

## **Лаборатория речевых и мультимодальных интерфейсов**

### **Руководитель лаборатории:**

Карпов Алексей Анатольевич, доктор технических наук, профессор - разработка речевых и мультимодальных человеко-машинных интерфейсов и интеллектуальных систем, karpov@iiias.spb.su.

### **Области исследований лаборатории**

Исследование и разработка методов естественного взаимодействия человека с компьютером. Автоматическое аудиовизуальное распознавание и понимание речи. Мультимодальные 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. Мультимодальное распознавание когнитивных нарушений.

**Общая численность:** 18 сотрудников.

### **Научные сотрудники и краткое наименование направления работ**

Карпов Алексей Анатольевич, главный научный сотрудник, доктор технических наук, профессор - разработка речевых мультимодальных человеко-машинных интерфейсов и интеллектуальных систем, karpov@iiias.spb.su.

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@iiias.spb.su.

Иванько Денис Викторович, старший научный сотрудник, кандидат технических наук - 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, биометрия, gumin.d@iias.spb.su.

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

Кагиров Ильдар Амирович, научный сотрудник - сбор и аннотирование баз данных малоресурсных языков РФ, корпусная лингвистика, исследование жестовых интерфейсов пользователя в сфере сервисной робототехники, формализация грамматических структур русского жестового языка, сбор и аннотирование баз данных русского жестового языка, kagirov@iias.spb.su.

Маркитантов Максим Викторович, научный сотрудник - автоматическое определение возраста и пола людей по аудио- и видеоинформации, многомодальная мягкая биометрия markitantov.m@iias.spb.su.

Рюмина Елена Витальевна, младший научный сотрудник - аффективные вычисления, цифровая обработка изображений, распознавание визуальных сигналов, автоматическое распознавание паралингвистических явлений, машинное обучение, нейронные сети, биометрические системы, человеко-машинные интерфейсы, gumina.e@iias.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@iiias.spb.su.

Ляксо Елена Евгеньевна, ведущий научный сотрудник, доктор биологических наук, профессор - паралингвистический анализ речи, определение психоэмоциональных состояний по речи, анализ детской речи, lyakso@gmail.com.

Крижановский Андрей Анатольевич, научный сотрудник, кандидат технических наук - корпусная лингвистика, индексирование, тезаурусы и машиночитаемые словари, лексикография, andrew.krizhanovsky@gmail.com.

Родионова Александра Павловна, научный сотрудник, кандидат филологических наук – анализ грамматики и фонетики карельского языка, sashenka22@yandex.ru.

Поволоцкая Анастасия Андреевна, младший научный сотрудник - экстралингвистический анализ разговорной речи, povolotskaya.a@iiias.spb.su.

Киселева Ксения Олеговна, младший научный сотрудник – анализ грамматики прибалтийско-финских языков, kiseleva.ksenia11@mail.ru.

## **Аспиранты**

Долгушин Михаил Дмитриевич, младший научный сотрудник «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 № 22-11-00321 «Интеллектуальная система многомодального распознавания аффективных состояний человека», 2022-2024.

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

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

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.

Рюмин Д.А. Грант-субсидия КНВШ Правительства Санкт-Петербурга для молодых кандидатов наук «Математические средства и интеллектуальная система для анализа психоэмоционального состояния и улучшения медицинской диагностики», 2024.

Иванько Д.В. Грант-субсидия КНВШ Правительства Санкт-Петербурга для молодых кандидатов наук «Разработка и исследование новых методов улучшения качества речевых аудиосигналов с использованием глубоких нейронных сетей», 2024.

Рюмина Е.В. Грант-субсидия КНВШ Правительства Санкт-Петербурга для молодых ученых «Интеллектуальная система мультимодального оценивания персональных качеств личности человека для автоматизации HR процессов», 2024.

Маркитантов М.В. Грант-субсидия КНВШ Правительства Санкт-Петербурга для молодых ученых «Разработка и исследование новых методов многомодального распознавания эмоций и сентимента с использованием нейросетевых технологий на основе современных темпоральных моделей и кросс модального внимания», 2024.

