УДК 004.048
DOI: 10.17586/0021-3454-2024-67-11-984-993

МЕТОДИКА СОЗДАНИЯ МНОГОМОДАЛЬНЫХ КОРПУСОВ ДАННЫХ ДЛЯ АУДИОВИЗУАЛЬНОГО АНАЛИЗА ВОВЛЕЧЕННОСТИ И ЭМОЦИЙ УЧАСТНИКОВ ВИРТУАЛЬНОЙ КОММУНИКАЦИИ

А. А. Двойникова*, А. А. Карпов

Санкт-Петербургский федеральный исследовательский центр Российской академии наук, СанктПетербург, Россия

** dvoynikova.a@iias.spb.su*

Аннотация. Представлена методика создания многомодальных корпусов данных, предназначенных для анализа поведенческих проявлений участников виртуальной коммуникации. Предложенная методика направлена на создание корпусов данных групповой коммуникации (более двух собеседников) с использованием систем телеконференций и учитывает особенности естественных проявлений поведенческих аспектов (вовлеченности и эмоций) участников разговора. Выделенные особенности составляют новизну предложенной методики. Методика состоит из трех основных этапов — подготовительного, записи и аннотирования данных. Методика была апробирована и валидирована при создании нового многомодального корпуса данных ENERGI, содержащего русскоязычные аудиовизуальные записи групповой коммуникации участников с помощью систем телеконференций. Созданный корпус предназначен для решения задач распознавания вовлеченности участников в коммуникацию, а также анализа проявления эмоций во время диалога. Предложенная методика является универсальной и может быть применима для сбора различных корпусов данных виртуальной коммуникации.

Ключевые слова: *методика создания корпусов данных, многомодальный корпус, анализ вовлеченности, анализ эмоций, аннотирование данных, виртуальная коммуникация*

Благодарности: работа выполнена в рамках бюджетной темы № FFZF-2022-0005.

Ссылка для цитирования: *Двойникова А. А., Карпов А. А.* Методика создания многомодальных корпусов данных для аудиовизуального анализа вовлеченности и эмоций участников виртуальной коммуникации // Изв. вузов. Приборостроение. 2024. Т. 67, № 11. С. 984–993. DOI: 10.17586/0021-3454-2024-67-11-984-993.

METHOD OF CREATING MULTIMODAL DATABASES FOR AUDIOVISUAL ANALYSIS OF ENGAGEMENT AND EMOTIONS OF VIRTUAL COMMUNICATION PARTICIPANTS

A. A. Dvoynikova*, A. A. Karpov

St. Petersburg Federal Research Center of the RAS, St. Petersburg, Russia

** dvoynikova.a@iias.spb.su*

Abstract: A method is presented for creating multimodal data bases designed to analyze behavioral manifestations of virtual communication participants. The proposed methodology is aimed at developing database of group communication (more than two interlocutors) using teleconference systems. The technique also takes into account the peculiarities of the natural manifestations of behavioral aspects (engagement and emotions) of the participants in the conversation. The identified features constitute the novelty of the proposed technique. The technique consists of three main stages — preparatory, recording, and annotation of data. The technique is tested and validated when creating a new multimodal data corpus ENERGI, containing Russian-language audiovisual recordings of group communication of participants using teleconferencing systems. The created corpus is designed to solve the problems of recognizing the involvement of participants in communication, as well as analyzing the manifestation of emotions during a dialogue. The proposed technique is universal and can be applied to collecting various corpora of virtual communication data.

Keywords: methodology for database creating, multimodal database, engagement analysis, emotion analysis, data annotation, virtual communication

Acknowledgments: the work was carried out within the framework of budget topic No. FFZF-2022-0005.

For citation: Dvoynikova A. A., Karpov A. A. Method of creating multimodal databases for audiovisual analysis of engagement and emotions of virtual communication participants. *Journal of Instrument Engineering*. 2024. Vol. 67, N 11. P. 984–993 (in Russian). DOI: 10.17586/0021-3454-2024-67-11-984-993.

Введение. В последнее годы сотрудники различных организаций, студенты и учащиеся стали активно использовать системы телеконференций. Во время виртуальной коммуникации понимание поведенческих аспектов собеседников затруднено. В процессе такого общения человеку сложно следить за невербальными проявлениями поведения сразу нескольких людей, к тому же системы телеконференций вносят дополнительные затруднения в виде ограниченного окном монитора видеоизображения собеседников. Автоматическая система распознавания невербальных поведенческих проявлений (эмоций собеседника и его вовлеченности в коммуникацию) может помочь анализировать групповую динамику разговора [1]. Для обучения вероятностных моделей такой системы необходимо иметь базу данных, состоящую из видеозаписей проявлений поведенческих аспектов для анализа вовлеченности и эмоций собеседников групповой виртуальной коммуникации.

