

Лаборатория речевых и многомодальных интерфейсов

Руководитель лаборатории:

Карпов Алексей Анатольевич, главный научный сотрудник, доктор технических наук, профессор, лауреат премии Правительства РФ в области науки и техники – разработка речевых и многомодальных человеко-машинных интерфейсов и интеллектуальных систем, karpov@iias.spb.su.

Области исследований лаборатории

Исследование и разработка методов естественного взаимодействия человека с компьютером. Автоматическое аудиовизуальное распознавание и понимание речи. Многомодальные интерфейсы. Интеллектуальные пространства и умные комнаты. Ассистивные технологии и системы информационной поддержки людей с ограниченными возможностями. Анализ русского жестового языка. Компьютерная паралингвистика. Аффективные вычисления. Распознавание психоэмоциональных состояний. Многомодальное распознавание когнитивных нарушений людей.

Общая численность: 13 сотрудников.

Научные сотрудники и краткое наименование направления работ

Кипяткова Ирина Сергеевна, старший научный сотрудник, кандидат технических наук, доцент – методы акустического и языкового моделирования на основе искусственных нейронных сетей для систем автоматического распознавания речи; разработка систем распознавания речи для малоресурсных языков России, kipyatkova@iias.spb.su.

Иванько Денис Викторович, старший научный сотрудник, кандидат технических наук / PhD – аудиовизуальное распознавание русской речи с применением микрофона и высокоскоростной видеокамеры, ivanko.d@iias.spb.su.

Рюмин Дмитрий Александрович, старший научный сотрудник, кандидат технических наук – человеко-машинные интерфейсы, цифровая обработка изображений, распознавание образов, автоматическое распознавание жестовых языков, автоматическое распознавание визуальной речи, многомодальные интерфейсы, машинное обучение, нейронные сети, биометрические системы, ryumin.d@iiias.spb.su.

Величко Алёна Николаевна, старший научный сотрудник, кандидат технических наук – методы автоматического выявления деструктивной паралингвистической информации в разговорной речи, velichko.a@iiias.spb.su.

Аксёнов Александр Александрович, старший научный сотрудник, кандидат технических наук – методы вычисления визуальных признаков для автоматического чтения речи по губам диктора, аудиовизуальное распознавание речи, виртуальные аватары, axyonov.a@iiias.spb.su.

Маркитантов Максим Викторович, старший научный сотрудник, кандидат технических наук / PhD – автоматическое определение возраста и пола людей по аудио- и видеоинформации, многомодальная мягкая биометрия, markitantov.m@iiias.spb.su.

Рюмина Елена Витальевна, научный сотрудник, кандидат технических наук* (диссертация защищена 11.12.2025) – аффективные вычисления, цифровая обработка изображений, распознавание визуальных сигналов, автоматическое распознавание паралингвистических явлений, машинное обучение, нейронные сети, биометрические системы, ryumina.e@iiias.spb.su.

Кагиров Ильдар Амирович, научный сотрудник – сбор и аннотирование баз данных малоресурсных языков РФ, корпусная лингвистика, исследование жестовых интерфейсов, формализация грамматических структур жестового языка, kagirov@iiias.spb.su.

Двойникова Анастасия Александровна, младший научный сотрудник – автоматическое распознавание эмоциональных состояний по текстовым данным, методы распознавания вовлеченности участников виртуальной коммуникации, dvoynikova.a@iiias.spb.su.

Долгушин Михаил Дмитриевич, младший научный сотрудник – автоматическое распознавание речи на основе малоресурсных данных, методы автоматического многомодального определения когнитивных нарушений людей, dolgushin.m@iias.spb.su.

Киселева Ксения Олеговна, младший научный сотрудник – анализ грамматики и речевых данных прибалтийско-финских языков, kiseleva.k@iias.spb.su.

Островский Артём Олегович, младший научный сотрудник – сквозной машинный перевод речи, нейросетевой синтез речи, artoym.ru@gmail.com.

