

AI-Enhanced Multimodal Communication: How to Build Rapport between Humans and Agents across Physical and Virtual Platforms

Elisabeth André

Universität Augsburg

Human-Centered Artificial Intelligence

Socially Interactive Artificial Intelligence

Artificial Intelligence that emulates communication styles inspired by human-human communication.

**Health Advisor
KRISTINA**



**Coaching Assistant for
Psychological Support**



**Robotic
Companion**



viva navel
robotics



European
Commission

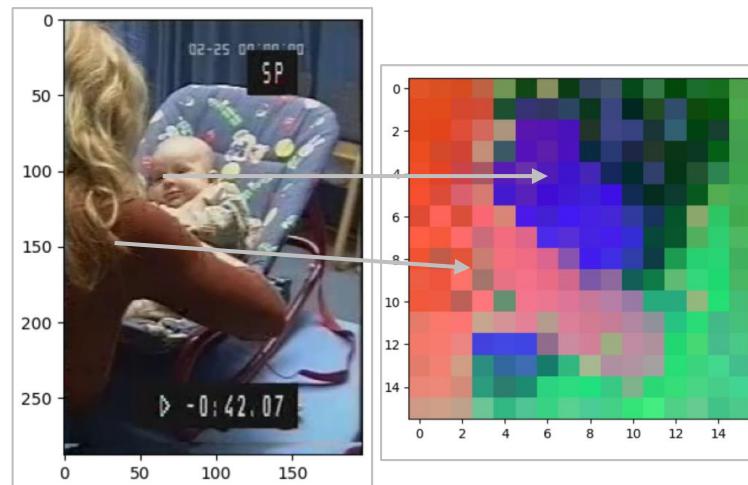


UNIA

Analysis of Social Interactions between Humans

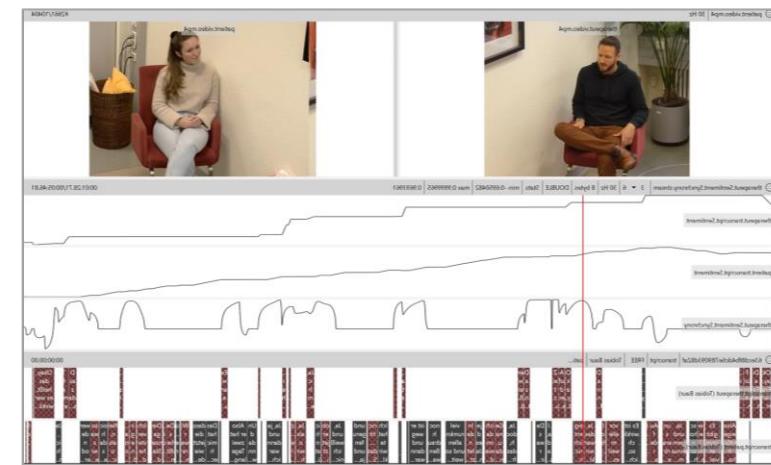
Artificial Intelligence for Human Behavior Analysis

SCHWAN: Effects of the COVID-19 Pandemic on Parenting and Child Development



2022 - 2025

TherapAI – Analysis of Emotional Expression in Psychotherapy Using AI



2022 - 2025

KodiLL: AI Feedback in Simulated Medical Consultation Sessions



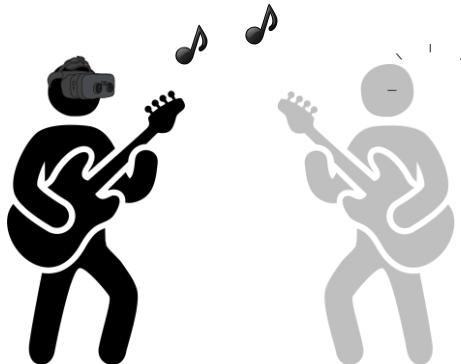
2021 - 2025



How enhance human-human communication across and via diverse xVR-settings?

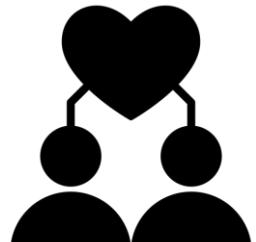
Connect people at different locations

→ Jamming in xVR



Group work in VR
→ Virtual Classroom

Perspective Taking in VR
→ Avatars representing (under-)privileged groups



Social Coaching in VR
→ Role Play with Virtual Agents



Connect people across
distances

01

Motivation

- COVID isolated musicians
- Jamming in-person impossible
- Existing remote solutions are audio-only
- XR/MR systems can foster co-presence
- What about MR remote jamming?

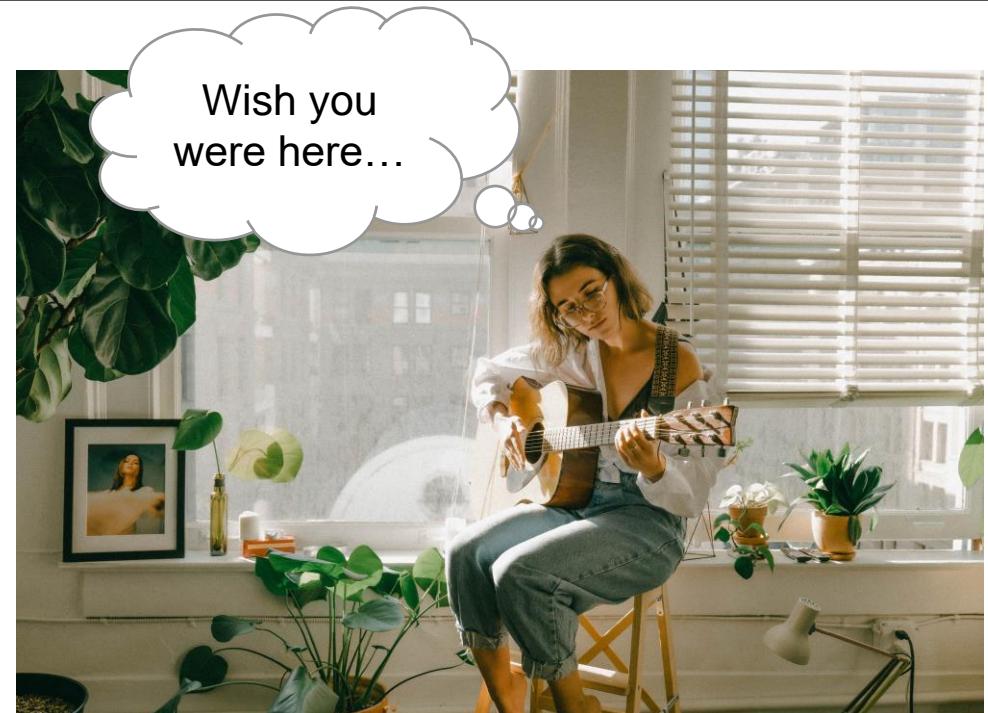


image credits: Chase Yi



3D Teleportation: Stream point clouds of jamming partners
(that are in different rooms) to each other

Virtual 3D Teleportation in Real-Time

Jamming Session from the Point of View of one Musician



Virtual 3D Teleportation in Real-Time

Jamming Session from the Point of View of Both Musicians



Remote Jamming in VR

- Stream point clouds of jamming partners (that are in different rooms) to each other

