



SPECOM 2025

27th International Conference on Speech and Computer
October 13-14th, 2025, Szeged, Hungary

Detailed Program

Program Overview – SPECOM 2025 Conference, Szeged, Oct. 13-14, 2025

CEST (UTC+2)	Day 1, 13 th October 2025, Monday	
08:00 – 09:00	Registration	
09:00 – 09:20	Opening Ceremony	
09:20 – 10:00	Keynote From Conversation to Conversational: Speech Synthesis and the Communicative Power of the Human Voice Éva Székely, KTH Royal Institute of Technology, Stockholm, Sweden	
10:00 – 11:00	Oral Session 1 Computational Paralinguistics Chair: Denis Ivanko	
11:00 – 11:20	Coffee break	
11:20 – 13:00	Oral Session 2 Automatic Speech Recognition Chair: Thomas Hain	
13:00 – 14:00	Lunch	
14:00 – 15:00	Oral Session 3 Speech Processing for Healthcare Chair: Branislav Gerazov	Poster Session 1 Pragmatics & Phonetics Chair: Katalin Mády
15:00 – 15:10	Short break	
15:10 – 16:10	Oral Session 4 Natural Language Processing Chair: Veronika Vincze	
17:00 – 18:00	Social program (Information Technology History Exhibition)	
19:30 – 22:00	Banquet	

Program Overview – SPECOM 2025 Conference, Szeged, Oct. 13-14, 2025

CEST (UTC+2)	Day 2, 14 th October 2025, Tuesday	
08:00 – 09:00	Registration	
09:00 – 10:00	Oral Session 5 Digital & Multimodal Speech Processing Chair: Irina Kipyatkova	Poster Session 2 Speech Synthesis & Healthcare Chair: Milan Sečujski
10:00 – 10:10	Short break	
10:10 – 10:50	Keynote Towards Responsible Multimodal Modeling for Mental Healthcare Heysem Kaya, Utrecht University, The Netherlands	
10:50 – 11:10	Coffee break	
11:10 – 12:50	Oral Session 6 Under-Resourced Languages Chair: Péter Mihajlik	
12:50 – 14:00	Lunch	
14:00 – 15:00	Oral Session 7 Speech Perception Chair: Vlado Delić	Poster Session 3 NLP & Multimodal Chair: Veronika Vincze
15:00 – 15:10	Short break	
15:10 – 16:10	Oral Session 8 Speaker Recognition Chair: Gábor Gosztolya	
16:10 – 16:30	Closing Ceremony	
17:30 – 18:00	Visit at the Szeged Town Hall	

Detailed Program – SPECOM 2025 Conference, Szeged, Oct. 13-14, 2025

Day 1, 13th October 2025, Monday

08:00 – 09:00 Registration

09:00 – 09:20 Opening Ceremony

09:20 – 10:00 Keynote

Chair: Gábor Gosztolya

Éva Székely: From Conversation to Conversational: Speech Synthesis and the Communicative Power of the Human Voice

10:00 – 11:00 Oral session: Computational Paralinguistics (3 presentations (1 online))

Chair: Denis Ivanko

10:00 – 10:20 Egor Kleshnev and Elena Lyakso: Emotions Manifestation by Adolescents With Intellectual Disabilities

10:20 – 10:40 Dániel Halmai and Gábor Gosztolya: Spoken Emotion Recognition Using Soft Labels

10:40 – 11:00 Kunjan Gajre, Rajnidhi Gupta, Ravindrakumar Purohit and Hemant Patil: NAMTalk: From Muscle Vibrations to Emotional Speech (*online*)

11:00 – 11:20 Coffee break

11:20 – 13:00 Oral session: Automatic Speech Recognition (5 presentations (1 online))

Chair: Thomas Hain

11:20 – 11:40 Jarod Duret, Salima Mdhaffar, Gaelle Laperriere, Ryan Whetten, Audrey Galametz, Catherine Kobus, Marion-Cécile Martin, Jo Oleiwan and Yannick Estève: In-Domain SSL Pre-Training and Streaming ASR: Application to Air Traffic Control Communications

11:40 – 12:00 Gergely Dobsinszki, Péter Mihajlik, Máté Kádár, Tibor Fegyó and Katalin Mády: Best Data is More Supervised Data – Even for Hungarian ASR

12:00 – 12:20 Yunus Emre Özköse and Ali Haznedaroglu: Enhancing Speech Recognition through Text-to-Speech and Voice Conversion Augmentation

12:20 – 12:40 Anton Polevoi, Alexander Kragin and Natalia Loukachevitch: Ground Truth-Free WER Prediction for ASR via Audio Quality and Model Confidence Features

12:40 – 13:00 Sara Pearsell, Oliver Niebuhr and Samuel Schmück: Evaluating the Performance of Several ASR Systems in Environmental and Industrial Noise (*online*)

13:00 – 14:00 Lunch

14:00 – 15:00 Oral session: Speech Processing for Healthcare (3 presentations (1 online))

Chair: Branislav Gerazov

14:00 – 14:20 Wing-Zin Leung, Heidi Christensen and Stefan Goetze: Text-to-Dysarthric-Speech Generation for Dysarthric Automatic Speech Recognition: Is Purely Synthetic Data Enough?

