Laboratory of Speech and Multimodal Interfaces

Head of the laboratory: Dr. Tech. Sci., Assoc.Prof. Alexey A. Karpov – development of speech and multimodal human-computer interfaces and systems, karpov@iias.spb.su, http://hci.nw.ru/en

Laboratory staff: 14 members.

Research activities – research and development of methods for natural human-computer interaction. Automatic audio-visual speech recognition and understanding. Multimodal user interfaces. Intelligent rooms and spaces. Assistive information technologies and systems for disabled people. Russian sign language research. Computational paralinguistics. Psycho-emotional states recognition.

Research fellows and brief information of the research-work direction

Senior researcher, Ph.D. – Irina S. Kipyatkova – methods for language and acoustic modeling based on artificial neural networks for automatic Russian speech recognition systems, kipyatkova@iias.spb.su.

Leading researcher, Dr. Bio. Sci., Prof. – Elena E. Lyakso – paralinguistic speech analysis, psycho-emotional states detection from speech, analysis of children's speech, lyakso@gmail.com.

Leading researcher, Ph.D. – Lev A. Stankevich – multimodal interfaces and robotic systems, stankevich_lev@inbox.ru.

Researcher – Denis V. Ivanko – audio-visual Russian speech recognition with the use of a microphone and a high-speed video camera, denis.ivanko11@gmail.com.

Researcher – Dmitry A. Ryumin – automatic recognition of gestures and elements of Russian sign language, dl_03.03.1991@mail.ru.

Junior researcher – Oksana V. Verkholyak – automatic recognition of speaker's emotional states using voice characteristics and tonality of the text of the statement, overkholyak@gmail.com.

Junior researcher (PhD student) – Alena N. Velichko – methods for automatic detection of destructive paralinguistic phenomena in conversational speech, velichko.a.n@mail.ru.

Junior researcher – Ildar A. Kagirov – formal representation of Russian sign language grammatical structures, collection and annotation of

Russian sign language databases, investigation of gesture based user interfaces in the area of service robotics, kagirov@iias.spb.su.

Junior researcher – Alexander A. Axyonov – visual features calculation methods for automatic lip-reading, a.aksenov95@mail.ru.

Junior researcher – Maxim V. Markitantov – automatic gender and age recognition from speech, m.markitantov@yandex.ru.

Programmers (master students)

Anastasia A. Dvoynikova – automatic recognition of emotional states by text data, anastasia.dvoynikova@gmail.com.

Elena V. Ryumina – automatic recognition of human emotions by facial expressions, ryumina_ev@mail.ru.

Nikita M. Markovnikov – methods for end-to-end continuous Russian speech recognition, niklemark@gmail.com.

Grants and projects

Karpov A. – Agreement N 14.616.21.0095 (ID: RFMEFI61618X0095) with the Ministry of Science and Higher Education of the Russian Federation, Special federal programme "Research and development in priority areas of development of the scientific and technological complex of Russia for 2014-2020" (Event 2.2), project "Multi-modal interface based on gestures, speech, and sign language for control of an assistive mobile information robot - AMIR", foreign partner: The University of West Bohemia, Pilsen, Czech Republic, 2018-2020 (jointly with A. Saveliev's laboratory).

Karpov A. – Project of the Russian Science Foundation N 18-11-00145 "Development and research of an intelligent system for complex paralinguistic analysis of speech", 2018-2020.

Karpov A. – Project of RFBR N 19-29-09081 mk "Software and infoware for intelligent analysis of video and audio information for assistive mobile systems in vehicles", 2019-2022 (jointly with A. Smirnov's laboratory)

Karpov A. – Project of RFBR N 16-37-60100-mol_a_dk "Development of a universal assistive information technology based on multimodal human-computer interfaces", 2016-2019

Kipyatkova I. – Project of RFBR N 18-07-01216-a "Development of an end-to-end continuous Russian speech recognition system using deep neural networks", 2018-2020.

Karpov A. – Project of RFBR N 18-07-01407-a "Automatic bimodal recognition of natural emotions in Russian speech", 2018-2020.

