Original research article
УДК 81-114.2
DOI: 10.18384/2949-5075-2024-2-40-51

SPEECH AND GESTURE ALIGNMENT IN IMMERSIVE COMMUNICATION: HUMAN-HUMAN VS. HUMAN-COMPUTER INTERACTION

V. Potekhin
Diplomatic Academy of the Ministry of Foreign Affairs of the Russian Federation, ulitsa Ostozhenka 53/2 build. 1, Moscow 119021, Russian Federation; e-mail: vadim.nebarsuk@gmail.com

Received by the editorial office 21.11.2023
Revised by the authors 07.02.2024
Accepted for publication 09.02.2024

Abstract

Aim. Based on a pilot multimodal study, this paper explores the alignment patterns of speech and gesture distribution elicited from participants in two regimes of immersive communication, Human–Human and Human–Computer interaction.

Methodology. A multimodal experiment was carried out in two interactional regimes, with the participants discoursing with a human partner and to a computer (with a computer stimulus exposure). Individual variance in discourse schemata in speech and gesture types distribution was established based on hierarchical cluster analysis that enabled to identify groups of participants exhibiting discursive recurrencies in immersive communication.

Results. The obtained results showed that the co-occurrence of gesture and discourse schemata helped reinforce the communicative intent in Human-Human communication. Using regression analysis, the study confirmed that individual variance in discourse schemata distribution was observed in the use of descriptive schemata irrespective of the interactional regime at play.

Research implications. The research findings suffice to assume that the extent of immersion in multimodal communication can be measured, predicted, and even deliberately preset. Overall, the article attests to the applicability of multimodal methodology to exploring immersion.

Keywords: discourse schemata, gesture functions, immersion, multimodality, multimodal behaviour

For citation:
Potekhin, V. O. (2024). Speech and gesture alignment in immersive communication: Human-Human vs. Human-Computer interaction. In: *Key Issues of Contemporary Linguistics*, 2, pp. 40–51. https://doi.org/10.18384/2949-5075-2024-2-40-51

Научная статья

АДАПТАЦИЯ РЕЧЕВОГО И ЖЕСТОВОГО ПОВЕДЕНИЯ В УСЛОВИЯХ ИММЕРСИВНОЙ КОММУНИКАЦИИ: РЕЖИМЫ ВЗАИМОДЕЙСТВИЯ «ЧЕЛОВЕК–ЧЕЛОВЕК» И «ЧЕЛОВЕК–КОМПЬЮТЕР»

Потехин В. О.

Дипломатическая академия Министерства иностранных дел Российской Федерации, 119021, г. Москва, ул. Остоженка 53/2, стр. 1, Российская Федерация; e-mail: vadim.nebarsuk@gmail.com

Поступила в редакцию 21.11.2023
После доработки 07.02.2024
Принята к публикации 09.02.2024

Аннотация

Цель. На основании пилотного полимодального исследования в статье изучаются модели распределения адаптации речи и жеста у участников эксперимента в условиях иммерсивной коммуникации, осуществляемой в двух режимах взаимодействия: «человек–человек» и «человек–компьютер».

Процедура и методы. Проведён полимодальный эксперимент, в ходе которого осуществлялась фиксация речевого и жестового поведения участников в процессе взаимодействия с другим человеком или с экраном компьютера (с представленным на нём динамическим стимулом). С помощью иерархического кластерного анализа, позволившего сгруппировать участников со схожими проявлениями иммерсивной коммуникации, установлено индивидуальное варьирование в распределении дискурсивных схем в речи и типов жестов.

Результаты. Полученные результаты демонстрируют, что коммуникативное намерение говорящего в режиме «человек–человек» было усилено с помощью синхронизированного использования жеста и речи. Проведённый регрессионный анализ показал, что индивидуальное различие в распределении дискурсивных схем наблюдалось в использовании дескриптивных схем вне зависимости от режима взаимодействия.

Теоретическая и/или практическая значимость. Полученные результаты позволяют предположить, что степень иммерсивности может быть измерена, спрогнозирована и даже намеренно предустановлена. В целом, данное исследование доказывает эффективность применения полимодального метода для исследования иммерсивности.

Ключевые слова: дискурсивные схемы, иммерсивность, полимодальность, полимодальное поведение, функциональные типы жестов

Для цитирования:

Потехин В. О. Адаптация речевого и жестового поведения в условиях иммерсивной коммуникации: режимы взаимодействия «человек–человек» и «человек–компьютер» // Вопросы современной лингвистики. 2024. № 2. С. 40–51. https://doi.org/10.18384/2949-5075-2024-2-40-51

Introduction

The growth of Human–Human and Human–Computer interaction has stimulated multiple studies exploring its major aspects or levels [1, p. 17], which are physical, cognitive, and affective [2, p. 139]. In the present study, we address the cognitive level of interaction which embraces the ways people construe information in communicating with each other and with computer. One of the possible methods of attesting the cognitive aspect of interaction is via conversational

practices [3] in a laboratory setting [4, p. 44–45]. Meanwhile, while this is mostly speech and gaze which are considered as two main sources of data, the multimodality of speech and gesture is largely neglected. However, it is the speech and gesture alignment that is thought to betray the communicative intent [5; 6; 7]. Consequently, in this study we address speech and gesture alignment with the view to disclose their specific alignment patterns which are typical of different interaction regimes in immersive communication or the communication which transforms its observer into its active participant [8; 9]. To proceed, we employ multimodal experiment and contrastive method, comparing their distribution in Human–Computer interaction and Human–Human interaction in immersive communication in two regimes, 1) with the participant exposed to a dynamic scene augmented by VR-technologies and presenting it in speech and gesture while viewing, 2) with the participant presenting the priorly shown dynamic scene to an active listener, expecting that under similar discourse circumstances we could determine the differences in selecting speech and gesture combinations.

