

*А. А. Зинина, А. А. Котов, Н. А. Аринкин,
Л. Я. Зайдельман, М. М. Цфасман*

A. Zinina, A. Kotov, N. Arinkin, L. Zaydelman, M. Tsfasman

НАПРАВЛЕНИЯ КОММУНИКАТИВНЫХ ДЕЙСТВИЙ В МУЛЬТИМОДАЛЬНОМ КОРПУСЕ REC¹

DIRECTIONS OF COMMUNICATIVE ACTIONS IN MULTIMODAL CORPUS REC

Аннотация. В статье описан новый блок разметки мультимодального корпуса REC (Russian Emotional Corpus) – разметка коммуникативно-значимых движений глаз (2 видео из подкорпуса эмоциональных интервью, общей продолжительностью 127 минут). Выделено 1875 аннотаций для движений глаз и 1415 аннотаций функциональной разметки. В работе показано, что направления взглядов и жестов важны в естественном общении: как в диалоге, так и при решении пространственных задач. На роботе Ф-2 смоделировано коммуникативное поведение в ситуации совместного с пользователем решения пространственной задачи (танграм) с использованием ориентированных жестов и направления взгляда. Взаимодействие пользователя с роботом было также записано для анализа коммуникативных ответов человека на ориентированные жесты робота (31 видеозапись, общей продолжительностью 310 минут).

Ключевые слова. мультимодальная коммуникация, ориентированные жесты, движения глаз, человеко-машинное взаимодействие.

Abstract. The article describes a new block in the marking of multimodal corpus REC (Russian Emotional Corpus): the marking of communicatively significant eye movements (2 videos from the emotional interviews subcorpus with a total duration of 127 minutes). There are 1875 annotations for eye movements and 1415 annotations of functional marking. The paper shows that eye gaze direction and pointing gestures are important in natural communication: both in dialogue and in spatial problem solving. We have modeled the communicative behavior of F-2 robot to solve spatial problems (tangram) together with the user. Oriented gestures and eye gaze directions are used in the robot behavior. Human – robot interaction was recorded to analyze the person's communicative responses to the robot's pointing gestures (31 videos with a total duration of 310 minutes).

Keyword. multimodal communication, oriented gestures, eye movements, human-machine interaction.

Введение

Интерфейс робота-компаньона должен быть «естественным» и простым для пользователя. Поэтому коммуникативное поведение такого робота должно быть максимально приближено к коммуникативному поведению человека. Кроме этого, поведение робота должно быть сложным и разнообразным, чтобы поддерживать более длительное и комфортное взаимодействие человека с роботом.

¹ Исследование выполнено при поддержке РФФИ, проект № 16-29-09601 офи_м.

В естественном коммуникативном взаимодействии собеседники используют сразу несколько каналов передачи информации и помимо естественной речи применяют целый комплекс невербальных средств: мимику, жесты, движения головы и тела, а также направление взгляда. Направление взгляда говорящего фиксируется слушающим, поэтому помимо зрительной функции обладает ещё и большой коммуникативной значимостью. В работе [Kibrik, Fedorova, 2018] рассматривается распределение внимания участников диалога: показывается, что внимание рассказчика в большей степени направлено на лицо говорящего и в меньшей степени — на его руки. Во многих работах, посвященных исследованию реальной коммуникации, отмечаются индивидуальные различия в поведении собеседников. Например, некоторые люди смотрят преимущественно в глаза, некоторые — преимущественно на рот собеседника, в то время как другие в различной степени распределяют взгляд между глазами и ртом [Kanan et al., 2015]. Более того, в эксперименте [Rogers et al., 2018] показывается, что субъективное восприятие зрительного контакта является продуктом взаимного взгляда на лицо (континуум между глазами и ртом собеседника), а не фактического взаимного зрительного контакта. Кроме этого, участники диалога часто субъективно переоценивают степень взаимного зрительного контакта [Gamer, Hecht, 2007]. Таким образом, если направление взгляда является существенным элементом коммуникации, то моделирование этого процесса важно как для анализа диалога, так и для создания привлекательных роботов, производящих ощущение зрительного контакта. Как было показано в [Häring, et. al., 2012] направленный взгляд робота повышает эффективность взаимодействия между роботом и пользователем.

