

# **Lecture Notes in Artificial Intelligence**

**11658**

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

Founding Editor

Jörg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

More information about this series at <http://www.springer.com/series/1244>

Albert Ali Salah · Alexey Karpov ·  
Rodmonga Potapova (Eds.)


# Speech and Computer


21st International Conference, SPECOM 2019  
Istanbul, Turkey, August 20–25, 2019  
Proceedings

*Editors*

Albert Ali Salah  
Utrecht University  
Utrecht, The Netherlands

Boğaziçi University  
Istanbul, Turkey

Rodmonga Potapova   
Moscow State Linguistic University  
Moscow, Russia

Alexey Karpov   
St. Petersburg Institute for Informatics  
and Automation of the Russian Academy  
of Sciences  
St. Petersburg, Russia

ISSN 0302-9743                      ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-030-26060-6              ISBN 978-3-030-26061-3 (eBook)  
<https://doi.org/10.1007/978-3-030-26061-3>

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

## **SPECOM 2019 Preface**

The International Conference on Speech and Computer (SPECOM) was established by the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS) and the Herzen State Pedagogical University of Russia thanks to the efforts of Prof. Yuri Kosarev and Prof. Rajmund Piotrowski.

In its long history, the SPECOM conference was organized alternately by SPIIRAS and by the Moscow State Linguistic University (MSLU) in their home cities. SPECOM 2019 was the 21st event in the series, organized by Boğaziçi University (Istanbul, Turkey), in cooperation with SPIIRAS and MSLU. The conference was sponsored by ASM Solutions Ltd. (Moscow, Russia) and supported by the International Speech Communication Association. The conference was held jointly with the 4th International Conference on Interactive Collaborative Robotics (ICR) – where problems and modern solutions of human–robot interaction were discussed – during August 20–25, 2019 at Boğaziçi University, one of the top research universities in Turkey, established in 1863.

During the conferences three invited talks were given by Prof. Hynek Hermansky (Julian S. Smith Professor of Electrical Engineering and the Director of the Center for Language and Speech Processing at the Johns Hopkins University in Baltimore, Maryland, USA and Research Professor at the Brno University of Technology, Czech Republic), Prof. Odette Scharenborg (Delft University of Technology, The Netherlands), and Prof. Erol Şahin (Computer Engineering Dept., Middle East Technical University, Ankara, Turkey).

It is often argued that in processing of sensory signals such as speech, engineering should apply knowledge of properties of human perception—both have the same goal of getting information from the signal. Prof. Hermansky's talk, entitled “If You Can't Beat Them, Join Them,” showed examples from speech technology that perceptual research can also learn from advances in technology. Since speech evolved to be heard and properties of hearing are imprinted on speech, engineering optimizations of speech technology often yield human-like processing strategies. Prof. Hermansky presented a model of human speech communication which suggests that redundancies introduced in speech production in order to protect the message during its transmission through a realistic noisy acoustic environment are being used by human speech perception for a reliable decoding of the message. That led to a particular architecture of an automatic recognition (ASR) system in which longer temporal segments of spectrally smoothed temporal trajectories of spectral energies in individual frequency bands of speech are used to derive estimates of the posterior probabilities of speech sounds. Combinations of these estimates in reliable frequency bands were then adaptively fused to yield the final probability vectors, which best satisfy the adopted performance monitoring criteria.

Speech recognition is the mapping of a continuous, highly variable speech signal onto discrete, abstract representations. In both human and automatic speech processing, the phoneme is considered to play an important role. Abstractionist theories of human

speech processing assume the presence of abstract, phoneme-like units that sequenced together constitute words, while many large vocabulary automatic speech recognition (ASR) systems use phoneme acoustic models. Prof. Scharenborg, in her talk entitled “The Representation of Speech in Human and Artificial Brain,” argued that phonemes might not be the unit of speech representation during human speech processing and that comparisons between humans and dynamic neural networks and cross-fertilization of the two research fields can provide valuable insights into the way humans process speech and thereby improve ASR technology. The present volume includes an invited paper by Prof. Scharenborg that discusses these issues at length.

Prof. Şahin’s talk, entitled “Animating Industrial Robots for Human–Robot Interaction,” discussed interesting interaction research for robot-assisted assembly operations in the production lines of factories. Prof. Şahin argued that as platforms that require fast and fine manipulation of parts and tools remain beyond the capabilities of robotic systems in the near future, robotic systems are predicted not to replace, but to collaborate with the humans working on the assembly lines to increase their productivity. He briefly summarized the vision and goals of a recent TUBITAK project, titled CIRAK, which aims to develop a robotic manipulator system that will help humans in an assembly task by handing them the proper tools and parts at the right time in a correct manner. Toward this end, Prof. Şahin shared his group’s recent studies on the creation of a commercial robotic manipulator platform, made more life-like by extensions and modifications to its look and behavior.