14:20 – 14:40 Mikhail Dolgushin, Daria Guseva and Alexey Karpov: Investigation of Explainable Multimodal Methods for Detecting Mental Disorders

14:40 – 15:00 Anna Shevlyakova, Vladimir Bochkarev and Stanislav Khristoforov: Colour Preferences in Schizophrenic Speech (*online*)

- 14:00 – 15:00** Poster session: **Pragmatics & Phonetics** (6 presentations (4 online))
Chair: Katalin Mády
- 14:00 – 15:00 Anastasia Ananeva, Anton Tomilov and Marina Volkova: Revealing the Hidden Temporal Structure of HubertSoft Embeddings based on the Russian Phonetic Corpus
- 14:00 – 15:00 Marie Fongaro, Barbara Gili Fivela, Maud Pélissier and Gabriel Hévr: Subtle Changes in L1 Stops of Late Salento Italian-French Bilinguals: An Acoustic Study Using AutoVOT Adapted for Italian and French
- 14:00 – 15:00 Alejandro López-García, María Alfaro-Contreras, Julien Meyer and Jose J. Valero-Mas: Whistler Identification in Whistled Spanish (Silbo): A Case Study (*online*)
- 14:00 – 15:00 Olga Mitrofanova, Polina Iurevtseva and Maxim Bakaev: What Do LLMs Know about Human Emotions? The Russian Case Study (*online*)
- 14:00 – 15:00 Abdelkader Seif El Islem Rahmani, Yasser Yahiaoui and Abdelghani Bouziane: Retention-Augmented Voice Assistant: A Lightweight Architecture for Stateful Interaction with Comprehensive Evaluation and Privacy-Preserving Design (*online*)
- 14:00 – 15:00 Rodmonga Potapova, Vsevolod Potapov, Ekaterina Karimova, Diana Smolskaya, Nikolay Bobrov, Leonid Motovskikh and Iurii Pozhilov: Cognitive Humor Processing in the Russian and English Internet Meme Chatting: EEG Study (*online*)
- 15:00 – 15:10** **Short break**
- 15:10 – 16:10** Oral session: **Nature Language Processing** (3 presentations (2 online))
Chair: Veronika Vincze
- 15:10 – 15:30 Natalia Bogdanova-Beglarian, Olga Blinova, Maria Khokhlova, Tatiana Sherstinova and Tatiana Popova: High-Frequency Multiword Units and the Typological Distribution of Multiword Units in Spoken Russian
- 15:30 – 15:50 Danil Tirskikh, Olesia Koroteeva, Yuri Matveev, Ekaterina Brovkina and Larisa Gonchar: Enhancing Retrieval Performance via LLM Hard-Negative Filtering (*online*)
- 15:50 – 16:10 Philine Kowol and Stefan Hillmann: Analyzing Web-Scraped and Generated Inputs for Automatic and Scalable Intent Classification (*online*)

Day 2, 14th October 2025, Tuesday

08:00 – 09:00 Registration

09:00 – 10:00 Oral session: Digital & Multimodal Speech Processing (3 presentations (1 online))

Chair: Irina Kipyatkova

09:00 – 09:20 *Alexander Zaburdaev, Denis Ivanko and Dmitry Ryumin*: CrossMP-SENet: Transformer-based Cross-Attention for Joint Magnitude-Phase Speech Enhancement

09:20 – 09:40 *Jason Clarke, Yoshihiko Gotoh and Stefan Goetze*: Ensembling Synchronisation-based and Face-Voice Association Paradigms for Robust Active Speaker Detection in Egocentric Recordings (online)

09:40 – 10:00 *Ali Alhejab, Tomáš Železný, Lamy Alkanhal, Ivan Gruber, Yazeed Al-harbi, Jakub Straka, Václav Javorek, Marek Hruží, Badriah Alkalifah and Ahmed Ali*: Saudi Sign Language Translation Using T5

09:00 – 10:00 Poster session: Speech Synthesis & Healthcare (6 presentations (2 online))

Chair: Milan Sečujski

09:00 – 10:00 *Jia-Lien Hsu and Pei-Wen Chien*: Adaptive Singing Voice Enhancement for Live Stages

09:00 – 10:00 *Evgeny Kostyuchenko*: Automated Assessment of Phrase Intelligibility for Russian Speech Based on Esophageal Voice

09:00 – 10:00 *Elena Lyakso, Olga Frolova, Anton Matveev, Petr Shabanov, Andrei Lebedev, Aleksandr Nikolaev, Egor Kleshnev, Severin Grechanyi and Ruban Nersisson*: Attention Deficit Hyperactivity Disorder: Identifying Approaches for Early Diagnosis, a Pilot Study

09:00 – 10:00 *Zhiyuan Xu and Joshua Reiss*: PinkVocalTransformer: Neural Acoustic-to-Articulatory Inversion based on the Pink Trombone