Ivanko D. – Project of RFBR N 18-37-00306-mol_a "Methods, models and algorithms of visual signals processing for lip-reading", 2018-2020.

Karpov A. – R&D contracts on "Development of voice control module software for a robotic exoskeleton for medical purposes" with the Volga State Technological University (VSTU, Yoshkar-Ola) within the framework of the integral project in the framework of the Government Statement No. 218, 2017-2019.

Karpov A. – R&D contract with Huawei Technologies Co. (Shenzhen, China), within the framework of the innovation research program HIRP Open, 2017-2019.

Karpov A. – R&D contract with "Mobile TeleSystems" (MTS, Moscow), 2019

Karpov A. – Contract with "ASM Solutions" LLC (Moscow), 2019.

University courses

ITMO University: Speech Recognition – A. Karpov.

SUAI University: Automated information management systems, Speech Recognition Technologies – I. Kipyatkova.

Scientific and organizational activity

Organization of 21th International Conference on Speech and Computer SPECOM-2019. http://specom.nw.ru/history/sites/2019. Istanbul, Turkey, 20–25 August 2019 – A. Karpov (General co-chair). Proceedings published in: Speech and Computer. Springer International Publishing Switzerland.: SPECOM 2019, LNAI 11658, 2019, 978 p. URL: https://link.springer.com/book/10.1007/978-3-030-26061-3.

International cooperation

Joint research and organization of scientific events in cooperation with the University of West Bohemia (Czech Republic), Bogazici University (Turkey), Namık Kemal University (Turkey), Leipzig University of Telecommunications (Germany), University of

Patras (Greece), Dresden University of Technology (Germany), Ulm University (Germany), United Institute of Information Problems of the National Academy of Sciences of Belarus, University of Aizu (Japan), University of Hertfordshire (Great Britain), Huawei Technologies company (China).

Participation in conferences and exhibitions

21st International Conference "Speech and Computer" SPECOM-2019, 20–25 August 2019, Istanbul, Turkey – Karpov A., Kipyatkova I., Kagirov I., Markitantov M. (conference co-organization).

4th International Conference on Interactive Collaborative Robotics ICR-2019, 20–25 August 2019, Istanbul, Turkey – Kagirov I., Karpov A.

17th International Conference on Pervasive Computing PerCom-2019, 11-15 March 2019, Kyoto, Japan – Ivanko D., Karpov A.

44th International Conference on Acoustics, Speech and Signal Processing, ICASSP-2019, 12-17 May 2019, Brighton, UK - Karpov A., Verkholyak O.

20th International Conference INTERSPEECH-2019, 15-19 September 2019, Graz, Austria – Karpov A.

19th International Conference on Discourse and Dialogue SIGDIAL-2019, 11-13 Septembre 2019, Stockholm, Sweden – Karpov A.

27th International Conference on Multimedia ACM Multimedia-2019, 21-25 October, Nice, France – Karpov A.

14th International Conference on Electromechanics and Robotics "Zavalishin's Readings", 17-20 April, Kurks, Russia – Ivanko D.

30th International Conference on Extreme Robotics ER-2019, 13-15 June 2019, St. Petersburg, Russia – D. Ryumin, I. Kagirov.

5th International Workshop on "Photogrammetric and computer vision techniques for video surveillance, biometrics and biomedicine" PSBB-2019, May 13-15, 2019, Moskow, Russia – D. Ryumin, D. Ivanko.

13th International Symposium on Intelligent Distributed Computing IDC 2019, 7-10 Oktober 2019, St. Peterburg, Russia – Velichko A., Kagirov I.

8th Seminar on Analysis of spoken Russian AP3-2019, St. Petersburg – Kagirov I., Velichko A., Markovnikov N.

3rd International Conference in Language Engineering and Applied Linguistics "R. Piotrowski's Readings 2019", St. Petersburg, November 27, 2019 - Karpov A., Dvoinikova A., Velichko A.

Huawei workshop on processing of acoustic data, speech and signals, November 30 - December 4, 2019, Minsk, Belarus - Markitantov M.V., Karpov A.A.

UNESCO International Conference “Language Technologies for All” (LT4All), December 4-6, 2019, Paris, France - A. Karpov.