Аспиранты

Долгушин Михаил Дмитриевич, «Методы и программная система многомодального распознавания когнитивных нарушений людей» (научный руководитель – д.т.н., проф. Карпов А.А.)

Гранты и проекты

Карпов А.А. Проект РНФ № 25-11-00319 ОНГ «Интеллектуальная система многомодального распознавания когнитивных нарушений людей», 2025-2027.

Иванько Д.В. Проект РНФ № 25-71-00093 «Разработка и исследование методов улучшения качества речевых аудиосигналов с использованием глубоких нейронных сетей», 2025-2027.

Рюмин Д.А. Проект РНФ № 24-71-00083 «Исследование и разработка интеллектуальной системы распознавания жестов для управления интерфейсами человеко-машинного взаимодействия», 2024-2026.

Аксёнов А.А. Проект РНФ № 24-71-00112 «Исследование и разработка системы для синтеза реалистичных движений губ цифровых аватаров в соответствии с произносимой речью», 2024-2026.

Кипяткова И.С. Проект РНФ № 24-21-00276, «Автоматическое многоязычное распознавание речи с переключением кодов (на примере русского и карельского языков)», 2024-2025.

Иванько Д.В. Проект РНФ № 23-71-01056 «Интеллектуальная система автоматического двухстороннего сурдоперевода на основе распознавания и синтеза аудиовизуальной и жестовой речи», 2023-2025.

Карпов А.А. Договор с Университетом ИТМО на выполнение составной части НИОКР в рамках Исследовательского центра в сфере ИИ «Сильный искусственный интеллект в промышленности» (3-я волна) в целях реализации программы развития центра, 2025-2026.

Иванько Д.В. Договор с ООО «Техкомпания Хуавэй» (Москва), 2024-2025.

Сотрудничество с ВУЗами

Карпов А.А. – Университет ИТМО, СПбГУ, РГПУ им. Герцена.
Рюмин Д.А., Иванько Д.В., Аксёнов А.А., Рюмина Е.В. – ВШЭ.
Двойникова А.А. – Университет ИТМО.
Кагиров И.А. – СПбГУ.

Международное сотрудничество

Карпов А.А. – со-организация международных конференций «Речь и Компьютер» SPECOM 2025 (Сегед, Венгрия) совместно с коллегами из Сегедского университета и SPECOM 2024 (Белград, Сербия) совместно с коллегами из Университета Нови-Сада.

Карпов А.А., Маркитантов М.В., Рюмина Е.В. – инициативные совместные исследования и публикации с коллегами из Ульмского университета (Германия) и Утрехтского университета (Нидерланды).

Иванько Д.В. – инициативные совместные исследования с коллегами из Чжэцзянского технологического университета (Китай).

Членство в российских и международных организациях, диссертационных советах

Карпов А.А. – член экспертного совета РНФ по конкурсам инициативных проектов; координатор подкомитета по Восточной Европе международной ассоциации по речевой коммуникации ISCA, член международных научных ассоциаций IEEE, ACM, EURASIP;

член редколлегии журналов «Информатика и автоматизация», «Информатика», «Речевые технологии», «Эргодизайн», «Multimodal Technologies and Interaction»; генеральный со-председатель международной конференции «Речь и Компьютер» SPECOM 2025 (Сегед, Венгрия); член программных комитетов ряда топовых международных конференций; член диссертационных советов при СПб ФИЦ РАН, Университете ИТМО и СПбПУ, ученых советов СПб ФИЦ РАН и СПИИРАН.

Рюмин Д.А., Кипяткова И.С., Кагиров И.А., Иванько Д.В., Маркитантов М.В., Величко А.Н., Аксёнов А.А., Рюмина Е.В. – члены организационного и программного комитетов международной конференции SPECOM, а также программных / технических комитетов ряда топовых международных конференций, в частности INTERSPEECH, ICASSP, CVPR, EMNLP, LREC и др.