09:00 – 10:00 *Sasangi Nayanathara, Randil Pushpananda, Inuri Harischandra and Thamira Weerakoon*: Enhancing Sinhala Text-to-Speech with End-to-End VITS Architecture (online)

09:00 – 10:00 *Anastasiia Sherban and Uliana Kochetkova*: Effectiveness of Tacotron2 for Intonation Model Synthesis in Russian (online)

10:00 – 10:10 Short break

10:10 – 10:50 Keynote

Chair: Alexey Karpov

Heysem Kaya: Towards Responsible Multimodal Modeling for Mental Healthcare

10:50 – 11:10 Coffee break

11:10 – 12:50 Oral session: Under-Resourced Languages (5 presentations (3 online))

Chair: Péter Mihajlik

11:10 – 11:30 *Vuk Stanojev, Tijana Nosek, Siniša Suzić, Darko Pekar, Vlado Delić and Milan Sečujski*: Improving Whisper-based Serbian ASR using Synthetic Speech

11:30 – 11:50 *Branislav Gerazov, Marcello Politi and Sébastien Bratières*: Arabic ASR on the SADA Large-Scale Arabic Speech Corpus with Transformer-based Models

11:50 – 12:10 *Vishwa Gupta and Gilles Boulianne*: Effect of Increased Temporal Resolution on Speech Recognition for French Quebec using Features from Speech Self-Supervised Learning Models (online)

12:10 – 12:30 *Anton Legchenko and Ivan Bondarenko*: Domain Knowledge and Language Embeddings for Low-Resource Multilingual Phoneme ASR (online)

12:30 – 12:50 *Rodmonga Potapova, Vsevolod Potapov, Tsend-Ayush Ganbaatar, Leonid Motovskikh and Nikolay Bobrov*: Sound and Colour in Phonosemantics: Perceptual and Acoustic Correlates of Mongolian Vowels (online)

12:50 – 14:00	Lunch
14:00 – 15:00	Oral session: Speech Perception (3 presentations) Chair: Vlado Delić
14:00 – 14:20	<i>Shree Harsha Bokkahalli Satish, Gustav Eje Henter and Éva Székely</i> : When Voice Matters: Evidence of Gender Disparity in Positional Bias of SpeechLLMs
14:20 – 14:40	<i>Mohammed Al-Radhi, Sadi Shurid and Géza Németh</i> : Prompting the Mind: EEG-to-Text Translation with Multimodal LLMs and Semantic Control
14:40 – 15:00	<i>George Close, Kris Hong, Thomas Hain and Stefan Goetze</i> : WhiSQA: Non-Intrusive Speech Quality Prediction using Whisper Encoder Features
14:00 – 15:00	Poster session: NLP & Multimodal Speech (6 presentations (3 online)) Chair: Veronika Vincze
14:00 – 15:00	<i>Anna Borzykh and Tatiana Shevchenko</i> : Rhythmic Diglossia Based on Discourse Types and Dialects of English: Australian and New Zealand Corpora
14:00 – 15:00	<i>Irina Kipyatkova, Kseniia Kiseleva, Mikhail Dolgushin and Ildar Kagirov</i> : Modeling Intra-Word Code-Switching for Karelian ASR
14:00 – 15:00	<i>José Luis Vázquez, Carlos Ulises Valdez, Marvin Aguero, Julio Mello, José Colbes and Sebastián Alberto Grillo</i> : Sector-Wise Backpropagation for Low-Resource Text Classification in Deep Models
14:00 – 15:00	<i>Vladimir Bochkarev, Andrey Achkeev and Anna Shevlyakova</i> : Estimation of the Genre Composition of the English Subcorpus of the Google Books Ngram (<i>online</i>)
14:00 – 15:00	<i>Aleksandra Maslenikova and Tatiana Popova</i> : Automatic Annotation of Discourse and Speech Formulas in Internet Communication: A Telegram Comment Corpus (<i>online</i>)
14:00 – 15:00	<i>Vera Evdokimova and Maria Maksimova</i> : Phonetic and Visual Characteristics of Cognitive Load (<i>online</i>)
15:00 – 15:10	Short break
15:10 – 16:30	Oral session: Speaker Recognition (4 presentations) Chair: Gábor Gosztolya
15:10 – 15:30	<i>Mohammed Hamzah Alsalihi and Dávid Sztahó</i> : Effect of Spoof Speech on Forensic Voice Comparison using Deep Speaker Embeddings
15:30 – 15:50	<i>Marina Volkova, Artem Chirkovskiy, Egor Ausev and Ekaterina Shangina</i> : Source Vendor Tracing of Audio Deepfakes
15:50 – 16:10	<i>Şule Bekiryazıcı, Cemal Hanılçı and Neyir Ozcan</i> : Enhancing Audio Replay Attack Detection with Silence-based Blind Channel Impulse Response Estimation
16:10 – 16:30	<i>Anton Yakovenko, Evgeny Bessonitsyn, Valeria Efimova and Mark Zaslavskiy</i> : Language-Specific Adaptation Strategies for Speaker Recognition using MobileNet
16:30 – 16:50	Closing Ceremony