Program of the 26th SPECOM conference, November 25-27, 2024 - Hotel Crowne Plaza, Belgrade, Serbia

CET, UTC+1	Day 0, 25th November 2024, Monday (hybrid mode)	
09:00 - 17:00	First Multi-Regional IEEE Standardisation Training Workshop – Why Standards Matter: How IEEE Standards Drive Industry Advancement and Foster Entrepreneurial Success	
CET, UTC+1	**Day 1, 26th November 2024, Tuesday (hybrid mode)**	
09:00 - 10:30	**Registration**	
10:30 - 12:30	**Opening Ceremony** (Hall Pacific&Atlantic) **TELFOR and SPECOM Conferences:** Introduction, Welcome Messages, Recognitions and Invited Presentation **Preserving Language Heritage Through Speech Technology: The Case of Upper Sorbian (Part 1)** Ivan Kraljevski, et al. Fraunhofer Institute IKTS, Dresden, Germany Alexey Karpov: Introduction of SPECOM conference and Program announcement	
12:30 - 14:00	**Welcome cocktail**	
14:00 - 14:30	**SPECOM Keynote Lecture 1** (Hall Baltic&Aegean) **Preserving Language Heritage Through Speech Technology: The Case of Upper Sorbian (Part 2)** Ivan Kraljevski, et al. Fraunhofer Institute IKTS, Dresden, Germany Chairperson: Vlado Delić	
14:30 - 16:30	**Oral Session 1** (Hall Baltic&Aegean) **Speech and Language Resources** Chairs: Branislav Gerazov and Siniša Suzić	**Poster session 1** (Before Mediterranean&Adriatic) **Affective Computing** Chairs: Yuri Matveev and Nikola Simić
16:30 - 17:00	**Coffee break**	
17:00 - 19:00	**Oral Session 2** (Hall Baltic&Aegean) **Computational Paralinguistics** Chairs: Wolfgang Minker and Branislav Popović	
20:00 - 23:00	**Banquet** (Restaurant "Velika Skadarlija")	
CET, UTC+1	**Day 2, 27th November 2024, Wednesday (hybrid mode)**	
09:00 - 10:30	**Oral Session 3** (Hall Baltic&Aegean) **Speaker Recognition** Chairs: Jacek Kudera and Nikša Jakovljević	
10:30 - 11:00	**Coffee break**	
11:00 - 11:40	**SPECOM Keynote Lecture 2** (Hall Baltic&Aegean) **Retrospective and Perspectives of TTS & STT Technology Development and Implementation for South Slavic Under-Resourced Languages.** Milan Sečujski, et al. FTS UNS, Novi Sad, Serbia Chairperson: Alexey Karpov	
11:40 - 13:00	**Oral Session 4** (Hall Baltic&Aegean) **Speech Synthesis and Perception** Chairs: Elena Lyakso and Milan Sečujski	**Poster session 2** (Before Mediterranean&Adriatic) **Speech Processing for Medicine** Chairs: László Tóth and Vuk Stanojev
13:00 - 14:00	**Lunch break** (Restaurant in Hotel Crowne Plaza)	
14:30 - 16:10	**Oral Session 5** (Hall Baltic&Aegean) **Automatic Speech Recognition** Chairs: Géza Németh and Irina Kipyatkova	**Poster session 3** (Before Mediterranean&Adriatic) **Natural Language Processing** Chairs: Gábor Gosztolya and Tijana Nosek
16:10 - 16:30	**SPECOM Closing Ceremony** (Hall Baltic&Aegean) Chairpersons: Alexey Karpov and Vlado Delić	
16:30 - 17:00	**Coffee break**	
	Social events (plan)	
17:00 - 21:00	Sightseeing tour of Belgrade by bus, including visits to the Nikola Tesla Museum and the Saint Sava Temple.	

