MultiMediate

Multi-modal Behaviour Analysis for Artificial Mediation

Recurring Grand Challenge at ACM International Conference on Multimedia

DFKI Saarbrücken: Philipp Müller, Jan Alexandersson

Augsburg University: Tobias Baur, Michael Dietz, Alexander Heimerl, Dominik Schiller, Elisabeth André

University of Stuttgart: Anna Penzkofer, Andreas Bulling

INRIA Sophia Antipolis: Michal Balazia, François Brémond

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

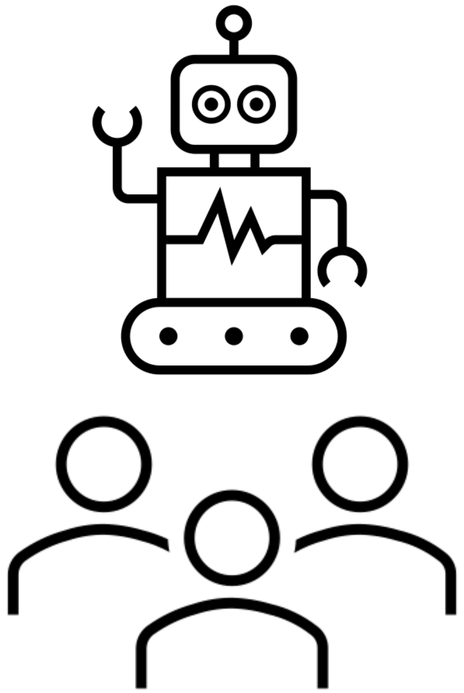
Inria

Artificial Mediators

Help to avoid and resolve conflicts

Keep engagement high

Summarize opinions



Balance participation

Help to keep the discussion focused

Give individual feedback on behavior

Human behavior **sensing** and **analysis** is **still inadequate** to realize this vision!

Enabling Artificial Mediation

Succession of tasks across the different iterations of **MultiMediate**

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

Inria

Challenge Tasks

MultiMediate'21:

- Eye Contact Detection
- Next Speaker Prediction

MultiMediate'22:

- Backchannel Detection
- Agreement Estimation from Backchannel

MultiMediate'23:

- Bodily Behaviour Recognition
- Engagement Estimation

MultiMediate'24:

- Multi-Domain Engagement Estimation

German Research Center for Artificial Intelligence
Universität Augsburg University
University of Stuttgart Germany
Inria

Challenge Tasks

MultiMediate'21:

- Eye Contact Detection
- Next Speaker Prediction

MultiMediate'22:

- Backchannel Detection
- Agreement Estimation from Backchannel

MultiMediate'23:

- **Bodily Behaviours Recognition**
- Engagement Estimation

MultiMediate'24:

