

## **Speech and Multimodal Interfaces Laboratory**

**Head of the Laboratory** – Dr. Tech. Sc., Assoc. Prof. Andrey Ronzhin – modeling natural interaction processes between human beings and computers. Scientific secretary of Dissertation Council for Doctor of Sciences Degree of SPIIRAS [ronzhin@iias.spb.su](mailto:ronzhin@iias.spb.su), <http://www.spiiras.nw.ru/speech>

**Laboratory staff** – 6 full-time members and 3 post-graduate students

### **Research Fellows**

PhD, senior researcher Alexey Karpov – research and development of multimodal user interfaces and information systems with speech dialogue, [karpov@iias.spb.su](mailto:karpov@iias.spb.su);

MSc, researcher Irina Kipyatkova – methods for generation of alternative phonemic transcriptions for vocabularies for Russian speech recognition, [kipyatкова@iias.spb.su](mailto:kipyatкова@iias.spb.su);

MSc, researcher Ildar Kagirov – methods and algorithms for automatic text processing for the linguistic processor of the Russian speech recognition system, [kagirov@iias.spb.su](mailto:kagirov@iias.spb.su);

MSc, researcher Maria Prischepa – development of personified human's models and methods of natural interaction with mobile information robot, [prischepa@iias.spb.ru](mailto:prischepa@iias.spb.ru)

Programmer Konstantin Baranov – development of software for mobile information robot, [baranov@iias.spb.ru](mailto:baranov@iias.spb.ru)

### **Post-graduate Students**

MSc, Sergey Glazkov – Development of multimodal interfaces for heterogeneous mobile devices. Supervisor - A. Ronzhin, [casper86@yandex.ru](mailto:casper86@yandex.ru);

MSc, researcher Victor Budkov – Development methods and models of audio-visual signal processing in web-systems for supporting distributed meetings, [budkov@iias.spb.su](mailto:budkov@iias.spb.su);

MSc, researcher Alexander Ronzhin – Combined methods and models of analysis of situation and meeting's participant's behavior in the smart, [ronzhinal@iias.spb.su](mailto:ronzhinal@iias.spb.su)

### **Graduated Dissertation**

A. Ronzhin – Methods and software for multi-channel distant speech processing and their use in interactive multimodal applications (Dr. Tech. Sc., Speciality “Mathematics and Software of Computers, Complexes and Computer Networks”).

### **Grants and Projects**

A. Karpov – grant of the President of Russia № 02.120.11.64898-MK «Development of a computer multi-modal system for audio-visual synthesis of spoken Russian speech and sign language of deaf people», 2010-2011.

R. Yusupov – project № 14.740.11.0357 «Development of principles and innovative information technologies for interaction of users with smart space», 2010-2012.

A. Ronzhin – project № П876 «Development of models, methods and toolkits of intelligent control for the mobile information enquiry robot», 2010-2012.

A. Ronzhin - project № П2360 “Development of methods for human-machine interaction and multimodal user interfaces for intelligent information systems”, The Russian Federal Targeted Program “Research and Research-Human Resources for Innovating Russia in 2009-2013”, 2009-2011.

A. Karpov - project № П2579 “Development of methods, models and algorithms for automatic recognition of audio-visual Russian speech”, The Russian Federal Targeted Program “Research and Research-Human Resources for Innovating Russia in 2009-2013”, 2009-2011.

A. Karpov - RFBR project № 08-08-00128-a “Modeling non-phonemic speech elements and creation of alternative transcriptions for spontaneous Russian speech recognition”, 2008-2010.

A. Ronzhin - RFBR project № 08-07-90002-CT\_a “Methods and multimodal interfaces for contactless communication of handicapped people with information inquiry systems”, 2009-2010.

A. Ronzhin - RFBR project № 10-08-00199-a “Multimodal adaptive interfaces for heterogeneous client devices in intelligent environments”, 2010-2012.

A. Ronzhin - ONIT RAS project № 4.2 “Development of means for universal multimodal access for Interactive TV”, 2009-2011.