This volume contains a collection of submitted papers presented at the conference, which were thoroughly reviewed by members of the Program Committee consisting of more than 100 top specialists, as well as an invited paper by Prof. Scharenborg. Each paper was reviewed, single blind, by two to four committee members (three reviewers on the average) and then discussed by the program chairs. In total, 57 papers were selected by the Program Committee for presentation at the SPECOM Conference. A total of 126 submissions were received and evaluated for SPECOM/ICR. The conference sessions were thematically organized, into Audio Signal Processing, Automatic Speech Recognition, Speaker Recognition, Computational Paralinguistics, Speech Synthesis, Sign Language and Multimodal Processing, and Speech and Language Resources. An increasing number of papers used deep neural network-based approaches across these themes.

We would like to express our gratitude to all authors for providing their papers on time, to the members of the Program Committee for their careful reviews and paper selection, and to the editors and correctors for their hard work in preparing this volume. Special thanks are due to Alen Demirel, Cem Tunçel, Bilge Yüksel, Hasan Küçük of BROS Group, our Conference Office, for their excellent work during the conference organization.

August 2019

Albert Ali Salah  
Alexey Karpov  
Rodmonga Potapova

# Organization

## General Co-chairs

Albert Ali Salah	Utrecht University and Boğaziçi University, The Netherlands/Turkey
Alexey Karpov	SPIIRAS Institute, Russia
Rodmonga Potapova	Moscow State Linguistic University, Russia

## Program Co-chairs

Heysem Kaya	Namık Kemal University, Turkey
Murat Saraçlar	Boğaziçi University, Turkey
Ebru Arisoy Saraçlar	MEF University, Turkey

## Program Committee

Shyam Agrawal, India	Ruediger Hoffmann, Germany
Lale Akarun, Turkey	Marek Hruz, Czech Republic
Rafet Akdeniz, Turkey	Kazuki Irie, Germany
Tanel Alumäe, Estonia	Rainer Jaeckel, Germany
Levent M. Arslan, Turkey	Oliver Jokisch, Germany
Ebru Arisoy Saraçlar, Turkey	Denis Jouvét, France
Elias Azarov, Belarus	Alexey Karpov, Russia
Peter Beim Graben, Germany	Heysem Kaya, Turkey
Marie-Luce Bourguet, UK	Andreas Kerren, Sweden
Christian-Alexander Bunge, Germany	Tomi Kinnunen, Finland
Eric Castelli, Vietnam	Irina Kipyatkova, Russia
Vladimir Chuchupal, Russia	Daniil Kocharov, Russia
Nicholas Cummins, Germany	Liliya Komalova, Russia
Vlado Delic, Serbia	Evgeny Kostyuchenko, Russia
Hasan Demir, Turkey	Ivan Kraljevski, Germany
Olivier Deroo, Belgium	Galina Lavrentyeva, Russia
Anna Esposito, Italy	Benjamin Lecouteux, France
Keelan Evanini, USA	Boris Lobanov, Belarus
Vera Evdokimova, Russia	Elena Lyakso, Russia
Nikos Fakotakis, Greece	Joseph Mariani, France
Mauro Falcone, Italy	Maria De Marsico, Italy
Vasiliki Foufi, Switzerland	Jindřich Matoušek, Czech Republic
Philip N. Garner, Switzerland	Yuri Matveev, Russia
Gábor Gosztolya, Hungary	Li Meng, UK
Tunga Gungor, Turkey	Peter Mihajlik, Hungary
Ivan Himawan, Australia	Iosif Mporas, UK

Bernd Möbius, Germany  
 Luděk Müller, Czech Republic  
 Satoshi Nakamura, Japan  
 Stavros Ntalampiras, Italy  
 Géza Németh, Hungary  
 Olga Perepelkina, Russia  
 Dimitar Popov, Bulgaria  
 Branislav Popović, Serbia  
 Rodmonga Potapova, Russia  
 Fabien Ringeval, France  
 Andrey Ronzhin, Russia  
 Paolo Rosso, Spain  
 Sakriani Sakti, Japan  
 Albert Ali Salah,

The Netherlands/Turkey

Murat Saraçlar, Turkey  
 Maximilian Schmitt, Germany  
 Friedhelm Schwenker, Germany  
 Vidhyasaharan Sethu, Australia  
 Milan Sečujski, Serbia