- **Multi-Domain Engagement Estimation**

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

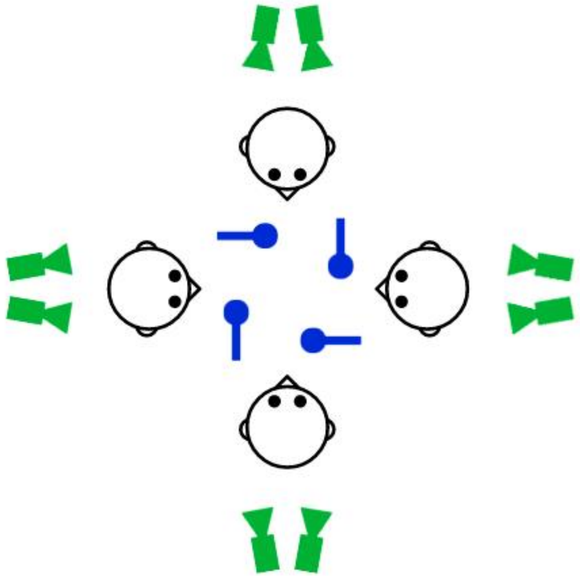
Inria

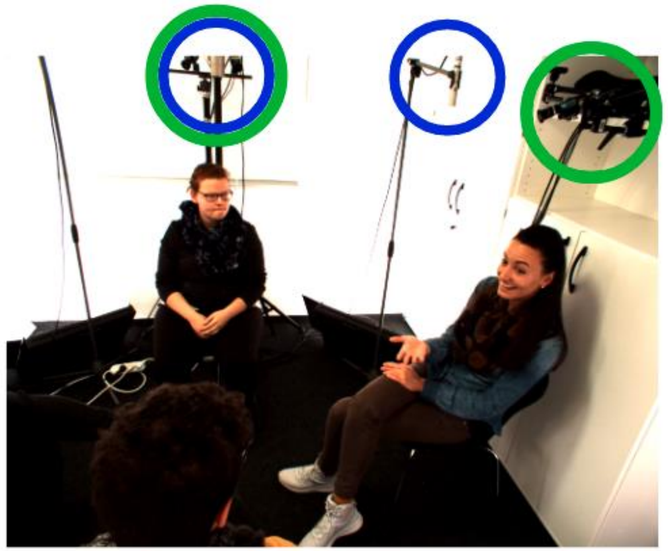
Dataset: MPIIGroupInteraction

22 group discussions

- 3 to 4 participants
- 20min duration each

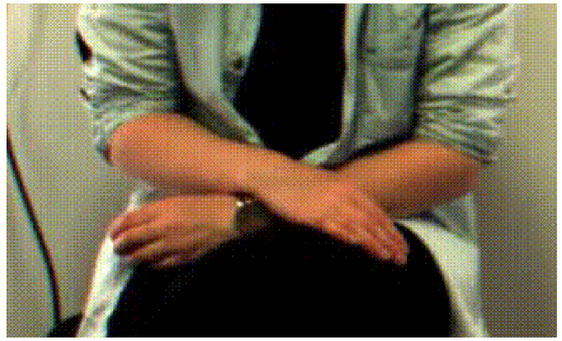
Evaluation on **unpublished test set** of 6 additional discussions



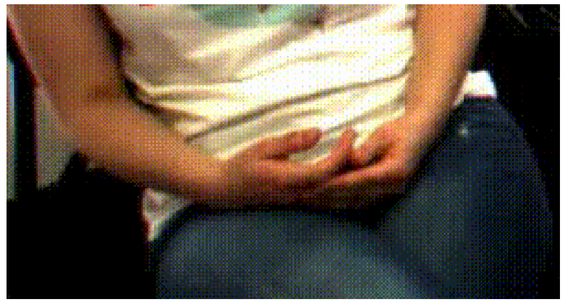


Müller, P., Huang, M. X., & Bulling, A. (2018). Detecting low rapport during natural interactions in small groups from non-verbal behaviour. IUI'18.