As known, multiple studies have addressed gesturing as part of multimodal (and also robotic) environment [10; 11; 12]; however, they mostly specify the structural or formal gesture characteristics while the discourse-dependent multimodal alignment of speech and gesture is seldom considered. We expect that the adopted Interactive Approach [13] will allow to specify selecting 1) the discourse schemata in speech, 2) the functional gestures which are mostly discourse dependent, 3) the speech and gesture alignment patterns or the patterns of their synchronized distribution.

The study elaborates two major hypotheses. First, we presume that the use of discourse schemata and gesture types will manifest significant differences in Human–Computer interaction accompanied by the stimulus exposure and in Human–Human interaction not accompanied by any stimulus exposure. Second, we expect that there exist differences in the individual distribution of discourse schemata and gesture types in Human–Computer and Human–Human interaction.

The work is structured as follows. First, we present Theoretical framework which comprises the studies on immersive communication and its features, the discourse types and discourse schemata in speech and also the studies on functional gesture types featuring the opportunities to explore speech and gesture alignments. In Methods and procedure section we describe the multimodal experiment design. In Results and Discussion section the results on the discourse schemata in speech, the functional gestures and the speech and gesture alignment patterns are specified. The research input is presented in the Final remarks.

Theoretical framework

When it comes to studying immersion, the research interest is typically geared toward the technical side of the matter, i.e. measuring the impact of VR[1]-/AR[2]-/MR[3]-technologies on the extent of the engagement with a dynamic stimulus [8; 9]. The present study adds to the technological dimension two more modalities (termed multimodal) that can be subjected to analysis, namely the user's discourse and gesture. Our belief is that while experiencing a dynamic scene augmented by VR-technologies, the user is being transformed from a passive observer into an active participant in an engrossing digital environment. This transformation, in turn, affects the user's multimodal behaviour. Viewing immersion as both a cognitive and discursive phenomenon, this paper offers a discourse-based method to analyzing it.

[1] VR: virtual reality, a computer-generated reality that projects the user into a 3D space by means of a headset.

[2] AR: augmented reality, any alteration made to a real-world image or video.

[3] MR: mixed reality, the merging of real and virtual worlds to produce new environments and visualizations where physical and digital objects co-exist in real time.

Importantly, the overall adopted approaches to exploring cognitive and discursive characteristics of immersive communication are the Interactive Approach [13] which allows to identify them in the interactive environment, here in Human–Computer and Human–Human interaction regimes, and the Integrative Functional Approach [14, p. 62–63]. Starting from the assumption that spontaneous communication may incorporate various discourse types, we expect immersive communication to exhibit the tendency toward varying discourse types distribution in the users' speech with and without stimulus exposure. Relying on the recognized notional discourse taxonomies [15; 16; 17], we differentiate between narrative, descriptive, argumentative, and expositive discourse types. This (rather) formal differentiation should not, however, come into conflict with the view of discourse as a continuum, which it essentially is, and, therefore, there is an ongoing overlap of distinct discourses. Nevertheless, setting discourses apart helps the researcher make sense of linguistic recurrencies typical of this or that discourse. At the same time, lexical features marking a specific discourse are incommensurable with this discourse as a whole for speech normally incorporates passages drawn from a variety of discourses [18, p. 8–9]. Apparently, there exist "larger-than-text" structures that conventionalize a discourse and facilitate our discursive expectations. These superstructures, which are independent of content or theme, are referred to as (discourse) schemata [19; 20]. Hence, each discourse type has an intrinsic set of schemata.

Conventional schemata for narrative passages are *setting*, *initiating events*, *characters' goals*, *attempts towards goals*, and *event outcomes* [21]. Descriptive passages are characterized by such schemata as *description/explanation*, *background information*, *elaboration*, *exemplification* [22]. The schemata for expositive discourse are as follows – *viewpoint formulation*, *viewpoint presentation*, *compare–contrast*, *cause–effect*, *problem–solution* [23]. Finally, schemata that characterize argumentative discourse are *standpoints at issue*, *starting points of discourse, argument advance*, *outcome presentation* [24]. It should be noted, however, that not all schemata have to be present in a discourse, some of them can be missing or re-arranged.

In the present study, we posit that immersion is likely to impact the user's multimodal behaviour (discourse and gesture use), whereas the factor of interaction regime (Human–Computer, Human–Human) will affect the variance in discourse schemata and gesture distribution. A large body of research has recently explored how discourse and gesture co-exist to reinforce the communicative intent [5; 6; 7]. It is now widely recognized that speech and gesture may form a single entity and organize our discourse. The co-occurrence of gesture and speech has received thorough scrutiny in the theory of growth points [25]. In terms of gesture, we will address the communicative functions of gesture manifested in gesture types [6]. In our study, four gesture functions will be considered – pragmatic, representational, deictic, and adaptive.

