Laboratory of Speech and Multimodal Interfaces

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

Laboratory staff: 12 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. Computational paralinguistics.

Research Fellows

PhD, senior researcher – Irina Kipyatkova – methods for language and acoustic modeling for automatic Russian speech recognition systems, kipyatkova@iias.spb.su

PhD, senior researcher – Anton Saveliev – cross-platform software and mobile services for teleconferencing, saveliev@iias.spb.su

Researcher – Vasilisa Verkhodanova – study of non-verbal phenomena in spontaneous speech, probabilistic models of speech disfluencies, verkhodanova@iias.spb.su

Junior researcher – Denis Ivanko – research and development of an automatic system of audiovisual Russian speech recognition with the use of a microphone and a high-speed video camera, ivanko@iias.spb.su

Junior researcher – Dmitry Ryumin – methods and software for recognition of elements of Russian sign language, dl_03.03.1991@mail.ru

Junior researcher – Alexander Denisov – methods and software for motion control of a robot with an anthropomorphic kinematic scheme, sdenisov93@mail.ru

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

Bachelor and Master Students

Bachelor students at the Department of Neuroinformatics and Robotics of SPb SUAI-SPIIRAS: 3 (thesis supervisor – Kipyatkova I.S.);

Master students at the Department of Speech Information Systems of the ITMO University: 2 (thesis supervisor – Karpov A.A.).

Grants and Projects

Karpov A.A. – Grant of the President of the Russian Federation № MD-254.2017.8 «Development and research of an automatic system for recognition of human's natural emotions from speech», 2017-2018.

Kipyatkova I.S. – Grant of the President of the Russian Federation № MK-1000.2017.8 «Development of a neural network based acoustic model for a Russian speech-to-text conversion system», 2017-2018.

Verkhodanova V.O. – Project of the Russian Foundation for Basic Research (RFBR) № 15-06-04465-a «Study of the acoustical cues differentiating phonational speech disfluencies in the spontaneous speech», 2015-2017.

Karpov A.A. – Project of RFBR № 15-07-04415-a «Models and methods of audio-visual signal processing for bimodal Russian speech recognition», 2015-2017.

Kipyatkova I.S. – Project of RFBR № 15-07-04322-a «Research of acoustic and language models based on artificial neural networks for an automatic large vocabulary Russian speech recognition system», 2015-2017.

Saveliev A.I. – Project of RFBR № 15-07-06774-a «Development of methods of multimedia data processing and exchange in peer-to-peer web application of multipoint videoconferencing», 2015-2017.

Karpov A.A. – Project of RFBR № 16-37-60100-mol_a_dk «Development of a universal assistive information technology based on multimodal human-computer interfaces», 2016-2019.

Kipyatkova I.S. – Grant-subsidy of the Committee on Science and Higher Education of the Government of St. Petersburg for young PhD researchers «Development and research of neural network hybrid acoustic models for the Russian speech recognition system», 2017.

Saveliev A.I. – Grant-subsidy of the Committee on Science and Higher Education of the Government of St. Petersburg for young PhD researchers «Development of a configuration method for the optimal arrangement of heterogeneous IoT-network modules», 2017.

Karpov A.A. – Project «Development of a technical project for the voice control module software for a robotic exoskeleton for medical purposes" under a contract with the Volga State Technological University (VSTU) within the framework of the integral project in the framework of the Government Statement No. 218 «On measures of state support for the development of cooperation between Russian higher education institutions and organizations, implementing complex projects for the creation of high-tech production», 2017.

Karpov A.A. – Contract with Huawei company, 2017-2018.

Karpov A.A. – Contract with ASM Solutions company, 2017.

University Courses

ITMO University: Speech recognition (Karpov A.A.)

SUAI University: Information systems; Automated information management systems (Kipyatkova I.S).