Карпов А.А. Договор с Университетом ИТМО на выполнение составной части НИОКР в рамках Исследовательского центра в сфере ИИ «Сильный искусственный интеллект в промышленности» в части разработки семейства алгоритмов гибридного и композитного ИИ,

алгоритмов генеративного дизайна и синтеза физических и цифровых объектов посредством перспективных методов ИИ, 2021-2024.

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

### **Сотрудничество с ВУЗаами**

Карпов А.А. – Университет ИТМО, СПбГУ

Двойникова А.А., Рюмина Е.В. – Университет ИТМО

Кагиров И.А. – СПбГУ

### **Международное сотрудничество**

Карпов А.А., Маркитантов М.В., Рюмина Е.В. – совместные исследования и публикации с Ульмским университетом (Германия), Утрехтским университетом (Нидерланды) и Университетом Нови-Сада (Сербия).

### **Членство в российских и международных организациях, диссертационных советах**

Карпов А.А. – член экспертного совета РФФИ по научным проектам; координатор подкомитета по Восточной Европе международной ассоциации по речевой коммуникации ISCA, член международных научных ассоциаций IEEE, ACM, EURASIP, IAPR; член редколлегии журналов «Информатика и автоматизация», «Речевые технологии», «Информатика», «Эргодизайн», «Multimodal Technologies and Interaction»; генеральный со-председатель международной конференции «Речь и Компьютер» SPECOM-2024 (Белград, Сербия); член программных комитетов международных конференций INTERSPEECH, ICASSP, LREC, IJCNN, ICMI, АСП, ПТИ, SLTU, TELFOR, O-COCOSDA, LT4All и др.; член диссертационных советов при СПб ФИЦ РАН и Университете ИТМО, ученых советов СПб ФИЦ РАН и СПИИРАН.

Кипяткова И.С. – член программных комитетов международных конференций INTERSPEECH, ICASSP, SPECOM, член ученого совета СПИИРАН.

Иванько Д.В. – член программных комитетов международных конференций INTERSPEECH, LREC, SPECOM, член ассоциации IEEE.

Рюмин Д.В. – член программных комитетов международных конференций INTERSPEECH, SPECOM.

### **Интеллектуальная собственность**

База данных «Речевая база данных с переключением кодов карельский-русский (KarRusCoS – Speech Database with Karelian-Russian Code-Switching)», авторы: Кипяткова И.С., Родионова А.П., Кагиров И.А., Крижановский А.А., дата регистрации: 06.12.2024, рег. номер: 2024625810.

Программа для ЭВМ «Интеллектуальная система автоматического двухстороннего сурдоперевода на основе распознавания и синтеза аудиовизуальной и жестовой речи», авторы: Иванько Д.В., Рюмин Д.А., дата регистрации: 22.10.2024, рег. номер: 2024684939.

Программа для ЭВМ «Интеллектуальная система многомодального анализа аффективных состояний человека (MASAI – Intelligent system for Multimodal Affective States Analysis)», авторы: Рюмин Д.А., Маркитантов М.В., Рюмина Е.В., Двойникова А.А., Карпов А.А., дата регистрации: 02.11.2024, рег. номер: 2024685861.

### **Награды, дипломы, стипендии**

Карпов А.А. – Юбилейная медаль «300 лет Российской академии наук» 08.02.2024 г. от Президента РАН.

Иванько Д.В., Рюмин Д.А. – Дипломы победителей конкурса грантов Санкт-Петербурга 2024 г. для молодых кандидатов наук от Правительства Санкт-Петербурга.

Маркитантов М.В., Рюмина Е.В. – Дипломы победителей конкурса грантов Санкт-Петербурга 2024 г. для молодых ученых от Правительства Санкт-Петербурга.

Рюмина Е.В. – Победитель конкурса на назначение стипендии имени С.П. Королёва для аспирантов на 2024-2025 учебный год.