Исследователи также выявили, что использование роботом ориентированных жестов, сопровождающих речь, способствует пониманию пространственной информации, увеличивает скорость и точность выполнения пространственных задач [Cabibihan, et. al., 2009]. В работе [Salem et al., 2012] показано, что испытуемые оценивают робота более позитивно, когда робот сопровождает речь жестами, даже если они семантически не соответствуют речи.

В лаборатории нейрокогнитивных технологий Курчатовского института мы разрабатываем робота Ф-2: персонального робота-компаньона, который способен поддерживать коммуникацию с человеком с помощью речи, жестов и мимики [Kotov et al, 2019]. Поведение робота-компаньона мы моделируем на основе анализа поведения людей

в реальных коммуникативных ситуациях на основе мультимодального корпуса REC (Russian Emotional Corpus). Корпус содержит размеченные в программе ELAN видеозаписи реальных эмоциональных диалогов ($n = 815$). В корпусе вручную размечаются речевые высказывания участников диалога. Для информанта размечаются движения его глаз, рук и губ. Для жеста или элемента мимики также отмечается коммуникативная функция в соответствии с типологией, представленной в [Котов, Зинина, 2015].

1. Исследование ориентированных жестов в корпусе

1.1. Направление взгляда

Стандартная разметка движений глаз учитывает значимые изменения направления линии взгляда (взгляд вбок, взгляд вверх), расширение глаз, прищур, подмигивание, закрытие глаз. Также в корпусе размечены поднятие бровей и движения носом. Такая разметка позволяет исследовать ключевые коммуникативные особенности информантов: например, глазодвигательное поведение во время фрустрации, радости, апелляции и др. Выделять паттерны типичные для начала или окончания разговора. Однако этой разметки недостаточно для моделирования сложного коммуникативного поведения робота-компаньона. Поэтому 2 видео из подкорпуса эмоциональных интервью (127 минут) были размечены с помощью дополненной разметки. Применявшаяся ранее в корпусе разметка направления взгляда была дополнена 7-ю новыми тегами: «справа» (от говорящего), «слева», «вверх», «вниз», «к собеседнику», «к объекту», «закрыты». Кроме этого, общий инвентарь коммуникативных функций был дополнен функциями, специфическими для глазодвигательного поведения: «размышление», «говорение», «фразовое ударение», «перечисление», «шутка», «слушание», «внимание на реакцию собеседника», «подражание направлению взгляда собеседника», «смягчение антисоциальной ситуации», «иконический взгляд», «сопровождение жеста», «взгляд на объект разговора». В общей сложности выделено 1875 аннотаций для движений глаз и 1415 аннотаций функциональной разметки. Результаты разметки представлены в таблице 1.

Исходя из полученных данных можно заключить, что информанты 36% всего времени находились в размышлении: отводили глаза в левую (38%) или в правую (23%) стороны. Во время разговора (20% от всех движений глаз) информанты в 100% случаев поддерживали зри-

Таблица 1. Функциональная разметка движений глаз X

Функция	Направление							сумма
	к собеседнику	вверх	слева	вниз	справа	закрыты	к объекту	
размышление	0	107	197	92	117	0	0	513
иконический взгляд	29	20	39	32	51	18	10	199
перечисление	1	3	11	4	5	0	0	24
шутка	2	3	5	1	7	0	0	18
смягчение антисоциальной ситуации	0	1	12	13	17	0	0	43
говорение	286	0	0	0	0	0	0	286
фразовое ударение	140	0	0	0	0	0	0	140
слушание	98	0	0	0	0	0	0	98
внимание на реакцию собеседника	68	0	0	0	0	0	0	68
подражание направлению взгляда собеседника	0	0	0	0	0	0	12	12
сопровождение жеста	0	0	0	7	3	0	0	10
взгляд на объект разговора	0	0	0	0	4	0	0	4
1415								

тельный контакт. Во время перечисления испытуемые в 46 % случаев смотрят налево, в 21 % — направо, в 17 % — вниз и в 13 % — вверх. Во время шутки взгляд респондентов также может быть направлен налево в 28 % случаев, направо — в 41 % случаев, вверх — в 17 % и вниз — в 6 %. Только иконический взгляд может выполняться в разных направлениях, к этому классу мы также отнесли демонстративное закрытие глаз (9 % случаев). Взгляд информанта в 100 % случаев направлен на собеседника при говорении, фразовом ударении, слушании, а также при обращении внимания на реакцию собеседника.