Methods and procedure

Recently, multimodal approach to speech and gesture has been widely applied in cognitive discourse studies. However, its potential to explore immersive communication has been paid scarce attention, if at all. In this study, we aim to test the applicability of multimodal methodology to exploring immersion.

Based on a pilot multimodal experiment, we will establish the contingency of speech and gesture elicited from participants in immersive communication that simulates their engagement in communicative discourse. We recruited 15 participants (85,8% female, middled-aged). The experiment was conducted in two stages. During the first stage which was Human–Computer interaction the participants were exposed to a dynamic visual stimulus and had to deliver online their verbal commentary as the dynamic scene was unfolding. The stimulus used was a video

footage that simulates (bodily) engagement (movement and object perception) of a viewer studying a dynamic image (AI-generated themes of Van Gogh's "Starry Night" and "Bedroom in Arles" presented as a single continuous image with alternating scenarios). During the second stage, which was Human–Human interaction without stimulus exposure, the participants had to retell their previous viewing experience to an active (interested) listener. At both stages, they were filmed, their voice recorded, and their movement captured on a camera. Thereafter, their speech and gesture use were annotated in ELAN software.

In terms of speech, we established speech patterns (manifested in discursive markers) typical of each discourse schema. The taxonomy of narrative discursive markers aligned with narrative discourse schemata [21] includes Story starters (***вижу*** *картину Ван Гога* / *I* ***can see*** *a Van Gogh painting, «****Лунная ночь****» / (Van Gogh's) "****Moonlit night****"*), Dynamic event (*синий цвет* ***двигался, переливался*** */ the blue colour was* ***in motion*** *and* ***iridescent***), Directed movement (*мы* ***приближаемся к*** *одному из домиков* / *we're* ***coming up to*** *one of the cabins*), Occurrence time (***затем*** / ***then, потом*** / ***after*** *(that),* ***далее*** / ***next***), Story endings (*картина* ***исчезла*** */ the painting* ***has*** *(now)* ***disappeared****, (у нас) полностью* ***растворилась*** *картинка* / *the image* ***has*** *completely* ***faded away***), etc.

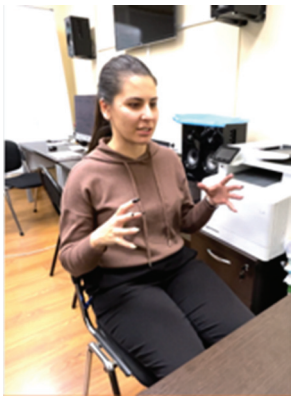
The taxonomy of descriptive discursive markers aligned with descriptive discourse schemata [22] includes Existential predicates (***выглядит*** *ещё более завораживающим* / *(it)* ***looks*** *even more intriguing,* ***выглядит*** *немного сюрреалистично* / *(it)* ***looks*** *somewhat surreal*), Spatial deixis (*шкаф стоял* ***слева*** */ the closet was* ***on the left,*** *чемодан был* ***в шкафу*** */ the suitcase was* ***in the closet***), Detailed elaboration (event, object, state) (***иду-иду*** */ I* ***keep going*** *(on and on),* ***Голубоватый*** *столик* / *a* ***bluish*** *small table*), Physical space presentation (***шкаф, кровать, стулья*** *жёлтые* / *(I can see) a* ***closet****, a* ***bed****, some yellow* ***chairs***), etc.

The taxonomy of expositive discursive markers aligned with expositive discourse schemata [23] includes Rhetorical question (***Что происходит?*** / ***What's going on?***), Linker (***собственно говоря*** / ***as a matter of fact, по-видимому*** / ***apparently***), Modality (***возможно*** / ***possibly, значит*** / ***that is***), Implicit comparison (***как бы*** / ***sort of, как будто бы*** / ***as if***), Evaluation (*достаточно* ***скудный*** */ fairly* ***modest***), Causation, consequence, and condition (***потому что*** / ***because, следовательно*** / ***consequently, поэтому*** / ***that's why***), Listener-oriented rhetorical patterns (*вот* ***знаете*** / ***you know***), self-oriented rhetorical patterns (***говорю*** / ***I say***), Hesitation (*поиск слов:* ***ммм, ааа*** */ word search:* ***errr, hmm, не знаю*** / ***dunno***), Metadiscourse (***такое ощущение*** / ***it feels like, ну, как-то так*** / ***that's about it***), etc.

The taxonomy of argumentative discursive markers aligned with argumentative discourse schemata [24] includes Volition (***хотелось бы****, чтобы картинка двигалась побыстрее* / ***I wish*** *the painting were moving a bit faster*), Uncertainty (***если я правильно помню*** / ***if my memory doesn't fail me***), Clarification (***есть такие видео на YouTube*** */ (so)* ***there're these videos on YouTube***), Drawing conclusions (*я* ***оказался прав*** / ***it turned out I was right***), etc.

The markers were further employed for studying speech and gesture behaviour in the participants' discourse.

