Face Off: External Tracking vs. Manual Control for Facial Expressions in Multi-User Extended Reality

MENSCH UND COMPUTER 25
DIGITAL DIVERSITY
31. AUGUST – 03. SEPTEMBER 2025 | Chemnitz

Katja Krug[1,3], Xiaoli Song[1], Wolfgang Büschel[2]

[1]Interactive Media Lab, Technische Universität Dresden; [2]VISUS, University of Stuttgart, Stuttgart, Germany; [3]Centre for Tactile Internet with Human-in-the-Loop (CeTI)

Motivation & Basic Idea

In XR, **expressive avatars** usually rely on **built-in sensors** in high-end HMDs. Since most devices lack these sensors, **workarounds** are needed to enable **facial expressions**.

- External methods, like **webcam-based tracking**, are often **unstable and unreliable**.
- Alternatives, like **manually triggered presets, don't reflect true expressions** and are **unnatural** to produce.

When built-in sensors are not available, do users **accept unstable tracking, or prefer manual presets?**

We studied 18 participants in dyads, comparing webcam-based tracking with manual triggering.

Figure 1: Tracked

Externally webcam-based tracked facial expressions transferred to an avatar.

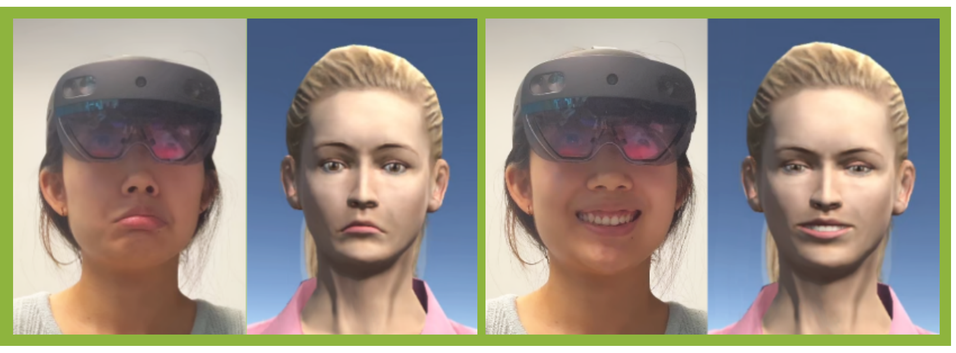


Figure 2: Manual

Preset facial expressions, manually triggered with a handheld controller.

Unstable truth or stable fakes: How would you rather express yourself in XR?

Setup & Study

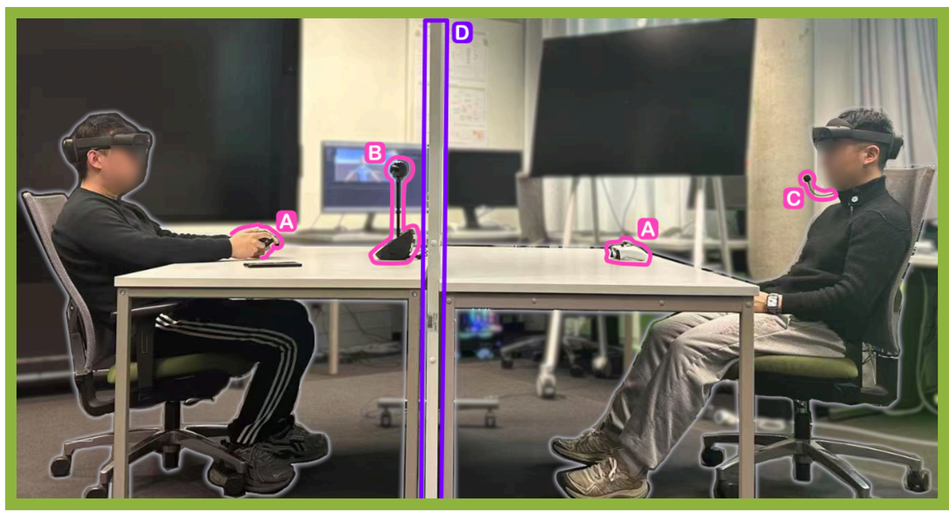


Figure 3: Setup

Participants sat across from each other, divided by a whiteboard.