На основе разметки была разработана компьютерная модель на языке Python, регулирующая глазодвигательное поведение робота.

В этой модели были смоделированы коммуникативные состояния, соответствующие функциям, перечисленным выше. Эффективность этой модели в ситуации человеко-машинного взаимодействия была доказана в экспериментальном исследовании [Цфасман и др., 2018]. Полученные результаты позволяют судить о вкладе направления взгляда в формирование положительного впечатления от робота, а также доказывают эффективность разработанной модели глазодвигательного поведения в ситуации человеко-машинного взаимодействия.

1.2. Указательные жесты

Базовая разметка движений рук включает разметку на четырех слоях: это активный и пассивный органы, способ выполнения движения и траектория. В корпусе встречается 1165 случаев, когда информанты указывают на объект, на себя или на оппонента. Как правило, указательные жесты выполняются указательным пальцем (59,1 %) или ладонью/кистью руки (20,3 %) (Рис. 1)

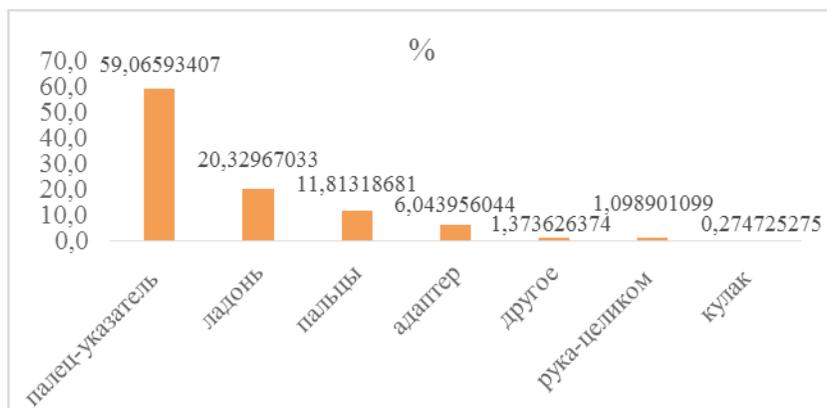


Рис. 1. Активные органы, с помощью которых выполняются указательные жесты (в процентах)

Указательные жесты обслуживают разные коммуникативные функции, описанные в [Котов, Зинина, 2015]. Как видно из Рис. 2, указательные жесты используются для апелляции (41,2 %) и сопровождают эмфазу (29,4 %), задействуются при ожидании обратной связи (8,8 %) и операциях с референтами (8,8 %).

