

doi: 10.17586/2226-1494-2024-24-5-834-842
УДК 004.4/.5

Создание и анализ многомодального корпуса данных для автоматического распознавания агрессивного поведения людей

Михаил Юрьевич Уздяев^{1✉}, Алексей Анатольевич Карпов²

^{1,2} Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

¹ uzdyaev.m@iias.spb.su, <https://orcid.org/0000-0002-7032-0291>

² karpov@iias.spb.su, <https://orcid.org/0000-0003-3424-652X>

Аннотация

Введение. Развитие цифровых систем коммуникации сопряжено с растущим количеством проявлений деструктивного поведения людей и необходимостью оперативного на него реагирования. Ввиду слабой формализации предметной области агрессии, наиболее перспективными методами распознавания деструктивного поведения являются методы, основанные на подходах машинного обучения, которые для эффективной работы требуют репрезентативных выборок релевантных данных. При создании корпусов поведенческих данных необходимо решить следующие проблемы: соответствие разметки данных корпуса реальному поведению; представленности поведения в однотипных ситуациях и в корпусе натурального поведения. Целью работы является разработка методики создания выборки многомодальных данных поведенческой агрессии человека, содержательно отражающей агрессию как явление и обеспечивающей релевантность данных. **Метод.** В работе описывается разработанная методика создания выборок многомодальных данных, содержащих спонтанное агрессивное поведение. В ходе содержательного анализа предметной области агрессивного поведения человека выделяются значимые атрибуты агрессии такие как явления (наличие субъекта и объекта агрессии, деструктивный характер агрессивного действия) и единицы анализа поведения (временные сегменты аудио и видео, на которых локализованы информанты); определяются типы регистрируемой агрессии (физическая и вербальная явные прямые); обосновываются критерии оценки агрессивного поведения каждого типа посредством введения перечня действий, однозначно определяющих каждый вид агрессии. Методика состоит из следующих этапов: сбор видео в открытом доступе в сети Интернет; выделение временных интервалов, на которых проявляется агрессия; локализация информантов на кадрах видео; транскрибирование реплик информантов; оценка актов физической и вербальной агрессии группой аннотаторов посредством разработанного алгоритма оценки поведения; вычисление согласованности оценок с помощью коэффициента Флейсса. **Основные результаты.** Для апробации методики создан и размечен группой аннотаторов аудиовизуальный корпус данных спонтанного агрессивного поведения русскоязычных информантов Audiovisual Aggressive Behavior in Online Streams (AVABOS). Корпус данных содержит видео- и аудиосегменты, на которых присутствует вербальная и физическая агрессии соответственно, проявляемые русскоязычными информантами в ходе онлайн-видеотрансляций. **Обсуждение.** Результаты согласованности разметки показали высокий уровень для физической агрессии ($\kappa = 0,74$) и средний уровень для вербальной ($\kappa = 0,48$), что подтверждает обоснованность разработанной методики. Корпус данных AVABOS может использоваться для решения задач автоматического распознавания агрессии человека. Помимо создания корпусов агрессивного поведения, методика также может использоваться для создания корпусов, содержащих другое поведение.

Ключевые слова

методика создания многомодального корпуса, методика оценки поведения, агрессивное поведение, распознавание агрессии, создание выборки данных, оценка согласованности разметки, коэффициент Флейсса

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда (№ 22-11-00321, <https://www.rscf.ru/project/22-11-00321/>).

Ссылка для цитирования: Уздяев М.Ю., Карпов А.А. Создание и анализ многомодального корпуса данных для автоматического распознавания агрессивного поведения людей // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 5. С. 834–842. doi: 10.17586/2226-1494-2024-24-5-834-842

© Уздяев М.Ю., Карпов А.А., 2024

Creation and analysis of multimodal corpus for aggressive behavior recognition

Mikhail Yu. Uzdiaev¹, Alexey A. Karpov²

^{1,2} St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

¹ uzdyae.v.m@iias.spb.su, <https://orcid.org/0000-0002-7032-0291>

² karpov@iias.spb.su, <https://orcid.org/0000-0003-3424-652X>

Abstract

The development of digital communication systems is associated with the increasing number of disruptive behavior incidents that require rapid response in order to prevent negative consequences. Due to weak formalization of human aggression, machine learning approaches are the most suitable for this area. Machine learning approaches require representative sets of relevant data for efficient aggression recognition. Datasets developing implies such problems as dataset labels relevance to the real behavior, the consistency of the situations, where behavior is manifested, and the naturalness of behavior. The purpose of this work is the development of an aggressive behavior datasets creation methodology that reflects the key aspects of aggression and provides relevant data. The work reveals the developed methodology for creation of multimodal datasets of natural aggression behavior. The analysis of human aggression subject area substantiates the key aspects of human aggression manifestations (the presence of subject and object of aggression, the destructiveness of the aggressive action), the behavior analysis units — the time intervals of audio and video with the localized informants, defines considering types of aggression (physical and verbal overt direct aggression), substantiates criteria for aggressive behavior assessment as a set of aggressive actions that define each aggression type. The methodology consists of the following stages: collecting video on the Internet, identifying time intervals where aggression is performed, localizing informants in video frames, transcribing informants' speech, collective labeling of physical and verbal aggression actions by a group of annotators (raters), assessing the reliability of annotations agreement using Fleiss' kappa coefficient. In order to evaluate the methodology a new audiovisual aggressive behavior in online streams corpus (AVABOS) was collected and labeled. The dataset contains audio and video segments that contains verbal and physical aggression correspondingly that manifested by Russian-speaking informants during online video streams. The results of interrater agreement reliability show substantial agreement for physical ($\kappa = 0.74$) and moderate agreement for verbal aggression ($\kappa = 0.48$) that substantiates the developed methodology. AVABOS dataset can be used in automatic aggression recognition tasks. The developed methodology can also be used for creating datasets with the other types of behavior.