Интеллектуальная собственность

Программа для ЭВМ «Программа автоматизированного сбора и обработки видеоданных с медиаплатформ», авторы: Аксёнов А.А., Маркитантов М.В., дата регистрации: 16.12.2025, рег. номер: 2025696157.

Программа для ЭВМ «Библиотека программных средств для обучения, тестирования и анализа нейросетевых моделей автоматического распознавания управляющих жестов человека», автор: Рюмин Д.А., дата регистрации: 05.12.2025, рег. номер: 2025694528.

Программа для ЭВМ «Интеллектуальная система адаптивного распознавания и интерпретации управляющих жестов человека», автор: Рюмин Д.А., дата регистрации: 21.11.2025, рег. номер: 2025692526.

Награды, дипломы, стипендии

Научно-педагогическая школа «Интеллектуальные технологии многомодального человеко-машинного взаимодействия» под руководством д.т.н., проф. Карпова А.А. признана победителем

конкурсного отбора и включена в Реестр ведущих научных и научно-педагогических школ Санкт-Петербурга.

Рюмина Е.В., Маркитантов М.В., Аксёнов А.А., Рюмин Д.А., Долгушин М.Д., Карпов А.А. (команда RAS) – 1-е место в конкурсе по распознаванию составных эмоциональных проявлений “Compound Expression Recognition Challenge” в рамках 9-го международного семинара-соревновании [ABAW](#) (Affective & Behavior Analysis in-the-wild) при ведущей международной конференции по компьютерному зрению ICCV 2025 (США).

Рюмина Е.В. – Победитель конкурса на назначение стипендии Президента РФ для аспирантов и адъюнктов на 2024-2025 учебный год.

Новые результаты исследований

1. Разработана система многомодального распознавания составных психоэмоциональных состояний пользователей, одновременно анализирующая мимику лица, визуальную сцену, автоматически распознанную речь и текстовые транскрипции речи посредством комплексирования нейросетевых моделей-трансформеров с кросс-модальными механизмами внимания и анализом результатов анализа сцены визуальной большой языковой моделью (Visual LLM), которая позволяет автоматически распознавать семь классов базовых эмоций (радость, грусть, страх, злость, отвращение, удивление и нейтральное состояние), а также составные психоэмоциональные состояния, когда одновременно доминируют две базовых эмоции; применение системы позволило победить в конкурсе в рамках 9-го международного соревнования по анализу естественных аффективных состояний и поведения пользователя ABAW при конференции по компьютерному зрению ICCV [21, 1, 5] (*представлен в РАН как важнейший результат).

2. Разработана система автоматического распознавания слитной карельской речи с поддержкой переключения кодов языков (с карельского на русский), использующая дообученную

интегральную нейросетевую модель на базе Wav2Vec 2.0 BERT и внешнюю языковую модель, обученную на карелоязычных текстах, аугментированных путем автоматического перевода случайным образом выбранных слов на русский и замены на слова с внутрисловным переключением кода языка, которая может использоваться для автоматического стенографирования речи на карельском языке, представляя практический интерес для полевых лингвистических исследований, направленных на изучение, поддержку и сохранение малоресурсного карельского языка, и способствуя общему развитию технологий распознавания и синтеза речи на малоресурсных языках РФ [15, 9, 10].

3. Разработана система распознавания динамических жестов рук, объединяющая оптическую информацию и данные о глубине с использованием адаптивного механизма согласования межмодальных представлений и обеспечения временной согласованности, что повышает робастность распознавания жестов к пространственным неоднозначностям, вариативности жестовых исполнений и видеOSHUMAM при ограниченном объеме обучающих данных, а нейросетевые модели на основе архитектур пространства состояний обеспечивают эффективное моделирование пространственно-временной динамики жестов на длинных интервалах при сниженной вычислительной сложности и высокой стабильности признаков, благодаря чему система может применяться в системах автоматического распознавания жестовых языков и бесконтактных интерфейсах человеко-машинного взаимодействия [17, 20].