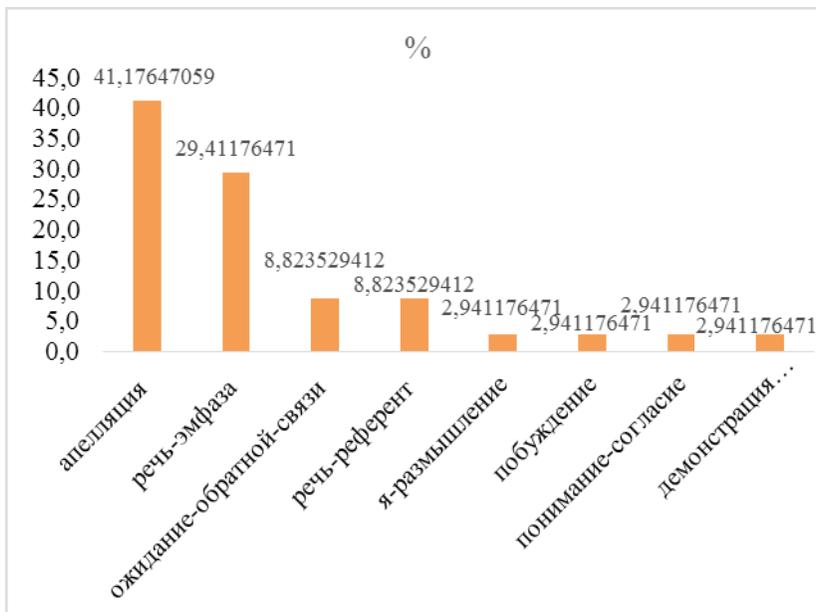


Рис. 2. Коммуникативные функции указательных жестов

1.3. Экспериментальное исследование указательных жестов робота

Для оценки влияния указательных жестов на формирование у пользователя положительного впечатления от робота был проведен эксперимент, в котором робот помогает испытуемому собирать головоломку (танграм) [Зинина и др., 2019]. В одном экспериментальном условии робот указывает человеку на необходимую фигуру головоломки (выполняет ориентированный жест), в другом — не указывает (выполняет неориентированные жесты). Экспериментальное взаимодействие робота и человека фиксировалось на две видеокамеры: первая записывала экспериментальную ситуацию сбоку, вторая записывала игровое поле и робота сверху. Взаимодействие пользователя с роботом было записано для анализа коммуникативных ответов человека на ориентированные жесты робота (31 испытуемый, 310 минут).

Разницу между экспериментальными условиями заметили 48,4% от всей выборки, из которых подавляющее большинство (73,3%) предпочло робота, демонстрирующего указательные жесты. Даже

если испытуемые не заметили разницы между двумя стратегиями жестового поведения робота (51,6 % от всей выборки отметили в анкете, что не увидели разницы в поведении), они, тем не менее, следовали указательным жестам робота и выбирали именно ту фигуру, на которую указал робот, в 78,5 % случаев. Следовательно, можно говорить об имплицитном влиянии ориентированных жестов робота на поведение пользователя, даже если испытуемые четко не осознавали это влияние. Поскольку оценки человека являются имплицитными и в половине случаев не проявляются в самоотчёте, реакции человека на поведение робота в эксперименте удобно изучать корпусными методами, записывая и размечая поведение испытуемых.

Выводы

Руководствуясь преимущественно прикладной задачей, мы развиваем корпус REC, при этом особое внимание уделяем ориентированным жестам и функциональному значению взглядов, ведь это ключевые особенности, способствующие повышению привлекательности робота для пользователя. Видеозаписи экспериментального исследования открывают возможности для изучения стратегий взаимодействия человека с роботом, например, реакций человека на одобрение роботом его действий, или, наоборот, на сообщение робота об ошибке человека. Анализ видеозаписей также позволяет исследовать разные стратегии взаимодействия человека с роботом, что важно как для теоретического анализа, так и для практического применения.

Литература

1. *Cabibihan J.J., So W.C., Nazar M., Ge S.S.* (2009), Pointing Gestures for a Robot Mediated Communication Interface // *Intelligent Robotics and Applications. ICIRA 2009. Lecture Notes in Computer Science*, Vol. 5928.
2. *Gamer M., Hecht H.* (2007), Are you looking at me? Measuring the cone of gaze. *Journal of Experimental Psychology*, Vol. 33, pp. 705–715.
3. *Håring M., Eichberg J., André E.* (2012), Studies on Grounding with Gaze and Pointing Gestures in Human-Robot-Interaction // *Social Robotics. ICSR 2012. Lecture Notes in Computer Science*, Vol. 7621.
4. *Kanan C., Bseiso D.N.F., Ray N.A., Hsiao J.H., Cottrell G.W.* (2015), Humans have idiosyncratic and task-specific scanpaths for judging faces // *Vision Research*, Vol. 108, pp. 67–76.
5. *Kibrik A.A., Fedorova O.V.* (2018), Language production and comprehension in face-to-face multichannel communication // *Компьютерная лингвистика и интеллектуальные технологии. Вып. 17 (24)*, с. 305–316.