Keywords

methodology for creating multimodal dataset, methodology for behavior assessment, aggressive behavior, aggression recognition, dataset creation, collective labeling, interrater reliability assessment, Fleiss' kappa coefficient

Acknowledgements

This work was supported financially by the Russian Science Foundation (project No. 22-11-00321, <https://www.rscf.ru/project/22-11-00321/>).

For citation: Uzdiaev M.Yu., Karpov A.A. Creation and analysis of multimodal corpus for aggressive behavior recognition. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 5, pp. 834–842 (in Russian). doi: 10.17586/2226-1494-2024-24-5-834-842

Введение

Развитие социкиберфизических систем, которые обеспечивают взаимодействие значительного числа людей и различных компьютерных и роботизированных систем, сопровождается увеличением количества случаев проявления деструктивного поведения, к которому, в первую очередь, относятся проявления агрессии людьми. Такое поведение зачастую приводит к негативным последствиям, включая нанесение вреда здоровью и гибели жертв актов агрессии. Для оперативного реагирования на такие случаи необходимо внедрение эффективных систем автоматического распознавания деструктивного поведения, основанных на методах машинного обучения, которые требуют больших репрезентативных корпусов данных, содержащих релевантные явления.

Необходимо обозначить основные проблемы, возникающие при создании корпусов многомодальных данных агрессивного поведения человека:

1) соответствие данных рассматриваемому типу поведения, которая выражается в выработке обоснованных критериев оценки поведения, куда вклю-

чены частные задачи определения типов агрессии, выделение существенных признаков для каждого типа, определения модальностей данных, в которых выражается агрессия;

2) однотипность ситуаций, в которых исполняется поведение, однотипность ситуаций может служить критерием для первичного отбора данных для их унификации;

3) естественность поведения, которая тесно связана с проблемой 1, так как естественность или спонтанность поведения напрямую влияет на соответствие данных реальному поведению;

4) выбор адекватных единиц анализа поведения, который связан с различением отдельных действий друг от друга и определением их временных границ, локализацией информантов, что особенно важно, когда участвуют несколько человек.

Целью исследования является разработка и апробация универсальной методики создания корпусов многомодальных данных, содержащих спонтанное агрессивное поведение русскоязычных информантов в ходе прямых видеотрансляций в сети Интернет. Поставлена задача преодоления проблем обоснования критериев

оценки поведения, недостаточного учета контекста ситуации и ненатуральности поведения.

Научная новизна предложенной методики состоит в выработке критериев оценки агрессивного поведения, которые представляют отдельные агрессивные действия и их существенные признаки. Выделены единицы анализа поведения: временные сегменты отдельных агрессивных действий с локализованными на них субъектами агрессии. В рамках апробации методики создан корпус многомодальных данных, содержащий спонтанное агрессивное поведение русскоязычных информантов.

Обзор методик создания корпусов поведенческих данных

Созданию выборок агрессивного поведения людей посвящен ряд научных работ. В работах [1, 2] рассмотрено создание выборок, содержащих многомодальное агрессивное поведение на вокзале и в железнодорожных вагонах. Эти работы имеют схожие методики создания выборок данных: категории оценки выводятся на основе анализа научных работ; анализируются выделенные временные сегменты видео (без локализации информантов) или реплики информантов; записывается наигранное поведение; разметка производится группой аннотаторов с вычислением степени согласованности их мнений. В работе [1] приведено содержательное раскрытие категорий оценки, тогда как в [2] такая информация отсутствует. В [3] рассмотрены аспекты вербальной агрессии в парном взаимодействии между профессиональным актером и испытуемым, на которого направлена агрессия. Здесь методика создания воспроизводит методику из [2]. В других работах [4, 5] исследовано наигранное агрессивное поведение на видео (другие модальности не рассматриваются). В ходе создания корпусов данных [4, 5] разметка выполнялась одним аннотатором, критерии оценки поведения не формулировались, информанты, проявляющие агрессию, на кадрах видео не выделялись. Работы [6, 7] посвящены созданию выборок спонтанного агрессивного поведения, записанного на камеры видеонаблюдения. Разметка также выполнялась одним аннотатором, содержательно критерии оценки поведения не раскрывались, не выполнялось выделение информантов на видео.

В [8] раскрыты аспекты методики создания выборок, содержащих вербальную агрессию в тексте на русском, испанском и татарском языках (другие модальности не рассматриваются). Работы интересны тем, что в них содержательно описаны критерии оценки агрессии, взятые на основании анализа известных научных работ; разметка выполнялась одним аннотатором. Для смежных задач анализа эмоций и настроений в тексте при создании корпусов данных также выполнялась разметка одним аннотатором [9, 10].

В работах [11–14] рассмотрено создание выборок эмоционального поведения. В них имеются следующие сходства: категории оценки берутся из научных работ без их содержательного раскрытия, аннотация выполняется путем субъективной оценки группой аннотато-

ров степени выраженности определенной категории с вычислением степени согласованности оценок. Условия ситуации обычно подробно описываются, однако не воспроизводятся при записи. Информанты специально не выделяются, единицами анализа служат временные сегменты, соответствующие действиям и репликам информантов.

Обобщая особенности методик создания известных корпусов данных, можно выделить следующие общие свойства: в большинстве работ критерии оценки поведения содержательно не раскрываются; разметка выполняется посредством субъективной оценки поведения, зачастую без раскрытия критериев оценки; отдельные информанты на видео не локализуются, т. е. неправильно выбираются единицы анализа, что в дальнейшем затрудняет обработку поведения, если на видео несколько человек по-разному ведут себя; часто поведение записывается в лабораторных условиях, где не воспроизводятся реальные условия ситуаций, что затрудняет перенос результатов этих выборок на реальные условия. В работах, описывающих выборки поведения в естественных условиях, не уделяется должное внимание контролю условий записи и рассмотрению поведения в других модальностях, отличных от видео. Кроме того, недостаточно представлены русскоязычные корпуса, для которых авторы разрабатывали методики оценки поведения.

