошибок и полноценно обучить нейросеть выявлять индивидуальные особенности личности

для помощи специалистам в принятии решений относительно кандидата.

**Список литературы:**

1. Calvo R.A. et al. (ed.). The Oxford handbook of affective computing. – Oxford Library of Psychology, 2015.
2. Coombs C.H. A reparameterization of the prisoner's dilemma game //Behavioral Science. – 1973. – Т. 18. – №. 6. – С. 424-428.
3. Goupil L. et al. Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature // Nature Communications. – 2021. – Vol. 12. – № 1. – P. 861.
4. Marcolla F., de Santiago R., Dazzi R. Novel Lie Speech Classification by using Voice Stress // Proceedings of the 12th International Conference on Agents and Artificial Intelligence. SCITEPRESS – Science and Technology Publications, 2020. – P. 742–749.
5. Schaefer A. et al. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers //Cognition and emotion. – 2010. – Т. 24. – №. 7. – С. 1153-1172.

**EXPERIENCE IN IDENTIFYING VERBAL AND NON-VERBAL MARKERS  
OF AGGRESSIVE BEHAVIOR FOR DIGITAL PROFILING**

**Hoffman O.O.<sup>1</sup>, Laushkina A.A.<sup>2</sup>, Andryunin K.M.<sup>2</sup>, Samigulin T.<sup>2</sup>,  
Garbuz A.O.<sup>1</sup>**

<sup>1</sup>Saint Petersburg State University,  
St. Petersburg, Russia

<sup>2</sup>St. Petersburg National Research University of Information Technologies,  
Mechanics and Optics, St. Petersburg, Russia

The article considers aggressiveness as one of the factors of unreliability of the applicant. It is assumed that the identification of verbal and nonverbal personality characteristics will help HR specialists in the selection process. The regularities of the manifestation and predisposition of the manifestation of aggressive behavior using psychodiagnostic techniques and the results of a neural network are determined.

*Keywords:* digital profiling, aggressiveness, verbal and non-verbal markers.

**Об авторах:**

Гофман Ольга Олеговна, кандидат психологических наук, доцент, ms.gofman@mail.ru.

Лаушкина Анастасия Александровна, ассистент факультета цифровых трансформаций, инженер мегафакультет трансляционных информационных технологий, nastasjalausckina@mail.ru.

Андрюнин Константин Маратович, магистрант, инженер мегафакультет трансляционных информационных технологий, andryunin.kostya@gmail.com.

Самигулин Тимур, инженер мегафакультет трансляционных информационных технологий, timursamigulin98@gmail.com.

Гарбуз Александра Олеговна, студент специалитета,  
garsan5401@yandex.ru.

**Об авторах на английском языке:**

Gofman Olga Olegovna, PhD in Psychology, associate professor,  
ms.gofman@mail.ru.

Laushkina Anastasia Alexandrovna, Assistant of the Faculty of Digital Transformation, engineer of the School of Translational Information Technologies, nastasjalausckina@mail.ru.

Andriunin Konstantin Maratovich, master's student, engineer of the School of Translational Information Technologies, andryunin.kostya@gmail.com.

Samigulin Timur, engineer of the School of Translational Information Technologies, timursamigulin98@gmail.com.

Garbuz Alexandra Olegovna, student of the specialty «Psychology of official activity» Saint-Petersburg University, garsan5401@yandex.ru).

Дата поступления в редакцию: 03.08.2023.

Дата принятия в печать: 28.08.2023.