

SLTU-2014

4th International Workshop
on Spoken Language Technologies
For Under-resourced Languages



Workshop Proceedings

14-16 May 2014

St. Petersburg, Russia

St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences

SLTU-2014

4th International Workshop
on Spoken Language Technologies
for Under-resourced Languages

SLTU-2014 Workshop Proceedings

14–16 May 2014

St. Petersburg, Russia

www.mica.edu.vn/sltu2014

St. Petersburg

2014

UDC 004.522

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences
(SPIIRAS)

SPIIRAS, 39, 14 line, St. Petersburg, 199178, Russia
<http://www.spiiras.nw.ru/speech>

Proceedings of the 4th International Workshop on Spoken Language Technologies for Under-
resourced Languages SLTU-2014. – St. Petersburg, Russia, 2014. – 268 p.

ISBN 978-5-8088-0908-6

This book of proceedings is a subject to copyright. All rights are reserved. Reproduction of this publication in any form by any electronic or mechanical means (including photocopying, recording or information storage and retrieval) is permitted only under the provisions of the Russian Federal Copyright Law and the prior permission in writing of SPIIRAS.

Edited by Speech and Multimodal Interfaces Laboratory of SPIIRAS
Printed by Publishing center GUAP

ISBN 978-5-8088-0908-6

© SPIIRAS, 2014
© Authors, 2014

SLTU-2014 Organizers



St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia



Laboratoire d'Informatique de Grenoble (LIG), France



Laboratoire Informatique d'Avignon (LIA), France



International Research Institute MICA, Vietnam



St. Petersburg National Research University of Information Technologies, Mechanics and Optics (University ITMO), Russia



“SPIIRAS-Service” Ltd., Russia

Workshop Chairs

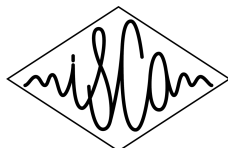
Alexey Karpov, SPIIRAS, St. Petersburg, Russia

Laurent Besacier, LIG, Grenoble, France

Pascal Nocera, LIA, Avignon, France

Eric Castelli, MICA, Hanoi, Vietnam

SLTU-2014 Sponsors



International Speech Communication Association (ISCA)



European Association for Signal Processing (EURASIP)



Russian Foundation for Basic Research (RFBR)



“Linfo” company (Moscow, Russia)

Official Service Agency



Monomax Congresses & Incentives

SLTU-2014 Workshop Committees

Scientific Committee

Etienne Barnard, North-West University, South Africa
Laurent Besacier, Laboratoire d'Informatique de Grenoble, France
Eric Castelli, International Research Institute MICA, Vietnam
Dirk Van Compernelle, KU Leuven, Belgium
Marelle Davel, North-West University, South Africa
Mark Gales, Cambridge University, UK
Oliver Jokisch, Leipzig University of Telecom., Germany
Alexey Karpov, SPIIRAS Institute, Russia
Kate Knill, Cambridge University, UK
Daniil Kocharov, St. Petersburg State University, Russia
Lori Lamel, LIMSI-CNSR, France
Haizhou Li, A-Star, Singapore
Yuri Matveev, University ITMO, Russia
Roger K. Moore, Sheffield University, UK
Pedro Moreno, Google, USA
Satoshi Nakamura, NAIST Institute, Japan
Pascal Nocera, Laboratoire Informatique d'Avignon, France
Francois Pellegrino, Université de Lyon, France
Andrey Ronzhin, SPIIRAS Institute, Russia
Yoshinori Sagisaka, Waseda University, Japan
Murat Saraclar, Bogazici University, Turkey
Ruhi Sarikaya, Microsoft, USA
Tanja Schultz, Karlsruhe Institute of Technology, Germany
Pavel Skrelin, St. Petersburg State University, Russia
Tan Tien Ping, Universiti Sains Malaysia, Malaysia
Isabel Trancoso, INESC ID, Portugal

Additional Reviewers

Willem Basson, North-West University, South Africa
William Hartmann, LIMSI-CNRS, France
Irina Kipyatkova, SPIIRAS Institute, Russia
Antoine Laurent, LIMSI-CNRS, France
Timur Pekhovsky, Speech Technology Center, Russia
Sakriani Sakti, NAIST Institute, Japan
Charl van Heerden, North-West University, South Africa
Werner Verhelst, Vrije Universiteit Brussel, Belgium