DETAILED PROGRAM OF THE SPECOM 2024 CONFERENCE

26th International Conference on Speech and Computer – SPECOM 2024
November 25-27, 2024, Hotel Crowne Plaza, Belgrade, Serbia
https://specom2024.ftn.uns.ac.rs/

25th November 2024, Monday, 09:00 - 17:00 (hybrid mode, CET, UTC+1)

First Multi-Regional IEEE Standardisation Training Workshop

Why Standards Matter: How IEEE Standards Drive Industry Advancement and Foster Entrepreneurial Success

Hybrid event, No Admission Charge https://cscn2024.ieee-cscn.org/workshop/standards
(If the registration concluded, you can ask for a few reserved places at specom2024@uns.ac.rs)

OPENING CEREMONY of TELFOR and SPECOM Conferences

Tuesday, November 26, 2024, 10:30-12:30 (hybrid mode, CET, UTC+1) Hall Pacific&Atlantic

Chairpersons: Representatives of Conferences: TELFOR & SPECOM

Introduction

Welcome Messages

Recognitions

Invited Presentation

1. "6G Opportunities on the Paradigm of AI and Sensing", Dr Konstantinos Manolakis, Nokia

2. "Preserving Language Heritage Through Speech Technology: The Case of Upper Sorbian – Part I", Ivan Kraljevski et al, Fraunhofer

Introduction of SPECOM conference and Program announcement – Alexey Karpov

Conclusions

SPECOM Keynote Lecture 1: Tuesday, November 26, 2024,
Part 1: 12:00-12:20 (CET, UTC+1) Hall Pacific&Atlantic
Part 2: 14:00-14:30 (CET, UTC+1) Hall Baltic&Aegean

Chairperson: Vlado Delić

Preserving Language Heritage Through Speech Technology: The Case of Upper Sorbian (Part 2). Ivan Kraljevski, Frank Duckhorn, Daniel Sobe, Constanze Tschoepe and Matthias Wolf; Fraunhofer IKTS - Dresden, Foundation for the Sorbian People - Bautzen, Brandenburg University of Technology - Cottbus, Germany

SPECOM Keynote Lecture 2: Wednesday, November 27, 2024, 11:00-11:40 (CET, UTC+1) Hall Baltic&Aegean

Chairperson: Alexey Karpov

Retrospective and Perspectives of TTS & STT Technology Development and Implementation for South Slavic Under-Resourced Languages. Milan Sečujski, Branislav Popović, Darko Pekar, Nikša Jakovljević, Edvin Pakoci, Siniša Suzić, Tijana Nosek, Nikola Simić, Vuk Stanojev and Vlado Delić: FTS University of Novi Sad, Serbia

1.1. **The ParlaSpeech Collection of Automatically Generated Speech and Text Corpora from Parliamentary Proceedings.** Nikola Ljubešić, Peter Rupnik and Danijel Koržinek; JSI and UoLj - Ljubljana, Slovenia and PJAIT - Warsaw, Poland

1.2. **Bulgarian Speech Resources in the CHILDES System.** Velka Popova and Dimitar Popov; KPUS - Shumen, Bugarska

1.3. **Harnessing Knowledge Distillation for Enhanced Text-to-Text Translation in Low-Resource Languages.** Manar Ouled Ahmed, Zuheng Ming and Alice Othmani; Declic AI Research, USPN and UPE - Paris, France

1.4. **Multiword Units in Russian Everyday Speech: Empirical Classification and Corpus-Based Studies.** Natalia Bogdanova-Beglarian, Olga Blinova, Maria Khokhlova, Tatiana Sherstinova and Tatiana Popova; SPSU and HSEU - Saint Petersburg, Russia

1.5. **ESC Corpus of Spoken Russian: Everyday Student Conversations Captured through Continuous Speech Recording in Natural Communicative Environments.** Tatiana Sherstinova and Irina Petrova; HSEU - Saint-Petersburg, Russia (remotely)

1.6. **OpenAV: Bilingual Dataset for Audio-Visual Voice Control of a Computer for Hand Disabled People.** Denis Ivanko, Dmitry Ryumin, Alexandr Axyonov, Alexey Kashevnik and Alexey Karpov; SPC RAS - Saint-Petersburg, Russia (remotely)

2.1. **A Cross-Multi-Modal Fusion Approach for Enhanced Engagement Recognition.** Denis Dresvyanskiy, Alexey Karpov and Wolfgang Minker; UU - Ulm, Germany, ITMO and SPC RAS - Saint-Petersburg, Russia

2.2. **Automatic Assessment of Signs of Alcohol Dependency Syndrome from Spontaneous Speech.** Gábor Gosztolya, Lázár András Bence, Ildiko Hoffmann, Otília Bagi, Fruzsina Fanni Farkas, Janka Gajdics, László Tóth and János Kálmán; HUN-REN–SZTE and UoS - Szeged, and HUN-REN - Budapest, Hungary