Методика создания корпусов многомодальных данных с агрессивным поведением человека

Общая схема предлагаемой методики, а также детали, специфичные непосредственно для оценки агрессивного поведения человека, приведены на рис. 1. Рассмотрим этапы предложенной методики.

Этап 1. Анализ предметной области

Выделение признаков ситуаций. В рамках работы рассмотрено агрессивное поведение в ходе онлайн-видеотрансляций в сети Интернет, которые характеризуются следующими атрибутами: на видео присутствуют несколько участников различного пола и возраста, которые коммуницируют как между собой, так и со зрителями трансляции при помощи различных средств коммуникации. Трансляции ведутся в большинстве случаев из помещений (жилые или офисные помещения, специально оборудованные студии и т. д.) Сами участники обычно располагаются перед камерой, сидя за столом, стоя в полный рост или лежа. В ходе трансляции могут возникать ситуации, когда один или несколько участников начинают проявлять физическую или вербальную агрессию, вызванную различными причинами: состоянием опьянения, личной неприязнью, конфликтными ситуациями и др.

Определение критериев оценки физической и вербальной агрессии. Определение модальностей. В рамках анализа предметной области определяются критерии оценки агрессивного поведения, единицы анализа, а также модальности данных, в которых представлено поведение. В настоящей работе агрессия рассмотрена как деструктивное поведение, исполняемое субъектом агрессии (агрессором) по отношению

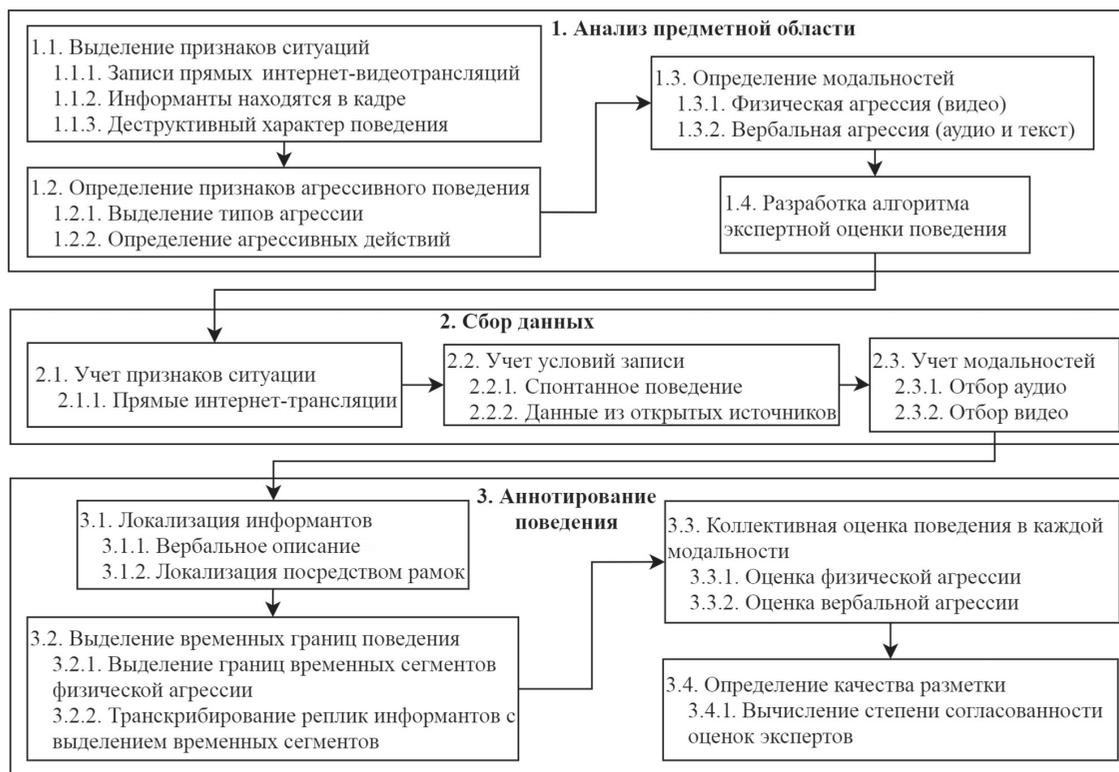


Рис. 1. Общая схема методики создания выборки данных поведенческой агрессии

Fig. 1. General scheme of aggressive behavior dataset creation methodology

к объекту агрессии (жертве), целью которого является нанесение ущерба жертве [15]. Выделим следующие характеристические признаки такого поведения, которые могут служить первичными критериями его оценки: обязательное присутствие субъекта агрессии; направленность поведения на внешний предмет (одушевленный или неодушевленный); поведение наносит ущерб объекту нападения. Помимо критериев оценки, данные признаки также служат для определения единиц анализа. Эти признаки определяют также единицы анализа при разметке выборки, которыми являются временные сегменты, на которых оценивается поведение агрессора, направленного на жертву. При этом обозначим необходимость обязательной локализации агрессора, без чего оценка поведения как агрессивного не может быть выполнена.

Оценка поведения внешним наблюдателем относится к методам не включенного внешнего наблюдения [16] за поведением, главной характеристикой которого является полное исключение влияния наблюдателя, что обеспечивает натуральность наблюдаемого поведения, а также объективный характер оценки. Признаками агрессивного поведения определим действия, однозначно относимые к тому или иному виду агрессии с учетом условий ситуации. Выделение перечня самих агрессивных действий основывается на содержательном анализе предметной области явной прямой физической и вербальной агрессии, а также статистического анализа результатов предварительного поискового наблюдения за агрессивным поведением русскоязычных пользователей в ходе прямых видеотрансляций [17]. В

методике в качестве критериев используются классы действий физической (захваты; кидание предметов; толчки; удары; укусы; обливание жидкостью; плевок; прочие акты физической агрессии) и вербальной (оскорбления; угрозы; грубые приказы; обвинения; поощрение насилия; прочие речевые акты) агрессии. Перечисленные классы агрессивных действий не исключают друг друга, т. е. информант может одновременно выполнять более одного действия. Подробное содержательное раскрытие критериев, а также примеры оценки поведения доступны в разработанной инструкции для разметки данных¹.