Organizing Committee

Alexey Karpov, SPIIRAS Institute, Russia
Irina Kipyatkova, SPIIRAS Institute, Russia
Daniil Kocharov, St. Petersburg State University, Russia
Yuri Matveev, University ITMO, Russia
Irina Podnozova, SPIIRAS Institute, Russia
Andrey Ronzhin, SPIIRAS Institute, Russia
Yaroslav Shalamaiko, SPIIRAS Service, Russia
Pavel Skrelin, St. Petersburg State University, Russia
Vladimir Vasiliev, University ITMO, Russia
Vasilisa Verkhodanova, SPIIRAS Institute, Russia
Rafael Yusupov, SPIIRAS Institute, Russia
Laurent Besacier, LIG, France
Eric Castelli, MICA, Vietnam
Pascal Nocera, LIA, France
Natalia Avdeenko, Monomax, Russia
Asya Gilmanova, Monomax, Russia
Anna Streltsova, Monomax, Russia

SLTU-2014 Chairmen's Message

Automatic speech processing for under-resourced languages is a very active field of research and development, which has experienced significant progress during the past decade. There are almost 7000 languages in the world and only for a small portion of them we have some spoken language technologies (speech recognition, text-to-speech synthesis, machine translation, etc.). Most languages from developing countries or minorities received only little attention so far and a lot of research in speech and language technologies are still needed for multilingual applications, especially for under-resourced languages.

We are very pleased to welcome you to the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU-2014) and to the Northern Capital of Russia – St. Petersburg. Three previous even-year SLTU Workshops were organized: SLTU'12 in Cape Town (South Africa), SLTU'10 in Penang (Malaysia), and SLTU'08 in Hanoi (Vietnam). During this Workshop we will discuss achievements and challenges in all areas related to processing under-resourced and endangered languages of the world: language resources development, acquisition, and representation including creation of dictionaries, language models, grammars, text and speech corpora; automatic speech recognition and synthesis for low-resourced languages and dialects; multilingual spoken language processing, machine translation and spoken dialogue systems.

In total, 40 high-quality papers have been selected from submissions for oral presentations at the Workshop and published in the SLTU-2014 Proceedings. Each paper was carefully reviewed by 2-4 Scientific Committee members, and the Committee did an excellent job in selecting the best submissions for SLTU-2014. The authors and co-authors of the papers originate from 4 continents (Europe, Asia, Africa and North America) and 22 countries (France, Germany, Russia, South Africa, Japan, UK, Romania, Vietnam, Czech Republic, Finland, Spain, Belgium, Estonia, Ireland, Hungary, Ukraine, USA, Portugal, Italy, Tunisia, Algeria and Qatar). We are delighted to find so much interest in this SLTU 2014 Workshop in St. Petersburg, Russia. We have observed an essential increase in the quantity and the quality of the papers compared to the previous SLTU Workshops and we have compiled an interesting 3-day scientific programme. Also 2 outstanding speakers have been invited to give keynote lectures: Satoshi Nakamura (Nara Institute of Science and Technology, Japan) and Mark Gales (University of Cambridge, UK). We would like to thank all the members of the Scientific Committee and authors for their great work in making this event a scientifically recognized International Workshop. Additionally to scientific activities we also provide a nice social programme, as well as excellent possibilities for acquaintance with cultural and historical valuables of St. Petersburg.

Special thanks also to all local organizers from SPIIRAS Institute, University ITMO, St. Petersburg State University, SPIIRAS-Service, and Monomax/Rustour service agency. Finally, we would like to extend our thanks to all our sponsors and supporters: ISCA association, EURASIP association, Russian Foundation for Basic Research (RFBR), Russian Academy of Sciences, "Linfo" company, etc.

Many thanks for coming to St. Petersburg and have an excellent SLTU-2014 Workshop!

Chairs of SLTU-2014 Workshop:

Alexey Karpov (SPIIRAS, Russia), *Laurent Besacier* (LIG, France),

Pascal Nocera (LIA, France) and *Eric Castelli* (MICA, Vietnam)

SLTU-2014
4th International Workshop
on Spoken Language Technologies
for Under-resourced Languages

SLTU-2014 Workshop Proceedings

Editor: Alexey Karpov

The workshop is supported by RFBR grant № 14-07-06012

Printed in Saint-Petersburg State University of
Aerospace Instrumentation (Russia)

Approved for publishing on 24.04.14.

Page format 60×84 1/8.

Number of copies 100. Order N 206.

SUAI University publisher house
190000, Saint-Petersburg, B. Morskaya, 67