Membership in Russian and International societies, editorial boards, etc.

Karpov A. – Expert of the RAS; scientific Ambassador of St. Petersburg; member of the European Association for Signal Processing (EURASIP), EURASIP Local Liaison Officer in Russia, member of the International Speech Communication Association (ISCA), member of the International Association for Pattern Recognition (IAPR); Editorial board member of the journals "SPIIRAS Proceedings" (St. Petersburg), "Speech Technologies" (Moscow) and "Informatics" (Minsk); Guest editor of the Journal on Journal on Multimodal User Interfaces (Springer), Speech Communication (Elsevier), Journal of Electrical and Computer Engineering (Hindawi); reviewer of several international journals IEEE/ACM Transactions on Audio, Speech and Language Processing; IEEE Transactions on Affective Computing; IEEE Transactions on Biomedical Engineering; IEEE Journal of Biomedical and Health Informatics; Neurocomputing; Computer Speech & Language; Speech Communication; IEEE Signal Processing Letters, Pattern Recognition Letters; Pattern Recognition; Language Resources and Evaluation; Soft Computing; Journal of Information Science; Acoustical Physics etc.; co-chair of the International Conference SPECOM series, technical/program committee member of the international conferences INTERSPEECH, ICASSP, ICPR, SLTU, SPECOM, Baltic HLT, HBU, SIU, DOGS, etc.

Kipyatkova I. – technical/programme committee member of the international conferences INTERSPEECH, ICASSP, SPECOM, ISNN, member of the organizing committee of the international conference SPECOM series; member of the International Speech Communication Association (ISCA).

Verkholyak O. – member of the International Speech Communication Association (ISCA), member of the Association for Computational Linguistics (ACL), member of the IEEE Young Professionals and IEEE Membership.

Ivanko D. - scientific committee member of the international conference LREC, IEEE Membership.

Intellectual property

- Patent of Russia for invention: Mobile autonomous robotic platform with block structure, No. 2704048 dated 10.23.2019, Savelyev A.I., Kharkov I.Yu., Pavlyuk N.A., Karpov A.A.

- Computer program: Software system for determining the gender and age of a speaker (GASpeakerRecognizer), No. 2019662952 dated 10.10.2019, Markitantov M.V., Karpov A.A.

- Computer program "Software for recording a gesture database using the Kinect v2 sensor", Ryumin D.A., Karpov A.A., Certificate No. 2019612755 of 02.27.2019

- Computer program “A software system for automatic speech recognition based on models using connectionist temporal classification”, Markovnikov N.M., Kipyatkova I.S., Certificate No. 2019615375 of April 25, 2019.

- Database “Corpora of continuous Russian speech for automatic speech recognition systems”, Kipyatkova I.S., Kagirov I.A., Karpov A.A., Certificate No. 2019620328 of February 27, 2019.

Recent Results

1. End-to-end encoder-decoder models with connectionist temporal classification using various types of neural networks, such as Highway, ResNet, DenseNet, DiracNet, and Transforme trained using data augmentation methods based on tempo and pitch perturbations and adding white noise have been developed for recognition of continuous Russian speech. This allows achieving higher recognition speed comparing to the standart speech recognition system. [13, 20, 24]

2. A multimodal method (color video stream and depth map) for recognizing static and dynamic one-handed gestures of the Russian sign language using a three-dimensional convolutional deep neural network with long short-term memory (LSTM) was developed, which allows to extract both short-term and long-term spatio-temporal characteristics of gestures [10, 15, 22, 31]

3. A method for recognizing emotions in conversational speech based on a hierarchical model of a recurrent neural network with long short-term memory (RNN-LSTM), as well as a data adaptation method that allows the efficient use of a cross-corpus experimental setup was developed and made it possible to obtain results that exceed the existing values in the literature. [1, 2, 5, 7]

4. A new approach to the recognition of the speaker's gender and age has been developed based on deep neural networks (DNN), trained and tested on the dataset of German speech aGender using both simple models and more complex ones based on different DNN topologies, including neural networks with fully connected and convolutional layers, which allowed to achieve the best recognition results in gender and age of the speaker by voice. [14, 27, 30]