Для анализа вовлеченности участников телеконференций существует несколько корпусов аудиовизуальных данных, например: NoXi [2], MEDICA [3], MHHRI [4], RECOLA [5], EngageWild [6], DAiSEE [7], Sümer [8], Whitehill [9], Psaltis [10] и др. Описание перечисленных корпусов представлено в работе [11]. Некоторые корпуса [6–10] представляют собой набор видеоданных, в которых участники молча смотрят видеоизображения на экране. Несколько баз данных [4, 8] отображают очное взаимодействие участников между собой без использования телеконференций. Участники всех указанных корпусов общаются между собой на английском, немецком или французском языке. Однако не существует подходящих баз данных русскоязычных диалогов. Также отсутствуют корпуса, содержащие групповую коммуникацию, все корпуса отображают общение между двумя людьми.

Многомодальных корпусов данных для моделирования и распознавания эмоций существует значительно больше [12], например: AFEW [13], Aff-Wild2 [14], IEMOCAP [15], Meld [16], CMU-MOSEI [17] и другие, в том числе единственный русскоязычный корпус RAMAS [18]. Большинство корпусов [15–18] содержат аудиовизуальные проявления эмоций, наблюдаемые в лабораторных условиях (по сценариям). Сбор данных о естественном проявлении эмоций участников является затруднительным, поэтому корпусов, содержащих такие данные, очень мало [13, 14]. Одна из основных проблем существующих эмоциональных аудиовизуальных корпусов заключается в некачественной разметке данных. Практически во всех корпусах аннотаторами данных являются люди без профильного образования в области психологии, которые размечают данные на основании субъективного взгляда на проявление эмоций. Известны только два корпуса (NoXi и RECOLA), в которых содержится разметка данных одновременно по вовлеченности и эмоциям собеседников во время коммуникации.

Таким образом, в связи с описанными выше проблемами, существует необходимость создания нового корпуса аудиовизуальных данных, который должен удовлетворять следующим критериям:

— содержать многомодальные данные (по крайней мере, видео- и аудиоданные) групповой коммуникации (двух и более людей);

— данные должны быть размечены по поведенческим аспектам (например, вовлеченности и эмоциям) коммуникации между людьми, разметка данных должна производиться экспертами в исследуемой области;

— поведенческие аспекты собеседников должны проявляться в натурных условиях, а также должен отсутствовать эффект Хоторна [19] при записи данных;

— собеседники должны вести спонтанный разговор на грамотном русском языке;

— групповая коммуникация должна происходить с использованием системы телеконференций.

Актуальность создания нового корпуса данных, соответствующего приведенным условиям, сопровождается задачей разработки универсальной методики, пригодной также и для создания других подобных корпусов.

Методика создания многомодального корпуса данных. На рис. 1 представлена методика создания многомодальных корпусов данных для многомодального анализа групповой виртуальной коммуникации собеседников.

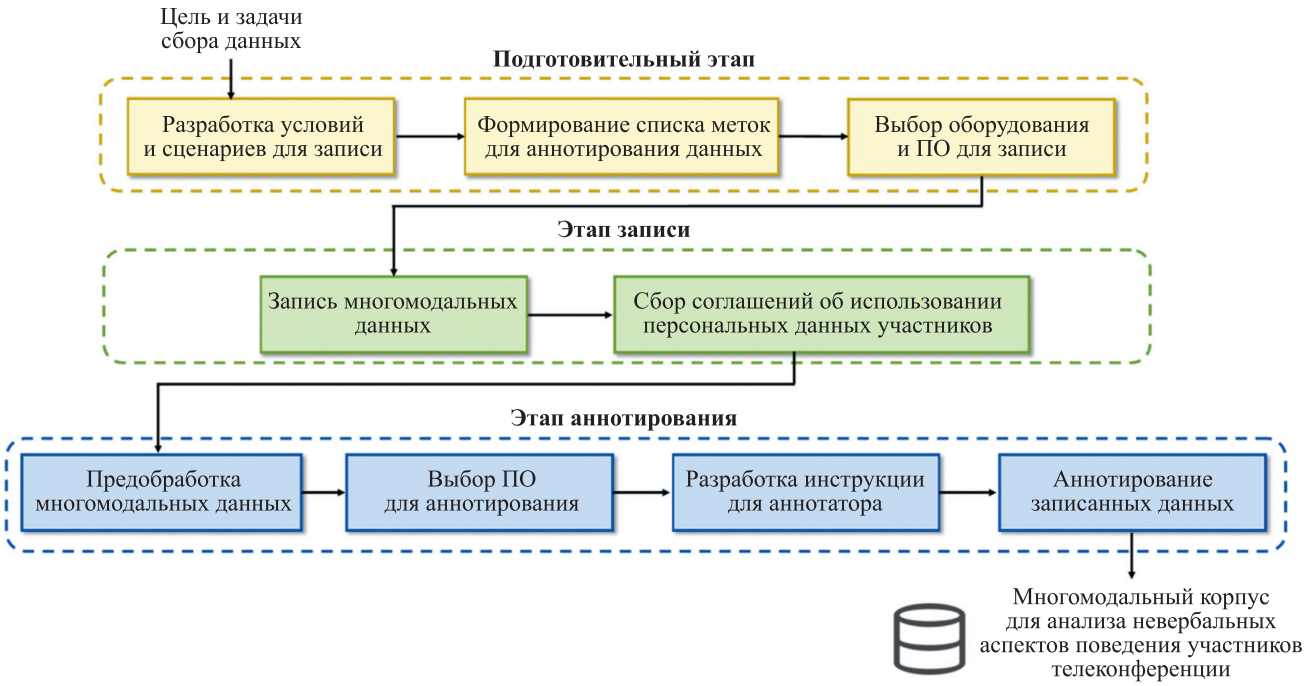

Рис. 1

Предложенная методика состоит из трех этапов: подготовительного, этапа записи и аннотирования данных.