4. Разработано математическое обеспечение и выполнена программная реализация системы аудиовизуального распознавания русской звучащей и жестовой речи, включая метод на основе регулируемых трансформеров и метод с межъязыковым переносом обучения, что составило основу комплексного программного обеспечения интеллектуальной системы автоматического двустороннего сурдоперевода состоящего из 4-х модулей (визуальное распознавание жестового

языка/речи, аудиовизуальное распознавание звучащей речи, синтез звучащей речи и синтез жестового языка/речи), которое может эффективно применяться для организации бесконтактной коммуникации между глухими и слышащими людьми [19].

5. Разработан метод многомодального определения наличия состояния депрессии у пользователя по анализу его аудио, видео и текстовой информации с использованием полуавтоматической разметки данных, позволившей улучшить качество данных, используемых при обучении моделей, а также уточнить временные метки и тексты высказываний, что, в совокупности с использованием преимущественно экспертных наборов признаков (OpenFace, Word2Vec, DenseNet и др.) и классических детерминированных методов машинного обучения (Catboost и др.), обладающих свойством интерпретируемости моделей, повысило точность автоматического распознавания депрессии, который может быть использован в экспертных системах многомодального распознавания когнитивных нарушений, в том числе, в первичном звене здравоохранения [13].

6. Собран и аннотирован новый исследовательский многомодальный русскоязычный корпус данных взаимодействия участников виртуальной коммуникации (ENERGI – ENgagement and Emotion Russian Gathering Interlocutors) длительностью 6,4 часа аудиовизуальных записей групповых телеконференций, содержащий экспертную разметку данных по 3-м классам вовлеченности участников, их эмоциональному возбуждению и валентности эмоций, а также 10-и классам коммуникативных жестов, преимущества которого заключаются в многомодальности и русскоязычности аудиовизуальных данных, сбалансированности распределения информантов и их групповой коммуникации в натуральных условиях профессиональных телеконференций [22].

Список публикаций:

Статьи, подготовленные совместно с зарубежными организациями:

1. Markitantov Maxim, Ryumina Elena, Kaya Heysem, Karpov Alexey. Multi-Modal Multi-Task Affective States Recognition Based on Label Encoder Fusion // In Proc. 26th International Conference INTERSPEECH 2025. ISCA, Rotterdam, Netherlands, 2025. pp. 3010-3014. DOI: 10.21437/Interspeech.2025-2060 (WoS, Scopus, CORE A)

2. Mamontov Danila, Karpov Alexey, Minker Wolfgang. A Multilingual Telegram Chatbot for Mental Health Data Collection // In Proc. 27th ACM International Conference on Multimodal Interaction ICMCI 2025. ACM, Canberra, Australia, 2025. pp. 794-796. DOI: 10.1145/3716553.3757095 (WoS, Scopus)

3. Karpov Alexey, Gosztolya Gabor. SPECOM 2025 Preface // Lecture Notes in Computer Science / Proc. 27th International Conference on Speech and Computer SPECOM 2025, Szeged, Hungary. 2026, vol. 16187 / 16188, pp. v-vi. DOI: 10.1007/978-3-032-07956-5 (Scopus)

Статьи, опубликованные в изданиях, индексируемых в WoS, Scopus:

4. Markitantov Maxim, Ryumina Elena, Karpov Alexey. Audio-Visual Occlusion-Robust Gender Recognition and Age Estimation Approach Based on Multi-Task Cross-Modal Attention // Expert Systems with Applications. 2026. 296A. ID 127473. DOI: 10.1016/j.eswa.2025.127473 (WoS, Scopus, Q1)

5. Ryumina Elena, Ryumin Dmitry, Axyonov Alexandr, Ivanko Denis, Karpov Alexey. Multi-Corpus Emotion Recognition Method based on Cross-Modal Gated Attention Fusion // Pattern Recognition Letters. 2025. 190. pp. 192-200. DOI: 10.1016/j.patrec.2025.02.024 (WoS, Scopus, Q1)