5. New types of informative features based on the color and geometry of the images of the lips for the audio-visual speech recognition system and lip-reading system were developed and implemented using computer vision libraries and various configurations of Hidden Markov Models, including coupled and multistream ones, which allowed to exceed the recognition accuracy values known in the literature for Russian speech. [8, 11, 18]

6. The multimedia database of vocabulary of the Russian sign language has been expanded and annotated, complex analysis has been conducted, which made it possible to clarify the classes of gestures relevant for automatic recognition and describe the variability of the gestures during spontaneous speech and is planned to be used in creating own gesture speech recognition system. [3, 23, 25]

7. A model of an intelligent interface for human-machine interaction was developed and justified from the usability point of view, which includes both voice control and a gesture interface and intended for use in assistive technologies (it was also implemented in the robotic medical exoskeleton and in the robotic trolley for supermarkets). [17, 21, 29, 32]

8. Methods of artificially increasing the amount of data using SMOTE and ADASYN techniques were studied, which positively affected the operation of the models; conducted experimental studies using neural networks and the principal component method made it possible to reduce

the dimension of data and to clear data from noise, which led to an increase in the accuracy of classification of false and true speech utterances. [16, 26]

Awards, certificates, scholarships

Karpov A.A. - Certificate “For significant merits in the field of science and many years of conscientious work”, the Ministry of Science and Higher Education of the Russian Federation, order No. 38 / pr, dated June 26, 2019

Publications

Papers prepared jointly with foreign organizations:

1. *Verkholyak O., Fedotov D., Kaya H., Zhang Y., Karpov A.* Hierarchical Two-Level Modelling of Emotional States in Spoken Dialog Systems. In Proc. 44th IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-2019, Brighton, UK, 2019, pp. 6700-6704. https://doi.org/10.1109/ICASSP.2019.8683240. (WoS, Scopus, SJR = 0.48)
2. *Kaya H., Fedotov D., Dresvyanskiy D., Doyran M., Mamontov D., Markitantov M., Akdag Salah A., Kavcar E., Karpov A., Salah A.A.* Predicting depression and emotions in the cross-roads of cultures, para-linguistics, and non-linguistics. In Proc. 9th ACM International Audio/Visual Emotion Challenge and Workshop AVEC’19, Nice, France, 2019, ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3347320.3357691. (Scopus)
3. *Ryumin D., Ivanko D., Kagirov I., Axyonov A., Karpov A., Zelezny M.* Human-Robot Interaction with Smart Shopping Trolley using Sign Language: Data Collection. In Proc. 2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019, Kyoto, Japan, 2019, pp. 949-954. https://doi.org/10.1109/PERCOMW.2019.8730886. (WoS, Scopus)

4. *Akhtiamov O., Siegert I., Karpov A., Minker W.* Cross-Corpus Data Augmentation for Acoustic Addressee Detection. In Proc. 20th ACL International Conference on Discourse and Dialogue SIGDial-2019, Stockholm, Sweden, 2019, pp. 274-283. https://www.aclweb.org/anthology/W19-5933 (Scopus)
5. *Fedotov D., Kim B., Karpov A., Minker W.* Time-Continuous Emotion Recognition Using Spectrogram Based CNN-RNN Modelling // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 93-102. https://doi.org/10.1007/978-3-030-26061-3_10 (Scopus, SJR = 0.28, Q2)
6. *Yu J., Markov K., Karpov A.* Speaking Style Based Apparent Personality Recognition // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 540-548. https://doi.org/10.1007/978-3-030-26061-3_55 (Q2, Scopus, SJR = 0.28)
7. *Verkholyak O.V., Kaya H., Karpov A.A.* Modeling short-term and long-term dependencies of the speech signal for paralinguistic emotion classification // SPIIRAS Proceedings, Issue 62, № 1, 2019, pp. 30-56. https://doi.org/10.15622/sp.18.1.30-56. (Q3, Scopus, SJR = 0.17)
8. *Ivanko D.V., Ryumin D.A., Karpov A.A., Zhelezny M.* Investigation of the influence of high-speed video data on the recognition accuracy of audiovisual speech // Information-Control Systems (Informatsionno-Upravliaiushchie Sistemy) [Information and Control Systems], No. 2, 2019, pp. 26-34. https://doi.org/10.31799/1684-8853-2019-2-26-34. (Scopus)
9. *Fedotov D.V., Verkholyak O.V., Karpov A.A.* Contextual continuous recognition of emotions in Russian speech using recurrent neural networks. Proceedings of the 8th Interdisciplinary Seminar “Analysis of Conversational Russian Speech” AR3-2019, St. Petersburg, St. Petersburg State University, 2019, pp. 96-99. (RSCI)