Different Rooms



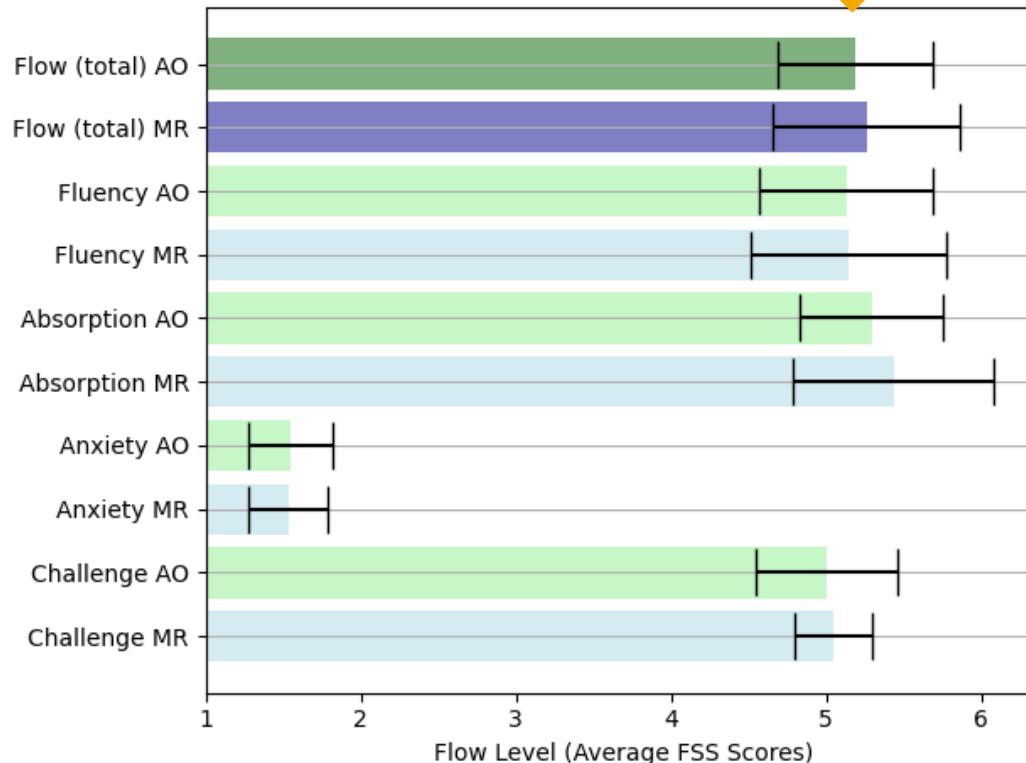
Laboratory Study

- **Experimental Setup:**
 - Pairs of musicians
 - Jamming in separated rooms
 - Audio-Only vs. Mixed Reality
 - Within-Subjects, N=24
- **Measures:**
 - **Questionnaires**
(Social Presence, Flow, Affect)
 - **Physiological** data analysis
 - **Qualitative** feedback