Next, we will provide specification of the gesture functions referred to previously. Pragmatic gestures are context-bound, and their meaning depends on the words they co-occur with. As it is illustrated in the example below (Figure 1), representational (or iconic) gestures establish an associative linkage between a hand movement (form) and the corresponding process (object). In the example taken from the stage without stimulus exposure (Figure 2), we can see an axis-like gesture employed to mark spatial orientation (deictic gesture). Finally, adaptors are self-oriented gestures employed to gain control of the communicative situation when



Gesture: representational

Speech: *да, она так* ***завлекает*** *немного человека внутрь* / *so, yeah, it* ***engages*** *you a bit inside of it*

Fig. 1 / Рис. 1. Representational gesture synchronized with the INITIATING EVENT narrative schema at the stage with stimulus exposure / Совместное использование репрезентирующего жеста с дискурсивной схемой нарратива ИСХОДНЫЕ СОБЫТИЯ на этапе с предъявлением стимула

Gesture: deictic

Speech: ***Видеосюжет****… был про знаковую картину Винсента Ван Гога* / ***The video****… was about Van Gogh's pivotal painting*

Fig. 2 / Рис. 2. Deictic gesture synchronized with the INITIATING EVENT narrative schema at the stage without stimulus exposure / Совместное использование дейктического жеста с дискурсивной схемой нарратива ИСХОДНЫЕ СОБЫТИЯ на этапе без предъявления стимула

Source: photographs (fig. 1, 2) from the author's personal archive.
Источник: фотографии (рис. 1, 2) из личного архива автора

the speaker is challenged. The participants' gesture use was annotated in ELAN software which enables to synchronize gestures and the discursive markers they co-occur with.

Results and discussion

Since two major hypotheses were tested, claiming that 1) the use of discourse schemata and gesture types will manifest significant differences in Human–Computer and Human–Human interaction, and that 2) the individual distribution of discourse schemata and gesture types in two interaction regimes displays differences, we will present and discuss the results in two subsections.

Discourse schemata and gesture types in Human–Computer and Human–Human interaction

In Figure 3 we present discourse schemata distribution in the participants' speech in Human–Computer interaction with stimulus exposure. Figure 4 shows discourse schemata distribution in Human–Human interaction without stimulus exposure.

Both Figure 3 and Figure 4 manifest variance in the use of descriptive, expositive and argumentative discourse schemata with Mean=40 with stimulus and Mean=45.2 without stimulus for description; Mean=37 with stimulus and Mean=35 without stimulus for exposition; and Mean=3.8 with stimulus and Mean=1 without stimulus for argumentation. Figure 5 and Figure 6 show gesture use activity during the two experimental stages (with and without stimulus).

As seen from Figure 5 and Figure 6, significant variance is observed in gesture use at both experimental stages with a significant increase in gesture activity at the stage without stimulus exposure. Three gesture types, pragmatic, representational, and deictic tend to be used significantly more frequently when the participants are not exposed to the stimulus. The obtained results show that discourse and gesture co-exist to reinforce a Human–Human interactional communicative intent rather than a Human–Computer interaction which proves the specificity of computer-oriented communication and the results obtained in [2; 3]. The results also suffice to claim that speech and gesture alignment patterns in both cases are stimulated by different types of growth points [25] mani-

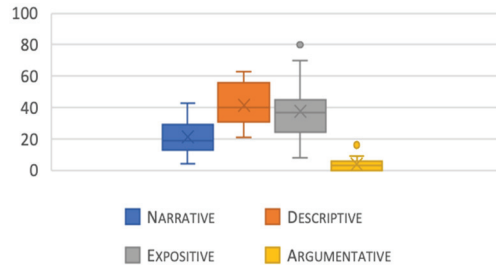


Fig. 3 / Рис. 3. Discourse schemata distribution with stimulus exposure / Распределение дискурсивных схем в речи на этапе с предъявлением стимула

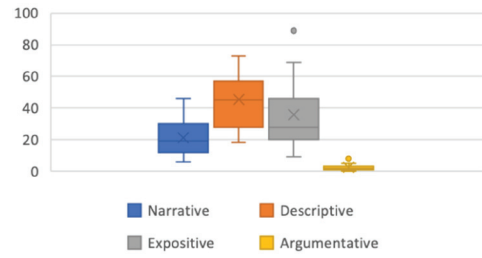


Fig. 4 / Рис. 4. Discourse schemata distribution without stimulus exposure / Распределение дискурсивных схем в речи на этапе без предъявления стимула

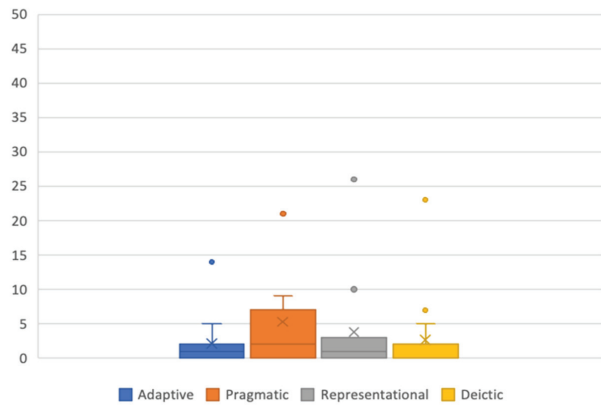


Fig. 5 / Рис. 5. Gesture activity among participants at the stage with stimulus / Распределение активности жеста на этапе с предъявлением стимула