В рамках настоящей работы рассмотрены видео-, аудио- и текстовая модальности ввиду того, что они наиболее полно отражают физическую и вербальную прямую явную агрессию (по классификации А. Басса [18]) в ходе прямых видеотрансляций.

Разработка алгоритма экспертной оценки физической и вербальной агрессии. На основании критериев агрессивных действий разработан алгоритм экспертной оценки, представляющий последовательность действий, которые должен выполнить аннотатор в ходе разметки при оценке неопределенных ситуаций.

— Если оценивается вербальная агрессия, выполняется приведение реплики информанта в утвердительную форму, если оно имеет вопросительную, сокращенную форму, форму косвенной речи или

¹ [Электронный ресурс]. Режим доступа: https://github.com/cafe1930/AVABOS/blob/main/labelling_instruction.pdf (дата обращения: 08.08.2024).

страдательного залога. Приведение высказывания к утвердительной форме снижает неопределенность при оценке реплики.

- Оценка ответного поведения жертвы. Демонстрация негативных эмоций и ответная агрессия свидетельствуют об ущербе жертве и об агрессивности действия агрессора.
- Проверка наличия актов агрессии информанта до или после оцениваемого сегмента, что свидетельствует в пользу агрессивности оцениваемого интервала.
- Оценка сопутствующего поведения агрессора. Демонстрация негативных эмоций агрессором, невербальное речевое поведение (высокий громкий голос), высокая двигательная активность свидетельствуют в пользу наличия агрессии.

Алгоритм служит для разрешения ситуаций, когда проявления агрессии носят неочевидный характер или слабо выражены путем раскрытия контекста оцениваемого действия.

Этап 2. Сбор данных. В многомодальный корпус данных поведенческой агрессии помещались видео с открытым доступом на видеохостинге YouTube, которые дополнительно фильтровались согласно критериям, описанным на этапе 1, «Выделение признаков ситуаций», чтобы все видео удовлетворяли условию однородности содержания. Открытый доступ видеоматериала означает публичную демонстрацию поведения, что не требует получения персонального согласия информантов на анализ публично демонстрируемого поведения.

Разметка. Выделение временных интервалов физической и вербальной агрессии. В рамках разметки актов физической агрессии первоочередной задачей является их временная локализация на видео. Выделены явно различимые границы начала и окончания как отдельных действий, так и последовательности однородных действий, если внутри этой последовательности невозможно отделить одно действие от другого. Дополнительно выделялись сегменты, где отсутствуют действия физической агрессии. В случае вербальной агрессии выполнялась полуавтоматическая диаризация и транскрибирование реплик с использованием моделей глубокого обучения Whisper [19, 20] с дальнейшей коррекцией вручную границ и текста реплик. Выделение временных сегментов физической агрессии

и транскрибирование реплик выполнялось посредством программного обеспечения (ПО) ELAN [21].

Локализация информантов на кадрах видео.

Выполнена локализация каждого информанта, участвующего в актах агрессии при помощи:

- структурированного вербального описания отличительных особенностей их внешности: пол, причёска, одежда, обувь и т. д.;
- локализации описанных информантов прямоугольными рамками с последующим полуавтоматическим трекингом посредством ПО Video-Label-Tracker¹ (рис. 2).

Этап 3. Процесс аннотирования данных. Аннотирование выполнялось с использованием ПО ELAN (рис. 3) группой из пяти аннотаторов (двух мужского и трех женского пола). Возрастом от 20 до 32 лет (средний возраст — 26,4 лет, стандартное отклонение — 4,6 года), три аннотатора имеют высшее техническое образование, один — степень кандидата технических наук и один является студентом. Каждый из аннотаторов должен был определить наличие либо отсутствие агрессии в речевых действиях и двигательной активности каждого локализованного информанта на всех выделенных для него временных сегментах. Процесс аннотирования выполнен на основании критериев оценки действий и алгоритма экспертной оценки физической и вербальной агрессии. При этом на каждом оцениваемом сегменте может содержаться более одного класса агрессивных действий. В ходе разметки аннотаторы должны выделить все возможные классы физической и вербальной агрессии, проявляемой информантом на оцениваемом сегменте.

Вычисление степени согласованности оценок аннотаторов. Для оценки согласованности разметки выбран коэффициент (κ) Флейсса [22]:

$$\kappa = \frac{P_a - P_e}{1 - P_e}, \quad (1)$$

где P_a — доля согласованных оценок, выполненных аннотаторами; P_e — доля ожидаемых случайных оценок. Выбор оценки обоснован тем, что она оценивает

¹ [Электронный ресурс]. Режим доступа: Video-Label-Tracker. <https://github.com/cafe1930/Video-Label-Tracker> (дата обращения: 08.08.2024).

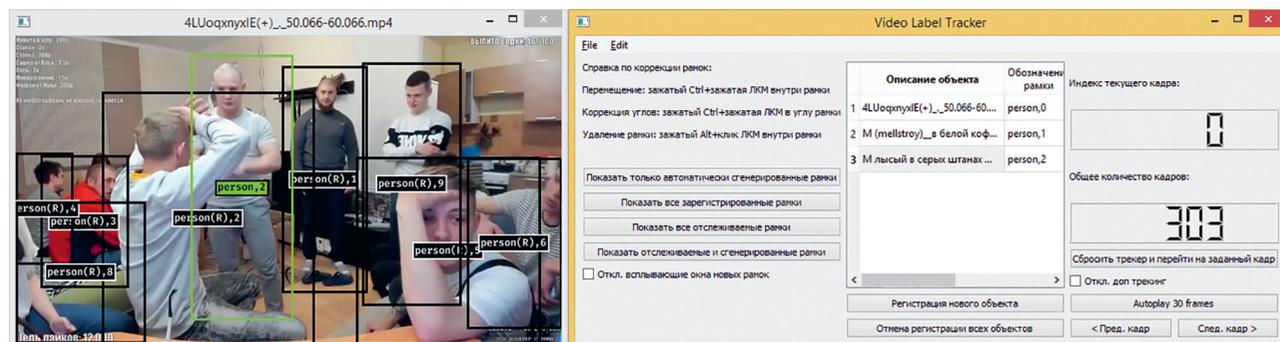


Рис. 2. Пример выполнения локализации информантов в программном обеспечении Video-Label-Tracker