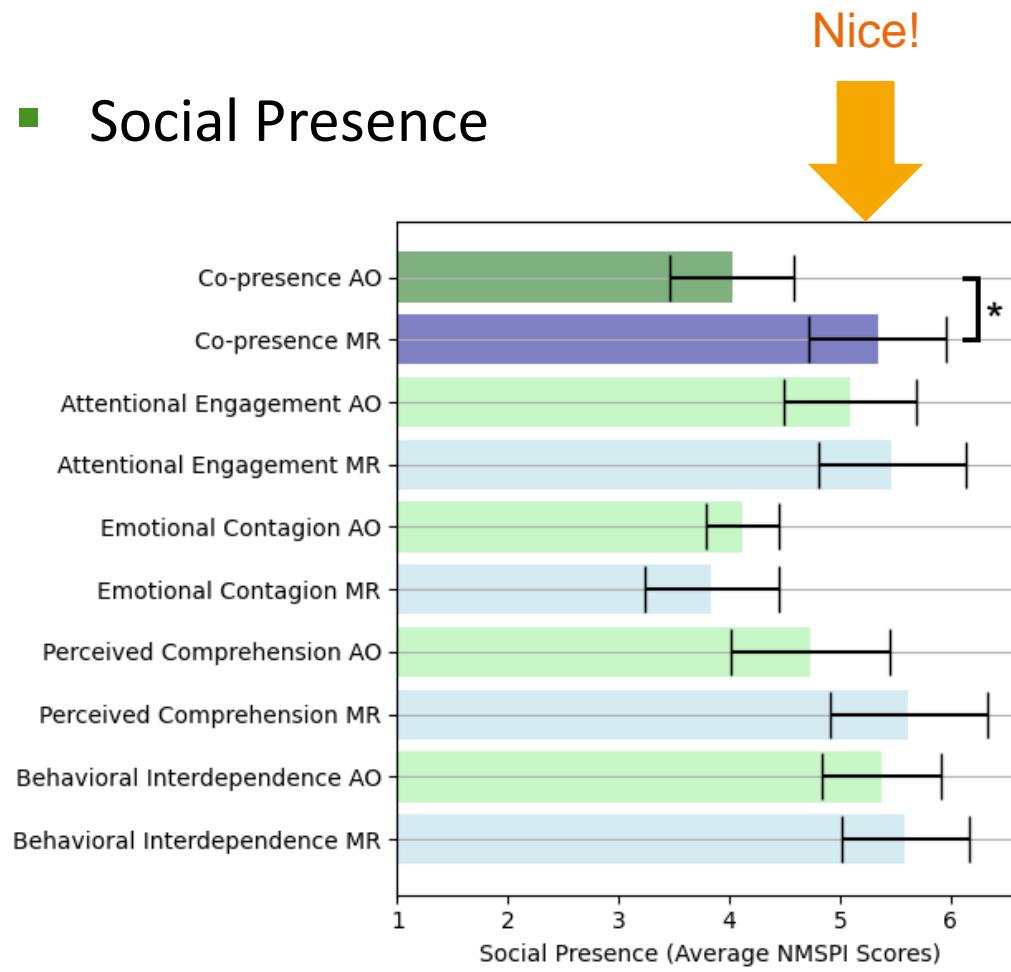


Questionnaire Results

- Flow



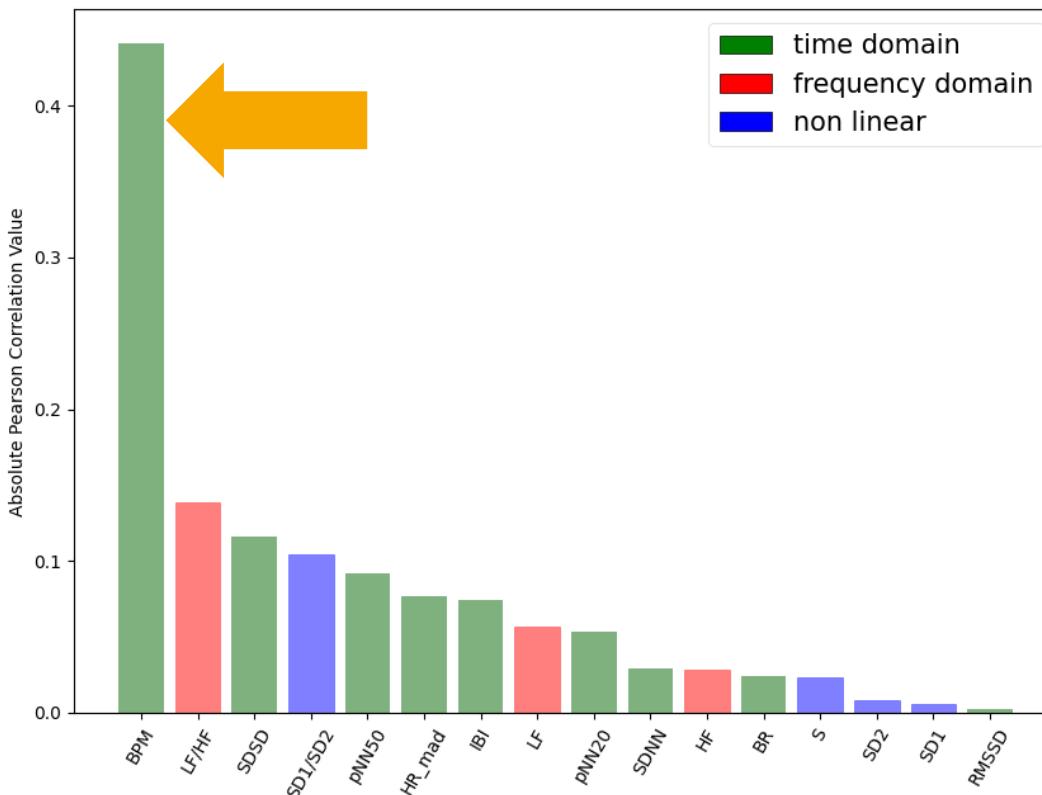
- Social Presence



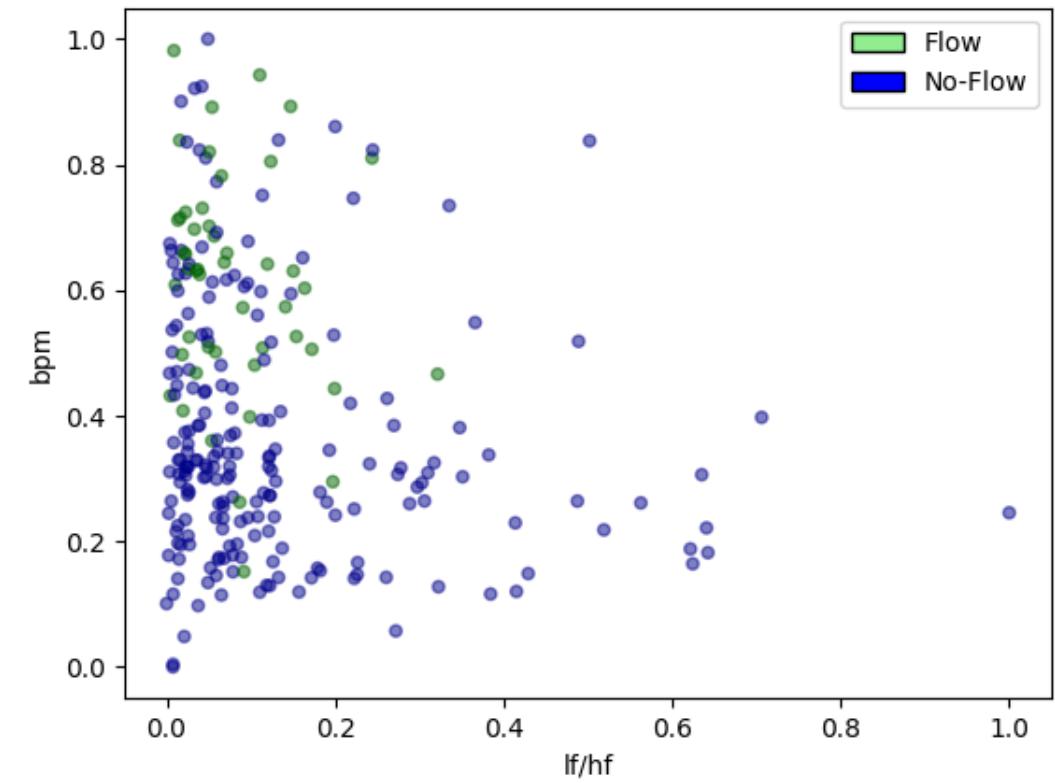
Physiological Data Analysis

- Flow has distinctive features that may enable real-time assessment

- Feature Correlations for MR



- Plot for MR



Qualitative Results: Feedback for Mixed Reality Jam

- positive

interesting/fascinating
enthusiasm/approval
better with MR
further interest/wish to use
improved communication

improved flow
fun/joyful
improved concentration
improved co-presence

- negative

trouble seeing instrument

better without MR
image latency
rouble seeing environment
rouble seeing partner
image stressful

HMD covers face
HMD limits movements
image too pixelated

struggling with MR (in general)

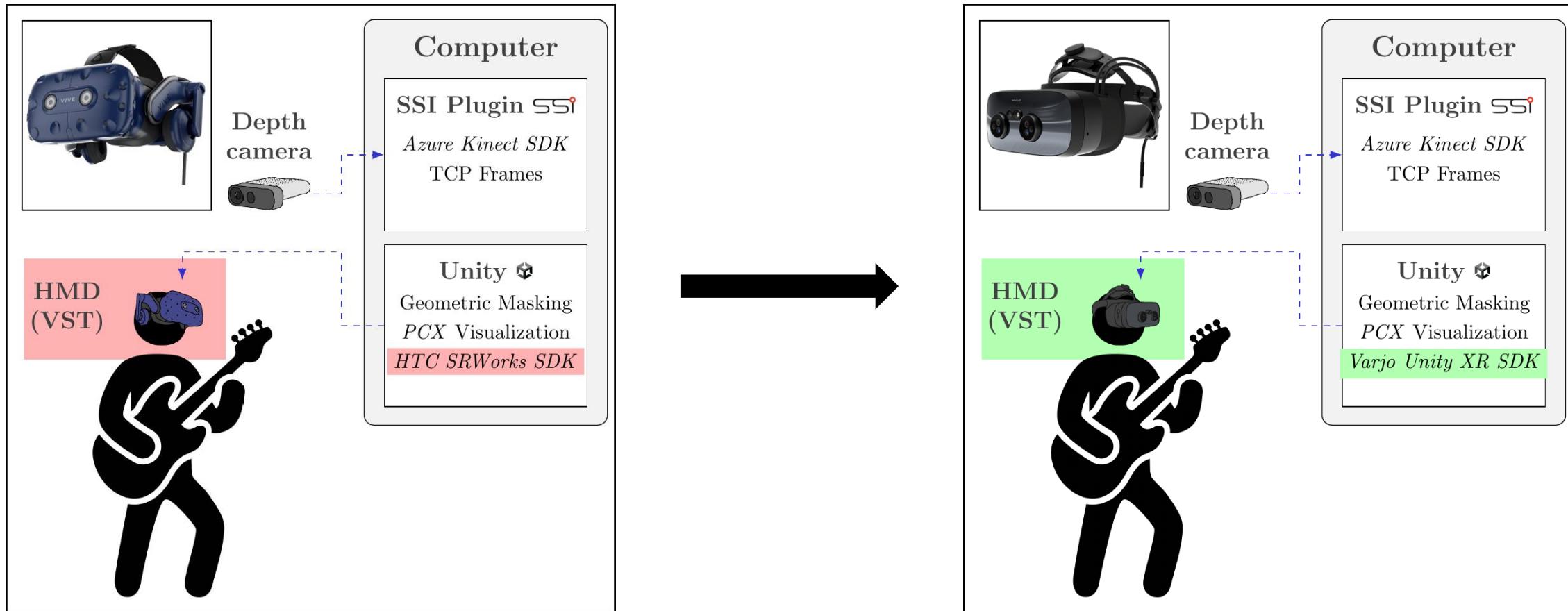
Conclusions

- Positive **affect increased** after jamming.
- **Co-presence higher** in Mixed Reality.
- **Physiological flow assessment** in real-time may be feasible.
- **Technical limitations** are a key issue.