Ingo Siegert, Germany  
 Vered Silber-Varod, Israel  
 Pavel Skrelin, Russia  
 Mikhail Stolbov, Russia  
 Tilo Strutz, Germany  
 György Szaszák, Hungary  
 Ivan Tashev, USA  
 Natalia Tomashenko, France  
 Laszlo Toth, Hungary  
 Isabel Trancoso, Portugal  
 Jan Trmal, USA  
 Liliya Tsirulnik, USA  
 Dirk Van Compernelle, Belgium  
 Vasilisa Verkhodanova, The Netherlands  
 Benjamin Weiss, Germany  
 Andreas Wendemuth, Germany  
 Matthias Wolff, Germany  
 Milos Zelezny, Czech Republic  
 Zixing Zhang, UK

## Additional Reviewers

Cem Rıfkı Aydın  
 Somnath Banerjee  
 Branko Brkljač  
 Buse Buz  
 Koray Çiftçi  
 Gretel Liz De la Peña Sarracén  
 Ali Erkan  
 Nikša Jakovljević  
 Atul Kumar

Alexander Leipnitz  
 Michael Maruschke  
 Iris Ouyang  
 Sergey Rybin  
 Siniša Suzić  
 Juan Javier Sánchez Junquera  
 Oxana Verkholyak  
 Celaledin Yeroglu



# Contents

The Representation of Speech and Its Processing in the Human Brain and Deep Neural Networks . . . . .	1
<i>Odette Scharenborg</i>	
A Detailed Analysis and Improvement of Feature-Based Named Entity Recognition for Turkish . . . . .	9
<i>Arda Akdemir and Tunga Güngör</i>	
A Comparative Study of Classical and Deep Classifiers for Textual Addressee Detection in Human-Human-Machine Conversations . . . . .	20
<i>Oleg Akhtiamov, Dmitrii Fedotov, and Wolfgang Minker</i>	
Acoustic Event Mixing to Multichannel AMI Data for Distant Speech Recognition and Acoustic Event Classification Benchmarking . . . . .	31
<i>Sergei Astapov, Gleb Svirskiy, Aleksandr Lavrentyev, Tatyana Prisyach, Dmitriy Popov, Dmitriy Ubskiy, and Vladimir Kabarov</i>	
Speech-Based L2 Call System for English Foreign Speakers. . . . .	43
<i>Mohammad Ateeq and Abualsoud Hanani</i>	
A Pattern Mining Approach in Feature Extraction for Emotion Recognition from Speech. . . . .	54
<i>Umut Avci, Gamze Akkurt, and Devrim Unay</i>	
Towards a Dialect Classification in German Speech Samples . . . . .	64
<i>Johanna Dobbriner and Oliver Jokisch</i>	
Classification of Regional Accent Using Speech Rhythm Metrics . . . . .	75
<i>Ghania Droua-Hamdani</i>	
PocketEAR: An Assistive Sound Classification System for Hearing-Impaired . . . . .	82
<i>Kamil Ekšteín</i>	
Time-Continuous Emotion Recognition Using Spectrogram Based CNN-RNN Modelling . . . . .	93
<i>Dmitrii Fedotov, Bobae Kim, Alexey Karpov, and Wolfgang Minker</i>	
Developmental Disorders Manifestation in the Characteristics of the Child's Voice and Speech: Perceptual and Acoustic Study. . . . .	103
<i>Olga Frolova, Viktor Gorodnyi, Aleksandr Nikolaev, Aleksey Grigorev, Severin Grechanyi, and Elena Lyakso</i>	