Conferences

19th International Conference "Speech and Computer" SPECOM-2017, 12-16 September 2017, Hatfield, Great Britain – Kipyatkova I.S., Verkhodanova V.O., Karpov A.A. (conference co-organization);

18th International Conference INTERSPEECH-2017, 20-24 August 2017, Stockholm, Sweden - Karpov A.A., Kipyatkova I.S.;

6th International Conference on Artificial Intelligence and Natural Language AINL-2017, September 20-23, 2017, St. Petersburg, Russia - Verkholyak O.V., Markovnikov N.M.;

2nd International Seminar "Photogrammetric and computer vision techniques for video surveillance, biometrics and biomedicine" PSBB-2017, May 15-17, 2017, Moscow, Russia - Ryumin D.A.;

IEEE International Conference "Processing of image and sound signals in the context of neurotechnologies" SPCN-2017, June 26-30, 2017, St. Petersburg, Russia - Verkholyak O.V., Karpov A.A.

2nd International Conference on Interactive Collective Robotics ICR-2017, September 13-14, 2017, Hatfield, Great Britain - Karpov A.A., Saveliev A.I., Ivanko D.V.;

International Scientific and Technical Conference "Zavalishin's Readings - 2017", April 18, 2017, St. Petersburg, Russia - Saveliev A.I., Denisov A.V.;

Scientific and Technical Seminar of Huawei 2nd Algorithm Workshop, June 19, 2017, St. Petersburg, Russia - Ivanko D.V.

Scientific and organizational activity

Organization and holding of the 19th International Conference "Speech and Computer" SPECOM-2017. http://specom.nw.ru/history/sites/2017 Hatfield (UK), September 12-16, 2017 – Karpov A.A. (chairman of the program committee). Proceedings published: Speech and Computer. Springer International Publishing Switzerland. A. Karpov et al. (Eds.): SPECOM 2017, LNAI 10458, 2017, 831 p. http://www.springer.com/gb/book/9783319664286

Organization and holding of the special session "Digital Revolution for Under-resourced Languages" at the 18th International Conference INTERSPEECH-2017, Stockholm, Sweden, August 20-24, 2017 – Karpov A.A. (co-organizer of the special session) http://ahclab.naist.jp/DigRevURL

Organization of a special issue of the International journal "Journal of Electrical and Computer Engineering" (indexed in Web of Science and Scopus, Hindawi, USA) – Karpov A.A. (invited editor) https://www.hindawi.com/journals/jece/si/324109

International Cooperation

Joint research and organization of scientific events in cooperation with the University of West Bohemia in Pilsen (Czech Republic), Bogazici University (Turkey), Namık Kemal University (Turkey), University of Patras (Greece), Leipzig University of Telecommunications (Germany), 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), the University of Hertfordshire (Great Britain), Huawei company (China).

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

Karpov A.A. – expert of the RAS; 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 "Speech Technology" (Moscow) and "Informatics" (Minsk); Guest editor of the Journal on Multimodal User Interfaces (Springer), Speech Communication (Elsevier) and 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; Speech Communication; Computer Speech & Language; 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, ISNN, HBU, etc.).

Kipyatkova I.S. – technical/programme committee member of the International Conferences INTERSPEECH, SPECOM, ISNN. Member of the organizing committee of the International Conference SPECOM series; Reviewer of Journal on Multimodal User Interfaces (Springer), Journal of Electrical Engineering and Computer Engineering (Hindawi).

Saveliev A.I. - Member of the organizing committees of international conferences SPECOM, ICR, Zavalishin's Readings.

Verkhodanova V.O. - Member of the program committee of the international conference SPECOM.

Intellectual Property registered

Programs and Data Bases

Patent RF for the invention № 2618389 issued on 03.05.2017: Karpov A.A., Ronzhin A.L. "A method for hands-free control of the mouse cursor".

Certificate on Database registration No. 2017621219 issued on 19.10.2017: Karpov A.A., Ivanko D.V., Ryumin D.A, Kipyatkova I.S. "Audiovisual corpus of the continuous Russian speech with high-speed video recordings (HAVRUS)".