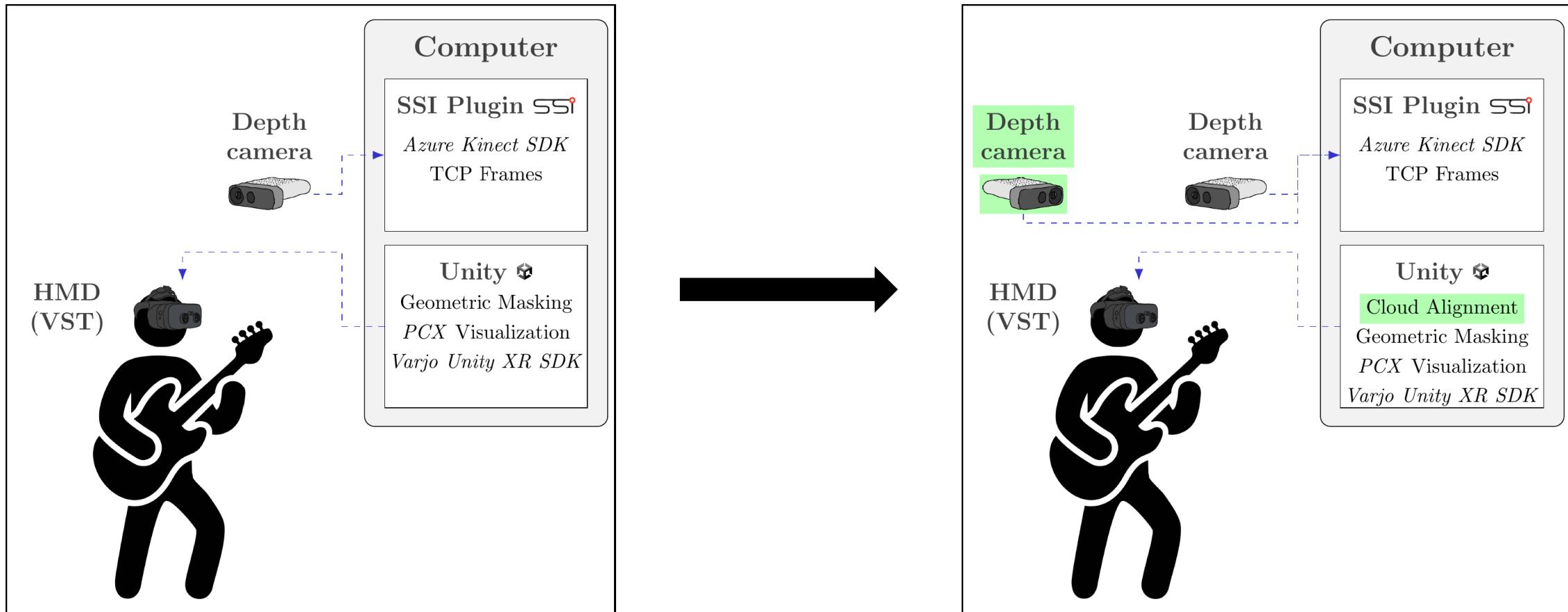
Mixed Reality

Upgrade



Dual Azure Kinect Setup

Upgrade



Comparison

BEFORE



AFTER



Group Work in VR

02

Virtual Situated Learning and Teaching with Avatars and Agents in Social Cyberspace



ViLearn_More aims to record and analyze cognitive and social-affective behavioral signals to explore regulatory mechanisms that modulate teaching/learning behavior in VR.



Bundesministerium
für Bildung
und Forschung



Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Technical
University
of Munich



Education in VR

Experiences from the original ViLearn project:

- The original Vilearn project applied situated learning (learning that takes place in the same context in which it is applied), with **successful learning results about VR while in VR** (Schroeter et al., 2021).
- **Enrichment through VR:** VR offers interactive learning in scenarios that are otherwise costly or unsafe
- **Social Dynamics:** Effective communication and peer interaction are vital, but VR can complicate these due to missing social cues
- **Sense of Community:** Despite challenges, some participants felt a stronger sense of community during VR collaborations.



analyse behavioural signals to explore **regulatory mechanisms** and **conversational dynamics** in educational VR environments

Data Collection Setup

▪ Experimental Setup

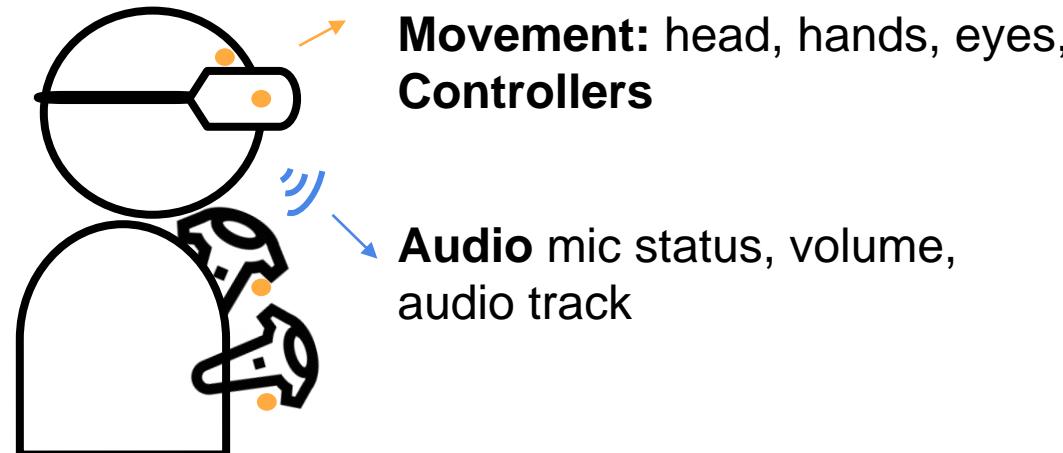
- Classroom seminars in **dyads** and **triads**
- 10 minutes conversation about seminar topic per group

Recording Session	Dyads	Triads	Participants
Dec 2023	4	3	17
May 2024	1	8	26
June 2024	4	3	17
November 2024	1	3	11
Total Data:	10	17	71



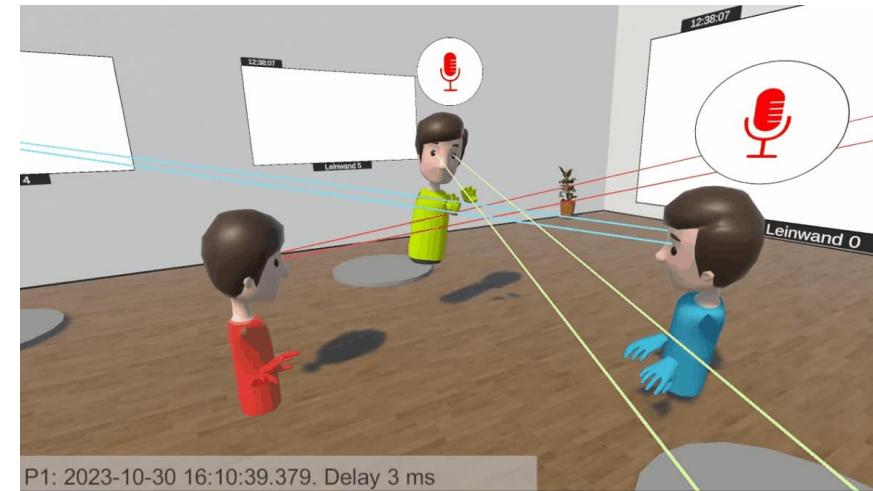
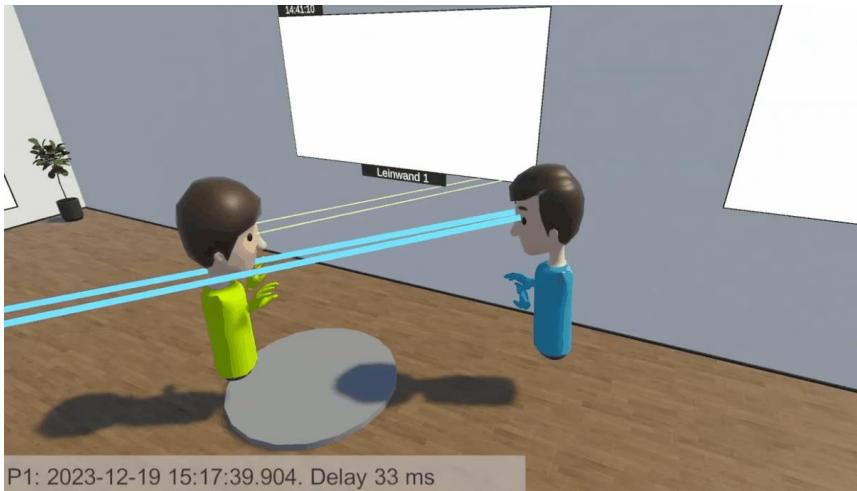
Hardware and Software