RUSLAN: Russian Spoken Language Corpus for Speech Synthesis. . . . .	113
<i>Lenar Gabdrakhmanov, Rustem Garaev, and Evgenii Razinkov</i>	
Differentiating Laughter Types via HMM/DNN and Probabilistic Sampling. . . . .	122
<i>Gábor Gosztolya, András Beke, and Tilda Neuberger</i>	
Word Discovering in Low-Resources Languages Through Cross-Lingual Phonemes . . . . .	133
<i>Fernando García-Granada, Emilio Sanchis, Maria Jose Castro-Bleda, José Ángel González, and Lluís-F. Hurtado</i>	
Semantic Segmentation of Historical Documents via Fully-Convolutional Neural Network . . . . .	142
<i>Ivan Gruber, Miroslav Hlaváč, Marek Hruží, and Miloš Železný</i>	
A New Approach of Adaptive Filtering Updating for Acoustic Echo Cancellation. . . . .	150
<i>Mahfoud Hamidia and Abderrahmane Amrouche</i>	
Code-Switching Language Modeling with Bilingual Word Embeddings: A Case Study for Egyptian Arabic-English. . . . .	160
<i>Injy Hamed, Moritz Zhu, Mohamed Elmahdy, Slim Abdennadher, and Ngoc Thang Vu</i>	
Identity Extraction from Clusters of Multi-modal Observations . . . . .	171
<i>Marek Hruží, Petr Salajka, Ivan Gruber, and Miroslav Hlaváč</i>	
Don't Talk to Noisy Drones – Acoustic Interaction with Unmanned Aerial Vehicles . . . . .	180
<i>Oliver Jokisch, Ingo Siegert, Michael Maruschke, Tilo Strutz, and Andrey Ronzhin</i>	
Method for Multimodal Recognition of One-Handed Sign Language Gestures Through 3D Convolution and LSTM Neural Networks . . . . .	191
<i>Ildar Kagirov, Dmitry Ryumin, and Alexandr Axyonov</i>	
LSTM-Based Kazakh Speech Synthesis . . . . .	201
<i>Arman Kaliyev</i>	
Combination of Positions and Angles for Hand Pose Estimation . . . . .	209
<i>Jakub Kanis, Zdeněk Krňoul, and Marek Hruží</i>	
LSTM-Based Language Models for Very Large Vocabulary Continuous Russian Speech Recognition System . . . . .	219
<i>Irina Kipyatkova</i>	

Svarabhakti Vowel Occurrence and Duration in Rhotic Clusters in French Lyric Singing . . . . .	227
<i>Uliana Kochetkova</i>	
The Evaluation Process Automation of Phrase and Word Intelligibility Using Speech Recognition Systems . . . . .	237
<i>Evgeny Kostuchenko, Dariya Novokhrestova, Marina Tirskaia, Alexander Shelupanov, Mikhail Nemirovich-Danchenko, Evgeny Choyzonov, and Lidiya Balatskaya</i>	
Detection of Overlapping Speech for the Purposes of Speaker Diarization . . .	247
<i>Marie Kunešová, Marek Hruží, Zbyněk Zajíc, and Vlasta Radová</i>	
Exploring Hybrid CTC/Attention End-to-End Speech Recognition with Gaussian Processes. . . . .	258
<i>Ludwig Kürzinger, Tobias Watzel, Lujun Li, Robert Baumgartner, and Gerhard Rigoll</i>	
Estimating Aggressiveness of Russian Texts by Means of Machine Learning . . . . .	270
<i>Dmitriy Levonevskiy, Dmitrii Malov, and Irina Vatamaniuk</i>	
Software Subsystem Analysis of Prosodic Signs of Emotional Intonation . . .	280
<i>Boris Lobanov and Vladimir Zhitko</i>	
Assessing Alzheimer's Disease from Speech Using the i-vector Approach . . .	289
<i>José Vicente Egas López, László Tóth, Ildikó Hoffmann, János Kálmán, Magdolna Pákáski, and Gábor Gosztolya</i>	
AD-Child.Ru: Speech Corpus for Russian Children with Atypical Development . . . . .	299
<i>Elena Lyakso, Olga Frolova, Arman Kaliyev, Viktor Gorodnyi, Aleksey Grigorev, and Yuri Matveev</i>	
Building a Pronunciation Dictionary for the Kabyle Language . . . . .	309
<i>Demri Lyes, Falek Leila, and Teffahi Hocine</i>	
Speech-Based Automatic Assessment of Question Making Skill in L2 Language . . . . .	317
<i>Eman Mansour, Rand Sandouka, Dima Jaber, and Abualsoud Hanani</i>	
Automatic Recognition of Speaker Age and Gender Based on Deep Neural Networks. . . . .	327
<i>Maxim Markitantov and Oxana Verkholiyak</i>	
Investigating Joint CTC-Attention Models for End-to-End Russian Speech Recognition. . . . .	337
<i>Nikita Markovnikov and Irina Kipyatkova</i>	