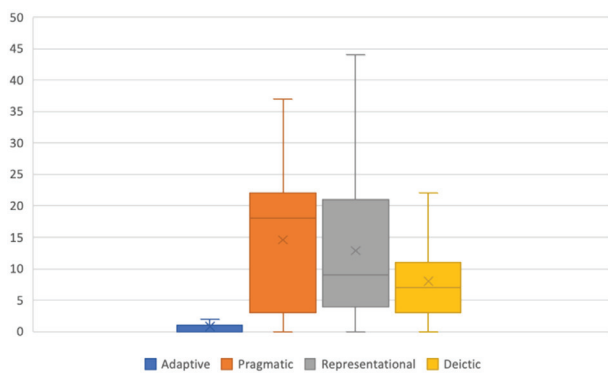


Fig. 6 / Рис. 6. Gesture activity among participants at the stage without stimulus / Распределение активности жеста на этапе без предъявления стимула

Source: box plots (fig. 3, 4, 5, 6) compiled by the author
Источник: диаграммы размаха (рис. 3, 4, 5, 6), составлено автором

fested in two regimes of immersive communication. Importantly, discourse schemata displaying four discourse types in speech [21; 22; 23; 24] serve as bad predictors of two scenarios of immersive communication (since no significant differences are observed in speech); whereas the gestures do display significant differences. To attest these differences, we explore the distribution of discourse schemata in speech and four gesture types in the individual probes of the experiment participants.

Individual variance in discourse schemata and gesture types distribution in Human–Computer and Human–Human interaction

To identify the variance in the multimodal behavior of individual participants, we applied hierarchical cluster analysis exploring the distribution recurrencies in the discourse schemata in the participants' speech during the two experimental stages. We hypothesized that there might be different alignment patterns within the participants groups. Using Jamovi software, we ran two hierarchical clustering analyses at both stages. The distribution of the participants in Human–Computer interaction with stimulus exposure was balanced: the resulting number of clusters equals three; each cluster shows recurring patterns in discourse use. In particular, the first cluster features a high incidence of expositive schemata. The second cluster shows lower discursive activity among all the discourse schemata under consideration. The third cluster is high in descriptive schemata. Interestingly, the participants' discursive activity in Human–Human interaction regime without stimulus exposure gained a different cluster distribution. The resulting number of clusters is two. The first cluster shows a high incidence of narrative, descriptive, and expositive discourse schemata, whereas the second cluster features the participants with a significant decrease in discourse activity overall (in particular, in narrative schemata).

Following this procedure, first, we determined the differences (using a series of ANOVA tests) in discourse schemata distribution in 1) clusters 1-3 in Human–Computer interaction, 2) clusters 4-5 in Human–Human interaction. The results show that in clusters 1-3 there are differences in the distribution of narrative and descriptive discourse schemata (with F=6.5 at p=0.039 for each type), whereas no difference was detected in the use of expositive and argumentative discourse. In clusters 4 and 5 significant difference is observed only in the distribution of descriptive discourse schemata (with F=7 at p=0.008). Therefore, we can claim that major differences between the participants' clusters in speech irrespective of the interaction regime lie in the use of descriptive discourse schemata. This suffices to prove that the use of Descriptive passages characterized by such schemata as description / explanation (as in ***выглядит*** *немного сюрреалистично / (it)* ***looks*** *somewhat surreal*), background information (as in ***шкаф, кровать, стулья*** *жёлтые / (I can see) a* ***closet****, a* ***bed****, some yellow* ***chairs***), elaboration (as in ***голубоватый*** *столик / a* ***bluish*** *small table*), exemplification (as in ***как в старые времена*** */* ***like in olden times***) [21] may serve as a good predictor of individual variance in speech irrespective of interaction regime in immersive communication. Importantly, no individual variance was observed in the use of expositive discourse schemata, which proves that this is notably the difference in expositive discourse use, both in overall distribution and in the individual participants' behavior distribution, which demarcates Human–Computer and Human–Human interaction. This means that expositive discourse schemata – viewpoint formulation (as in ***возможно*** */* ***possibly****,* ***значит*** */* ***that is***), viewpoint presentation (as in *достаточно* ***скудный*** */ fairly* ***modest***), compare–contrast (as in ***в отличие от*** *картины /* ***as opposed to*** *the painting*), cause–effect (as in ***следовательно*** */* ***therefore***), problem–solution (as in ***всё-таки*** *открылось окно /* ***finally*** *the window opened*) [23] serve as rigid predictors of immersive communication on the whole specifying the speech of each and ev-

ery participant irrespective of the interaction regime.

Second, we determined five gesture profiles of five participants' clusters expecting that there might be significant differences in these profiles. Using a series of ANOVA tests, we found that none of the gesture types displays significant variance among clusters 1-3 and 4-5, which means that gesture behavior in each of the five clusters is not determined by the discourse schemata distribution. Consequently, whereas overall gesture activity is dependent on the interaction regime, Human–Computer or Human–Human [10; 11], the gesture distribution is not determined by it. Presumably, gestural distribution is mostly dependent on the discourse schemata distribution rather than on the interaction type or the discourse type on the whole.