- **VR Hardware:** HP Reverb G2 Omnicept Edition
- **Windows Computer:** XMG Pro 17
- **Software:** Unity Engine version 2020.3 running University of Würzburg Vilearn software
- **Eye Data:** Tobii VR4 Eye Tracker in the headset



Data Collection Pipeline in Unity

- **Data Collection Pipeline in Unity**
 - Eye tracking data (blinks, gaze)
 - Movement (head, hands)
 - Voice



Importance of Blinks in Social Interactions

- Blinks are not only physiological, for humidifying the eye, but 48% - 71% of blinks are associated with **communicative aspects** [Ford et al., 2013] (setting: dyadic video-mediated scenario)
- **Blink duration** impacts communicative behaviour: participants unconsciously responded with **shorter answers to interlocutor's longer blinks** [Homke et al., 2018] (setting: dyadic video-mediated scenario)
- **Blink Sync:**
 - can predict **interest level** (higher synchrony = higher interest) [Nakano et al., 2019] (setting: individual video watching scenario)
 - **problem-solving performance** in dyads interactions in videocalls (higher synchrony = better performance) [Hoffmann et al., 2023];
 - might indicate **effective social communication** [Nakano et al., 2011]

Blinks in Virtual Environments

- Although valuable in social interaction, blinks are **little studied** in social interactions in VE
- VEs can have a visual fatigue leading to an **increased blink rate** [Lopes et al., 2020]
- **Higher blink rate** was recorded for participants exploring an environment in VR compared to on 2D screen [Dennison et al., 2016], similar higher blink rate while playing Flappy Bird in VR compared to 2D screen [Marshev, 2021].
- Additionally, the blinks information within **virtual learning environments** may indicate:
 - **Engagement**
 - **Task performance**
 - **Interest level**
- Providing valuable information on classroom interactions.

Key Findings on Blinking in our Study

- **Blink Rate:** Similar to face-to-face (F2F) for dyads; lower for triads, consistent with previous studies who report higher blink rate for people engaged in a conversation
- **Synced Blinks:** Higher in dyads (0.61) than triads (0.51), though not significant; blink rate may influence syncing—further studies needed.
- **Temporal Structure:** Synced blinks show similar timing in dyads and triads, findings on dyads consistent with F2F studies.



Given the similarities to conversation in F2F, blinks data from VR could indicate effective social communication patterns, problem solving and interest level.

Perspective Taking in VR

03

Artificial Intelligence for Assessment (AI FORA)

■ Objectives:

- To better **understand AI-based decision-making** processes in the **allocation of public services** and **generate improvement suggestions**.
- Better understanding of (culturally influenced) **values** and their resulting implications.
- Better understanding of **contextual factors**.
- Better understanding of **emerging developments**.



VolkswagenStiftung

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ



cress

UNIA

Artificial Intelligence for Assessment (AI FORA)

- AI for social assessment usually trained on **data sets from the real world**
- **Bias** and social injustices are often **reproduced**
- Such bias may be more felt by affected/underprivileged persons
- **Perception gap** between privileged and underprivileged groups



VolkswagenStiftung

JOHANNES GUTENBERG
UNIVERSITÄT MAINZ



cress

|UNI

Conduct co-creative workshops to design AI solutions that are regarded as “fair”

- Previous studies: Experiences in **VR** can increase **empathy** via embodiment and perspective-taking
- Goal: **Foster empathy** for underprivileged groups prior to co-creative workshops.

Can **VR** help to build **empathy** among stakeholders?



VR Perspective Taking

A participant embodies a persona representing a differently privileged demographic group via VR full-body tracking, acclimating in a mirrored elevator before facing a biased AI system.



Biased AI System Disadvantaging Individuals Based on Gender and Age in Credit Decisions

Types of Embodiment

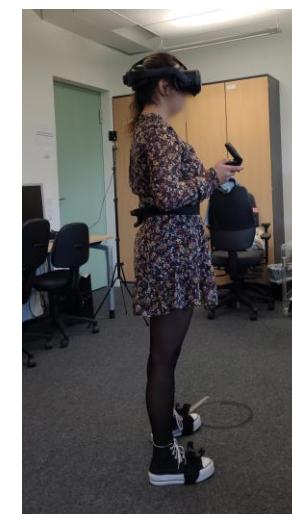
- **ME—Mental Embodiment:**

- Participants were given profiles of four personas (in printed form).
- Participants role-played printed profiles of four personas while being assessed by a Wizard of Oz credit scoring AI in a lab.



- **VE—Virtual Embodiment:**

- Participants received their digital profiles in VR before entering a virtual lab replica.
- Participants embodied virtual avatars of the four personas in VR using motion capture.



Experimental Setting: Unfair AI Decisions Based on Gender and Age

- **Perspective-Taking** in VR for fostering empathy for underprivileged groups
- Hope: Participants will be **affected** just as an underprivileged group
- This was a **preparation stage** for our co-creativity workshops



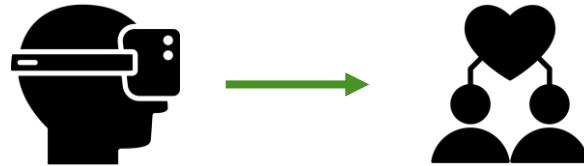
Privileged



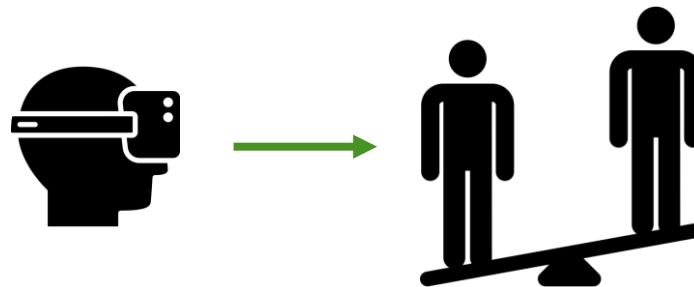
Underprivileged

Main Findings

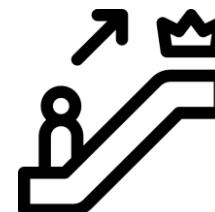
- VR can increase the sense of empathy towards the embodied persona.



- VR can intensify the perception of AI bias.



- Perception of the AI bias is lower when privileged characters are embodied in VR.



Social Coaching with Autonomous Agents

04

Social Coaching with a Virtual Job Interviewer

User takes part in a gamified job interview simulation lead by a virtual recruiter.