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

## **Новые результаты исследований**

Разработан новый многомодальный метод и программная система MASAI автоматического распознавания эмоционально-окрашенной аудиовизуальной речи для 6 базовых эмоций и нейтрального состояния, а также сентимента (негативный, нейтральный, позитивный) на основе анализа аудио, видео и текстовой информации посредством комплексирования современных нейросетевых моделей и кросс-модального механизма внимания; предложенные методы многомодального распознавания позволили повысить показатели эффективности (точности и F-меры) автоматического распознавания эмоций и сентимента в сравнении с известными методами, что открывает новые перспективы для автоматизированных систем человеко-машинного взаимодействия, распознавания вовлеченности коммуникаторов и анализа пользовательского контента [1, 3].

Разработан новый аудиовизуальный метод и программная система AVCER автоматического распознавания 12 составных эмоций (радостно-удивленный, испуганно-удивленный, печально-удивленный, отвращенно-удивленный, разгневанно-удивленный, печально-напуганный, печально-разгневанный и т.д.) на основе машинного обучения без ознакомления с данными с использованием современных нейросетевых моделей (Wav2Vec 2.0 и ResNet-50 с LSTM моделью); объединение модальностей и решение о распознавании составных эмоций принимается на основе попарной суммы взвешенных распределений вероятностей базовых эмоций, что может найти применение при диагностике психоэмоциональных расстройств, выявлении стрессового поведения и учете эмпатии при человеко-машинном взаимодействии [2].

Разработано новое открытое математическое и программное обеспечение автоматического многомодального оценивания 5 персональных качеств личности человека (ПКЛЧ) по модели OCEAN, которое выполняет анализ гетерогенных (экспертных и

нейросетевых) признаков из мультимедийных записей, включающих аудио, видео и текстовую информацию, с последующим агрегированием признаков посредством управляемой сиамской нейронной сети, что позволяет улучшить значения показателей средней точности распознавания и коэффициента корреляции согласованности (Concordance Correlation Coefficient, CCC) ПКЛЧ в сравнении с известными методами, и имеет практическое применение в автоматическом ранжировании кандидатов на вакансии [7, 8, 18].

Усовершенствован метод и программная система аудиовизуального распознавания речи водителей автомобилей, основанный на использовании регулируемого трансформера с механизмом итеративного уточнения модели, многократного кодирования и ансамбля классификаторов, а также стратегии пространственно-временного объединения аудио- и видеoinформации, позволяющий эффективно синхронизировать модальности, учитывать их взаимное влияние и адаптироваться к сложным акустическим и визуальным условиям, включая шумное окружение, слабую освещенность и высокую степень окклюзии лиц, что повышает точность и робастность распознавания речевых команд, и делает систему более устойчивой к внешним помехам, обеспечивая эффективное бесконтактное управление ассистивными транспортными системами [6, 17].

Разработана система автоматического распознавания речи для малоресурсного карельского языка (ливвиковское наречие) на основе предварительно обученных современных многоязычных нейросетевых моделей Wav2Vec 2.0 и Whisper с дополнительной интеграцией внешней языковой модели, позволяющая достичь более высоких значений показателя точности распознавания речи по сравнению с базовыми модульными подходами, предназначенная для преобразования карельской речи в текст, что имеет важное значение для изучения речевого наследия и сохранения карельского языка, являющегося одним из официальных языков Республики Карелия, а также развития технологий автоматической обработки речи на различных малоресурсных языках РФ [20].