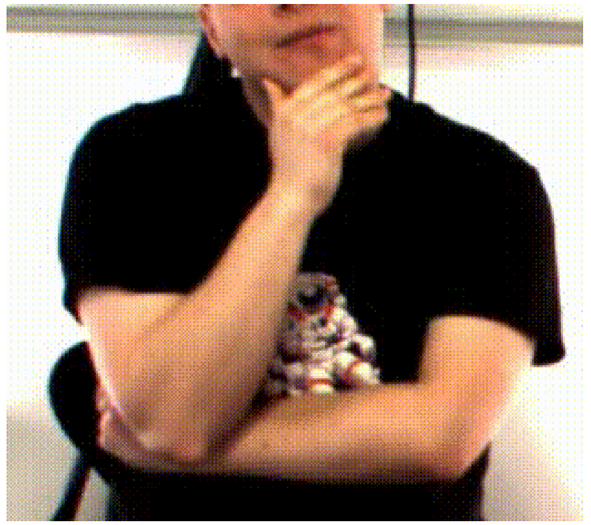
Task: Bodily Behaviours Recognition



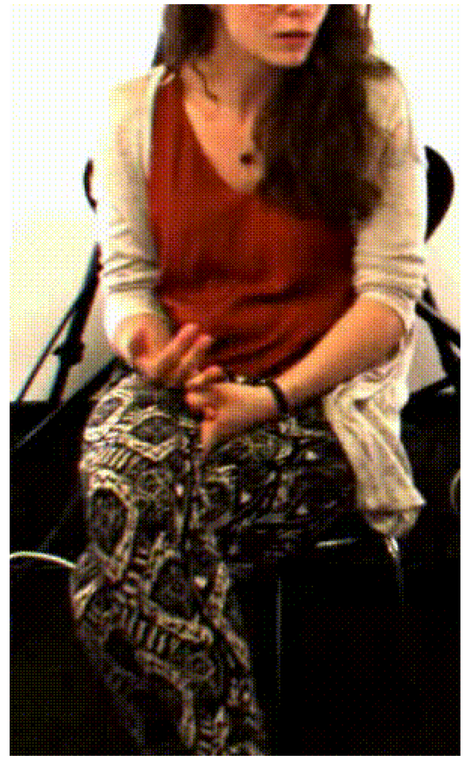
Fumbling



Fumbling, Gesturing



Face touching,
Arms crossed



Gesticulating, Shrugging,
Legs crossed, Fumbling

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

Inria

Balazia, M., Müller, P., Tánczos, Á. L., Liechtenstein, A. V., & Bremond, F. (2022). Bodily behaviors in social interaction: Novel annotations and state-of-the-art evaluation. ACM MM'22.

Task: Bodily Behaviours Recognition

Based on the **Ethological Coding System for Interviews** (Troisi, 1999)

15 behaviour classes, 26 hours of human behaviour annotated

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

Inria

Balazia, M., Müller, P., Tánczos, Á. L., Liechtenstein, A. V., & Bremond, F. (2022). Bodily behaviors in social interaction: Novel annotations and state-of-the-art evaluation. ACM MM'22.

Task: Bodily Behaviours Recognition

Input:
64 frames from 3 views

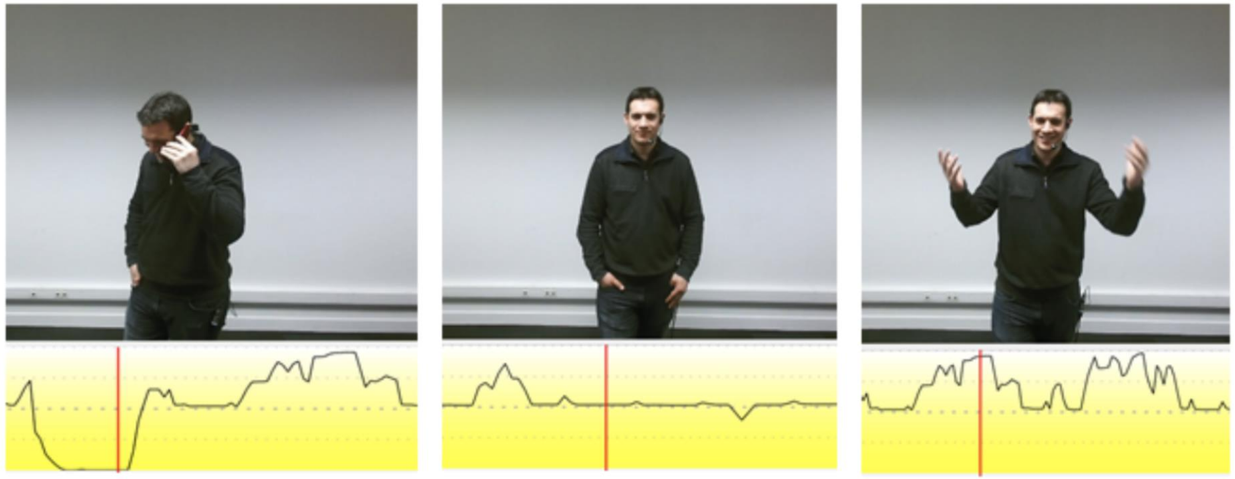
Output:
Behaviour Labels (multi-label)

Metric:
Mean Average Precision
(Macro Average)

Method	MAP
Trivial Baseline	0.24
Video Swin Transformer, average across views	0.56
gkdx2 (from 2023) [3rd]	0.626
jdy203 [2nd]	0.63
gkdx2 [1st]	**0.63**

gkdx2: Kun Li, Dan Guo, Guoliang Chen, Feiyang Liu, Meng Wang: Data Augmentation for Human Behavior Analysis in Multi-Person Conversations. ACM MM'23.

