

MINT 2026

**Workshop on Multimodal Interaction in Face-to-Face
Dialogue (MINT)**

Proceedings of the Workshop

October 29, 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 979-8-89176-482-8

Introduction

Welcome to MINT, the first workshop on *Multimodal Interaction in Face-to-Face Dialogue*, co-located with EMNLP 2026 and taking place on October 29, 2026, in Budapest, Hungary.

This first edition brings together researchers from computational linguistics, natural language processing, computer vision, and cognitive science to explore multimodal verbal and non-verbal behaviours in face-to-face communication. The workshop covers models that integrate verbal and non-verbal cues; empirical studies of multimodal communication; datasets combining synchronised speech, video, and motion; methods for evaluating multimodal understanding and generation; and applications such as embodied conversational agents, social robots, and virtual avatars. With these themes, we want to connect the study of human communication with the development of AI systems.

We received 33 direct submissions and 2 ARR commitments, of which 18 were accepted as archival papers and four as non-archival papers. We thank all the authors for their contributions and the program committee members for dedicating their time and expertise to shaping the program. We are also excited to welcome our two invited speakers, Vera Demberg (Saarland University) and Judith Holler (Radboud University and Max Planck Institute for Psycholinguistics), and thank them for sharing their research with us.

Finally, we are grateful to the Multimodal Language Department (MLD) at the Max Planck Institute for Psycholinguistics for sponsoring this event. We look forward to a stimulating exchange of ideas and thank everyone for being part of the first MINT workshop.

The MINT organisers: Esam Ghaleb, Ece Takmaz, Diego Frassinelli, Bulat Khaertdinov, Raquel Fernández, Zerrin Yumak, Asli Ozyurek.

Sponsors

The MINT organizers are grateful for the support from the Multimodal Language Department (MLD) at Max Planck Institute for Psycholinguistics (www.mpi.nl).

Program Committee

Program Chairs

Esam Ghaleb, Max Planck Institute for Psycholinguistics
Ece Takmaz, Utrecht University
Diego Frassinelli, LMU Munich
Bulat Khaertdinov, Maastricht University
Raquel Fernández, University of Amsterdam
Zerrin Yumak, Utrecht University
Asli Ozyurek, Max Planck Institute for Psycholinguistics / Radboud University

Program Committee

Iya Burenko, TU Dresden
Hendrik Buschmeier, Bielefeld University
Kerstin Fischer, University of Southern Denmark
Patrick Healey, Queen Mary University of London
Gustav Eje Henter, KTH Royal Institute of Technology
Hayley Hung, TU Delft
Hyewon Jang, Göteborg University
Ruben Janssens, Ghent University
Casey Kennington, Boise State University
Taras Kucherenko, National Library of Sweden
Lanmiao Liu, Max Planck Institute for Psycholinguistics
Rachel McDonnell, Trinity College Dublin
Chinmaya Mishra, Max Planck Institute for Psycholinguistics
Patrizia Paggio, University of Malta
Catherine Pelachaud, Sorbonne University
David Schlangen, University of Potsdam
Gabriel Skantze, KTH Royal Institute of Technology
Hugh Mee Wong, Utrecht University

Invited Speakers

Vera Demberg, Saarland University & Max Planck Institute for Informatics
Judith Holler, Radboud University & Max Planck Institute for Psycholinguistics

Table of Contents

<i>Multi-modal, Multi-Speaker Dialogue Turns Modeling with Multi-modal LLMs</i> Lukas Hilgert and Jan Niehues	1
<i>From Output Metrics to Cognitive Process: Disfluency and Lexical Entrainment for Evaluating Real-Time AI-Mediated Communication</i> Wanqing Psyche He and Susan Fussell	10
<i>Cadence-Stratified Multimodal Perception for Real-Time Child–Robot Interaction</i> Marwan Mohammed, Javad Zarrin, Salma ElSayed and Vera Kempe	22
<i>Multimodal Code-Mixed Hindi-English Dataset For Humor Recognition in Conversations</i> Abhishek Singh Bisht, Rakesh Kumar and Amrita Chaturvedi	33
<i>Heads That Ask: Prosody of Questionhood in Whispering and Mutual (In)visibility of Speakers</i> Nasim Mahdinazhad Sardhaei, Marzena Zygis and Hamid Sharifzadeh	40
<i>Exploring Multimodal Turn-Taking Cues in Face-to-Face Conversation using Voice Activity Projection</i> Willem Berner, Julio Cesar Cavalcanti, Kalle Åström and Gabriel Skantze	50
<i>Automatic Annotation of Co-speech Gesture for Multimodal Communication Research – A Framework and a Case Study</i> Ilya Burenko, Peter Uhrig and Anna Wilson	59
<i>Same Motions, Different Winners: The Gesture-Generation Leaderboard and Its Human Agreement Are Configuration-Tunable</i> Deep Gandhi, Ali Asaria and Tony Salomone	72
<i>Behavioral Alignment and the Engagement for Non-Speaking Participants in Online Meetings</i> Trung Duc Anh Dang, Zhiqing Peng and Patrizia Paggio	78
<i>Including Non-Manual Markers in Studies of American Sign Language Dialogue</i> Jillian Stebbins, Amber Hamilton, Emily Nover, Stephanie Patterson, Allison Fitch, Patrick Boudreault, Raja Kushalnagar and Cecilia O Alm	89
<i>Do Motion Tokenizers for Co-Speech Gesture Generation Encode Gesture Semantics?</i> Varsha Suresh, Divij Jain, Jia Liu, M. Hamza Mughal and Vera Demberg	96
<i>Looking for Objects and Looking for Help: Robot Gaze Elicits Human Cooperation</i> Anna Manaseryan, Thi Lan Anh Bui and Casey Kennington	105
<i>Speaker gaze and disfluencies during storytelling toward a robot or a human: The role of partner type and temperamental shyness</i> Nils Frederik Tolksdorf and Adriana Hanulíková	117
<i>FriendBench: Benchmarking Dyadic Familiarity Inference in Humans and Multimodal Large Language Models</i> Jeffrey M. Girard, Jason Zheng, Jacqueline Romero Vertino, Antony D’Avirro and Benjamin Peloquin	127
<i>MentaLED: A Multimodal Conversational Agent for Mental Health Support Using Linguistic and Paralinguistic Speech Cues</i> Changhyun Kim, Dahye Kim, Eunseon Seong, Harim Lee, Sangyool Lee and Dong-Kyu Chae	144

<i>Annotating and Recognizing the Palm Up Open Hand Gesture in Dutch</i>	
Fenna Berger, Alan Cienki and Lucia Donatelli	150
<i>The Hidden Cost of Personality: How Conditioning Methods Disrupt Gesture Generation in Small Language Models</i>	
Tova Radhe and Alessio Galatolo	160
<i>Gaze as Evidence for Common Grounding: A Cross-Corpus Analysis of MapTask and MUNDEX</i>	
Nan Li, Albert Gatt and Massimo Poesio	175