Most Participative Demo
Award at UMAP



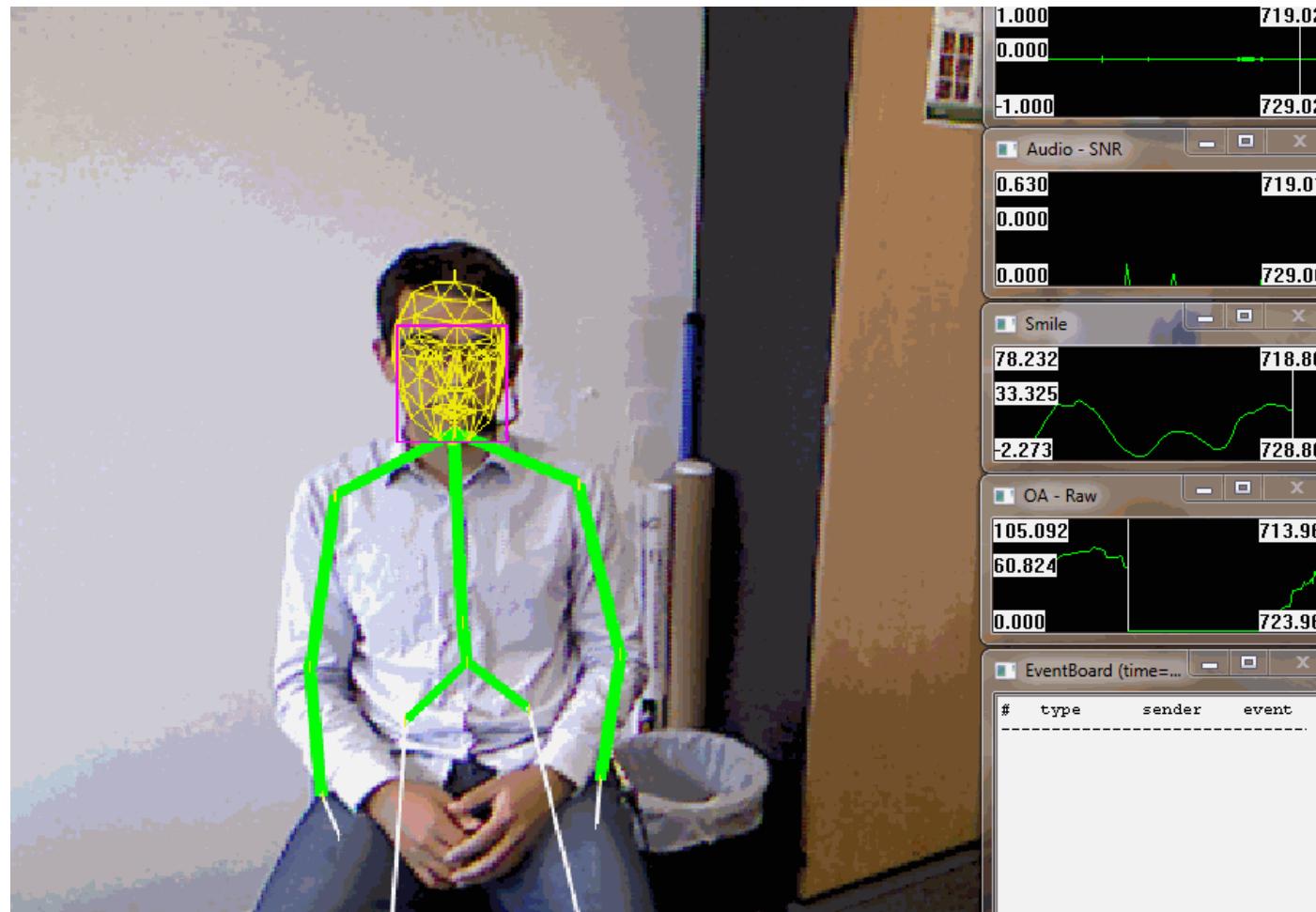
Exercises for Social Coaching

Physical Games Cards give Hints on How to Behave



Job Interview Training with a Virtual Character

During the game, sensors record and analyze the user's behavior in real time.