A. Karpov - Grant of the Human Capital Foundation: “Multimodal assistive system based on technologies of Russian speech recognition and technical vision”, 2009-2010.

I. Kipyatkova – Young Scientist Grant from the Committee on Science and Higher Education of the Administration of St. Petersburg “Development of a software module of language modeling for a dictation system of conversational Russian speech”, 2010.

A. Karpov – PhD Grant from the Committee on Science and Higher Education of the Administration of St. Petersburg “Development of an intelligent system for contactless work with a computer by means of speech input and head gestures”, 2010.

A. Ronzhin – PhD Grant from the Committee on Science and Higher Education of the Administration of St. Petersburg “Support technologies based on audiovisual processing methods for hybrid E-meetings in intelligent”, 2010.

Al. Ronzhin – Student Grant from the Committee on Science and Higher Education of the Administration of St. Petersburg “Development of system for supporting and photographing meeting participants based on distributed camera system”, 2010.

A. Karpov – Contract for research «Development of software for speech processing in a diving dialogue system» with «Oceanpribor» concern, 2010.

A. Karpov - RFBR grant № 10-07-09260-моб\_3 for participation in the 20th International Conference ICPR-2010 (Istanbul, Turkey), 2010.

A. Ronzhin - RFBR grant № 10-08-16018-моб\_3\_poc for participation in the 3rd International Conference ICCTS-2010 (Ulan-Ude, Russia), 2010.

Al. Ronzhin - RFBR grant № 09-08-09210-моб\_3 for participation in the 6th International Workshop on Multimodal Interfaces eINTERFACE'10 (Amsterdam, the Netherlands), 2010.

### **University Courses**

SPSUAI: Neural network and expert systems (A. Ronzhin).

SPSUAI: Methods of artificial intelligence (A. Ronzhin).

SPSUAI: Robots and mechatronic systems control (A. Ronzhin)

### **Organized Scientific Events**

The 4th Interdisciplinary Workshop “Analysis of Conversational Russian Speech”, St. Petersburg, 26-27 August 2010.

### **Participation in Conferences**

The 6th International Summer Workshop on Multimodal Interfaces eINTERFACE'10, Amsterdam, the Netherlands, 12 July-6 August 2010 - Al. Ronzhin, A. Karpov.

The 3rd Conference on Smart Spaces (ruSMART'10) in conjunction with 10th International Conference on Next Generation Wired/Wireless Advanced Networking NEW2AN 2010, 22–25 August 2010, St. Petersburg, Russia - A. Ronzhin, Al. Ronzhin.

The 11th International Conference Interspeech-2010, Makuhari, Chiba, Japan, 26-30 September 2010 - A. Karpov.

The 20th International Conference on Pattern Recognition ICPR-2010, Istanbul, Turkey, 23-26 August 2010 - A. Karpov.

The 3rd International Conference ICCTS-2010, Ulan-Ude, Russia, 6-11 September 2010 - A. Ronzhin.

The 5th International Conference GSSCP-2010, Ulan Bator, Mongolia, 13-17 September 2010 - A. Ronzhin.

The 20th International Conference on Computer Graphics and Vision GRAPHICON-2010, St. Petersburg, Russia, 20-24 September 2010 - Al. Ronzhin.

The 13th International Conference on Text, Speech and Dialogue TSD-2010, Brno, Czech Republic, 6-10 September 2010 - A. Karpov.

The 12th International Conference “Regional Informatics”, St. Petersburg, Russia, 20-22 October 2010 - I. Kipyatkova, M. Prischepa, V. Budkov.

The 2nd International Conference ICUMT-2010, Moscow, Russia, 18-20 October 2010 - V. Budkov.

The 8th International Conference Fruct-2010, Lappeenranta, Finland, 9-12 November 2010 - Al. Ronzhin, S. Glazkov.

The 10th International Conference PRIA-2010, St. Petersburg, Russia, 5-12 December 2010 - V. Budkov, A. Karpov.