Подготовительный этап *заключается в разработке сценариев*, содержащих описание условий проведения экспериментов, данные о количестве участников и их взаимодействии. Важным условием для записи корпусов данных является естественное поведение дикторов в натурных условиях.

Далее приведены примеры сценариев для групповой коммуникации людей в телеконференции:

1) лекции — особенностью данного сценария является наличие главного диктора, который активно что-то рассказывает, объясняет: например, лекция преподавателя для группы студентов;

2) семинары — все собеседники должны активно вести разговор; примеры семинара — коллективный мозговой штурм, обсуждение работы ученика учителем и самим учеником;

3) просмотр видеоконтента — данный сценарий представляет человекомашинное взаимодействие, при котором человек просматривает видеоинформацию на экране: например, просмотр студентом онлайн-курса.

На подготовительном этапе необходимо также обозначить *список меток для аннотирования*, которые должны соответствовать целям и задачам корпуса данных. Одним из возможных вариантов анализа поведенческих аспектов групповой коммуникации является распознавание проявлений степени вовлеченности участников в разговор и их эмоций. Для того чтобы выделить свойства проявлений степени вовлеченности был произведен опрос экспертов. Выбор экспертов осуществлялся по принципу наличия высшего психологического образования, при этом они являлись действующими педагогами в различных предметных областях. Были опрошены три эксперта, после чего их ответы были агрегированы. Исходя из полученных ответов было сформировано определение вовлеченности как степени сфокусированности внимания человека на происходящем. На основе мнений экспертов и информации, представленной в работах [6, 8, 9], были сформированы свойства проявлений (метки аннотирования) степени вовлеченности участников в телеконференцию (табл. 1).

Также были выделены свойства степени активации и валентности эмоций участников групповой коммуникации, которые представлены в табл. 2.

Таблица 1

Свойство проявления вовлеченности	Степень вовлеченности		
	высокая	средняя	низкая
	диктор или активный слушатель	пассивный слушатель	слушатель, не участвующий в коммуникации
Речь	Активно говорит, речь эмоциональнее, чем обычно. Задает вопросы и дает развернутые ответы. Наблюдается звуковая паравербалика, смех	Дает пассивные краткие ответы	Общается с другими людьми (вне конференции)
Взгляд	Взгляд направлен в монитор	Смотрит в монитор, иногда отводит взгляд в сторону	Бегающий взгляд, часто отводит взгляд от монитора
Жесты	Жестикулирует, кивает/покачивает головой. Осуществляет движение запястья в сторону экрана	Закрывает рот рукой, подпирает голову рукой	Зевает, потягивается
Мимика	Приоткрыт рот, легкая улыбка, движение бровями	Мышцы лица расслаблены	Глаза закрыты
Поза тела	Корпус тела наклонен вперед (к экрану)	Закрытая поза тела/рук (скрещивание рук, руки вместе)	Ложится на стол, корпус тела повернут в сторону, крутится на стуле
Другое	—	Делает записи/пометки	Использует другие предметы (телефон, другой монитор), участник покидает область видимости камеры

Таблица 2

Свойство проявления эмоций	Степень проявления эмоций		
Валентность	Негативная	Нейтральная	Позитивная
	Выражение мимикой таких эмоций, как грусть, отвращение, удивление отрицательное, страх, гнев	Мышцы лица расслаблены	Выражение мимикой таких эмоций, как радость, удивление положительное
Активация	Низкая	Средняя	Высокая
	Отсутствие эмоций	Незначительное проявление эмоций (улыбка)	Яркое выражение эмоций (смех)

В случаях когда участник телеконференции не включил камеру и микрофон, анализ характеристик проявления его эмоций становится невозможным. Поэтому в таких случаях выделяется отдельный класс — „нет аннотации“. Также для эффективного анализа степени вовлеченности участников телеконференции необходимо учитывать второстепенные метки — жесты коммуникации. Категория „жесты“ включает в себя следующие метки: рука подпирает голову, подбородок; рука закрывает глаза; зевота; кивок и покачивание головой; скрещенные руки; рука зажимает нос; нет жестов.

