

SALMA 2026

Speech and Audio Language Models Workshop (2nd Edition)

Proceedings of the Workshop

October 28, 2026

©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-496-5

Preface

Welcome to the second edition of the Speech and Audio Language Models Workshop (SALMA 2026), held in Budapest, Hungary, on October 28, 2026, co-located with EMNLP 2026.

Speech and audio are among the most natural, information-dense, and temporally grounded forms of human communication, yet they remain under-explored within mainstream large language model research. SALMA focuses on approaches that treat speech and audio as first-class linguistic modalities, covering the architectural design, training paradigms, data creation, alignment, and evaluation of large audio-language models across speech, non-speech sounds, music, and mixed audio. Following the inaugural edition at ICASSP 2025, this second edition brings the NLP and speech/audio communities together around shared architectures, objectives, benchmarks, and evaluation practices.

This year, we received 41 direct submissions and 7 papers committed through ACL Rolling Review (ARR). We accepted 19 papers (6 oral and 13 poster presentations), an acceptance rate of 40% over 47 papers (excluding one withdrawn submission). Of these, 13 papers appear in these proceedings, and 6 non-archival papers are presented at the workshop only. The accepted papers cover a broad range of topics, including the safety of audio-language models, multilingual and low-resource speech recognition and understanding, code-switched speech, speech-to-speech translation, expressive speech synthesis, neural audio codecs, audio reasoning benchmarks, and accessibility applications.

We are grateful to the 49 members of our program committee for their careful and timely reviews, to all the authors who submitted their work, and to the EMNLP 2026 workshop chairs and the ACL publication chairs for their support.

We hope you enjoy the workshop!

The SALMA 2026 Organizers

Santosh T.Y.S.S., Sreyan Ghosh, Ramani Duraiswami, Shinji Watanabe, Siddhant Arora, Sonal Kumar, Yossi Adi, Dinesh Manocha, and Leibny Paola Garcia Perera

Organizing Committee

Organizers

Santosh T.Y.S.S., Amazon, USA
Sreyan Ghosh, Google DeepMind, USA
Ramani Duraiswami, University of Maryland, USA
Shinji Watanabe, Carnegie Mellon University, USA
Siddhant Arora, Carnegie Mellon University, USA
Sonal Kumar, University of Maryland, USA
Yossi Adi, Hebrew University of Jerusalem, Israel
Dinesh Manocha, University of Maryland, USA
Leibny Paola Garcia Perera, Johns Hopkins University, USA

Program Committee

Program Committee

Yossi Adi, Hebrew University of Jerusalem and Meta
Sivaji Bandyopadhyay, Jadavpur University
Songjun Cao, Tencent
Hao Chen, Harbin Institute of Technology
David Dale, FAIR at Meta
Pavel Denisov, Fraunhofer IAIS
Kunal Dhawan, NVIDIA
Keelan Evanini, NBME
Marco Gaido, Fondazione Bruno Kessler
Yuan Gan, University of Tokyo
Yiwei Guo, Shanghai Jiaotong University
Mark A. Hasegawa-Johnson, University of Illinois, Urbana Champaign
Yejin Jeon, McGill University and Mila - Quebec Artificial Intelligence Institute
Zengrui Jin, Tsinghua University
Casey Kennington, Boise State University
Ondrej Klejch, University of Edinburgh, University of Edinburgh
Myoung-Wan Koo, Sogang University
Rishabh Kumar, Indian Institute of Technology Bombay
Tsz Kin Lam, NatWest Group
Phuong-Hang Le, Université Paris Dauphine - PSL
Sang-Hoon Lee, Ajou University
Bohan Li, Shanghai Jiaotong University
Weiyue Li, Harvard University
Xiang Li, Shenzhen Research Institute of Big Data
Xiquan Li, Shanghai Jiaotong University
Hexin Liu, Nanyang Technological University
Jie Liu, Nanjing University
Siqi Ouyang, CMU, Carnegie Mellon University
Yu-Chi Pai, CMU, Carnegie Mellon University
Ankita Pasad, NVIDIA
Rohit Paturi, Amazon / AWS AI Labs
Puyuan Peng, University of Texas at Austin
S Sakshi, University of Maryland, College Park
Leda Sari, Otter AI
Christophe Servan, AMIAD
Shuichiro Shimizu, Keio University
Aghilas Sini, LIUM, Université du Mans
Zhe Su, Carnegie Mellon University
Rupak Vignesh Swaminathan, Amazon
Hao Tang, University of Edinburgh
Pankaj Wasnik, Sony Research India
Chen Xu, Harbin Engineering University
Xuenan Xu, Shanghai Artificial Intelligence Laboratory
Liumeng Xue, Hong Kong University of Science and Technology
Shaofei Xue, Alibaba Group
Yazheng Yang, The University of Hong Kong