The 4th Saint-Petersburg Congress «Professional education, science, innovations in XXI century», St. Petersburg, Russia, 18-19 November 2010 - A. Karpov.

The All-Russian conference “Zavalishin’s readings”, St. Petersburg, Russia, 14 March 2010 – V. Budkov, Al. Ronzhin.

The 5th All-Russian Conference “Next-generation Systems and Control Tasks”, Dombai, Russia, 5-9 April 2010 - A. Karpov.

The 13th International Student Olympiad on Automatic Control, St. Petersburg, Russia, 26-28 May 2010 - Al. Ronzhin.

The 7th International Conference Fruct-2010, St. Petersburg, Russia, 26-30 April 2010 - V. Budkov.

The 12th Session of the Russian Acoustic Society, Moscow, Russia, 15-17 June 2010 - A. Ronzhin.

### **International Cooperation**

A. Ronzhin - member of the committee on East Europe of the International Speech Communication Association (ISCA), member of the academy of navigation and motion control, Chairman of the International Conference “Speech and Computer” SPECOM, member of the European Association “OpenInterface”.

A. Karpov - 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 of Pattern Recognition (IAPR), member of the European Association “OpenInterface”.

International Cooperation with University of West Bohemia in Pilsen (Czech Republic), Bogazici University in Istanbul (Turkey), Dresden University of Technology (Germany), United Institute of Information Problems of the National Academy of Sciences of Belarus, Universidad Nacional Autonoma de Mexico (Mexico), Institute of Computer Science University of Bialystok (Poland), International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences and Ministry of Education and Science of Ukraine, Universidad Autónoma Metropolitana Unidad Iztapalapa de México (Mexico), University of Aizu (Japan).

### **Scientific awards**

Diploma for winner of grants of the President of Russia for state support of young Russian scientists in 2010 – A. Karpov.

Diploma for winner of a competition “St. Petersburg Grants 2010 for young scientists” from the Administration of St. Petersburg – I. Kipyatkova;

Diploma for winner of a competition “St. Petersburg Grants 2010 for young PhDs” from the Administration of St. Petersburg – A. Ronzhin;

Diploma for winner of a competition “St. Petersburg Grants 2010 for young PhDs” from the Administration of St. Petersburg – A. Karpov;

Diploma for winner of a competition “St. Petersburg Grants 2010 for students” from the Administration of St. Petersburg – Al. Ronzhin.

### **Research Activities**

Research and development of methods for natural human-computer interaction. Automatic spontaneous speech recognition and understanding. Multimodal user interfaces. Intelligent rooms and spaces. Biometric security systems. Assistive information systems for disabled people.

### **Recent Results**

1. Methodical and mathematical support for development of interactive multimodal applications and configuration of software/hardware resources based on multichannel audio and video signal processing modules, which carry out multistream tasks on distributed computer units

and allow developers to create cross platform services for organization natural interaction between users taking into account communication constraints of a user, client device, environments and applied domain, is proposed [2,6,9].

2. A multifunctional system for video surveillance based on a distributed camera system, which includes five IP-cameras and technologies of video-stream processing for detection and motion tracking of participants, registration of them during a meeting in the intelligent meeting room has been elaborated. The system notifies the intelligent meeting room about current locations of participants and automatically prepares a report with face photos of all participants [11,17,22].

3. A model of automatic extra-large vocabulary conversational Russian speech recognition system has been developed. Recognition vocabulary capacity consists of up to 2 million word-forms. Computer-aided tools for statistical modeling of Russian speech are applied in this recognition model. Preliminary experiments on recognition of sentences of continuous speech from the collected multi-speaker Russian speech corpus have been conducted. [4,16,18].

4. A model of bimodal automatic recognition of audio-visual Russian speech has been developed. It uses the mathematical apparatus of Coupled Hidden Markov Models of the first order, which allows to make fusion of vector streams from audio and video speech modalities from on the level of states of joint probability model. The recognition model takes into account time discrepancies (asynchrony) between streams of corresponding speech items – phonemes and visemes. Such asynchrony is peculiar to conversational natural speech and the model makes information fusion considering individual weights of modality informativity in dependence on acoustical conditions [1,3,8].