Следующее действие подготовительного этапа — *выбор оборудования и программного обеспечения (ПО)* для записи многомодальных данных. На сегодняшний день существует большое количество систем телеконференций: Telegram (https://telegram.org/), Яндекс Телемост (https://telemost.yandex.ru/), Webinar.ru (https://webinar.ru), Microsoft Teams (https://webinar.ru),

Skype (https://www.skype.com/), Zoom (https://zoom.us/) и др. Сравнив технические особенности и возможности лицензионных версий перечисленных систем, можно сделать вывод, что система Zoom является наиболее подходящей для групповых видеозвонков и записи базы данных по предложенным сценариям. Лицензионная версия Zoom позволяет совершать групповые переговоры длительностью 30 ч для 100 присутствующих, сохранять субтитры речи на русском языке. Основное преимущество Zoom — сохранение отдельных аудиодорожек конференции для каждого диктора, с помощью этой функции происходит автоматическая диаризация дикторов. Также в Zoom обеспечивается автоматическое шумоподавление во время диалогов участников.

Этап записи данных состоит из двух блоков действий: запись многомодальных данных и сбор соглашений об использовании персональных данных участников. Запись производится в условиях, соответствующих сценариям и с применением выбранного технического обеспечения. При записи корпуса данных необходимо использовать реальные кейсы и ситуации, например такие, как совещание, образовательный процесс студентов и другие, с использованием программного обеспечения, привычного для всех участников телеконференции. При таких условиях участники чувствуют себя естественно и у них не возникает эффекта Хоторна [19] — эффекта наблюдателя, при котором участник изменением своего поведения реагирует на факт осознания того, что за ним наблюдают.

Этап аннотирования данных состоит из блока предобработки данных, выбора ПО для аннотирования, создания инструкции для аннотаторов и собственно аннотирования данных. *Предобработка записанных данных* необходима для построения качественных моделей распознавания коммуникативного поведения собеседников. Предобработка данных включает в себя диаризацию дикторов, а также выделение области видеоконтента каждого участника телеконференции. С помощью хэширования кадров, вычисления каскадов Хаара [20] и преобразования Хафа [21] можно выделить область видеоконтента отдельно для каждого участника групповой коммуникации в Zoom. Результаты такой предобработки показаны на рис. 2: *а* — исходные видеоданные, *б* — обработанные видеокадры.

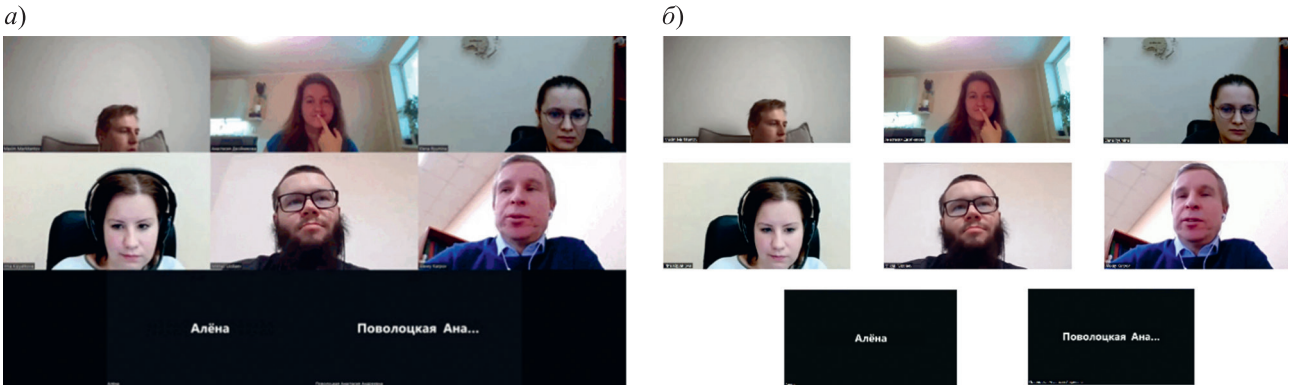

Рис. 2

Диаризация дикторов происходит автоматически во время записи данных с помощью модулей ПО Zoom.

Выбор ПО для аннотирования данных и создание инструкций для аннотаторов являются важными шагами для получения качественной аннотации данных. Аннотирование данных можно проводить с помощью ПО ELAN [22]; пример использования ПО ELAN (скриншот диалогового окна) показан на рис. 3.

Аннотаторами (для данных, содержащих поведенческие проявления собеседников) должны быть носители русского языка, которые прошли тесты на эмоциональный интеллект (опросник эмоционального интеллекта ЭмИн [23], видеотест на распознавание эмоций [24]) с результатом выше 7 баллов по 10-балльной шкале.

Последний этап методики создания многомодального корпуса — ***аннотирование данных*** — является самым трудоемким, так как требует значительных временных затрат для разметки данных вручную, а также высокой концентрации и внимания аннотаторов.