Fig. 2. An example of informants localization using Video-Label-Tracker program

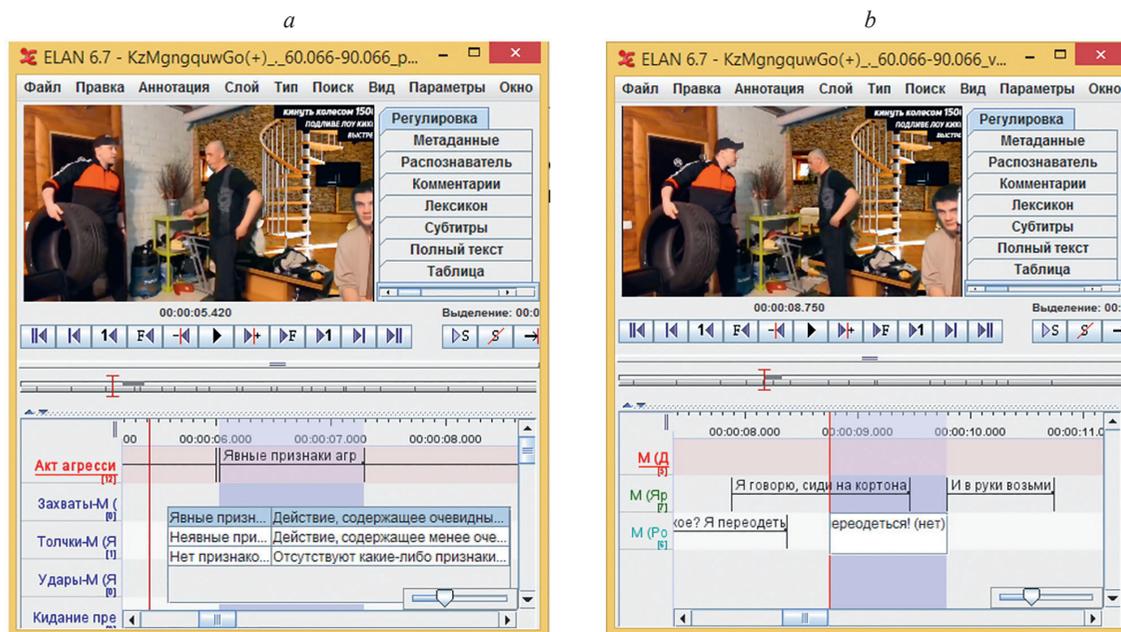


Рис. 3. Разметка физической (а) и вербальной (б) агрессии в программном обеспечении ELAN
 Fig. 3. Labeling of physical (a) and verbal (b) aggression using ELAN annotation tool

согласованность нескольких аннотаторов, которые разметили все экземпляры данных в корпусе.

Ввиду того, что каждый оцениваемый временной сегмент или реплика в случае оценивания физической или вербальной агрессии может содержать более одного класса действий, вычислялась согласованность разметки для физической и вербальной агрессии в целом; сочетаний классов агрессивных действий, присваиваемых разными аннотаторами одним и тем же сегментам; отдельных классов агрессивных действий физической и вербальной агрессии.

Результаты разметки корпуса данных

Предложенная методика была апробирована на практике, в результате чего создан аудиовизуальный корпус данных поведенческой агрессии в ходе онлайн-трансляций Audiovisual Aggressive Behavior in

Online Streams (AVABOS) [23]. Технические характеристики корпуса, приведены в табл. 1.

В связи с тем, что выполнялась коллективная разметка, класс агрессивного действия определялся путем мажоритарного голосования: присваивался класс агрессивного действия на основании совпадений оценок минимум трех аннотаторов из пяти. Статистика по количеству и длительности отдельных действий для каждого подвида вербальной и физической агрессии приведена в табл. 2. Поскольку на каждом оцениваемом сегменте данных может быть более одного типа агрессивных действий, в табл. 2 также приводится время и количество сегментов, где присутствует колонка «Наличие» и отсутствует колонка «Отсутствие» физической и вербальной агрессии.

Результаты оценок согласованности посредством коэффициента Флейсса (1) при совместной разметке физической и вербальной агрессии пятью аннотатора-

Таблица 1. Технические характеристики корпуса данных AVABOS

Table 1. Technical statistics of AVABOS

Параметр	Значение
Общая длительность данных, с	9490
Количество информантов	231 (167 мужчин, 64 женщины)
Количество видеофайлов	222
Общий объем данных, ГБ	1,6
Формат видео	mp4
Разрешение видео, пикселей	от 480 × 270 до 1280 × 720
Частота видеок кадров, кадров/с	от 20 до 30
Формат аудиофайлов	M4A
Квантование аудиосигнала, бит	16
Частота дискретизации аудиосигнала, Гц	44 100

Таблица 2. Статистика данных по длительности и количеству агрессивных действий
Table 2. Time and number of actions statistics

Физическая агрессия										
Класс	Наличие	Отсутствие	Захваты	Толчки	Удары	Кидание предметов	Укусы	Обливание жидкостью	Плевки	Прочие акты физической агрессии
Длительность, с	3626	18 678	2704	258	550	14	10	8	0	37
Количество	2492	9723	1761	179	471	18	9	4	0	21
Вербальная агрессия										
Класс	Наличие	Отсутствие	Оскорбления	Угрозы	Грубые приказы	Обвинения	Поощрение насилия	Прочие речевые акты		
Длительность, с	4807	2581	1132	512	1089	1520	317	223		
Количество	2699	1870	598	251	696	734	175	129		