Certificate on Software Registration No. 2017618845 issued on 10.08.2017: Karpov A.A., Ryumin D.A., Ivanko D.V., Kipyatkova I.S., Budkov V.Y. "Audio-visual Russian speech recognition system with a microphone and a high-speed video camera (AVSpeechRecognition)".

Certificate on Software Registration No. 2017661398 issued on 11.10.2017: Saveliev A.I., Karasev E.Y. "Component for managing user accounts in the Internet-communication system."

Certificate on Software Registration No. 2017661405 issued on 12.10.2017: Saveliev A.I., Karasev E.Y. "Server data management and storage system of the peer-to-peer videoconferencing application".

Certificate on Software Registration No. 2017661406 issued on 12.10.2017: Saveliev AI, Karasev E.Y. "Software platform for management and configuration of modular, communication web applications."

Recent Results

1. A multi-modal system for audio-visual continuous Russian speech recognition (AVSpeechRecognition) using high-speed video data has been developed. It simultaneously analyses audio from a microphone and video from a high-speed camera (using JAI Pulnix camera with 200 fps at the resolution of 640x480) and fuses information with Coupled Hidden Markov Models (CHMM). Experimental setup and collected multimodal database HAVRUS have allowed us to explore the impact brought by the high-speed video recordings with various frames per second (fps) starting from standard 25 fps up to high-speed 200 fps, as well as to increase the word recognition accuracy and to improve robustness of speech recognition in conditions of dynamic acoustic noises with SNR below 10dB [10, 19].

2. Acoustic models based on artificial neural networks with different architectures have been developed and studied. Time Delay Neural Networks (TDNN), Convolutional Neural Networks, Recurrent Convolutional Neural Networks, Long Short-Term Memory (LSTM) and Bi-directional LSTM models were applied. The developed models were embedded in our very large vocabulary automatic Russian speech recognition system (the vocabulary size is more than 150K word-forms) [8, 16].

3. Novel acoustic feature extraction methods for automatic recognition of speaker's emotional states from speech using deep recurrent neural networks with Long Short-Term Memory allowing for long-range time dependency modelling and capturing time structure of the signal were proposed in order to enhance feature representation of the time signal. Furthermore, combination techniques were developed to merge the new feature extraction methods with existing ones, showing effectiveness of joint representation in comparison with standalone approaches [17, 2].

4. A prototype of a computer system for automatic recognition of manual gestures using Microsoft Kinect 2.0 has been developed. The prototype is able to recognize continuous fingerspelling gestures, as well as sequences of digits in the Russian and Kazakh sign languages. In the Russian sign language, there are 33 letters, which are demonstrated in the form of static gestures. In Kazakh sign language, Russian fingerspelling alphabet is supplemented with additional 9 letters, which are shown dynamically (the vocabulary now contains only 52 manual gestures). A visual database of isolated sign gestures using Kinect 2.0 has been collected, it contains recordings of 2 persons (male and female) demonstrating gestures of the sign language, and every demonstrator has repeated each of 52 gestures at least 30 times [7].

5. Architectures, algorithms and software for automatic processing and transmission of multimodal information streams in video conferencing systems have been developed. A peer-to-peer approach is proposed for organizing the process of exchanging multimodal information between videoconferencing participants. The proposed approach has made it possible to reduce the amount of data transferred in the process of video conferencing, as well as to reduce the consumption of server resources and distributed heterogeneous devices, which enables efficient transmission and processing of speech and multimodal data [25].

6. The computer system for paralinguistic analysis of natural speech has been improved. The system uses a set of state-of-the-art methods for extracting multiple informative features from audio signals, filtering and multi-level data normalization, machine learning and classification based on artificial neural networks such as Extreme Learning Machines. This system took the first place (winner) at the 9th INTERSPEECH Computational Paralinguistics Challenge ComParE-2017 in "Snoring Sub-Challenge" (Stockholm, Sweden) [5].