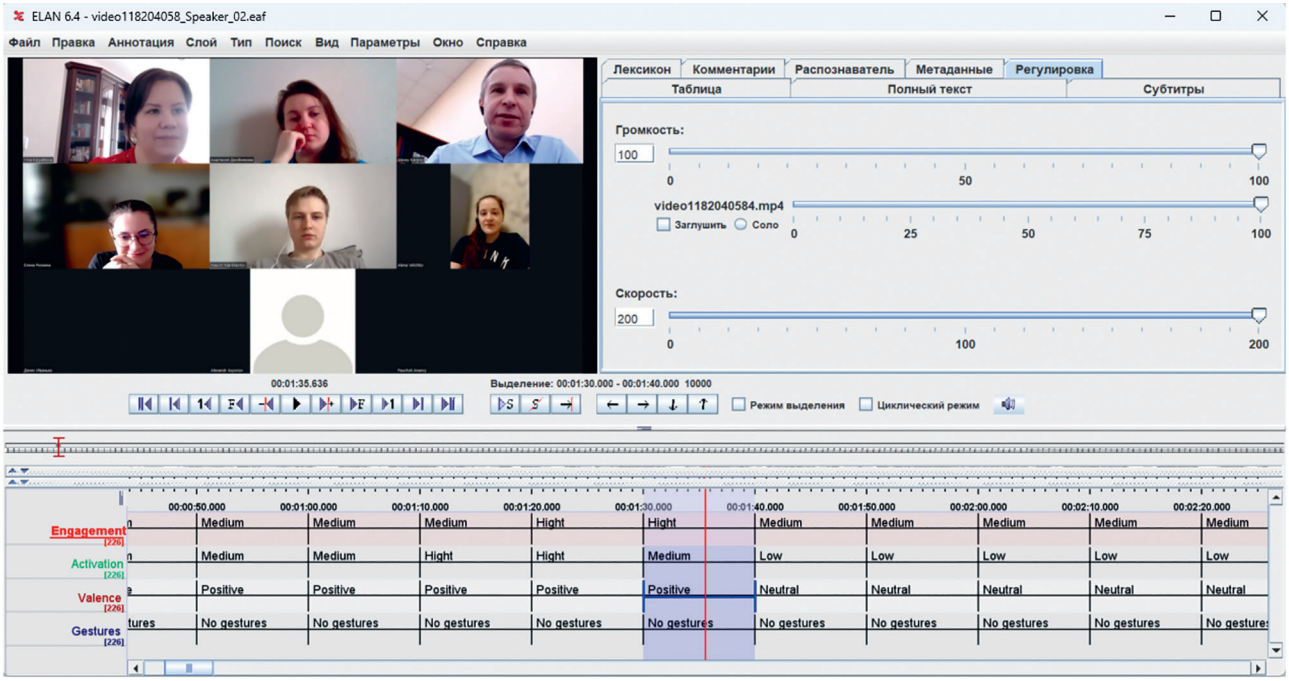

Рис. 3

Корпус данных ENERGI. С помощью описанной выше методики создания корпусов данных был собран многомодальный корпус ENgagement and Emotion Russian Gathering Interlocutors — ENERGI [25]. ENERGI содержит русскоязычные аудиовизуальные записи групповой коммуникации людей с помощью систем телеконференций. Данный корпус предназначен для решения задач распознавания вовлеченности участников в коммуникацию, а также анализа проявления эмоций во время беседы. Примеры кадров корпуса ENERGI представлены на рис. 2 и 3. Ниже приведены характеристики корпуса ENERGI.

Общее количество участников	75
Среднее число участников телеконференции	≈ 7
Общее количество телеконференций	88
Средняя продолжительность телеконференции, мин	≈ 40
Общая длительность видеоконтента, ч	≈ 58,5
Общий объем данных, Гб	≈ 115
Формат видеозаписи	mp4
Разрешение видеокадра	От 640 × 360 до 3120 × 2080
Частота кадров, кадр/с	25
Формат аудиоданных	m4a
Частота дискретизации аудиоданных, Гц	32000

На сегодняшний день аннотировано 6 ч видеозаписей корпуса ENERGI. На рис. 4 показано распределение размеченных данных по классам вовлеченности, на рис. 5 — распределение данных по классам эмоций.

Анализ рисунков показывает, что распределение данных является несбалансированным. Преобладание низкой активации эмоций и нейтральной валентности обусловливается

спецификой сценария корпуса данных. Одним из условий при создании корпуса был сбор данных в естественных условиях и отсутствии эффекта Хоторна. В естественных условиях при решении рабочих задач люди не так активно проявляют эмоции, чаще они находятся в нейтральном состоянии. Именно этим объясняется полученный дисбаланс классов эмоций.