Papers published in editions, indexed by WoS, Scopus:

10. *Ryumin D., Kagirov I., Ivanko D., Axyonov A., Karpov A.* Automatic detection and recognition of 3D manual gestures for human-machine interaction. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS Archives 42(2/W12), 2019, pp. 179-183. https://doi.org/10.5194/isprs-archives-XLII-2-W12-179-2019. (Scopus, SJR = 0.31)
11. *Ivanko D., Ryumin D., Karpov A.* Automatic lip-reading of hearing impaired people // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives 42(2/W12), 2019, pp. 97-101. https://doi.org/10.5194/isprs-archives-XLII-2-W12-97-2019. (Scopus, SJR = 0.31)
12. *Kipyatkova I.* LSTM-Based Language Models for Very Large Vocabulary Continuous Russian Speech Recognition System // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 219-226. (Q2, Scopus, SJR = 0.28)
13. *Markovnikov N., Kipyatkova I.* Investigating Joint CTC-Attention Models for End-to-End Russian Speech Recognition // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 337-347. (Q2, Scopus, SJR = 0.28)
14. *Markitantov M., Verkholyak O.* Automatic Recognition of Speaker Age and Gender Based on Deep Neural Networks // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 327-336. (Q2, Scopus, SJR = 0.28)
15. *Kagirov I., Ryumin D., Axyonov A.* Method for Multimodal Recognition of One-Handed Sign Language Gestures Through 3D Convolution and LSTM Neural Networks // Lecture Notes in Computer Science, Springer LNAI 11658, SPECOM 2019, 2019, pp. 191-200. (Q2, Scopus, SJR = 0.28)

16. *Velichko A., Budkov V., Kagirov I., Karpov A.* Applying Ensemble Learning Techniques and Neural Networks to Deceptive and Truthful Information Detection Task in the Flow of Speech // Studies in Computational Intelligence, Springer SCI 868, IDC 2019, 2020, pp. 457-466. https://doi.org/10.1007/978-3-030-32258-8_54 (Scopus, SJR = 0.183)
17. *Kagirov I., Karpov A., Kipyatkova I., Klyuzhev K., Kudryavcev A., Kudryavcev I., Ryumin D.* Lower Limbs Exoskeleton Control System Based on Intelligent Human-Machine Interface // Studies in Computational Intelligence, Springer SCI 868, IDC 2019, 2020, pp. 477-482. https://doi.org/10.1007/978-3-030-32258-8_56 (Scopus, SJR = 0.183)
18. *Ivanko D., Ryumin D., Kipyatkova I., Axyonov A., Karpov A.* Lip-reading Using Pixel-based and Geometry-based Features for Multimodal Human-Robot Interfaces // Smart Innovation, Systems and Technologies, Springer, vol. 154, Zavalishin's Readings 2019, 2020, pp. 477-486. https://doi.org/10.1007/978-981-13-9267-2_39. (Scopus, SJR = 0.156)
19. Ryumin D., Ivanko D., Kagirov I., Axyonov A., Karpov A. Vision-Based Assistive Systems for Deaf and Hearing Impaired People // In: Favorskaya M., Jain L. (eds) Computer Vision in Advanced Control Systems-5. Intelligent Systems Reference Library, Springer, vol. 175, 2020, pp. 197-224. https://doi.org/10.1007/978-3-030-33795-7_7 (Scopus, SJR = 0.163, Q3)
20. *Markovnikov N.M., Kipyatkova I.S.* The study of methods for constructing coder-decoder models for recognizing Russian speech // Information Management Systems (Informatsionno-Upravliaiushchie Sistemy), No. 4, 2019, pp. 45–53. https://doi.org/10.31799/1684-8853-2019-4-45-53 (Scopus)
21. *Kagirov I.A., Karpov A.A., Kipyatkova I.S., Klyuzhev K.S., Kudryavtsev A.I., Kudryavtsev I.A., Ryumin D.A.* Intelligent interface for controlling the robotic medical exoskeleton of the lower extremities Remotion // Aerospace and Environmental