6. Kotov A. A., Arinkin N. A., Zaydelman L. Y., Zinina A. A. (2019), Linguistic Approaches to Robotics: From Text Analysis to the Synthesis of Behavior, Language, Music and Computing // LMAC 2017. Communications in Computer and Information Science, Vol. 943, pp. 207–214.
7. Rogers S. L., Speelman C. P., Guidetti O., Longmuir M. (2018), Using dual eye tracking to uncover personal gaze patterns during social interaction // Scientific Reports, Vol. 8, Article number: 4271
8. Salem M., Kopp S., Wachsmuth I. (2012), Generation and Evaluation of Communicative Robot Gesture // International Journal of Social Robotics, Vol. 4(2), pp. 201–2017.
9. Зинина А. А., Аринкин Н. А., Зайдельман Л. Я., Котов А. А. (2019), Роль ориентированных жестов при коммуникации робота с человеком // Компьютерная лингвистика и интеллектуальные технологии. В печати.
10. Котов А. А., Зинина А. А. (2015), Функциональный анализ невербального коммуникативного поведения. Компьютерная лингвистика и интеллектуальные технологии. Вып. 14. Т. 1. М.: РГГУ, с. 299–310.
11. Цфасман М. М., Аринкин Н. А., Зайдельман Л. Я., Зинина А. А., Котов А. А. (2018), Разработка глазодвигательной коммуникативной системы робота Ф-2 на основе мультимодального корпуса. М.: Институт психологии РАН, с. 1328–1330.

References

1. Cabibihan JJ., So W.C., Nazar M., Ge S.S. (2009), Pointing Gestures for a Robot Mediated Communication Interface // Intelligent Robotics and Applications. ICIRA 2009. Lecture Notes in Computer Science, Vol. 5928.
2. Gamer M., Hecht H. (2007), Are you looking at me? Measuring the cone of gaze. Journal of Experimental Psychology, Vol. 33, pp. 705–715.
3. Häring M., Eichberg J., André E. (2012), Studies on Grounding with Gaze and Pointing Gestures in Human-Robot-Interaction // Social Robotics. ICSR 2012. Lecture Notes in Computer Science, Vol. 7621.
4. Kanan C., Bseiso D.N.F., Ray N.A., Hsiao J.H., Cottrell G.W. (2015), Humans have idiosyncratic and task-specific scanpaths for judging faces // Vision Research, Vol. 108, pp. 67–76.
5. Kibrik A. A., Fedorova O. V. (2018), Language production and comprehension in face-to-face multichannel communication // Компьютерная лингвистика и интеллектуальные технологии. Вып. 17 (24), с. 305–316.
6. Kotov A. A., Arinkin N. A., Zaydelman L. Y., Zinina A. A. (2019), Linguistic Approaches to Robotics: From Text Analysis to the Synthesis of Behavior, Language, Music and Computing // LMAC 2017. Communications in Computer and Information Science, Vol. 943, pp. 207–214.
7. Rogers S. L., Speelman C. P., Guidetti O., Longmuir M. (2018), Using dual eye tracking to uncover personal gaze patterns during social interaction // Scientific Reports, Vol. 8, Article number: 4271
8. Salem M., Kopp S., Wachsmuth I. (2012), Generation and Evaluation of Communicative Robot Gesture // International Journal of Social Robotics, Vol. 4(2), pp. 201–2017.

9. *Zinina A. A., Arinkin N. A., Zajdel'man L. Ya., Kotov A. A. (2019), Rol' orientirovanny'x zhestov pri kommunikacii robota s chelovekom [The role of oriented gestures during robot's communication to a human] // Komp'yuternaja lingvistika i intellektual'nye tehnologii [Computer linguistics and intellectual technologies]. In press.*
10. *Kotov A., Zinina A. (2015), Funkcional'nyj analiz neverbal'nogo kommunikativnogo povedenija [Functional analysis of nonverbal communicative behavior]. Komp'yuternaja lingvistika i intellektual'nye tehnologii [Computer linguistics and intellectual technologies] Issue 14. Vol. 1. Moscow.: RSUH, pp.299–310.*
11. *Tsfasman M. M., Arinkin N. A., Zajdel'man L. Ya., Zinina A. A., Kotov A. A. (2018), Razrabotka glazodvigatel'noj kommunikativnoj sistemy' robota F-2 na osnove mul'timodal'nogo korpusa [Developing eye gaze system for the F-2 robot based on a multimodal corpus]. M.: Institute of Psychology of RAS, pp.1328–1330.*

Зинина Анна Александровна

Курчатовский институт, Москва, Россия

Zinina Anna

Kurchatov Institute, Moscow, Russia

E-mail: Zinina_aa@nrcki.ru