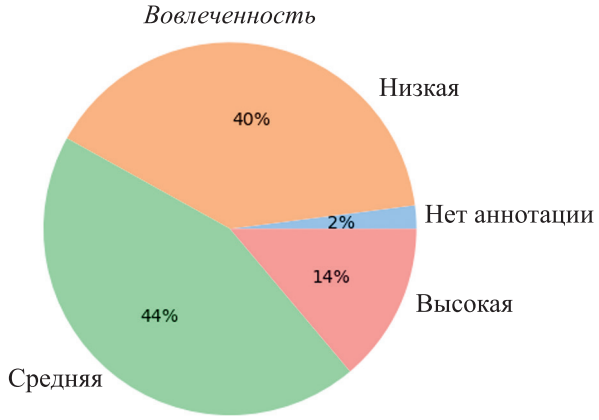

Рис. 4

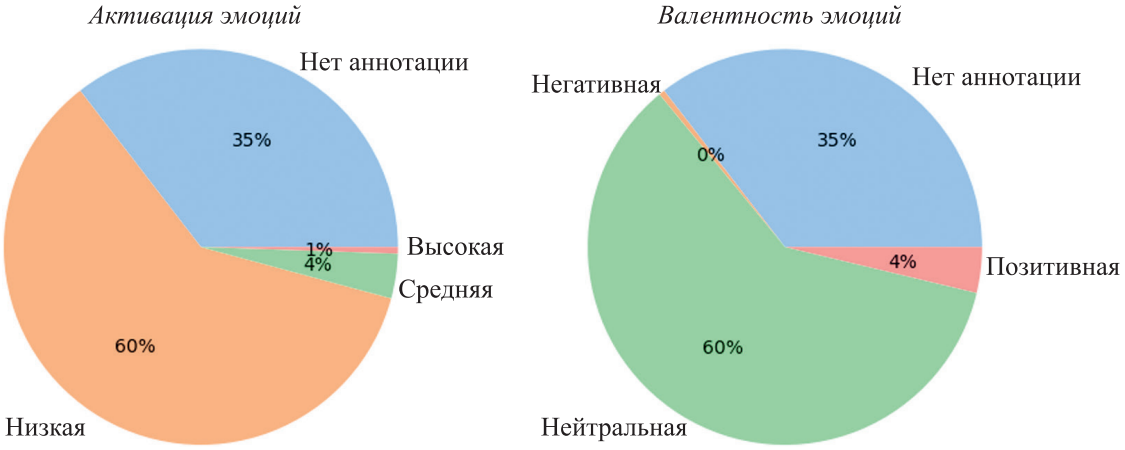

Рис. 5

Заключение. Предложена новая методика создания многомодальных корпусов данных для анализа поведенческих проявлений участников групповой виртуальной коммуникации. Примерами поведенческих аспектов собеседников являются их вовлеченность в разговор и естественные эмоции. Выделены свойства классов вовлеченности, активации и валентности эмоций. Предложенная методика является универсальной и может быть использована для создания различных корпусов данных виртуальной коммуникации. Новизна методики заключается в сборе данных групповой коммуникации, а также в том, что проявление поведенческих аспектов участников происходит в натурных условиях. Методика была апробирована при создании нового многомодального корпуса данных ENERGI, который содержит русскоязычные данные групповой виртуальной коммуникации онлайн-участников.

Направление дальнейших исследований — продолжение аннотирования данных корпуса. Новый корпус данных ENERGI будет применим для разработки системы анализа вовлеченности и эмоций участников виртуальной коммуникации.

СПИСОК ЛИТЕРАТУРЫ

1. *Ткаченя А. В., Давыдов А. Г., Киселёв В. В., Хитров М. В.* Классификация эмоционального состояния диктора с использованием метода опорных векторов и критерия Джини // Изв. вузов. Приборостроение. 2013. Т. 56, № 2. С. 61–66.