Overall, the study proves the efficiency of the adopted Interactive Approach [13] and the Integrative Functional Approach [14] which allowed to specify speech and gesture distribution and individual variance in Human–Computer interaction accompanied by the stimulus exposure and in Human–Human interaction not accompanied by any stimulus exposure. The study also shows that the discourse taxonomies [15; 16; 17] may serve as a reliable instrument to explore information construal in speech since they allow to mark the complete body of the speech produce of participants. While the results do not attest to the efficiency of identifying discourse schemata and functional gesture types alignment, they invariably prove that more detailed research is necessary to disclose the alignment among single schemata and not among overall discourse types. Finally, the results have shown that the cognitive research of Human-Computer and Human-Human interaction can shed light both on the conversation practices [3] in each of these regimes and on their multimodal specificity, which proves the importance of this aspect of interaction studies formulated in [2].

Final remarks

The adopted Interactive Approach [13] and the Integrative Functional Approach [14] may be applicable to contrastive discourse studies exploring a multimodal potential of communication as modulated by VR-/AR-/MR-technologies. Newly emerging media of discourse delivery require an evitable consideration of the impact of the technology factor on the cognitive and language faculties of individuals exposed to various (dynamic) stimuli. Varying the extent of the stimulus exposure cannot but give rise to ethical concerns when an individual's language and body use develop a digital extension in virtual environments and can be manipulated for personal gain. Besides discourse and gesture alignment, another multimodal dimension worth considering is gaze. In the studies to come, we intend to specify gesture, discourse, and gaze alignment patterns in the two interactional regimes in question.

REFERENCES

1. Efimenko, T. N. & Ivanova, Yu. E. (2020). Verbal and nonverbal means of communication as functional operators of the meaning in English business discourse (exemplified in public speeches). In: *Bulletin of Moscow Region State University. Series: Linguistics*, 1, 15–28. DOI: 10.18384/2310-712X-2020-1-15–28 (in Russ.).
2. Karray, F., Alemzadeh, M., Saleh, J. A. & Arab, M. N. (2008). Human-computer interaction: Overview on state of the art. In: *International Journal of Smart Sensing and Intelligent Systems*, 1 (1), 137–159. DOI: 10.21307/ijssis-2017-283.
3. Valtolina, S., Matamoros, R. A. & Epifania, F. (2023). Methods for Evaluating Conversational Agents' Communicability, Acceptability and Accessibility Degree. In: *Human-Computer Interaction – INTERACT 2023. Proceedings of 19th IFIP TC13 International Conference. Part II.* Cham: Springer. pp. 372–382. DOI: 10.1007/978-3-031-42283-6_21 (Series: Lecture Notes in Computer Science (LNCS). Vol. 14143).

4. Dittrikh, A. G. (2023). The modern fields of pragmalinguistics research. In: *Key Issues of Contemporary Linguistics*, 4, 42–52. DOI: 10.18384/2949-5075-2023-4-42-52 (in Russ.).
5. Mcneill, D., Levy, E. T. & Duncan, S. D. (2014). Gesture in Discourse. In: Tannen, D., Hamilton, H. E. & Schiffrin, D. (eds.). *The Handbook of Discourse Analysis*. Chichester, UK: John Wiley & Sons, Inc. pp. 262–289. DOI: 10.1002/9781118584194.ch12.
6. Iriskhanova, O. K. & Cienki, A. (2018). The semiotics of gestures in cognitive linguistics: contribution and challenges. In: *Issues of Cognitive Linguistics*, 4 (57), 25–36. DOI: 10.20916/1812-3228-2018-4-25-36.
7. Kiose, M., Leonteva, A. & Agafonova, O. (2022). Aesthetic multimodality of speech and gesture: Towards its functional framework. In: *Languages and Modalities*, 2, 1–17. DOI: 10.3897/lamo.@.78840.
8. Apostolopoulos, J. G., Chou, P. A., Culbertson, B., Kalker, T., Trott, M. D. & Wee, S. (2012). The road to immersive communication. In: *Proceedings of IEEE*, 100 (4), 974–990. DOI: 10.1109/JPROC.2011.2182069.
9. Shilina, M. G. & Wirth, J. (2021). Immersive technologies in media: Towards the concept of generative mediatization? In: *RUDN Journal of Studies in Literature and Journalism*, 26 (4), 672–680. DOI: 10.22363/2312-9220-2021-26-4-672-680.
10. Holzapfel, H., Nickel, K. & Stiefelhagen, R. (2004). Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3D pointing gestures. In: *ICIM '04: Proceedings of the 6th International Conference on Multimodal Interfaces*. New York: Association for Computing Machinery. pp. 175–182. DOI: 10.1145/1027933.1027964.
11. Gunes, H. & Piccardi, M. (2007). Bi-modal emotion recognition from expressive face and body gestures. In: *Journal of Network and Computer Applications*, 30 (4), 1334–1345. DOI: 10.1016/j.jnca.2006.09.007.
12. Kotov, A., Zinina, A., Zaidelman, L. & Arinkin, N. (2021). Generative Classification of Elements of Nonverbal Communication. In: *Computational Linguistics and Intellectual Technologies: papers from the Annual International Conference "Dialogue" (2021)*, 20S, 1095–1105. Moscow: Russian State University for the Humanities. DOI: 10.28995/2075-7182-2021-20-1095-1105 (in Russ.).
13. Khan, T., Barthel, H., Amrhein, K., Ebert, A. & Liggesmeyer, P. (2015). An Interactive Approach for Inspecting Software System Measurements. In: *Human-Computer Interaction – INTERACT 2015. Proceedings of 15th IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT' 2015)*. Cham: Springer. pp. 1–8. DOI: 10.1007/978-3-319-22698-9_1 (Series: Lecture Notes in Computer Science (LNCS). Vol. 9298).
14. Malakhova, V. L. (2022). Analysis of the formation of English discourse multimodal sense space from the perspective of integrative functional methodology. In: *Bulletin of the Moscow Region State University. Series: Linguistics*, 1, 60–69. DOI: 10 .18384/2310-712X-2022-1-60-69 (in Russ.).
15. Werlich, E. (1976). *A Text Grammar of English*. Heidelberg: Quelle and Meyer.
16. Longacre, R. E. (1983). *The Grammar of Discourse*. New York: Plenum.
17. Virtanen, T. (1992). Issues of text typology: Narrative – a 'basic' type of text? In: *Text – Interdisciplinary Journal for the Study of Discourse*, 12 (2), 293–310. DOI: https://doi.org/10.1515/text.1.1992.12.2.293.
18. Kibrik, A. A. (2009). Modus, genre and other parameters of discourse classification. In: *Topics in the study of language*, 2, 3–21 (in Russ.).
19. Van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
20. Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. In: *Text – Interdisciplinary Journal for the Study of Discourse*, 8 (3), 243–281. DOI: https://doi.org/10.1515/text.1.1988.8.3.243.
21. Mandler, J. M. & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. In: *Cognitive Psychology*, 9 (1), 111–151. DOI: https://doi.org/10.1016/0010-0285(77)90006-8.
22. MacSaveny, T. (2010). Towards a Description of Descriptive Discourse. In: *GIALens*, 4 (3). URL: https://www.diu.edu/documents/gialens/Vol4-3/Tim-MacSaveny-Description-of-Descriptive-Discourse.pdf (accessed: 21.10.2023).
23. Nippold, M. A. & Scott, C. M. (eds.) (2010). *Expository discourse in children, adolescents, and adults: Development and disorders*. New York, NY: Psychology Press.