Awards, certificates, scholarships

1. Certificate of winner of the grant of the President of the Russian Federation for state support of young Russian scientists – doctors of sciences in the field of knowledge "Technical and engineering sciences" (Certificate No. MK-254.2017.8 of the Council on grants of the President of the Russian Federation) - Karpov A.A.

2. Certificate of winner of the grant of the President of the Russian Federation for state support of young Russian scientists – candidates of sciences in the field of knowledge "Technical and engineering sciences" (Certificate No. MK-1000.2017.8 of the Council on grants of the President of the Russian Federation) - Kipyatkova I.S.
3. Diploma of the winner of the St. Petersburg grant competition in 2017 for young PhD from the Government of St. Petersburg - Kipyatkova I.S.
4. Diploma of the winner of the St. Petersburg grant competition in 2017 for young PhD from the Government of St. Petersburg - Saveliev A.I.
5. The winner (1st place) of the INTERSPEECH Computational Paralinguistics Challenge (ComParE-2017) in the "Snoring Sub-Challenge", Stockholm, Sweden, August 2017 - Karpov A.A.

References

Papers published in journals and editions, indexed by WoS, Scopus

1. Kipyatkova I., Karpov A. A Study of Neural Network Russian Language Models for Automatic Continuous Speech Recognition Systems // Automation and Remote Control, Springer, Vol. 78, No. 5, 2017, pp. 858-867. https://doi.org/10.1134/S0005117917050083 (WoS JCR=0,492, Scopus SJR=0,34, Q2)
2. Kaya H., Salah A., Karpov A., Frolova O., Grigorev A., Lyakso E. Emotion, Age, and Gender Classification in Children's Speech by Humans and Machines // Computer Speech and Language, Elsevier, 2017, Vol. 46, pp. 268-283. https://doi.org/10.1016/j.csl.2017.06.002 (WoS JCR=1,900, Scopus SJR=0,168, Q1)
3. Petrovsky A., Wan W., Rosa-Zurera M., Karpov A. Signal Processing Platforms and Algorithms for Real-life Communications and Listening to Digital Audio // Journal of Electrical and Computer Engineering, Hindawi, Volume 2017, 2017, Article ID 2913236, http://dx.doi.org/10.1155/2017/2913236 (WoS, Scopus SJR=0,168, Q3)
4. Basov O., Kipyatkova I., Saveliev A. Multimodal Subscriber Interfaces for Infocommunication Systems // Computing and Informatics, Slovak Academy of Sciences, Vol. 36, 2017, pp. 908-924. http://dx.doi.org/10.4149/cai_2017_4_908 (WoS, Scopus SJR=0,253, Q3)
5. Kaya H., Karpov A. Introducing Weighted Kernel Classifiers for Handling Imbalanced Paralinguistic Corpora: Snoring, Addressee and Cold. In Proc. INTERSPEECH-2017, Stockholm, Sweden, ISCA, 2017, pp. 3527-3531. http://isca-speech.org/archive/Interspeech_2017/pdfs/0653.PDF