A: Controllers
B: Web-Cam
C: Microphone
D: Divider

Figure 4: Participant Perspective

Participants could see their conversation partners' 3D avatar through their HMD.

During the study, two participants wearing **AR HMDs** sat face-to-face at a table, divided by a whiteboard. They **could not see each** other's physical **bodies** and instead could **only see** each other's virtual **avatars** in front of them.

We used a **within-subject** design with **two conditions**:

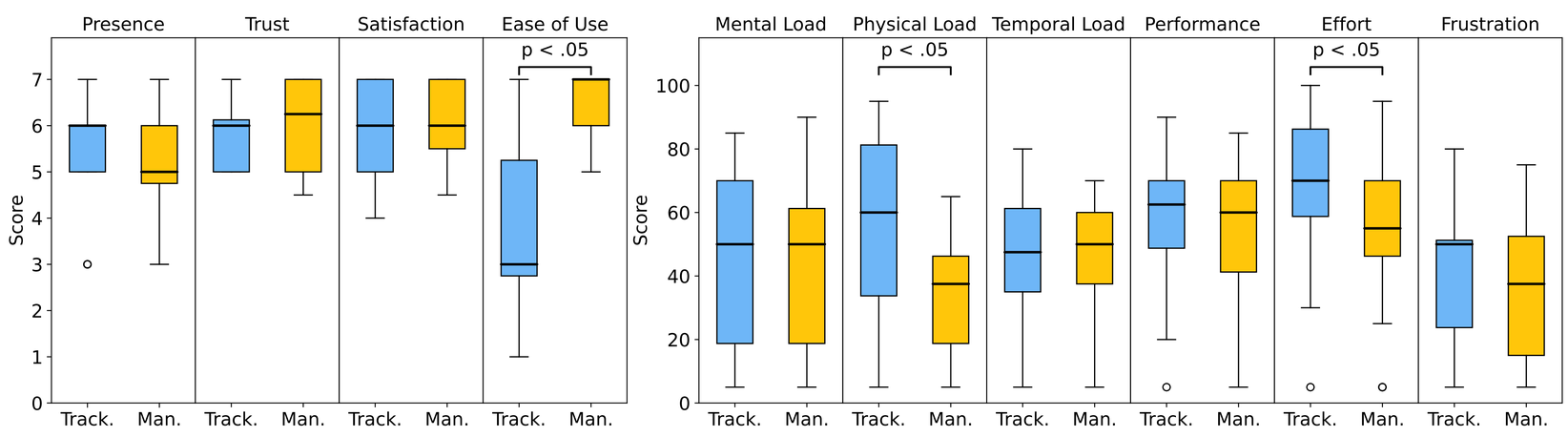
- **Real-time webcam tracking** and
- **Manually triggered preset expressions**.

The **task** was a **dialogue-based** adaptation of *"Who am I?"*, where guessers identified their hidden character through **non-verbal avatar cues**. Participants alternated between the roles of guesser and answerer.

We measured **social interaction** (social presence, interpersonal trust, communication satisfaction) and **usability** (task load, preference, ease of use).

Results

Social interaction remained unaffected —
but **manual presets were clearly preferred**, being **significantly easier and less effortful** than unstable tracking.



Contact Information

Katja Krug *katjakrug@acm.org* **Xiaoli Song** *xiaoli.song@mailbox.tu-dresden.de* **Wolfgang Büschel** *bueschel@acm.org*

TECHNISCHE UNIVERSITÄT DRESDEN
INTERACTIVE MEDIA LAB DRESDEN
Universität Stuttgart
VISUS VISUALISIERUNGSINSTITUT DER UNIVERSITÄT STUTTGART
CeTI Centre for Tactile Internet with Human-in-the-Loop