24. van Eemeren, F. H. (2010). *Strategic maneuvering. Extending the pragma-dialectical theory of argumentation*. Amsterdam-Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/aic.2 (Series: Argumentation in Context. Vol. 2).
25. McNeill, D. & Duncan, S. D. (2000). Growth points in thinking for speaking. In: McNeill, D. (ed.). *Language and Gesture*. Cambridge: Cambridge University Press. pp. 141–161.

ЛИТЕРАТУРА

1. Ефименко Т. Н., Иванова Ю. Е. Вербальные и невербальные средства коммуникации как операторы смысла в англоязычном деловом дискурсе (на материале публичных выступлений) // Вестник Московского государственного областного университета. Серия: Лингвистика. 2020. № 1. С. 15–28. DOI: 10.18384/2310-712X-2020-1-15-28.
2. Karray F., Alemzadeh M., Saleh J. A., Arab M. N. Human-computer interaction: Overview on state of the art // International Journal of Smart Sensing and Intelligent Systems. 2008. Vol. 1. Iss. 1. P. 137–159. DOI: 10.21307/ijssis-2017-283.
3. Valtolina S., Matamoros R. A., Epifania F. Methods for Evaluating Conversational Agents' Communicability, Acceptability and Accessibility Degree // Human-Computer Interaction – INTERACT 2023. Proceedings of 19th IFIP TC13 International Conference (York, UK, August 28 – September 1, 2023), Part II. Cham: Springer, 2023. P. 372–382. DOI: 10.1007/978-3-031-42283-6_21 (Series: Lecture Notes in Computer Science (LNCS). Vol. 14143).
4. Диттрих А. Г. Современные области прагмалингвистических исследований // Вопросы современной лингвистики. 2023. № 4. С. 42–52. DOI: 10 18384/2949-5075-2023-4-42-52.
5. Mcneill D., Levy E. T., Duncan S. D. Gesture in Discourse // The Handbook of Discourse Analysis / D. Tannen, H. E. Hamilton and D. Schiffrin (eds.). Chichester, UK: John Wiley & Sons, Inc., 2014. P. 262–289. DOI: 10.1002/9781118584194.ch12.
6. Ирисханова О. К., Ченки А. The semiotics of gestures in cognitive linguistics: contribution and challenges (Семиотика жестов в когнитивной лингвистике: достижения и актуальные вопросы) // Вопросы когнитивной лингвистики. 2018. № 4. С. 25–36. DOI: 10.20916/1812-3228-2018-4-25-36.
7. Kiose M., Leonteva A., Agafonova O. Aesthetic multimodality of speech and gesture: Towards its functional framework // Languages and Modalities. 2022. No. 2. P. 1–17. DOI: 10.3897/lamo.@.78840.
8. Apostolopoulos J. G., Chou P. A., Culbertson B., Kalker T., Trott M. D., Wee S. The road to Immersive Communication // Proceedings of IEEE. 2012. Vol. 100. No. 4. P. 974–990. DOI: 10.1109/JPROC.2011.2182069.
9. Шилина М. Г., Вирт Ю. Immersive technologies in media: Towards the concept of generative mediatization? // Вестник Российского университета дружбы народов. Серия: Литературоведение. Журналистика. 2021. Т. 26. № 4. С. 672–680. DOI: 10.22363/2312-9220-2021-26-4-672-680.
10. Holzapfel H., Nickel K., Stiefelhagen R. Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3D pointing gestures // ICIM '04: Proceedings of the 6th International Conference on Multimodal Interfaces. New York: Association for Computing Machinery, 2004. P. 175–182. DOI: 10.1145/1027933.1027964.
11. Gunes H., Piccardi M. Bi-modal emotion recognition from expressive face and body gestures // Journal of Network and Computer Applications. 2007. Vol. 30. Iss. 4. P. 1334–1345. DOI: 10.1016/j.jnca.2006.09.007.
12. Котов А., Зайдельман Л., Зинина А., Аринкин Н. Генеративная классификация элементов невербальной коммуникации // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной международной конференции «Диалог» (2021). Москва: РГГУ, 2021. Вып. 20 S. С. 1095–1105. DOI: 10.28995/2075-7182-2021-20-1095-1105.
13. Khan T., Barthel H., Amrhein K., Ebert A., Liggesmeyer P. An Interactive Approach for Inspecting Software System Measurements // Human-Computer Interaction – INTERACT 2015. Proceedings of 15th IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT' 2015). Cham: Springer, 2015. P. 1–8. DOI: 10.1007/978-3-319-22698-9_1 (Series: Lecture Notes in Computer Science (LNCS). Vol. 9298).
14. Малахова В. Л. Анализ формирования мультимодального смыслового пространства английского дискурса через призму интегративной функциональной методологии // Вестник Мо-