5. A dialog model for interaction between the mobile information robot and users, which includes the following modes: speech dialog, escorting to a point of interest, and uses a speech keyword spotting system for determination of goods name or a service, about which the robot should output information or show a route, has been developed [5,13,20].

6. The proposed logical-time model for compilation of multimedia content of web-based E-meeting system takes into account a current situation inside the meeting room, states of audio, video and presentation equipment as well as constraints of user mobile devices, and provides both distant control of equipment and support for distributed participants [9,14,28].

### **Intellectual Properties**

1. Certificate on Software Registration № 2010617640 of the Federal Service for Intellectual Property, Patents and Trademarks (Rospatent) on 18 November 2010: A. Ronzhin, A. Karpov, I. Kipyatkova. Multimodal user interface for an intelligent information system.
2. Certificate on Database Registration (proposal № 2010620586, November 2010) of the Federal Service for Intellectual Property, Patents and Trademarks (Rospatent): A. Karpov, A. Ronzhin, I. Kipyatkova. Multimedia corpus of audio-visual Russian speech - RusAVSpeechCorpus.
3. Certificate on Software Registration (proposal № 2010616189, November 2010) of the Federal Service for Intellectual Property, Patents and Trademarks (Rospatent): A. Karpov. Computer software for synchronization, capturing and combining of audio-visual speech modalities - RusAVSpeechRecorder.

### **References**

1. A.A. Karpov. Audio-visual speech interface for systems of control and notification // Proceedings of Southern Federal University. Technical sciences. – Taganrog, 2010. № 3 (104), pp. 218-222. (in Rus.).
2. R.M. Yusupov, A.L. Ronzhin. From Smart Devices to Smart Space // Herald of the Russian Academy of Sciences, MAIK Nauka, Volume 80, Number 1, pp. 45-51. 2010. (in Rus.).
3. A.A. Karpov, L.I. Tsirulnik, M. Zelezny. Development of a Computer System "Talking Head" for Text-to-Audiovisual-Speech Synthesis // Information technologies. – Moscow: New technologies, № 8, vol. 9, 2010, pp. 13-18. (in Rus.).