6. Akhtiamov O., Sidorov M., Karpov A., Minker W. Speech and Text Analysis for Multimodal Addressee Detection in Human-Human-Computer Interaction. In Proc. INTERSPEECH-2017, Stockholm, Sweden, ISCA, 2017, pp. 2521-2525. http://isca-speech.org/archive/Interspeech_2017/pdfs/0501.PDF
7. Ryumin D., Karpov A. Towards Automatic Recognition of Sign Language Gestures using Kinect 2.0. In Proc. 19th International Conference on Human-Computer Interaction HCII-2017, Vancouver, Canada, Springer LNCS vol. 10278, 2017, pp. 89-104. https://link.springer.com/chapter/10.1007/978-3-319-58703-5_7
8. Kipyatkova I. Experimenting with Hybrid TDNN/HMM Acoustic Models for Russian Speech Recognition. In Proc. 19th International Conference on Speech and Computer SPECOM-2017, Hatfield, UK, Springer LNCS 10458, 2017, pp. 362-369. https://link.springer.com/chapter/10.1007/978-3-319-66429-3_35
9. Verkhodanova V., Shapranov V., Kipyatkova I. Hesitations in Spontaneous Speech: Acoustic Analysis and Detection. In Proc. 19th International Conference on Speech and Computer SPECOM-2017, Hatfield, UK, Springer LNCS vol. 10458, 2017, pp. 398-406. https://link.springer.com/chapter/10.1007/978-3-319-66429-3_39
10. Ivanko D., Karpov A., Kipyatkova I., Ryumin D., Saveliev A., Budkov V., Ivanko Dm., Železný M. Using a High-Speed Video Camera for Robust Audio-Visual Speech Recognition in Acoustically Noisy Conditions. In Proc. 19th International Conference on Speech and Computer SPECOM-2017, Hatfield, UK, Springer LNCS vol. 10458, 2017, pp. 757-766.: https://link.springer.com/chapter/10.1007/978-3-319-66429-3_76
11. Akhtiamov O., Pugachev A., Karpov A., Sidorov M., Minker W. Are You Addressing Me? Multimodal Addressee Detection in Human-Human-Computer Conversations. In Proc. 19th International Conference on Speech and Computer SPECOM-2017, Hatfield, UK, Springer LNCS vol. 10458, 2017, pp. 152-161. https://link.springer.com/chapter/10.1007/978-3-319-66429-3_14
12. Hlaváč M., Gruber I., Železný M., Karpov A. Semi-automatic Facial Key-point Dataset Creation. In Proc. 19th International Conference on Speech and Computer SPECOM-2017, Hatfield, UK, Springer LNCS vol. 10458, 2017, pp. 662-668. https://link.springer.com/chapter/10.1007/978-3-319-66429-3_66
13. Gruber I., Hlaváč M., Železný M., Karpov A. Facing Face Recognition with ResNet: Round One. In Proc. 2nd International Conference on Interactive Collaborative Robotics ICR-2017, Hatfield, UK, Springer LNCS vol. 10459, 2017, pp. 67-74. https://link.springer.com/chapter/10.1007/978-3-319-66471-2_8
14. Kryuchkov B., Syrkin L., Usov V., Ivanko D., Ivanko Dm. Using Augmentative and Alternative Communication for Human Robot Interaction during Maintaining Habitability of a Lunar Base. In Proc. 2nd International Conference on Interactive Collaborative Robotics ICR-2017, Hatfield, UK, Springer LNCS vol. 10459, 2017, pp. 95–104.