6. Shilov Nikolay, Ponomarev Andrew, Ryumin Dmitry, Karpov Alexey. Generative Adversarial Framework with Composite Discriminator for Organization and Process Modelling—Smart City Cases // Smart Cities. 2025. 8(2). ID 38. DOI: 10.3390/smartcities8020038 (WoS, Scopus, Q1)

7. Ryumina Elena, Axyonov Alexandr, Abdulkadirov Timur, Koryakovskaya Darya, Ryumin Dmitry. Cross-Lingual Bimodal Emotion

Recognition with LLM-Based Label Smoothing // Big Data and Cognitive Computing. 2025. 9. 285. DOI: 10.3390/bdcc9110285 (WoS, Scopus, Q1)

8. *Карпов А.А., Попова П.К., Попов В.В.* XXVI Международная конференция “Речь и Компьютер” SPECOM-2024 // Известия Российской академии наук. Серия литературы и языка. 2025. 84(3). С. 119-126. DOI: 10.31857/S1605788025030124 (Scopus Q1, RSCI)

9. *Кагиров И.А., Киселева К.О., Кипяткова И.С.* Анализ переключения кодов в речи носителей карельского языка // Terra Linguistica. 2025. 2. 16. С. 22–40. DOI: 10.18721/JHSS.16202 (Scopus Q1)

10. *Кипяткова И.С., Кагиров И.А., Долгушин М.Д.* Применение предварительно обученных многоязычных моделей для распознавания карельской речи // Информатика и автоматизация. 2025. 24(2). С. 604-630. DOI: 10.15622/ia.24.2.9 (Scopus, RSCI)

11. *Долгушин М.Д., Карпов А.А.* Аналитический обзор речевых и многомодальных методов распознавания когнитивных нарушений людей // Информатика и автоматизация. 2025. 24(6). С. 1683-1720. DOI: 10.15622/ia.24.6.6 (Scopus, RSCI)

12. *Аксёнов А.А., Рюмина Е.В., Рюмин Д.А.* Метод генерации анимации цифрового аватара с речевой и невербальной синхронизацией на основе бимодальных данных // Научно-технический вестник информационных технологий, механики и оптики. 2025. 25(4). С. 651-662. DOI: 10.17586/2226-1494-2025-25-4-651-662 (Scopus, RSCI)

13. *Величко А.Н., Карпов А.А.* Многомодальный подход к определению депрессии с использованием полуавтоматической разметки данных и детерминированных методов машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. 2025. 25(6). С. 1107–1116. DOI: 10.17586/2226-1494-2025-25-6-1107-1116 (Scopus, RSCI)

14. *Dolgushin Mikhail, Guseva Daria, Karpov Alexey.* Investigation of Explainable Multimodal Methods for Detecting Mental Disorders // Lecture Notes in Computer Science / Proc. 27th International Conference on Speech and Computer SPECOM 2025, Szeged, Hungary. 2026. 16187. pp. 173-187. DOI: 10.1007/978-3-032-07956-5_12 (WoS, Scopus)

15. *Kipyatkova Irina, Kiseleva Kseniia, Dolgushin Mikhail, Kagirov Ildar*. Modeling Intra-word Code-Switching for Karelian ASR // Lecture Notes in Computer Science / Proc. 27th International Conference on Speech and Computer SPECOM 2025, Szeged, Hungary. 2026. 16188. pp. 104-117. DOI: 10.1007/978-3-032-07959-6_8 (WoS, Scopus)

16. *Zaburdaev Alexander, Ivanko Denis, Ryumin Dmitry*. CrossMP-SENet: Transformer-Based Cross-Attention for Joint Magnitude-Phase Speech Enhancement // Lecture Notes in Computer Science / Proc. SPECOM 2025, Szeged, Hungary. 2026. 16188. pp. 174-188. DOI: 10.1007/978-3-032-07959-6_13 (WoS, Scopus)