4. I.S. Kipyatkova, A.A. Karpov. Automatic processing and statistical analysis of news text corpus for language modeling of the Russian speech recognition system // Information-control systems. – SPb.: GUAP, № 4(47), 2010, pp. 2-8. (in Rus.).
5. M.V. Prischepa, V.Yu. Budkov, A.L. Ronzhin. Intelligent control system for mobile information-inquiry robot // Information-control systems, 2010, № 6, pp. 2-6. (in Rus.).
6. A.L. Ronzhin, A.A. Karpov. Development of interactive applications with multimodal interfaces // Proceedings of Tomsk State University of Control Systems and Radioelectronics, № 1 (21), part 1, 2010, pp. 124-127. (in Rus.).
7. R.M. Yusupov, A.L. Ronzhin. From Smart Devices to Smart Space // Herald of the Russian Academy of Sciences, MAIK Nauka, Volume 80, Number 1, pp. 63-68.
8. A.A. Karpov. Automatic recognition of audio-visual Russian speech with usage of an asynchronous model // Information-measuring and control systems. – Moscow: Radiotechnology, № 7, vol. 8, 2010, pp. 91-96. (in Rus.).
9. A.L. Ronzhin, V.Yu. Budkov, A. Karpov. Multichannel System of Audio-Visual Support of Remote Mobile Participant at E-Meeting // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2010, LNCS 6294, 2010, pp. 62–71.
10. A. Ronzhin, A. Karpov, I. Kipyatkova, M. Zelezny. Client and Speech Detection System for Intelligent Infokiosk. // Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231, 2010, pp. 560–567.
11. A.L. Ronzhin, M.V. Prischepa, A.A. Karpov. A Video Monitoring Model with a Distributed Camera System for the Smart Space // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2010, LNCS 6294, 2010, pp. 102–110.
12. A. Karpov, A. Ronzhin, K. Markov, M. Zelezny. Viseme-Dependent Weight Optimization for CHMM-Based Audio-Visual Speech Recognition. In Proc. INTERSPEECH-2010 International Conference, ISCA Association, Makuhari, Japan, 2010, pp. 2678-2681.
13. V.Yu. Budkov, M.V. Prischepa. Dialog model for mobile information robot control // Artificial intellect, Donetsk: GUIII, № 4, 2010, pp. 78-81. (in Rus.).
14. A.L. Ronzhin, V.Yu. Budkov, D. Ganbat, R. Naydandorj, B. Doljinsuren. Approaches to registration and analysis of situation in lecture auditory in distant learning systems // Proceeding of MUST. №4/213, MUST, 2010. (in Rus.).
15. P. Campr, E. Dikici, M. Hruz, A. Kindiroglu, Z. Krnoul, A.L. Ronzhin, H. Sak, D. Schorno, L. Akarun, O. Aran, A. Karpov, M. Saraclar, M. Zelezny. Automatic Fingersign to Speech Translator. In Proc. 6-th Summer Workshop on Multimodal Interfaces eINTERFACE'2010, Amsterdam, The Netherlands, 2010, pp. 69-82.
16. A. Karpov, A. Ronzhin, I. Kipyatkova, A.L. Ronzhin, L. Akarun. Multimodal Human Computer Interaction with MIDAS Intelligent Infokiosk. In Proc. 20-th International Conference on Pattern Recognition, IAPR Association, Turkey, Istanbul, 2010, pp. 3862-3865.
17. A.L. Ronzhin, M.V. Prischepa, Budkov V. Yu., A.A. Karpov, A.L. Ronzhin. Distributed System of Video Monitoring for the Smart Space. In. Proc. GraphiCon'2010. Saint-Petersburg, Russia, 2010 pp. 207-214. (in Rus.).
18. I. Kipyatkova. Investigation of statistical n-gram language models for continuous Russian speech recognition with an extra-large vocabulary. In. Proc. of 4-th Interdisciplinary Workshop "Conversational Russian Speech Analysis", (AP<sup>3</sup> - 2010) – SPb.: GUAP, 2010, pp. 33-40, (in Rus.).
19. S. Glazkov. Development of interactive applications with multimodal interface for heterogeneous mobile devices. In. Proc. of 4-th Interdisciplinary Workshop "Conversational Russian Speech Analysis", (AP<sup>3</sup> - 2010) – SPb.: GUAP, 2010, pp. 51-56, (in Rus.).
20. M. Prischepa, V. Budkov. Development of interactive model of speech communication of mobile information robot with mall visitors. In. Proc. of 4-th Interdisciplinary Workshop "Conversational Russian Speech Analysis", (AP<sup>3</sup> - 2010) – SPb.: GUAP, 2010, pp. 46-50, (in Rus.).