Jinming Zhao, Amazon
Shilin Zhou, Soochow University
Yi Zhu, Reality Defender

Table of Contents

<i>TurnStyle: Suggesting Style Prompts for Instruction-Guided Text-to-Speech Using Bimodal Dialogue Contexts</i>	
Jack McLaughlin and Cecilia O Alm	1
<i>Nuha-Speech: Building General-Purpose Arabic Speech-LLMs</i>	
Yingzhi Wang, Reem Alhazzani and Muhammad AL-Qurishi	11
<i>When How You Speak Changes What Models Refuse: Investigating Prosody-Conditioned Safety Divergence in Audio-Language Models</i>	
Akash Das and Ishan Roy	18
<i>Gold-Preserving Synthetic Augmentation for Verbatim Swedish Automatic Speech Recognition</i>	
Giuseppe Chiriatti, Dimitrios Kokkinakis, Flavio Bertini, Ricardo Muñoz Sánchez and Herbert Lange	38
<i>Pragmatic Semantic Communication of Scene Description for Blind and Low-Vision Users</i>	
Peng Wu, Ruxiao Chen and Shuaishuai Guo	45
<i>Lost in Speech: Trilingual Spoken Hallucination Detection Across Audio and Transcripts</i>	
Meruyert Aristombayeva, Jason S Lucas, Chaewan Chun and Dongwon Lee	53
<i>Improving Urdu ASR with Fine-Tuned Small Language Models for Post-ASR Error Correction</i>	
Sajjad Haider, Solat J. Sheikh, Zuha Aqib, Farah Inayat and Zehra Ahmed	74
<i>SEABED: SouthEast Asian Benchmark for Evaluating Audio Reasoning</i>	
Harshit Rajgarhia, Asif Shaik, Rachuri Lokesh, Sushanta Kumar Pani and Abhishek Mukherji	85
<i>Am I Audible Now? Dissecting Prosody in Neural Audio Codecs</i>	
Nikunj Gupta, Shreshth Rai, Kartik Sharma and Dinesh K Vishwakarma	121
<i>Echoes in Plato’s Cave: Measuring Global and Local Alignment Between Speech and Language Representations</i>	
Alef Iury Siqueira Ferreira, Lucas Rafael Stefanel Gris, F S de Oliveira, Arlindo Rodrigues Galvão Filho and Anderson Da Silva Soares	134
<i>CoSE-E: A Benchmark for Code-switched Speech Evaluation in Enterprise Settings</i>	
Shama Gupta, Hoang H Nguyen, Chelsea Huang, Lindsay Devon Brin and Fanny Riols	156
<i>Context-Aware Multimodal Claim Verification in Spoken Dialogues</i>	
Chaewan Chun, Delvin Ce Zhang and Dongwon Lee	180
<i>Lipi-Ghor: A Large-Scale, Diarized, Multi-Domain Bengali Speech Corpus and a Scalable Pipeline for Building It</i>	
Sanjid Hasan, Ahm Fuad, Risalat Labib and Bayazid Hasan Simon	190