17. *Ryumin Dmitry, Egorova Angelina*. MoDeG-Prompt: Depth-Enhanced Multimodal Gesture Recognition with Dynamic Cross-Modal Prompting for Few-Shot Learning // Communications in Computer and Information Science / Proc. 28th International Conference “Internet and Modern Society” IMS 2025. St. Petersburg, 2026. 2672, pp. 347-360. DOI: 10.1007/978-3-032-05144-8_26 (Scopus)

18. *Axyonov Alexandr, Dolgushin Mikhail, Ryumin Dmitry*. NeRF-LipSync: A Diffusion Model for Speech-Driven and View-Consistent Lip Synchronization in Digital Avatars // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences / Proc. 6th International Workshop PSBB 2025. Moscow, 2025. XLVIII-2/W9-2025. pp. 25-31. DOI: 10.5194/isprs-archives-XLVIII-2-W9-2025-25-2025 (WoS, Scopus)

19. *Ivanko Denis, Ryumin Dmitry*. Intelligent System for Automatic Bidirectional Sign Language Translation Based on Recognition and Synthesis of Audiovisual and Sign Speech // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences / Proc. PSBB 2025. Moscow, 2025. XLVIII-2/W9-2025. pp. 131-136. DOI: 10.5194/isprs-archives-XLVIII-2-W9-2025-131-2025 (WoS, Scopus)

20. *Ryumina Elena, Ryumin Dmitry, Ivanko Denis*. G-MAE: Gesture-aware Masked Autoencoder for Human-Machine Interaction // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences / Proc. PSBB 2025. Moscow, 2025. XLVIII-2/W9-2025. pp. 241-248. DOI: 10.5194/isprs-archives-XLVIII-2-W9-2025-241-2025 (WoS, Scopus)

21. *Ryumina Elena., Markitantov Maxim., Axyonov Alexandr., Ryumin Dmitry, Dolgushin Mikhail, Karpov Alexey.* Zero-Shot Multimodal Compound Expression Recognition Approach using Off-the-Shelf Large Visual-Language Models // Proc. IEEE/CVF International Conference on Computer Vision (ICCV 2025) Workshops, 9th ABAW Workshop, USA, 2025. pp. 71-79. (WoS, Scopus)

Статьи, опубликованные в отечественных изданиях, индексируемых в РИНЦ:

22. *Двойникова А.А., Величко А.Н., Карпов А.А.* Многомодальный корпус данных взаимодействия участников виртуальной коммуникации ENERGI // Известия высших учебных заведений. Приборостроение. 2025. 68(12). С. 1011–1019. DOI: 10.17586/0021-3454-2025-68-12-1011-1019 (RSCI, ВАК, РИНЦ)

23. *Аксёнов А.А.* Многомодальная генерация речи, мимики и жестов в цифровых аватарах: современные методы и перспективы // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. 2025. 4. С. 155-182. DOI: 10.17308/sait/1995-5499/2025/4/155-182 (RSCI, ВАК, РИНЦ)

24. *Смолянинова А.В., Павлова Т.А., Доровских И.В., Карпов А.А., Долгушин М.Д., Краснослободцева Л.А., Сейку Ю.В., Болдаков Д.Ю.* Распознавание речи искусственным интеллектом у пациентов, страдающих легким когнитивным расстройством и деменцией // Психиатрия и психофармакотерапия. 2025. 4. С. 50-54. DOI: 10.62202/2075-1761-2025-27-4-50-54 (RSCI, ВАК, РИНЦ)

25. *Киселева К.О., Кипяткова И.С.* Переключение кода с карельского на русский: морфонологические аспекты внутрисловного копирования русских единиц в речи носителей карельского языка // Труды 11-го междисциплинарного семинара «Анализ разговорной русской речи» AP³-2025, Санкт-Петербург. 2025. С. 43-47. (РИНЦ)

26. *Долгушин М.Д., Гусева Д.Д., Карпов А.А.* Исследование методов автоматизированной диагностики когнитивных нарушений на материале видеointервью // Труды 11-го междисциплинарного семинара «Анализ разговорной русской речи» AP³-2025, Санкт-Петербург. 2025. С. 25-30. (РИНЦ)