## Список публикаций:

*Статьи, подготовленные совместно с зарубежными организациями:*

1. *Dresvyanskiy D., Markitantov M., Yu J., Kaya H., Karpov A.* Multi-modal Arousal and Valence Estimation under Noisy Conditions // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). ABAW-2024. 2024. pp. 4773-4783. DOI: 10.1109/CVPRW63382.2024.00480 (WoS, Scopus)

2. *Ryumina E., Markitantov M., Ryumin D., Kaya H., Karpov A.* Zero-Shot Audio-Visual Compound Expression Recognition Method based on Emotion Probability Fusion // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). ABAW-2024. 2024. pp. 4752-4760. DOI: 10.1109/CVPRW63382.2024.00478 (WoS, Scopus)

3. *Dresvyanskiy D., Karpov A., Minker W.* A Cross-Multi-modal Fusion Approach for Enhanced Engagement Recognition // Lecture Notes in Computer Science, SPECOM-2024. 2024, vol. 15300, pp. 3-17. DOI: 10.1007/978-3-031-78014-1\_1 (Scopus)

4. *Mamontov D., Zepf S., Karpov A., Minker W.* Cross-Cultural Automatic Depression Detection Based on Audio Signals // Lecture Notes in Computer Science, SPECOM-2024. 2024, vol. 15299. pp. 309-323. DOI: 10.1007/978-3-031-77961-9\_23 (Scopus)

5. *Karpov A., Delić V.* SPECOM 2024 Preface // Proc. 26th International Conference on Speech and Computer SPECOM 2024, Belgrade, Serbia, Springer LNCS, vol. 15299 / 15300, 2024, pp. v-vi. DOI: 10.1007/978-3-031-78014-1 (Scopus)

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

6. *Ryumin D., Axyonov A., Ryumina E., Ivanko D., Kashevnik A., Karpov A.* Audio-visual speech recognition based on regulated transformer and spatio-temporal fusion strategy for driver assistive systems // Expert Systems with Applications. Elsevier. 2024, vol. 252, ID 124159. DOI: 10.1016/j.eswa.2024.124159 (WoS, Scopus Q1)

7. *Ryumina E., Markitantov M., Ryumin D., Karpov A.* OCEAN-AI framework with EmoFormer cross-hemiface attention approach for

personality traits assessment // Expert Systems with Applications. Elsevier. 2024, vol. 239, ID 122441. DOI: 10.1016/j.eswa.2023.122441 (WoS, Scopus Q1)

8. *Ryumina E., Markitantov M., Ryumin D., Karpov A.* Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment // Pattern Recognition Letters. Elsevier. 2024, vol. 185, pp. 45-51. DOI: 10.1016/j.patrec.2024.07.004 (WoS, Scopus Q1)

9. *Othman W., Kashevnik A., Ali A., Shilov N., Ryumin D.* Remote Heart Rate Estimation Based on Transformer with Multi-Skip Connection Decoder: Method and Evaluation in the Wild // Sensors. 2024. 24. pp. 775. DOI: 10.3390/s24030775 (WoS, Scopus Q1)

10. *Двойникова А.А., Кагиров И.А., Карпов А.А.* Метод распознавания сентимента и эмоций в транскрипциях русскоязычной речи с использованием машинного перевода // Информатика и автоматизация. 2024, Т. 23, № 4, С. 1173-1198. DOI: 10.15622/ia.23.4.9 (Scopus, Перечень ВАК, РИНЦ)

11. *Поволоцкая А.А., Карпов А.А.* Аналитический обзор методов автоматического анализа экстралингвистических компонентов спонтанной речи // Информатика и автоматизация. 2024, Т. 23, № 1, С. 5-38. DOI: 10.15622/ia.23.1.1 (Scopus, Перечень ВАК, РИНЦ)

12. *Иванько Д.В., Рюмин Д.А.* Автоматический сурдоперевод: обзор нейросетевых методов распознавания и синтеза звучащей и жестовой речи // Научно-технический вестник информационных технологий, механики и оптики. 2024, Т. 24, № 5. С. 669-686. DOI: 10.17586/2226-1494-2024-24-5-669-686 (Scopus, Перечень ВАК, РИНЦ)

13. *Уздяев М.Ю., Карпов А.А.* Создание и анализ многомодального корпуса данных для автоматического распознавания агрессивного поведения людей // Научно-технический вестник информационных технологий, механики и оптики. 2024, Т. 24, № 5. С. 834-842. DOI: 10.17586/2226-1494-2024-24-5-834-842 (Scopus, Перечень ВАК, РИНЦ)