Таблица 3. Значения степени согласованности оценок аннотаторов к для физической и вербальной агрессии
Table 3. Values of interrater reliability κ for physical and verbal aggression assessment

Для физической агрессии										
Общее	Сочетания	Отсутствие	Захваты	Кидание предметов	Толчки	Удары	Укусы	Обливание жидкостью	Плевки	Прочие акты физической агрессии
0,74	0,68	0,74	0,72	0,53	0,41	0,67	0,87	0,54	0,25	0,09
Для вербальной агрессии										
Общее	Сочетания	Отсутствие	Оскорбления	Угрозы	Грубые приказы	Обвинения	Поощрение насилия	Прочие речевые акты		
0,48	0,48	0,48	0,64	0,69	0,63	0,58	0,50	0,13		

ми приведены в табл. 3. Колонка «Общее» содержит оценки согласованности при разметке наличия или отсутствия агрессии по всему корпусу данных, колонка «Сочетания» содержит оценки согласованности совпадения присваиваемых аннотаторами различных подвидов физической и вербальной агрессии для одних и тех же интервалов, колонка «Отсутствие» содержит оценки согласованности при разметке не агрессивных действий.

Обсуждение

В результате анализа согласованности совместной разметки физической агрессии были получены значения оценок степени согласованности в целом ($\kappa = 0,74$). Такое значение можно интерпретировать как высокий уровень согласованности оценок (диапазон $0,61 \leq \kappa \leq 0,80$ согласно [24, 25]). Также высокий уровень согласованности был достигнут при оценке совпадения сочетаний классов агрессивных действий (0,68), отсутствия агрессии (0,74), отдельных классов «удары» (0,67) и «захваты» (0,72). Совместная разметка класса «толчки» (0,41) показала средний уровень согласованности оценок (диапазон $0,41 \leq \kappa \leq 0,60$ согласно [24, 25]). Таким образом, для классов действий, представленных в корпусе данных наибольшим коли-

чеством (табл. 3), был достигнут высокий или средний уровень согласованности; для мало представленных классов получены высокий уровень согласованности («укусы»), средний («обливание жидкостью») и низкий («плевки», при том, что в корпусе отсутствуют сегменты, размеченные этим классом хотя бы тремя аннотаторами, а также «прочие действия») уровни.

Согласованность разметки вербальной агрессии показала более низкие результаты по сравнению с физической агрессией. Так, общая оценка согласованности при разметке вербальной агрессии (0,48) интерпретируется как средний уровень. Также средний уровень согласованности достигнут при разметке реплик, содержащих вербальную агрессию (0,48), совпадение сочетаний оценок классов агрессивных действий (0,48), классов «поощрение насилия» и «обвинения». При этом классы «оскорбления», «угрозы» и «грубые приказы» достигли высокого уровня согласованности. Отметим, что классы вербальной агрессии представлены в корпусе данных более равномерно, чем классы физической агрессии (табл. 2).

Лучшая, по сравнению с вербальной агрессией, согласованность разметки физической агрессии может быть обусловлена наличием у вербальной агрессии форм, отличных от утвердительной, наличием в корпусе реплик, содержащих нецензурные междометия, а также

различий в интерпретации реплик аннотаторами. Также следует отметить, что в данной работе выполнена оценка спонтанного поведения информантов, что также могло снизить согласованность. В других известных работах, где выполнялась совместная разметка агрессии [1, 2] оценивалось наигранное поведение. В работах, посвященных смежной теме создания корпусов эмоционального поведения [11, 14], были достигнуты схожие показатели согласованности с теми, которые получены в настоящей работе. Необходимо подчеркнуть, что разметка выполнялась специалистами с техническим образованием, которые размечали исходя из выведенных критериев оценки агрессии и алгоритму оценки. Исходя из этого, достигнутые результаты согласованности свидетельствуют о целесообразности применения в разметке выработанных критериев оценки агрессии, выделения единиц анализа, использования разработанного алгоритма оценки.

Заключение

В работе представлена новая методика создания корпусов многомодальных данных для анализа спонтанного агрессивного поведения, отличающаяся содержательным раскрытием критериев оценки физической и вербальной агрессии, алгоритмом экспертной оцен-

ки, позволяющим размечать неочевидные примеры агрессии, выполнением совместной разметки группой аннотаторов спонтанного поведения.

Предложенная методика апробирована на практике, в результате чего создан аудиовизуальный корпус данных поведенческой агрессии AVABOS, содержащий спонтанное агрессивное поведение русскоязычных информантов в ходе прямых видеотрансляций в сети Интернет. Совместная разметка показала высокий уровень согласованности оценок аннотаторов при разметке физической агрессии и средний уровень при разметке вербальной агрессии. При этом при разметке отдельных классов как физической, так и вербальной агрессии также был достигнут высокий уровень согласованности оценок аннотаторов. Результаты оценок степени согласованности подтвердили целесообразность разработки настоящей методики.

Подход может быть использован для создания выборок как агрессивного, так и иного поведения с учетом коррекции на специфику другого типа поведения. Размеченный корпус AVABOS в дальнейшем планируется применять для создания и обучения нейросетевых моделей, предназначенных для одно- и многомодального моделирования и автоматического распознавания случаев агрессивного поведения людей.