сковского государственного областного университета. Серия: Лингвистика. 2022. № 1. С. 60–69. DOI: 10.18384/2310-712X-2022-1-60-69.
15. Werlich E. A Text Grammar of English. Heidelberg: Quelle and Meyer, 1976. 315 p.
16. Longacre R. E. The Grammar of Discourse. New York: Plenum, 1983. 423 p.
17. Virtanen T. Issues of text typology: Narrative – a 'basic' type of text? // Text – Interdisciplinary Journal for the Study of Discourse. 1992. Vol. 12. No. 2. P. 293–310. DOI: https://doi.org/10.1515/text.1.1992.12.2.293.
18. Кибрик А. А. Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. 2009. № 2. С. 3–21.
19. Van Dijk T. A. Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980. 317 p.
20. Mann W. C., Thompson S. A. Rhetorical Structure Theory: Toward a functional theory of text organization // Text – Interdisciplinary Journal for the Study of Discourse. 1988. Vol. 8. No. 3. P. 243–281. DOI: https://doi.org/10.1515/text.1.1988.8.3.243.
21. Mandler J. M., Johnson N. S. Remembrance of things parsed: Story structure and recall // Cognitive Psychology. 1977. Vol. 9. Iss. 1. P. 111–151. DOI: https://doi.org/10.1016/0010-0285(77)90006-8.
22. MacSaveny T. Towards a Description of Descriptive Discourse // GIALens. 2010. Vol. 4. No. 3 [Электронный ресурс]. URL: https://www.diu.edu/documents/gialens/Vol4-3/Tim-MacSaveny-Description-of-Descriptive-Discourse.pdf (дата обращения: 21.10.2023).
23. Expository discourse in Children, Adolescents, and Adults: Development and Disorders / M. A. Nippold, C. M. Scott (eds.). New York, NY: Psychology Press, 2010. 336 p.
24. van Eemeren F. H. Strategic maneuvering. Extending the pragma-dialectical theory of argumentation. Amsterdam-Philadelphia: John Benjamins, 2010. 308 p. DOI: https://doi.org/10.1075/aic.2 (Series: Argumentation in Context. Vol. 2).
25. McNeill D., Duncan S. D. Growth points in thinking for speaking // Language and Gesture / ed. D. McNeill. Cambridge: Cambridge University Press, 2000. P. 141–161.

INFORMATION ABOUT THE AUTHOR

Vadim O. Potekhin (Moscow) – Senior Lecturer, Department of European Languages, Diplomatic Academy of the Ministry of Foreign Affairs of the Russian Federation;
ORCID: 0009-0003-2370-5097; e-mail: vadim.nebarsuk@gmail.com

ИНФОРМАЦИЯ ОБ АВТОРЕ

Потехин Вадим Олегович (Москва) – старший преподаватель кафедры европейских языков Дипломатической академии Министерства иностранных дел Российской Федерации;
ORCID: 0009-0003-2370-5097; e-mail: vadim.nebarsuk@gmail.com

FOR CITATION

Potekhin V. O. Speech and gesture alignment in immersive communication: Human-Human vs. Human-Computer interaction. In: *Key Issues of Contemporary Linguistics*, 2024, no. 2, pp. 40–51.
DOI: 10.18384/2949-5075-2024-2-40-51

ПРАВИЛЬНАЯ ССЫЛКА НА СТАТЬЮ

Потехин В. О. Адаптация речевого и жестового поведения в условиях иммерсивной коммуникации: режимы взаимодействия «человек–человек» и «человек–компьютер» // Вопросы современной лингвистики. 2024. № 2. С. 40–51.
DOI: 10.18384/2949-5075-2024-2-40-51