14. *Величко А.Н., Карпов А.А.* Подход к распознаванию депрессии по речи человека с использованием полуавтоматической разметки данных // Информационно-управляющие системы. 2024, № 4,

С. 2-11. DOI: 10.31799/1684-8853-2024-4-2-11 (Scopus, Перечень ВАК, РИНЦ)

15. *Капустя К.Л., Князтькова И.С., Кагиров И.А.* Аналитический обзор интегральных моделей и стратегий распознавания речи на основе архитектуры трансформер // Информационно-управляющие системы. 2024, № 5, С. 2-15. DOI: 10.31799/1684-8853-2024-5-2-15 (Scopus, Перечень ВАК, РИНЦ)

16. *Бубеев Ю.А., Сыркин Л.Д., Поляков А.В., Усов В.М., Карпов А.А., Иванов А.В.* Применение телеконсультаций для оценки личностного адаптационного потенциала и выявления показаний к оказанию квалифицированной психологической помощи работникам в удаленных регионах РФ // Авиакосмическая и экологическая медицина. 2024. Т. 58. С. 5-16. DOI: 10.21687/0233-528X-2024-58-1-5-16 (Scopus, Перечень ВАК, РИНЦ)

17. *Axyonov A., Ryumin D., Ivanko D., Kashevnik A., Karpov A.* Audio-Visual Speech Recognition In-The-Wild: Multi-Angle Vehicle Cabin Dataset and Attention-Based Approach // In Proc. 49th IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-2024, IEEE Press, Seoul, Korea, 2024, pp. 8195-8199. DOI: 10.1109/ICASSP48485.2024.10448048 (WoS, Scopus)

18. *Ryumina E., Ryumin D., Karpov A.* OCEAN-AI: Open Multimodal Framework for Personality Traits Assessment and HR-processes Automatization. In Proc. 25th International Conference INTERSPEECH-2024. ISCA, Kos, Greece, 2024, pp. 3630-3631 (WoS, Scopus, CORE A)

19. *Ivanko D., Ryumin D., Axyonov A., Kashevnik A., Karpov A.* OpenAV: Bilingual Dataset for Audio-Visual Voice Control of a Computer for Hand Disabled People // Lecture Notes in Computer Science, SPECOM-2024, 2024, vol. 15299, pp. 163-173. DOI: 10.1007/978-3-031-77961-9\_12 (Scopus)

20. *Kipyatkova I., Kagirow I., Dolgushin M., Rodionova A.* Towards a Livvi-Karelian End-to-End ASR System // Lecture Notes in Computer Science, SPECOM-2024, 2024, vol. 15299, pp. 57-68. DOI: 10.1007/978-3-031-77961-9\_4 (Scopus)

21. *Guseva D., Mitrofanova O., Dolgushin M.* Human and Machine Keyphrase Perception in Russian Text and Speech // Lecture Notes in Computer Science, SPECOM-2024, 2024, vol. 15299, pp. 265-280. DOI: 10.1007/978-3-031-77961-9\_20 (Scopus)

22. *Kosulin K., Karpov A.* A Survey of Masked Face Recognition Methods and Corpora/Data // Springer Geography. IMS-2022. 2024, pp. 27-37. DOI: 10.1007/978-3-031-50609-3\_3 (Scopus)

23. *Ivanko D., Ryumin D., Markitantov M.* End-to-End Visual Speech Recognition for Human-Robot Interaction // AIP Conference Proceedings. 2024, vol. 3021, pp. 82-90. DOI: 10.1063/5.0197720 (Scopus)

*Статьи, опубликованные в отечественных изданиях, индексируемых в РИНЦ:*

24. *Двойникова А.А., Карпов А.А.* Методика создания многомодальных корпусов данных для аудиовизуального анализа вовлеченности и эмоций участников виртуальной коммуникации // Известия вузов. Приборостроение. 2024, Т. 67, № 11, С. 984–993. DOI: 10.17586/0021-3454-2024-67-11-984-993 (RSCI, Перечень ВАК, РИНЦ)