2. *Cafaro A., Wagne, J., Baur T., Dermouche S., Torres Torres M*. et al. The NoXi database: multimodal recordings of mediated novice-expert interactions // Proc. of the 19th ACM Intern. Conf. on Multimodal Interaction. 2017. P. 350–359. DOI: 10.1145/3136755.313678.
3. *Guhan P., Agarwal M., Awasthi N., Reeves G., Manocha D.* et al. ABC-Net: Semi-Supervised Multimodal GAN-based Engagement Detection using an Affective, Behavioral and Cognitive Model // arXiv preprint arXiv:2011.08690. 2020.
4. *Celiktutan O., Skordos E., Gunes H*. Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement //IEEE Trans. on Affective Computing. 2017. Vol. 10, N 4. P. 484–497. DOI: 10.1109/TAFFC.2017.2737019.
5. *Ringeval F., Sonderegger A., Sauer J., Lalanne D*. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions // Proc. of the 10th IEEE Intern. Conf. and Workshops on Automatic Face and Gesture Recognition. 2013. P. 1–8. DOI: 10.1109/FG.2013.6553805.
6. *Kaur A., Mustafa A., Mehta L., Dhall A*. Prediction and localization of student engagement in the wild // Digital Image Computing: Techniques and Applications (DICTA). 2018. P. 1–8. DOI: 10.1109/DICTA.2018.8615851.
7. *Gupta A., D'Cunha A., Awasthi K., Balasubramanian V*. DAiSEE: Towards user engagement recognition in the wild // arXiv preprint arXiv:1609.01885. 2016.
8. *Sümer Ö., Goldberg P., D'Mello S., Gerjets P., Trautwein U., Kasneci E.* Multimodal engagement analysis from facial videos in the classroom // IEEE Trans. on Affective Computing. 2021. Vol. 14, N 2. P. 1012–1027. DOI: 10.1109/TAFFC.2021.3127692.
9. *Whitehill J., Serpell Z., Lin Y. C., Foster A., Movellan J. R*. The faces of engagement: Automatic recognition of student engagementfrom facial expressions // IEEE Trans. on Affective Computing. 2014. Vol. 5, N 1. P. 86–98. DOI: 10.1109/TAFFC.2014.2316163.
10. *Psaltis A., Apostolakis K. C., Dimitropoulos K., Daras P*. Multimodal student engagement recognition in prosocial games // IEEE Trans. on Games. 2017. Vol. 10, N 3. P. 292–303. DOI: 10.1109/TCIAIG.2017.2743341.
11. *Двойникова А. А., Кагиров И. А., Карпов А. А*. Аналитический обзор методов автоматического распознавания вовлеченности пользователя в виртуальную коммуникацию // Информационно-управляющие системы. 2022. № 5(120). С. 12–22. DOI: 10.31799/1684-8853-2022-5-12-22.
12. *Двойникова А. А., Маркитантов М. В., Рюмина Е. В., Уздяев М. Ю., Величко А. Н.* и др. Анализ информационного и математического обеспечения для распознавания аффективных состояний человека // Информатика и автоматизация. 2022. Т. 21, № 6. С. 1097–1144. DOI: 10.15622/ia.21.6.2.
13. *Dhall A., Goecke R., Gedeon T*. Collecting large, richly annotated facial-expression databases from movies // Journal of Latex Class Files. 2007. Vol. 6, N 1.
14. *Kollias D., Zafeiriou S.* Aff-wild2: Extending the aff-wild database for affect recognition // arXiv preprint arXiv:1811.07770. 2018.
15. *Busso C., Bulut M., Lee C. C., Kazemzadeh A., Mower E*. et al. IEMOCAP: Interactive emotional dyadic motion capture database // Language Resources and Evaluation. 2008. Vol. 42, N 4. P. 335–359. DOI: 10.1007/s10579-008-9076-6.
16. *Poria S., Hazarika D., Majumder N., Naik G., Cambria E*. et al. Meld: A multimodal multi-party dataset for emotion recognition in conversations // Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. P. 527–536.
17. *Zadeh A. B., Liang P. P., Poria S., Cambria E., Morency L. P.* Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph // Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. 2018. P. 2236–2246. DOI: 10.18653/v1/P18-1208.
18. *Perepelkina O., Kazimirova E., Konstantinova M*. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing // Proc. of the Intern. Conf. on Speech and Computer. 2018. P. 501–510. DOI: 10.1007/978-3-319-99579-3_52.
19. *Jones S. R. G*. Was there a Howthorne effect? // American Journal of Sociology. 1992. Vol. 98, N 3. P. 451–468.
20. *Viola P., Jones M*. Rapid Object Detection using a Boosted Cascade of Simple Features // Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. (CVPR). 2001. Vol. 1. P. I–I. DOI: 10.1109/CVPR.2001.990517.
21. Pat. 3069654 USA. Method and means for recognizing complex patterns / *P. V. C. Hough.* 1962 [Электронный ресурс]: https://patents.google.com/patent/US3069654.
22. *Lausberg H., Sloetjes H.* Coding gestural behavior with the NEUROGES-ELAN system // Behavior Research Methods. 2009. Vol. 41, N 3. P. 841–849. DOI: 10.3758/BRM.41.3.841.
23. *Люсин Д. В*. Новая методика для измерения эмоционального интеллекта: опросник ЭмИн // Психологическая диагностика. 2006. Т. 4. С. 3–22.

24. *Люсин Д. В., Овсянникова В. В.* Измерение способности к распознаванию эмоций с помощью видеотеста // Психологический журнал. 2013. Т. 34, № 6. С. 82–94.

25. Свид. о рег. № 2023624954. База данных проявлений вовлеченности и эмоций русскоязычных участников телеконференций (ENERGI — ENgagement and Emotion Russian Gathering Interlocutors) / *А. А. Двойникова, А. А. Карпов.* 25.12.2023.

СВЕДЕНИЯ ОБ АВТОРАХ

Анастасия Александровна Двойникова — СПбФИЦ РАН, СПИИРАН, лаборатория речевых и многомодальных интерфейсов; мл. научный сотрудник;
E-mail: dvoynikova.a@iias.spb.su

Алексей Анатольевич Карпов — д-р техн. наук, профессор; СПбФИЦ РАН, СПИИРАН, лаборатория речевых и многомодальных интерфейсов; руководитель лаборатории;
E-mail: karpov@iias.spb.su

Поступила в редакцию 23.07.24; одобрена после рецензирования 02.08.24; принята к публикации 23.09.24.