2.3. **RAG and Few-Shot Prompting in Emotional Text Generation.** Elizaveta Vologina, Anastasia Matveeva, Olesia Makhnytkina, Yuri Matveev and Nursaule Burambayeva; ITMO - Saint-Petersburg, Russia and LNGENU - Astana, Kazakhstan

2.4. **Automatic Detection of Irony Based on Acoustic Features and Facial Expressions.** Uliana Kochetkova, Pavel Skrelin, Vera Evdokimova, Nikolai Borisov, Pavel Scherbakov, Petr Fedkin and Rada German; SPSU - Saint-Petersburg, Russia (remutely)

2.5. **Sentiment Analysis for Egyptian Arabic-English Code-Switched Data using Traditional Neural Models and Advanced Language Models.** Ahmed Sherif and Caroline Sabty; GIU - Cairo, Egypt (remutely)

2.6. **An Enhanced Compact Convolution Transformer for Age, Gender and Emotion Detection in Egyptian Arabic Speech.** Aya Abdalla, Nada Sharaf and Caroline Sabty; GIU - Cairo, Egypt (remutely)

3.1. **Voice Cloning and Mismatch Conditions in Forensic Automatic Speaker Recognition.** Jacek Kudera, Miriam Coccia, Sharifeh Fadaeijouybari, Till Preidt, Akshay Ranjan and Angelika Braun; TU - Trier, Germany

3.2. **On the Influence of CNN-based Feature Learning Modules in Neural Speaker Verification Framework.** Jahangir Alam and Md Shahidul Alam; CRIM - Montreal, Quebec, Canada (remotely)

3.3. **Transformation of Emotional Speech to Anger Speech to Reduce Mismatches in Testing and Enrollment Speech for Speaker Recognition System.** Shalini Tomar and Shashidhar G. Koolagudi; NIT - Surathkal, Karnataka, India (remotely)

3.4. **Investigating Data Requirements for Hindi Speaker Recognition: A Comparative Study with English.** Parth Khadse, Sabyasachi Chandra, Puja Bharati, Debolina Pramanik, G Satya Prasad, Aniket Aitawade and Shyamal Kumar Das Mandal; IIT - Kharagpur, West Bengal, India (remotely)

3.5. **Practical Evaluation and Validation of Methods for Automatic Speaker Identification (as Applied to Various Languages).** Rodmonga Potapova, Vsevolod Potapov and Irina Kuryanova; MSLU, LMSU and IAML - Moscow, Russia (remotely)

4.1. **End-to-End Speech Synthesis for the Serbian Language Based on Tacotron.** Tijana Nosek, Siniša Suzić, Milan Sečujski, Vuk Stanojev, Darko Pekar and Vlado Delić; FTS UNS - Novi Sad, Serbia

4.2. **ChildTinyTalks (CTT): A Benchmark Dataset and Baseline for Expressive Child Speech Synthesis.** Shaimaa Alwaisi, Mohammed Salah Al-Radhi and Géza Németh; BUTE - Budapest, Hungary

4.3. **Assessment of Children's Ability to Manifest Emotions in Facial Expressions, Voice and Speech by Humans, Automatic, and on a Likert Scale.** Elena Lyakso, Olga Frolova, Anton Matveev, Aleksandr Nikolaev and Ruban Nersisson; SPSU - Saint-Petersburg, Russia and VIT - Vellore, India

4.4. **In Pursuit for the Best Error Metric for Optimisation of Articulatory Vowel Synthesis.** Branislav Gerazov, Paul Konstantin Krug, Daniel Van Niekerk, Anqi Xu, Peter Birkholz and Yi Xu; FEEIT - Skopje, North Macedonia, TUD - Dresden, Germany and UCL - London, UK

5.1. **Comparison of Well- and Lower-Resourced Self-Training in ASR.** Yue Luo and Péter Mihajlik; BUTE and HUN-REN - Budapest, Hungary

5.2. **Towards a Livvi-Karelian End-to-End ASR System.** Irina Kipyatkova, Ildar Kagirov, Mikhail Dolgushin and Alexandra Rodionova; SPC RAS - Saint-Petersburg, Russia