Литература

1. Lefter I., Rothkrantz L.J.M., Burghouts G.J. A comparative study on automatic audio-visual fusion for aggression detection using meta-information // *Pattern Recognition Letters*. 2013. V. 34. N 15. P. 1953–1963. <https://doi.org/10.1016/j.patrec.2013.01.002>
2. Lefter I., Burghouts G.J., Rothkrantz L.J.M. An audio-visual dataset of human-human interactions in stressful situations // *Journal on Multimodal User Interfaces*. 2014. V. 8. N 1. P. 29–41. <https://doi.org/10.1007/s12193-014-0150-7>
3. Lefter I., Jonker C.M., Tuentje S.K., Veling W., Bogaerts S. NAA: A multimodal database of negative affect and aggression // *Proc. of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2017. P. 21–27. <https://doi.org/10.1109/ACII.2017.8273574>
4. Sernani P., Falcionelli N., Tomassini S., Contardo P., Dragoni A.F. Deep learning for automatic violence detection: Tests on the AIRTLab dataset // *IEEE Access*. 2021. V. 9. P. 160580–160595. <https://doi.org/10.1109/ACCESS.2021.3131315>
5. Ciampi L., Foszner P., Messina N., Staniszewski M., Gennaro C., Falchi F., Serao G., Cogiel M., Golba D., Szczęśna A., Amato G. Bus violence: An open benchmark for video violence detection on public transport // *Sensors*. 2022. V. 22. N 21. P. 8345. <https://doi.org/10.3390/s22218345>
6. Perez M., Kot A.C., Rocha A. Detection of real-world fights in surveillance videos // *Proc. of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2019. P. 2662–2666. <https://doi.org/10.1109/ICASSP.2019.8683676>
7. Cheng M., Cai K., Li M. RWF-2000: An open large scale video database for violence detection // *Proc. of the 25th International Conference on Pattern Recognition (ICPR)*. 2021. P. 4183–4190. <https://doi.org/10.1109/ICPR48806.2021.9412502>
8. Potapova R., Komalova L. On principles of annotated databases of the semantic field “aggression” // *Lecture Notes in Computer Science*. 2014. V. 8773. P. 322–328. https://doi.org/10.1007/978-3-319-11581-8_40
9. Апанасович К.С., Махныткина О.В., Кабаров В.И., Далеvская О.П. RuPersonaChat: корпус диалогов для персонализации разговорных агентов // *Научно-технический вестник информационных технологий, механики и оптики*. 2024. Т. 24. № 2. С. 214–221. <https://doi.org/10.17586/2226-1494-2024-24-2-214-221>

References

1. Lefter I., Rothkrantz L.J.M., Burghouts G.J. A comparative study on automatic audio-visual fusion for aggression detection using meta-information. *Pattern Recognition Letters*, 2013, vol. 34, no. 15, pp. 1953–1963. <https://doi.org/10.1016/j.patrec.2013.01.002>
2. Lefter I., Burghouts G.J., Rothkrantz L.J.M. An audio-visual dataset of human-human interactions in stressful situations. *Journal on Multimodal User Interfaces*, 2014, vol. 8, no. 1, pp. 29–41. <https://doi.org/10.1007/s12193-014-0150-7>
3. Lefter I., Jonker C.M., Tuentje S.K., Veling W., Bogaerts S. NAA: A multimodal database of negative affect and aggression. *Proc. of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2017, pp. 21–27. <https://doi.org/10.1109/ACII.2017.8273574>
4. Sernani P., Falcionelli N., Tomassini S., Contardo P., Dragoni A.F. Deep learning for automatic violence detection: Tests on the AIRTLab dataset. *IEEE Access*, 2021, vol. 9, pp. 160580–160595. <https://doi.org/10.1109/ACCESS.2021.3131315>
5. Ciampi L., Foszner P., Messina N., Staniszewski M., Gennaro C., Falchi F., Serao G., Cogiel M., Golba D., Szczęśna A., Amato G. Bus violence: An open benchmark for video violence detection on public transport. *Sensors*, 2022, vol. 22, no. 21, pp. 8345. <https://doi.org/10.3390/s22218345>
6. Perez M., Kot A.C., Rocha A. Detection of real-world fights in surveillance videos. *Proc. of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 2662–2666. <https://doi.org/10.1109/ICASSP.2019.8683676>
7. Cheng M., Cai K., Li M. RWF-2000: An open large scale video database for violence detection. *Proc. of the 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 4183–4190. <https://doi.org/10.1109/ICPR48806.2021.9412502>
8. Potapova R., Komalova L. On principles of annotated databases of the semantic field “aggression”. *Lecture Notes in Computer Science*, 2014, vol. 8773, pp. 322–328. https://doi.org/10.1007/978-3-319-11581-8_40
9. Apanasovich K.S., Makhnytkina O.V., Kabarov V.I., Dalevskaya O.P. RuPersonaChat: a dialog corpus for personalizing conversational agents. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 2, pp. 214–221. (in Russian). <https://doi.org/10.17586/2226-1494-2024-24-2-214-221>