Medicine (Aviakosmicheskaia i Ekologicheskaia Meditsina), No. 5, 2019, pp. 92-98. https://doi.org/10.21687/0233-528X-2019-53-5-92-98 (Scopus, SJR = 0.23, Q3)

Papers published in editions, indexed by Russian Science Citation Index (RSCI):

22. *Kagirov I.A., Tolstoy I.M., Savelyev A.I., Karpov A.A.* Gesture control of a collaborative robot // Robotics and technical cybernetics. vol. 7, No. 2, 2019, P. 139-144. https://doi.org/10.31776/RTCJ.7208
23. *Kagirov I.A., Manueva Yu.S.* Development of a predicate model for the selection of lexical meanings when translating into Russian sign language based on the semantic dictionary of VA Aces and automatic dialing systems "Dialing" // Scientific Bulletin of Nizhny Novgorod State Technical University, No. 1 (74), 2019, P. 41-60. https://doi.org/10.17212/1814-1196-2019-1-41-60.
24. *Markovnikov N.M., Kipyatkova I.S.* The study of methods for constructing integrated recognition systems for Russian speech without extracting features. Proceedings of the 8th interdisciplinary seminar "Analysis of spoken Russian speech" AR3-2019, St. Petersburg, St. Petersburg State University, 2019, pp. 71-77.
25. *Kagirov I.A.* The project of the specialized corpora of the Russian sign language for use in human-machine interfaces. Proceedings of the 8th Interdisciplinary Seminar "Analysis of Conversational Russian Speech" AR3-2019, St. Petersburg, St. Petersburg State University, 2019, pp. 39-45.
26. *Velichko A.N., Budkov V.Yu.* Development of a prototype system for the automatic determination of false and true information in speech. Proceedings of the 8th interdisciplinary seminar "Analysis of spoken Russian speech" AR3-2019, St. Petersburg, St. Petersburg State University, 2019, pp. 17-20.
27. *Markitantov M.V., Karpov A.A.* Automatic recognition of age and gender by voice based on deep neural networks // Information-

measuring and control systems, Vol. 17, No. 5, 2019. pp. 76-83. https://doi.org/10.18127/j20700814-201905-10

28. *Aksenov A.A., Ryumin D.A., Kagirov I.A.* The method of multimodal recognition of one-handed gestures // Information-measuring and control systems, V. 17, No. 5, 2019. pp. 84-92. https://doi.org/10.18127/j20700814-201905-11
29. *Kagirov I.A., Karpov A.A., Kipyatkova I.S., Klyuzhev K.S., Kudryavtsev A.I., Kudryavtsev I.A., Ryumin D.A.* Control of a robotic medical exoskeleton through an intelligent interface. Proceedings of the international scientific and technical conference "Extreme Robotics", 2019, pp. 462-468. https://doi.org/10.31776/ConfER.30.2019
30. *Markitantov M.* An analytical review of systems for automatic speaker's age recognition by voice. In Proc. VIII Congress of Young Scientists, St. Petersburg, ITMO, Vol. 3, 2019, pp. 246-251.
31. *Ryumin D., Aksenov A.* Method for the automatic detection and recognition of 3D one-handed gestures. In Proc. VIII Congress of Young Scientists, St. Petersburg, ITMO, Vol. 4, 2019, pp. 64-69.
32. *Karpov A.A., Sergeev S.F., Lakhin O.I., Mikchailyuk M.V, Kryuchkov B.I., Usov V.M.* Human-Robot Interaction in a Manned Space Flight: An Ontological Approach // Manned Spaceflight, No. 4 (33), 2019, pp. 70-91. https://doi.org/10.34131/MSF.19.4.70-91