21. A.L. Ronzhin, V.Yu. Budkov. Speaker diarization system based on algorithm for voice activity detection in multichannel audio stream. In: Proc. XII session of the Russian Acoustic Society, Moscow, 2010, Vol. 3, pp. 77-81, (in Rus.).
22. A.L. Ronzhin, M.V. Prischepa, A.A. Karpov. Multimodal Control System for a Smart Room // In Proc. 13<sup>th</sup> International Student Olympiad on Automatic Control, Saint-Petersburg, Russia, 2010, pp. 115-119.
23. R.M. Yusupov, B.V. Sokolov, A.L. Ronzhin. Development of interactive applications for multimodal control the intelligent meeting room equipment. In Proc. Int. Conf. GSSCP-2010, Mongolia, 2010, pp. 200-203, (in Rus.).
24. A.L. Ronzhin, V.Yu. Budkov. Logic-temporal model of compilation multimedia web-interface for distributed meeting organization. In: Proc. III Int. Conf. Infocommunication and computation technologies and systems, Russia, 2010, pp. 240-244, (in Rus.).
25. V. Budkov, M. Prischepa, A. Ronzhin, A. Karpov. Multimodal Human-Robot Interaction. In Proc. 2nd International Conference ICUMT-2010, Moscow, Russia, IEEE Xplore 2010.
26. I.S. Kipyatkova. Continuous Russian speech recognition system with usage of an extra-large vocabulary. Proceedings of XII Saint-Petersburg International conference "Regional Informatics 2010" (RI-2010), Saint-Petersburg, 2010, pp. 360, (in Rus.).
27. V.Yu. Budkov, M.V. Prischepa. Dialogue interaction model of mobile information robot. Proceedings of XII Saint-Petersburg International conference "Regional Informatics 2010", Saint-Petersburg, 2010, pp. 36-37, (in Rus.).
28. V.Yu. Budkov, A.L. Ronzhin. System based on web application for support of distributed events. Proceedings of XII Saint-Petersburg International conference "Regional Informatics 2010", Saint-Petersburg, 2010, pp. 37, (in Rus.).
29. S. Glazkov. Development of interactive applications with multi-modal interfaces for mobile devices. In Proc. 8th Conference of Open Innovations Framework Program FRUCT – Finland, Lappeenranta, 2010, pp. 53-57.
30. A.L. Ronzhin, M. Prischepa, V. Budkov. Web-system for mobile participant support. In Proc. 8th Conference of Open Innovations Framework Program FRUCT – Finland, Lappeenranta, 2010, pp. 154-159.
31. M. Prischepa, A.L. Ronzhin, A. Karpov. Multifunction model for video monitoring of the smart room based on Internet camera system. In Proc. Zavalishin's readings – SPb.: GUAP, 2010, pp. 168-176, (in Rus.).
32. V. Budkov, A. Ronzhin. Algorithm of speech activity detection in the multichannel system of speaker journaling. In Proc. Zavalishin's readings – SPb.: GUAP, 2010, pp. 36-43, (in Rus.).
33. D. Ganbat, A. Ronzhin, R. Naydandorj, V.Yu. Budkov, M.V. Prischepa. Development of web-system for providing learning services for remote mobile users // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 21-34. (in Rus.).
34. R.M. Yusupov, A.L. Ronzhin. On activities of the scientific-educational center "Intelligent space technologies" // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 247-248. (in Rus.).
35. I.S. Kipyatkova, A.A. Karpov. An Analytical Survey of Large Vocabulary Russian Speech Recognition Systems // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 7-20. (in Rus.).
36. I.S. Kipyatkova, A.A. Karpov. Development and Research of a Statistical Russian Language Model // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 35-49. (in Rus.).
37. I.S. Kipyatkova, A.A. Karpov. Experiments on Continuous Russian Speech Recognition with Usage of an Extra-Large Vocabulary // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 63-74. (in Rus.).
38. M.V. Prischepa, V.Yu. Budkov, A.L. Ronzhin. Development of interactive television system with multimodal access // Proceedings of SPIIRAS. Issue 12, Saint-Petersburg, 2010, pp. 50-62. (in Rus.).

39. V. Budkov, M. Prischepa, A. Ronzhin. Development of the Mobile Information-Reference Robot's Dialog Model. In Proc. 10-th International Conference on Pattern Recognition and Image Analysis: New Information technologies, PRIA-2010, St. Petersburg, Russia, 2010, Vol. 2, pp. 253-257.
40. A. Kindiroglu, H. Yalcin, O. Aran, M. Hruz, P. Campr, L. Akarun, A. Karpov. A Multi-Lingual Fingerspelling Recognition for Handicapped Kiosk. In Proc. 10-th International Conference on Pattern Recognition and Image Analysis: New Information technologies PRIA-2010, St. Petersburg, Russia, 2010, Vol. 2, pp. 33-36.