Job Interview Training with a Virtual Character

A scoring system tracks the quality of the user's behavior,



Explanations of Assessment Processes for Social Coaching

Job Interview Training

Open Body Posture, leaning forward, orientation towards interlocutor



Nothing specific found

Leaning backwards, crossing arms, looking away from interlocutor



Detected Voice Activity



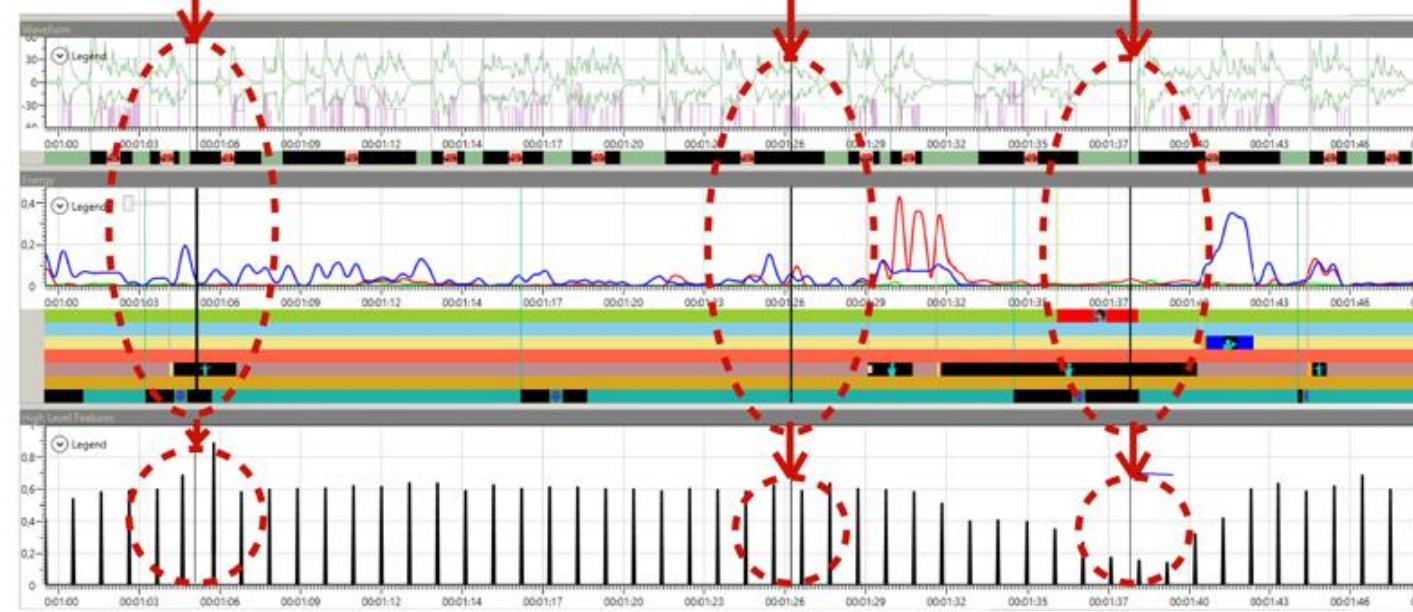
Movement Energy



Detected Social Cues



High Level Information on „Engagement“



High

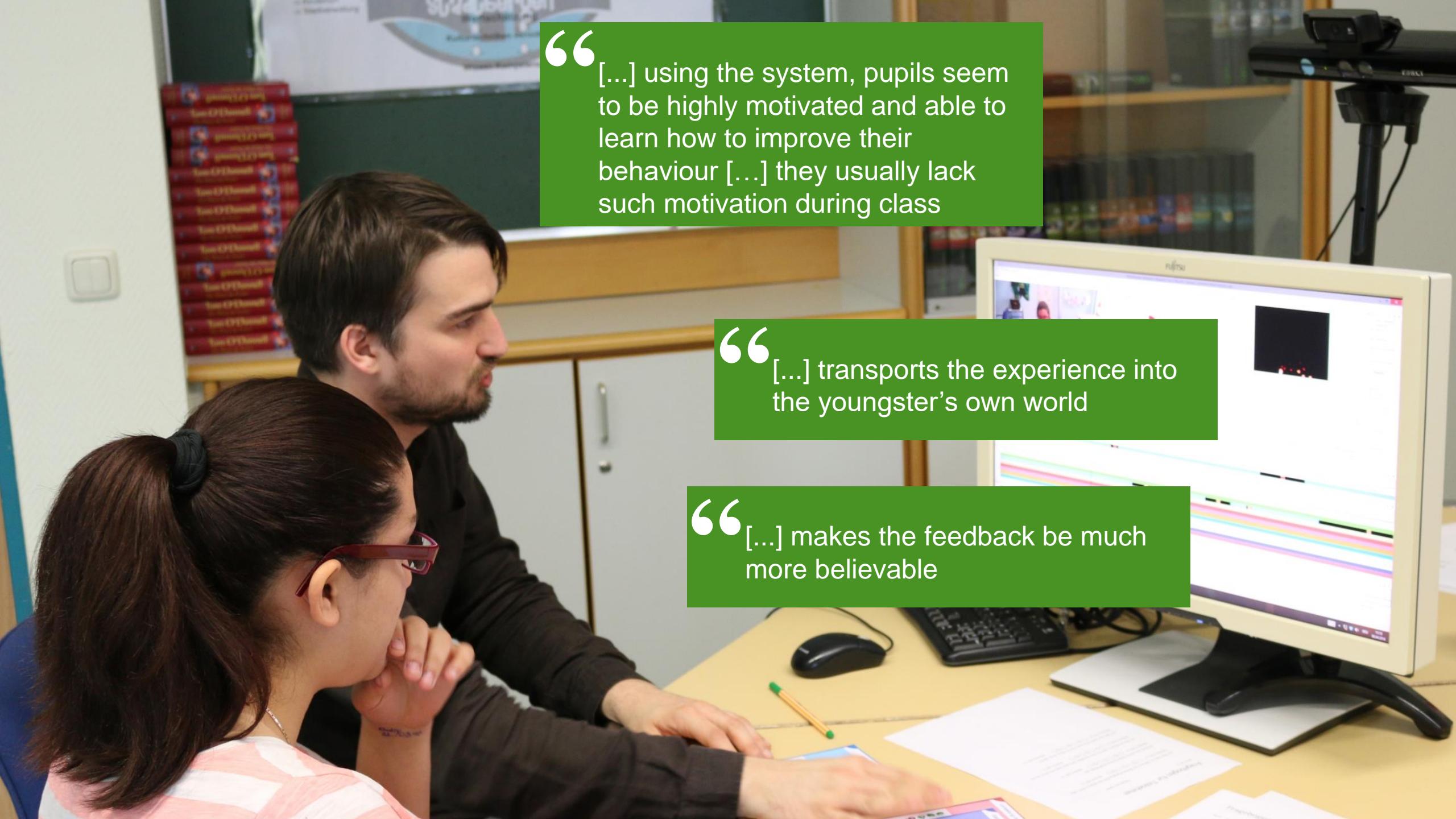
Medium

Low

“ [...] using the system, pupils seem to be highly motivated and able to learn how to improve their behaviour [...] they usually lack such motivation during class

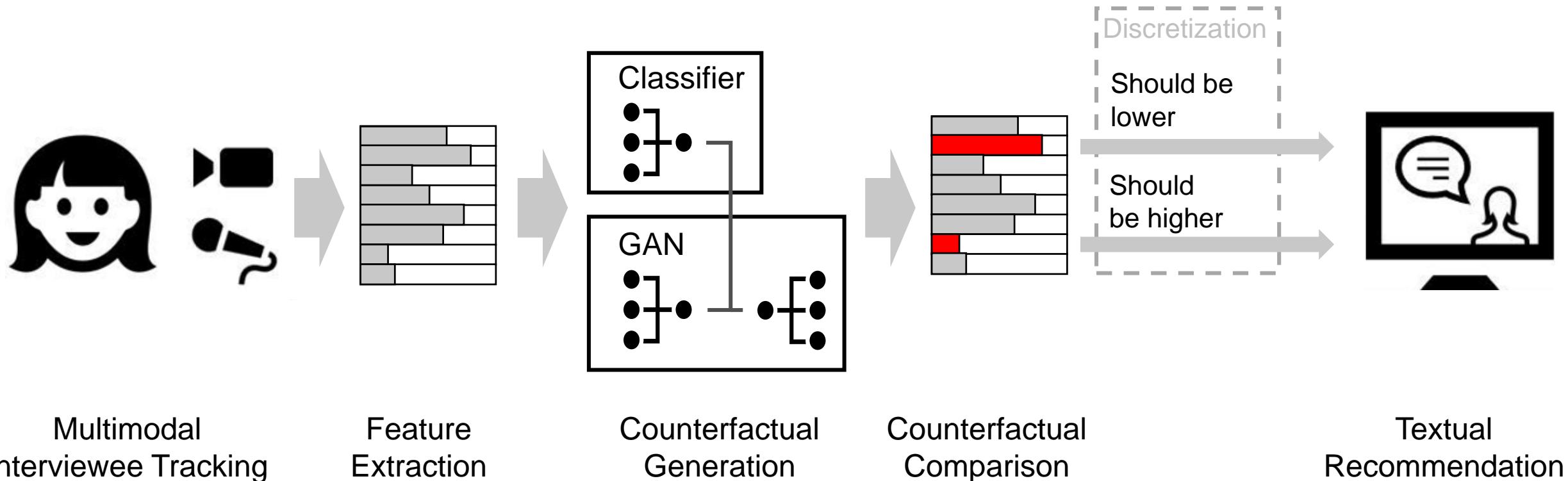
“ [...] transports the experience into the youngster's own world

“ [...] makes the feedback be much more believable



“GAN I hire you?” - A System for Personalized Virtual Job Interview Training

- Counterfactual reasoning for generating verbal feedback about observed social behavior
- Use of counterfactual explanations, explaining to a user that a modified version of her/his social behavior would have led to a better behavior rating



“GAN I hire you?” - A System for Personalized Virtual Job Interview Training

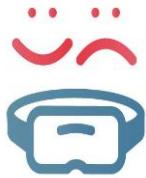
Verbal Feedback Provided to the User via the Virtual Coach



MITHOS: Mixed Reality Training for Teachers

MITHOS offers an engaging MR training experience enriched with virtual agents, providing an immersive and dynamic platform for simulating conflict situations in educational settings.