15. Pugachev A., Akhtiamov O., Karpov A., Minker W. Deep Learning for Acoustic Addressee Detection in Spoken Dialogue Systems. In Proc. 6th International Conference on Artificial Intelligence and Natural Language AINL-2017, St. Petersburg, Communications in Computer and Information Science, Springer, vol. 789, pp. 45-53. https://link.springer.com/chapter/10.1007/978-3-319-71746-3_4

16. Markovnikov N., Kipyatkova I., Karpov A., Filchenkov A. Deep neural networks in Russian language recognition. In Proc. 6th International Conference on Artificial Intelligence and Natural Language AINL-2017, St. Petersburg, Springer, Communications in Computer and Information Science, vol. 789, pp. 54-67. https://link.springer.com/chapter/10.1007/978-3-319-71746-3_5

17. Verkholyak O., Karpov A. Combined feature representation for emotion classification from Russian speech. In Proc. 6th International Conference on Artificial Intelligence and Natural Language AINL-2017, St. Petersburg, Communications in Computer and Information Science, Springer, vol. 789, pp. 68-73. https://link.springer.com/chapter/10.1007/978-3-319-71746-3_6

18. Vatamaniuk I., Budkov V., Kipyatkova I., Karpov A. Methods and Algorithms of Audio-Video Signal Processing for Analysis of Indoor Human Activity. In: Favorskaya M., Jain L. (eds.) Computer Vision in Control Systems-4. Intelligent Systems Reference Library, Springer, vol. 136. 2018, pp. 139-173.: https://doi.org/10.1007/978-3-319-67994-5_6

19. Ryumin D., Karpov A. Parametric representation of the speaker's lips for multimodal sign language and speech recognition. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. In Proc. ISPRS International Workshop "Photogrammetric and computer vision techniques for video Surveillance, Biometrics and Biomedicine" PSBB-2017, Moscow, 2017, pp. 155-161. https://doi.org/10.5194/isprs-archives-XLII-2-W4-155-2017

20. Kryuchkov B., Usov V., Tchertopolokhov V., Ronzhin A., Karpov A. Simulation of the "cosmonaut-robot" system interaction on the lunar surface based on methods of machine vision and computer graphics. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. In Proc. ISPRS International Workshop "Photogrammetric and computer vision techniques for video Surveillance, Biometrics and Biomedicine" PSBB-2017, Moscow, 2017, pp. 129-133. https://doi.org/10.5194/isprs-archives-XLII-2-W4-129-2017

Papers published in Russian journals and editions, indexed by RCSI

21. Kipyatkova I., Karpov A. Research of neural network models of the Russian language for automatic speech recognition systems // Avtomatika i Telemekhanika, vol. 78, No. 5, 2017, pp. 110-122.

22. Velichko A., Budkov V., Karpov A. Analytical Survey of Computational Paralinguistic Systems for Automatic Recognition of Deception in Human Speech // Informatsionno-Upravliaiushchie Sistemy, No. 5, 2017, pp. 30-41.

23. Kryuchkov B., Karpov A., Usov V., Chertopolokhov V. Multi-level monitoring of the gesture control of a mobile robot with out-of-ship activities on the lunar surface. Proceedings of the XIX International Conference "Problems of Control and Modeling in Complex Systems" PUMSS-2017, Samara, 2017, pp. 153-159.

24. Budkov V., Saveliev A., Basov O., Ronzhin A. Corpus of Russian speech for the study of the truth of the transmitted message // Proceedings of the 7th Interdisciplinary Workshop "Analysis of Conversational Russian Speech" AP3 - 2017, St. Petersburg, 2017, pp. 21-25.

25. Karasev E., Saveliev A., Malov D. Managing audio and video streams in peer-to-peer videoconferencing applications. Proceedings of the 10th Multi-Conference MCU-2017, vol. 3, 2017, pp. 94-96.

Other publications

26. Tampel I., Karpov A. Automatic speech recognition. Tutorial - Spb: University of ITMO, 2017, 152 p.

27. Verkholyak O., Karpov A. Combining utterance-level and frame-level feature representations for emotion classification from speech. In Proc. IEEE International Symposium "Video and Audio Signal Processing in the Context of Neurotechnologies", SPCN-2017, 2017, pp. 31.

28. Syrkin L., Zuykova A., Karpov A., Usov V. Application of an alternative method of communication for everyday interaction of a person with reduced physical capacity and a robot assistant. Proceedings of the Conference "Cognitive Research at the Present Stage" KISE-2017, Kazan, 2017.

29. Velichko A., Sokolov B., Karpov A., Budkov V. A brief review of the methods used in paralinguistic analysis of speech. Collection of reports of the 70th international student scientific conference GUAP. Part 2. Technical Sciences, St. Petersburg: GUAP, 2017, pp. 51-53.

30. Saveliev A. Development of a configuration method for the optimal arrangement of heterogeneous modules of the IoT network. Materials of the XXII St. Petersburg Assembly of Young Scientists and Specialists, 2017, p. 143.

31. Kipyatkova I. Development and research of neural network hybrid acoustic models for the Russian speech recognition system. Materials of the XXII St. Petersburg Assembly of Young Scientists and Specialists, 2017, p. 201.