Engagement Estimation



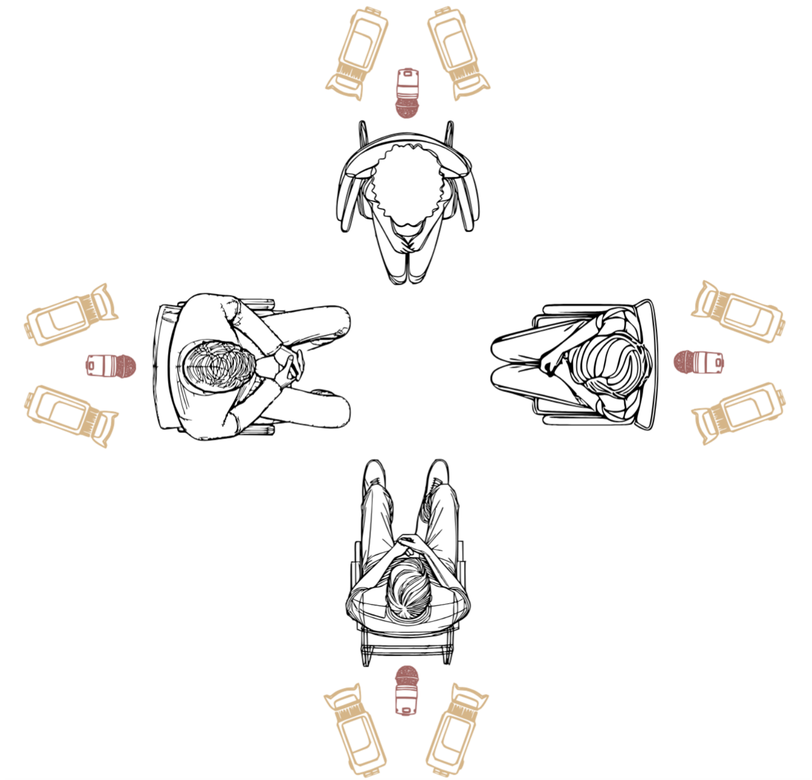
Low Engagement | Medium Engagement | High Engagement

Knowing how **engaged** users are is crucial for **artificial mediators**

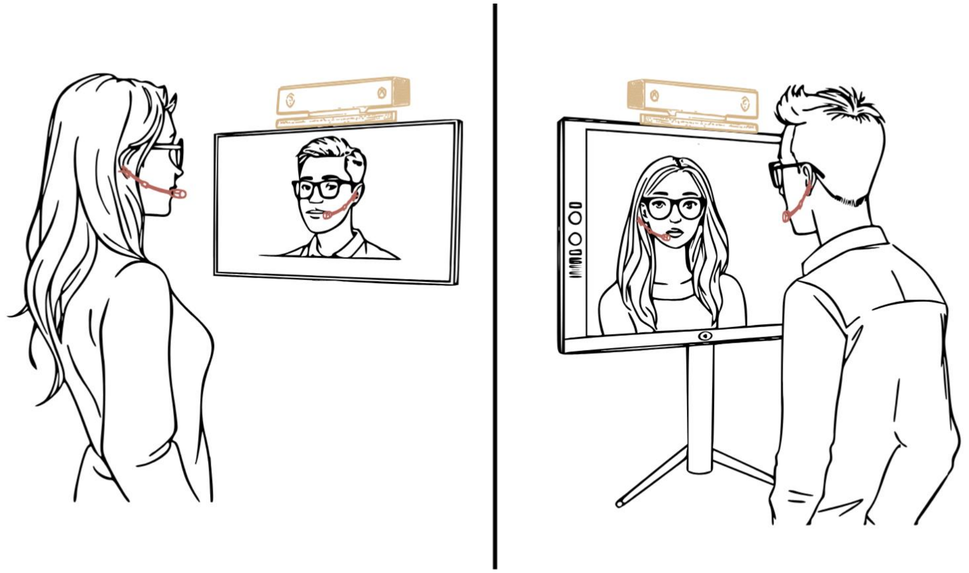
Engagement is a complex, **context-** and **culture-dependent** phenomenon

Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., ... & Valstar, M. (2017). The NoXi database: multimodal recordings of mediated novice-expert interactions. ICMI'17.