Author Clustering with and Without Topical Features . . . . .	348
<i>Polina Panicheva, Olga Litvinova, and Tatiana Litvinova</i>	
Assessment of Syllable Intelligibility Based on Convolutional Neural Networks for Speech Rehabilitation After Speech Organs Surgical Interventions . . . . .	359
<i>Evgeny Kostuchenko, Dariya Novokhrestova, Svetlana Pekarskikh, Alexander Shelupanov, Mikhail Nemirovich-Danchenko, Evgeny Choyazonov, and Lidiya Balatskaya</i>	
Corpus Study of Early Bulgarian Onomatopoeias in the Terms of CHILDES . . . . .	370
<i>Velka Popova and Dimitar Popov</i>	
EEG Investigation of Brain Bioelectrical Activity (Regarding Perception of Multimodal Polycode Internet Discourse) . . . . .	381
<i>Rodmonga Potapova, Vsevolod Potapov, Nataliya Lebedeva, Ekaterina Karimova, and Nikolay Bobrov</i>	
Some Peculiarities of Internet Multimodal Polycode Corpora Annotation . . . .	392
<i>Rodmonga Potapova, Vsevolod Potapov, Liliya Komalova, and Andrey Dzhunkovskiy</i>	
New Perspectives on Canadian English Digital Identity Based on Word Stress Patterns in Lexicon and Spoken Corpus . . . . .	401
<i>Daria Pozdeeva, Tatiana Shevchenko, and Alexey Abyzov</i>	
Automatic Speech Recognition for Kreol Morisien: A Case Study for the Health Domain . . . . .	414
<i>Nuzhah Gooda Sahib-Kaudeer, Baby Gobin-Rahimbux, Bibi Saamiyah Bahsu, and Maryam Farheen Aasiyah Maghoo</i>	
Script Selection Using Convolutional Auto-encoder for TTS Speech Corpus . . . . .	423
<i>Meysam Shamsi, Damien Lolive, Nelly Barbot, and Jonathan Chevelu</i>	
Pragmatic Markers Distribution in Russian Everyday Speech: Frequency Lists and Other Statistics for Discourse Modeling . . . . .	433
<i>Natalia Bogdanova-Beglarian, Tatiana Sherstinova, Olga Blinova, and Gregory Martynenko</i>	
Curriculum Learning in Sentiment Analysis . . . . .	444
<i>Jakub Sido and Miloslav Konopik</i>	
First Minute Timing in American Telephone Talks: A Cognitive Approach . . . . .	451
<i>Tatiana Shevchenko and Tatiana Sokoreva</i>	

Syntactic Segmentation of Spontaneous Speech: Psychological and Cognitive Aspects . . . . .	459
<i>Anton Stepikhov, Anastassia Loukina, and Natella Stepikhova</i>	
Dual-Microphone Speech Enhancement System Attenuating both Coherent and Diffuse Background Noise . . . . .	471
<i>Mikhail Stolbov and Quan Trong The</i>	
Reducing the Inter-speaker Variance of CNN Acoustic Models Using Unsupervised Adversarial Multi-task Training . . . . .	481
<i>László Tóth and Gábor Gosztolya</i>	
Estimates of Transmission Characteristics Related to Perception of Bone-Conducted Speech Using Real Utterances and Transcutaneous Vibration on Larynx . . . . .	491
<i>Teruki Toya, Peter Birkholz, and Masashi Unoki</i>	
Singing Voice Database . . . . .	501
<i>Liliya Tsurulnik and Shlomo Dubnov</i>	
How Dysarthric Prosody Impacts Naïve Listeners' Recognition . . . . .	510
<i>Vass Verkhodanova, Sanne Timmermans, Matt Coler, Roel Jonkers, Bauke de Jong, and Wander Lowie</i>	
Light CNN Architecture Enhancement for Different Types Spoofing Attack Detection . . . . .	520
<i>Marina Volkova, Tseren Andzhukaev, Galina Lavrentyeva, Sergey Novoselov, and Alexander Kozlov</i>	
Deep Neural Network Quantizers Outperforming Continuous Speech Recognition Systems . . . . .	530
<i>Tobias Watzel, Lujun Li, Ludwig Kürzinger, and Gerhard Rigoll</i>	
Speaking Style Based Apparent Personality Recognition . . . . .	540
<i>Jianguo Yu, Konstantin Markov, and Alexey Karpov</i>	
Diarization of the Language Consulting Center Telephone Calls . . . . .	549
<i>Zbyněk Zajíc, Josef V. Psutka, Lucie Zajícová, Luděk Müller, and Petr Salajka</i>	
NN-Based Czech Sign Language Synthesis . . . . .	559
<i>Jan Zelinka, Jakub Kanis, and Petr Salajka</i>	
Re-evaluation of Words Used in Speech Audiometry . . . . .	569
<i>Aleksandar Živanović, Vlado Delić, Siniša Sužić, Ivana Sokolovac, and Maja Marković</i>	
<b>Author Index . . . . .</b>	<b>579</b>