REFERENCES

1. Tkachenya A.V., Davydov A.G., Kiselev V.V., Khitrov M.V. *Journal of Instrument Engineering*, 2013, no. 2(56), pp. 61–66. (in Russ.)
2. Cafaro A., Wagne J., Baur T., Dermouche S., Torres Torres M. et al. *Proc. of the 19th ACM Intern. Conf. on Multimodal Interaction*, 2017, pp. 350–359, DOI: 10.1145/3136755.313678.
3. Guhan P., Agarwal M., Awasthi N., Reeves G., Manocha D. et al. *arXiv preprint arXiv:2011.08690*, 2020.
4. Celiktutan O., Skordos E., Gunes H. *IEEE Transactions on Affective Computing*, 2017, no. 4(10), pp. 484–497, DOI: 10.1109/TAFFC.2017.2737019.
5. Ringeval F., Sonderegger A., Sauer J., Lalanne D. *Proc. of the 10th IEEE Intern. Conf. and Workshops on Automatic Face and Gesture Recognition,* 2013, pp. 1–8, DOI: 10.1109/FG.2013.6553805.
6. Kaur A., Mustafa A., Mehta L., Dhall A. *2018 Digital Image Computing: Techniques and Applications* (DICTA), 2018, pp. 1–8, DOI: 10.1109/DICTA.2018.8615851.
7. Gupta A., D'Cunha A., Awasthi K., Balasubramanian V. *arXiv preprint arXiv:1609.01885*, 2016.
8. Sümer Ö., Goldberg P., D'Mello S., Gerjets P., Trautwein U., Kasneci E. *IEEE Transactions on Affective Computing*, 2021, no. 2(14), pp. 1012–1027, DOI: 10.1109/TAFFC.2021.3127692.
9. Whitehill J., Serpell Z., Lin Y.C., Foster A., Movellan J.R. *IEEE Transactions on Affective Computing*, 2014, no. 1(5), pp. 86–98, DOI: 10.1109/TAFFC.2014.2316163.
10. Psaltis A., Apostolakis K. C., Dimitropoulos K., Daras P. *IEEE Transactions on Games*, 2017, no. 3(10), pp. 292–303, DOI: 10.1109/TCIAIG.2017.2743341.
11. Dvoynikova A.A., Kagirov I.A., Karpov A.A. *Information and Control Systems*, 2022, no. 5(120), pp. 12–22, DOI: 10.31799/1684-8853-2022-5-12-22. (in Russ.)
12. Dvoynikova A.A., Markitantov M.V., Ryumina E.V., Uzdyaev M.Yu., Velichko A.N. et al. *Informatics and Automation*, 2022, no. 6(21), pp. 1097–1144, DOI: 10.15622/ia.21.6.2 (in Russ.)
13. Dhall A., Goecke R., Gedeon T. *Journal of latex class files*, 2007, no. 1(6).
14. Kollias D., Zafeiriou S. *arXiv preprint arXiv:1811.07770*, 2018.
15. Busso C., Bulut M., Lee C.C., Kazemzadeh A., Mower E. et al. *Language Resources and Evaluation*, 2008, no. 4(42), pp. 335–359, DOI: 10.1007/s10579-008-9076-6.
16. Poria S., Hazarika D., Majumder N., Naik G., Cambria E. et al. *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 527–536.
17. Zadeh A.B., Liang P.P., Poria S., Cambria E., Morency L.P. *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*, 2018, pp. 2236–2246, DOI: 10.18653/v1/P18-1208.
18. Perepelkina O., Kazimirova E., Konstantinova M. *Proc. of the Intern. Conf. on Speech and Computer*, 2018, pp. 501–510, DOI: 10.1007/978-3-319-99579-3_52.
19. Jones S.R.G. *American Journal of sociology*, 1992, no. 3(98), pp. 451–468.
20. Viola P., Jones M. *Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (CVPR), 2001, vol. 1, pp. I-I, DOI: 10.1109/CVPR.2001.990517.
21. Patent USA 3069654, *Method and means for recognizing complex*, P.V.C. Hough, Priority 1962.
22. Lausberg H., Sloetjes H. *Behavior research methods*, 2009, no. 3(41), pp. 841–849, DOI: 10.3758/BRM.41.3.841
23. Lyusin D.V. *Psychological diagnostics*, 2006, vol. 4, pp. 3–22. (in Russ.)
24. Lyusin D.V., Ovsyannikova V.V. *Psychological journal*, 2013, no. 6(34), pp. 82–94. (in Russ.)
25. Certificate of registration of the database 2023624954, *Baza dannykh proyavleniy vovlechennosti i emotsiy russkoyazychnykh uchastnikov telekonferentsiy (ENERGI — ENgagement and Emotion Russian Gathering Interlocutors)* (Database of Manifestations of Engagement and Emotions of Russian-Speaking Participants in Teleconferences (ENERGI - ENgagement and Emotion Russian Gathering Interlocutors)), A.A. Dvoynikova, A.A. Karpov, Priority 25.12.2023. (in Russ.)

DATA ON AUTHORS

Anastasia A. Dvoynikova — St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Laboratory of Speech and Multimodal Interfaces, Junior Researcher;
E-mail: dvoynikova.a@iias.spb.su

Alexey A. Karpov — Dr. Sci., Professor; St. Petersburg Federal Research Center of the RAS, St. Petersburg Institute for Informatics and Automation of the RAS, Laboratory of Speech and Multimodal Interfaces; Head of the Laboratory;
E-mail: karpov@iias.spb.su

Received 23.07.24; approved after reviewing 02.08.24; accepted for publication 23.09.24.