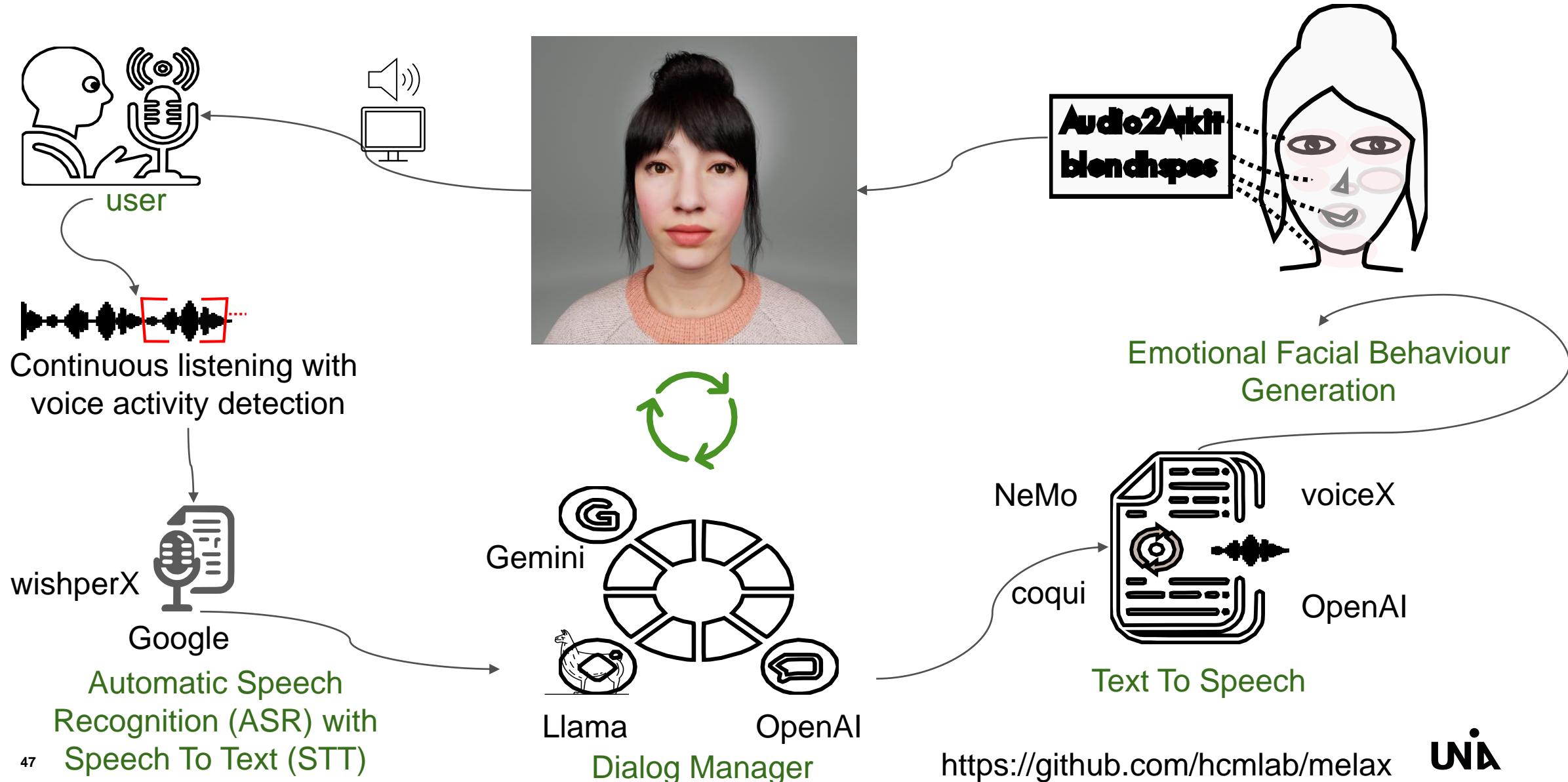
MITHOS: Mixed Reality Training for Teachers



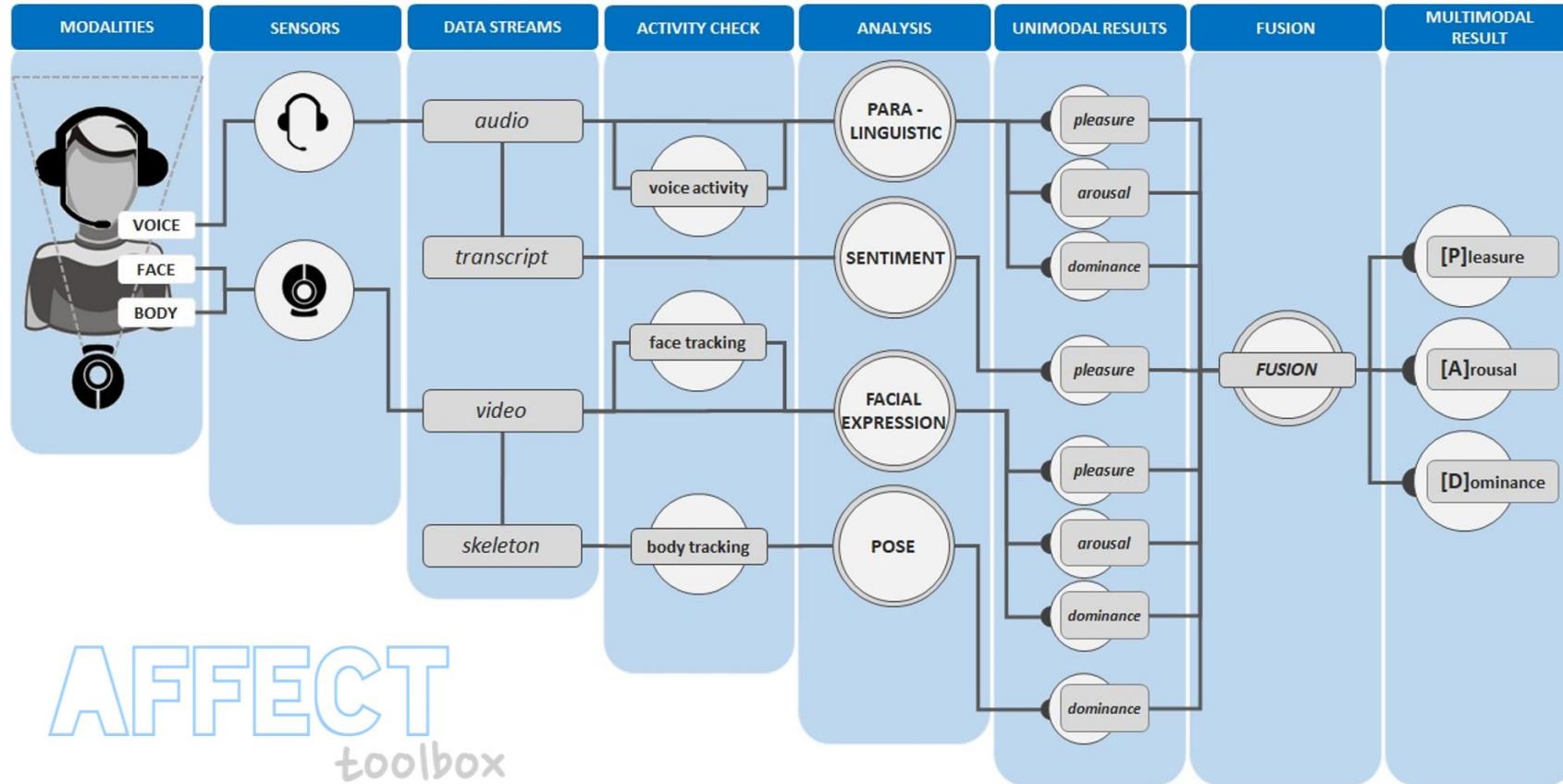
Tools for Developing Customized Virtual Agents

05

MeLaX: Framework for Building Virtual Agents



AffectToolbox - Accessible Real-Time Affect Recognition for HCI Applications



audiovisual sensory devices (e.g. webcams) provide easy means to generate all considered data streams (i.e. audio, transcript, video and skeleton data)

activity checks trigger the machine learning based analysis of respective modalities

uni-modal results of applied affect recognition models are represented by a subset of pleasure, arousal and/or dominance scores

unimodal emotional cues are the input for an event-driven **fusion** algorithm, which deduces a coherent affective state, represented in the continuous PAD emotional space

AFFECT
toolbox

AFFECT
toolbox

Conclusions

06

Conclusions

- **Increasing Connectedness: “Wish You Were Here”**
 - Bringing people together virtually, regardless of geographical barriers.
- **Virtual Collaboration: Meeting Avatars in Shared Spaces**
 - Enabling individuals to gather and interact as avatars in virtual classrooms
- **Improving Empathy: Walking in Others' Shoes**
 - VR simulations allowing users to experience different perspectives.
- **Training with Autonomous Agents**
 - Preparing individuals for real-world scenarios through interaction with VR agents.

Conclusions

- xVR has the potential to **connect people** across physical distances and diverse communication abilities.
- xVR provides **enhanced interaction capabilities**, but still suffers from **reduced communication channels** which may affect the social dynamics of the conversations.
- **Understanding the dynamics of interaction** is essential:
 - Learning from **Human-Human Interaction in Real-World-Settings**
 - **Human-to-Human in xVR**
 - **Human-to-Agent in xVR**



Universität Augsburg
Fakultät für Angewandte
Informatik

Vielen Dank für Ihre Aufmerksamkeit!



Elisabeth André
Human-Centred Artificial Intelligence
Universität Augsburg
andre@hcm-lab.de
<https://hcm-lab.de>



<https://github.com/hcmlab/>