Multi-Domain Engagement Estimation Challenge



MPIIGroupInteraction (Müller et al., 2018)



NoXI Corpus (Cafaro et al., 2017)

Cafaro, A., Wagner, J., Baur, T., Dermouche, S., Torres Torres, M., Pelachaud, C., ... & Valstar, M. (2017). The NoXi database: multimodal recordings of mediated novice-expert interactions. ICMI'17.

Müller, P., Huang, M. X., & Bulling, A. (2018). Detecting low rapport during natural interactions in small groups from non-verbal behaviour. IUI'18.

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

Inria

Multi-Domain Engagement Estimation Challenge

Training Data	Validation Data	Test Data
NoXI-train French, German, English	NoXI-val French, German, English	NoXI-test French, German, English
		NoXI (new languages) Indonesian, Arabic, Spanish, Italian
	MPIIGroupInteraction-val German	MPIIGroupInteraction-test German

Metric: Combined Concordance Correlation Coefficient (CCC)

German Research Center for Artificial Intelligence

Universität Augsburg University

University of Stuttgart Germany

Inria

Müller, P., Balazia, M., Baur, T., Dietz, M., Heimerl, A., Penzkofer, A., ... & Bulling, A. (2024). MultiMediate'24: Multi-Domain Engagement Estimation. ACM MM'24.

Task: Multi-Domain Engagement Estimation

Baseline Experiments

Features	NOXI		NOXI (Add. Languages)		MPIIGroupInteraction		Combined
	Val CCC	Test CCC	Val CCC	Test CCC	Val CCC	Test CCC	Test CCC
Video							
OpenFace 2.0	0.81	0.28	-	0.13	**0.09**	0.00	0.14
OpenPose	0.83	0.48	-	0.41	0.01	0.06	0.32
CLIP	**0.88**	0.48	-	0.38	-0.01	0.06	0.31
Voice							
eGemaps v2	0.77	0.56	-	0.47	0.00	**0.15**	0.39
w2vbert2	0.77	**0.64**	-	**0.51**	0.05	0.09	**0.41**
Text							
XLM RoBERTa	0.62	0.40	-	0.29	0.00	0.00	0.23

Method	CCC NoXi base	CCC NoXi Additional Languages	CCC MPII GroupInteraction	CCC combined
Challenge Baseline	0.64	0.51	0.09	0.41
Syntax [4th]	0.72	0.69	0.50	0.64
HFUT-LMC [3rd]	**0.76**	0.67	0.49	0.64
AI-lab [2nd]	0.69	0.72	0.54	0.65
USTC-IAT-United [1st]	0.72	**0.73**	**0.59**	**0.68**

Syntax: Deepak Kumar, Surbhi Madan, Pradeep Singh, Abhinav Dhall, Balasubramanian Raman: Towards Engagement Prediction: A Cross-Modality Dual-Pipeline Approach using Visual and Audio Features. ACM MM'24.
HFUT-LMC: Jia Li, Yangchen Yu, Yin Chen, Yu Zhang, Peng Jia, Yunbo Xu, Ziqiang Li, Meng Wang, Richang Hong: DAT: Dialogue-Aware Transformer with Modality-Group Fusion for Human Engagement Estimation. ACM MM'24.

German Research Center for Artificial Intelligence
Universität Augsburg University
University of Stuttgart Germany
Inria

The Future of MultiMediate

Expand multi-domain engagement estimation challenge

Introduce metrics that **measure fairness** of challenge solutions

Mailing List

Evaluation remains possible for all tasks [1].

Stay tuned by signing up for the **mailing list** [2]!

German Research Center for Artificial Intelligence
Universität Augsburg University
University of Stuttgart Germany
Inria

[1] https://multimediate-challenge.org
[2] https://multimediate-challenge.org/SignUp/