5.3. **Benchmarking Whisper under Diverse Audio Transformations and Real-time Constraints.** Sergei Katkov, Antonio Liotta and Alessandro Vietti; FUBB - Bolzano, Italy

5.4. **Advances in OpenASR21 Evaluation with Increased Temporal Resolution for Speech Self-Supervised Learning Models.** <u>Vishwa Gupta</u>; CRIM - Montreal, Quebec, Canada (remotely)

5.5. **Pre-Training and Adverse Audio Samples for Data-Efficient Wake Word Detection.** Manuel Torralbo Lezana, Ariane Méndez Amuchategi, Maia Aguirre Pascual de Zulueta and <u>Arantza Del Pozo Echezarreta</u>; BRTA - Donostia, San Sebastián, Spain (remotely)

Poster session 1: Affective Computing   (9 posters (5 remotely) with 5 min oral introduction)

Tuesday, November 26, 2024, 14:30-16:30 (hybrid mode, CET, UTC+1) Hall in front of Mediterranean&Adriatic

Chairpersons: Yuri Matveev and Nikola Simić

P1.1. **Emotion Recognition by Vocalizations of Nonhuman Primates: Human and Automatic Classification.** <u>Olga Frolova</u>, Anton Matveev, Elena Lyakso, Tamara Kuznetsova and Inna Golubeva; SPSU and PIP RAS - Saint-Petersburg, Russia

P1.2. **Multimodal Emotion Recognition using Compressed Graph Neural Networks.** <u>Tijana Đurkić</u>, Nikola Simić, Siniša Suzić, Dragana Bajović, Zoran Perić and Vlado Delić; FTS UNS - Novi Sad, Serbia

P1.3. **Utilizing Speaker Models and Topic Markers for Emotion Recognition in Dialogues.** <u>Olesia Makhnytkina</u>, Yuri Matveev, Alexander Zubakov and Anton Matveev; ITMO - Saint-Petersburg, Russia

P1.4. **How Children Recognize Emotions from Video and Audio.** <u>Elena Lyakso</u>, Olga Frolova, Aleksandr Nikolaev, Severin Grechanyi, Yulia Filatova and Ruban Nersisson; SPSU - Saint-Petersburg, Russia and VIT - Vellore, India

P1.5. **Neurophysiological Correlates of Textual Modulation in Visual Stimuli: An Experimental Study of Russian and English Memes.** Rodmonga Potapova, Vsevolod Potapov, Ekaterina Karimova, Leonid Motovskikh and <u>Nikolay Bobrov</u>; MSLU, LMSU and IHNAN - Moscow, Russia (remotely)

P1.6. **Influence of Linguistic and Sociolinguistic Factors on Speech Rate Perception.** <u>Anastasia Ananeva</u> and Uliana Kochetkova; SPSU - Saint-Petersburg, Russia (remotely)

P1.7. **MMHS: Multimodal Model for Hate Speech Intensity Prediction.** Aman Goel and <u>Abhishek Poswal</u>; CMU - Pittsburgh, PA, USA and HKUST - Hong Kong (remotely)

P1.8. **DuFCALF: Instilling Sentience in Computerized Song Analysis.** <u>Himadri Mukherjee</u>, Matteo Marciano, Ankita Dhar and Kaushik Roy; WBSU - Berunanpukhuria, West Bengal, and SNU India, and NYU - Abu Dhabi, UAE (remotely)

P1.9. **AutoMode-ASR: Learning to Select ASR Systems for Better Quality and Cost.** Ahmet Gündüz, <u>Yunsu Kim</u>, Kamer Ali Yüksel, Mohamed Al-Badrashiny, Thiago Castro Ferreira and Hassan Sawaf; aiXplain Inc. - San Jose, CA, USA (remotely)

Poster session 2: Speech Processing for Medicine   (9 posters (4 remotely) with 5 min oral introduction)

Wednesday, November 27, 2024, 11:40-13:00 (hybrid mode, CET, UTC+1) Hall in front of Mediterranean&Adriatic