10. Hassoun Al-Jawad M.M., Alharbi H., Almkhtar A.F., Alnawas A.A. Constructing twitter corpus of Iraqi Arabic Dialect (CIAD) for sentiment analysis // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22. № 2. С. 308–316. <https://doi.org/10.17586/2226-1494-2022-22-2-308-316>
11. Busso C., Bulut M., Lee C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S., Narayanan S.S. IEMOCAP: Interactive emotional dyadic motion capture database // Language Resources and Evaluation. 2008. V. 42. N 4. P. 335–359. <https://doi.org/10.1007/s10579-008-9076-6>
12. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing // Lecture Notes in Computer Science. 2018. V. 11096. P. 501–510. https://doi.org/10.1007/978-3-319-99579-3_52
13. Ringeval F., Sonderegger A., Sauer J., Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions // Proc. of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). 2013. P. 1–8. <https://doi.org/10.1109/FG.2013.6553805>
14. Busso C., Parthasarathy S., Burmania A., AbdelWahab M., Sadoughi N., Provost E.M. MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception // IEEE Transactions on Affective Computing. 2017. V. 8. N 1. P. 67–80. <https://doi.org/10.1109/TAFFC.2016.2515617>
15. Ениколопов С.Н. Понятие агрессии в современной психологии // Прикладная психология. 2001. № 1. С. 60–72.
16. Groth-Marnat G., Wright A.J. Handbook of Psychological Assessment. John Wiley & Sons, 2016. 824 p.
17. Uzdiaev M., Vatamaniuk I. Investigation of manifestations of aggressive behavior by users of sociocyberphysical systems on video // Lecture Notes in Networks and Systems. 2021. V. 231. P. 593–604. https://doi.org/10.1007/978-3-030-90321-3_49
18. Buss A.H. The Psychology of Aggression. John Wiley & Sons, 1961. 307 p. <https://doi.org/10.1037/11160-000>
19. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision // International conference on machine learning (PMLR). 2023. V. 202. P. 28492–28518.
20. Plaquet A., Bredin H. Powerset multi-class cross entropy loss for neural speaker diarization // Proc. of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 2023. P. 3222–3226. <https://doi.org/10.21437/Interspeech.2023-205>
21. Lausberg H., Sloetjes H. Coding gestural behavior with the NEUROGES-ELAN system // Behavior Research Methods. 2009. V. 41. N 3. P. 841–849. <https://doi.org/10.3758/BRM.41.3.841>
22. Fleiss J.L. Measuring nominal scale agreement among many raters // Psychological Bulletin. 1971. V. 76. N 5. P. 378–382. <https://doi.org/10.1037/h0031619>
23. Уздяев М.Ю., Карпов А.А. Аудиовизуальный корпус данных поведенческой агрессии в ходе онлайн трансляций (Audiovisual Aggressive Behavior in Online Streams dataset – AVABOS). Свидетельство о государственной регистрации базы данных № 2022623239. 2022.
24. Landis J.R., Koch G.G. The measurement of observer agreement for categorical data // Biometrics. 1977. V. 33. N 1. P. 159–174. <https://doi.org/10.2307/2529310>
25. Fleiss J.L., Levin B., Paik M.C. Statistical Methods for Rates and Proportions. John Wiley & Sons, 2013. 800 p.
10. Hassoun Al-Jawad M.M., Alharbi H., Almkhtar A.F., Alnawas A.A. Constructing twitter corpus of Iraqi Arabic Dialect (CIAD) for sentiment analysis. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, vol. 22, no. 2, pp. 308–316. <https://doi.org/10.17586/2226-1494-2022-22-2-308-316>
11. Busso C., Bulut M., Lee C., Kazemzadeh A., Mower E., Kim S., Chang J.N., Lee S., Narayanan S.S. IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 2008, vol. 42, no. 4, pp. 335–359. <https://doi.org/10.1007/s10579-008-9076-6>
12. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. *Lecture Notes in Computer Science*, 2018, vol. 11096, pp. 501–510. https://doi.org/10.1007/978-3-319-99579-3_52
13. Ringeval F., Sonderegger A., Sauer J., Lalanne D. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *Proc. of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–8. <https://doi.org/10.1109/FG.2013.6553805>
14. Busso C., Parthasarathy S., Burmania A., AbdelWahab M., Sadoughi N., Provost E.M. MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception. *IEEE Transactions on Affective Computing*, 2017, vol. 8, no. 1, pp. 67–80. <https://doi.org/10.1109/TAFFC.2016.2515617>
15. Enikolopov S.N. The concept of aggression in the contemporary psychology. *Prikladnaja psihologija*, 2001, no. 1, pp. 60–72. (in Russian)
16. Groth-Marnat G., Wright A.J. *Handbook of Psychological Assessment*. John Wiley & Sons, 2016, 824 p.
17. Uzdiaev M., Vatamaniuk I. Investigation of manifestations of aggressive behavior by users of sociocyberphysical systems on video. *Lecture Notes in Networks and Systems*, 2021, vol. 231, pp. 593–604. https://doi.org/10.1007/978-3-030-90321-3_49
18. Buss A.H. *The Psychology of Aggression*. John Wiley & Sons, 1961, 307 p. <https://doi.org/10.1037/11160-000>
19. Radford A., Kim J.W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust speech recognition via large-scale weak supervision. *International conference on machine learning (PMLR)*, 2023, vol. 202, pp. 28492–28518.
20. Plaquet A., Bredin H. Powerset multi-class cross entropy loss for neural speaker diarization. *Proc. of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2023, pp. 3222–3226. <https://doi.org/10.21437/Interspeech.2023-205>
21. Lausberg H., Sloetjes H. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 2009, vol. 41, no. 3, pp. 841–849. <https://doi.org/10.3758/BRM.41.3.841>
22. Fleiss J.L. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 1971, vol. 76, no. 5, pp. 378–382. <https://doi.org/10.1037/h0031619>
23. Uzdiaev M.Iu., Karpov A.A. Audiovisual Aggressive Behavior in Online Streams dataset – AVABOS. *Certificate of state registration of the database 2022623239*, 2022. (in Russian)
24. Landis J.R., Koch G.G. The measurement of observer agreement for categorical data. *Biometrics*, 1977, vol. 33, no. 1, pp. 159–174. <https://doi.org/10.2307/2529310>
25. Fleiss J.L., Levin B., Paik M.C. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, 2013, 800 p.

Авторы

Уздяев Михаил Юрьевич — младший научный сотрудник, Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, [sc 57211179414](https://orcid.org/0000-0002-7032-0291), <https://orcid.org/0000-0002-7032-0291>, uzdyae.m@iias.spb.su

Карпов Алексей Анатольевич — доктор технических наук, профессор, руководитель лаборатории, Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, [sc 57219469958](https://orcid.org/0000-0003-3424-652X), <https://orcid.org/0000-0003-3424-652X>, karpov@iias.spb.su

Authors

Mikhail Yu. Uzdiaev — Junior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, [sc 57211179414](https://orcid.org/0000-0002-7032-0291), <https://orcid.org/0000-0002-7032-0291>, uzdyae.m@iias.spb.su

Alexey A. Karpov — D.Sc., Professor, Head of Laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, [sc 57219469958](https://orcid.org/0000-0003-3424-652X), <https://orcid.org/0000-0003-3424-652X>, karpov@iias.spb.su

Статья поступила в редакцию 12.08.2024
Одобрена после рецензирования 24.08.2024
Принята к печати 27.09.2024

Received 12.08.2024
Approved after reviewing 24.08.2024
Accepted 27.09.2024



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»