Chairpersons: László Tóth and Vuk Stanojev

P2.1. **Investigating the Utility of wav2vec 2.0 Hidden Layers for Detecting Multiple Sclerosis.** <u>Gábor Gosztolya</u>, László Tóth, Veronika Svindt, Judit Bóna and Ildiko Hoffmann; HUN-REN–SZTE and UoS - Szeged, and HUN-REN and ELTE - Budapest, Hungary

P2.2. **Cross-Cultural Automatic Depression Detection based on Audio Signals.** <u>Danila Mamontov</u>, Sebastian Zepf, Alexey Karpov and Wolfgang Minker; UU - Ulm, Germany, ITMO and SPC RAS - Saint-Petersburg, Russia

P2.3. **Detecting Depression from Audio Data.** Mary Idamkina and Andrea Corradini; UoL - Liverpool, UK and MCI - Innsbruck, Austria

P2.4. **Exploring MetaConformer for Speech Enhancement.** Lukas Foerner and Maximilian Dauner; UHA - Augsburg and MUAS - Munich, Germany

P2.5. **Approach to Assessing the Quality of Syllable Pronunciation by Patients in the Process of Speech Rehabilitation Based on Comparison with Healthy Speakers.** German Egle, Dariya Novokhrestova, Svetlana Tomilina and Evgeny Kostyuchenko; TUSUR - Tomsk, Russia

P2.6. **Binary and Multiclass Classification of Dysphonia Using Whisper Encoder and One-Dimensional Convolutional Neural Network.** Dosti Aziz and David Sztahó; BUTE - Budapest, Hungary and UoS - Sulaymaniyah, Iraq (remutely)

P2.7. **A Comparative Study for Contextualized Spoken Answer Classification in German Medical Questionnaires.** Philipp L. Harnisch, Daniel Schuhmann and Stefan Hillmann; TUB - Berlin, Germany (remotely)

P2.8. **Depression Classification using Token Merging-based Speech Spectrotemporal Transformer.** Lokesh Kumar, Kumar Kaustubh and S R Mahadeva Prasanna; IIT - Dharwad, Karnataka, India (remotely)

P2.9. **Integration of Short-Term and Long-Term Harmonic Peaks in a Two-Level Discriminative Weight Training Framework for Voice Activity Detection.** Yingwei Tan; VMIT - Beijing, China (remotely)

Poster session 3: Natural Language Processing   (9 posters (5 remotely) with 5 min oral introduction)

Wednesday, November 27, 2024, 14:30-16:10 (hybrid mode, CET, UTC+1) Hall in front of Mediterranean&Adriatic

Chairpersons: Gábor Gosztolya and Tijana Nosek

P3.1. **Bias Unveiled: Enhancing Fairness in German Word Embeddings with Large Language Models.** Yasser Saeid and Thomas Kopinski; SWUAS - Meschede, Germany (remotely)

P3.2. **How to detect imbalances in the Google Books Ngram corpus?** Valery Solovyev and Anna Ivleva; KFU - Kazan, Russia

P3.3. **Ancient Egyptian Hieroglyphic Texts Structure Identification.** Radek Marik, Renata Landgráfová and Jiří Liška; CTU and CU - Prague, Czech Republic

P3.4. **Multidimensional Rhythm: Comparing Rhythmic Properties of Australian and New Zealand Monologues.** Anna Borzykh and Tatiana Shevchenko; MSLU - Moscow, Russia

P3.5. **Separating Party Conversation by Applying Contrastive Learning Methodology.** Anandakumar Singaravelan and Jia-Lien Hsu; FJCU - New Taipei City, Taiwan

P3.6. **Predicting the Valence Rating of Russian Words Using Various Pre-Trained Word Embeddings.** Vladimir Bochkarev, Andrey Savinkov and Anna Shevlyakova; KFU - Kazan, Russia (remutely)

P3.7. **Cross-Lingual Summarization of Speech-to-Speech Translation: A Baseline 115.** Pranav Karande, Balaram Sarkar and Chandresh Kumar Maurya; IITI - Indor, Madhya Pradesh, India (remotely)

P3.8. **Human and Machine Keyphrase Perception in Russian Text and Speech.** Daria Guseva, Olga Mitrofanova and Mikhail Dolgushin; SPSU and SPC RAS - Saint-Petersburg, Russia (remotely)

P3.9. **Conformer LLM - Convolution Augmented Large Language Models.** Prateek Verma; US - Stanford, California, USA (remotely)