

## ***Supplementary Material***

The supplementary material provides an overview of the supplementary video content in section 1, detailed survey information in section 2, and additional setup details in section 3. Relevant methodology and data will be made available for research purposes.

### **1 SUPPLEMENTARY VIDEO**

The supplementary video demonstrates the generated gestures and interactions, including:

1. An overview of state-of-the-art VR-based interaction and gesture generation using generative models.
2. Example interactions illustrating the four methods, each represented under one example condition: HEA, NEA, and DV.
3. A comparison of gesture quality across all four methods, focusing exclusively on the HEA condition.
4. A side-by-side comparison of a representative method under both HEA and NEA conditions, highlighting the differences between these two emotions.
5. A reconstruction sequence generated from a real human driving video.

### **2 SURVEY**

#### **2.1 Survey Questionnaire**

The full study took approximately one hour per participant, though the exact duration varied depending on individual interaction times. At the beginning of each session, participants completed an example interaction to become familiar with the environment. The complete questionnaire is provided in table S1.

#### **2.2 Participant Qualitative Feedback**

Below, we present selected qualitative responses from participants, denoting EMAGE, TalkSHOW, PIXIE+DECA, and AMUSE+FaceFormer as M1, M2, M3, and M4 respectively, and participants as P.

- **P6 (M1):** “High sound volume. Much better experience when you can feel the intensity of interaction.”
- **P19 (M4):** “The movement made the interaction feel more alive, with gestures being more prominent.”
- **P31 (M1):** “Way too much energy, looked a bit like they were over-exaggerating.”
- **P2 (M3):** “Both the facial and body expression is more natural this time, and more aligned with the sound.”
- **P17 (M4):** “The movement of the avatar seemed to match the conversation very well.”
- **P12 (M2):** “Movements were sluggish and unsync with voice cues, which felt odd.”
- **P48 (M3):** “The last series of videos had more distinguishable gestures, which kept it engaging.”
- **P9 (M1):** “Movements were sluggish and unsync with voice cues, which felt odd.”
- **P10 (M2):** “I could only perceive the movements of the mouth clearly; facial gestures felt static.”
- **P16 (M1):** “Eyes were very static. Lips didn’t seem to move naturally.”
- **P30 (M3):** “Movement with mouth was realistic, but the expressions still felt stiff.”
- **P42 (M1):** “Strange posture and eye movement, but nice hand gestures.”

Participants provided mixed feedback regarding arousal, with some appreciating the increased intensity of gestures (P6, M1; P19, M4), while others found the energy excessive (P31, M1). For animation naturalness, improved synchronization was acknowledged (P2, M3; P17, M4), although sluggish or unsynchronized movements were also noted (P12, M2). Feedback on diversity was positive, with P48 (M3) highlighting that varied animations kept interactions engaging. However, participants raised concerns about facial movements, citing static or unsynchronized gestures (P9, M1; P10, M2) and stiffness in expressions (P16, M1; P30, M3), despite generally favorable responses to hand gestures (P42, M1). Overall, body gestures appeared more natural than facial animations, which require further refinement.

### 3 EXPERIMENTAL SETUP DETAILS

#### 3.1 SMPL-X Model

Table S2 and fig. S1 (left) illustrate the SMPL-X joints and their corresponding indices. The mesh typically represents a person with a height of 1.7 m, weight of 60 kg, and 55 joints (J).

#### 3.2 Lighting and Setup in Blender

The outdoor scene in fig. S1 (right) features a patio with seating, greenery, and a pool. The lighting simulates an evening ambiance, enhanced by a fire source near the sofa. The Blender lighting setup uses a High Dynamic Range Image (HDRI) for ambient illumination, while point lights illuminate the fire area. The camera-to-chair distance was consistently maintained at 1.5 meters, ensuring uniformity across sessions. The only variations arose from participants' torso heights and preferred viewing angles.

**Table S1. Questions for Evaluating Perceptual Study Across VR Conditions**

Condition	Questions
Pre	Did you play any video game in the past 12 months? How was your experience with virtual environments? Write positive and negative experiences for these interactions.
HEA	I felt like I was interacting with a real person. Their facial expressions looked natural/realistic. Their body movements looked natural/realistic. Rate your enjoyment during the interaction. Was the interaction warm and comfortable? What was your perceived emotion of the avatar's animation?
NEA	I felt like I was interacting with a real person. Their facial expressions looked natural/realistic. Their body movements looked natural/realistic. Rate your enjoyment during the interaction. Was the interaction warm and comfortable? What was your perceived emotion of the avatar's animation?
DV	Do you perceive diversity in the two animations presented?
Post	Did you feel the strongest sense of closeness to your conversational partner? Did the interaction with the avatar feel mostly like a real person? The avatar's facial expressions looked the most natural/realistic. The avatar's body movements looked the most natural/realistic. Provide qualitative feedback about your overall VR interaction experience.



Figure S1: Left: **SMPL-X joints**. Right: **Blender render of an outdoor scene**.

Table S2: **SMPL-X Joint Names with Indices**.

Index	Joint Name
0	Pelvis
1	Left Hip
2	Right Hip
3	Spine1
4	Left Knee
5	Right Knee
6	Spine2
7	Left Ankle
8	Right Ankle
9	Spine3
10	Left Foot
11	Right Foot
12	Neck
13	Left Collar
14	Right Collar
15	Head
16	Left Shoulder
17	Right Shoulder

Index	Joint Name
18	Left Elbow
19	Right Elbow
20	Left Wrist
21	Right Wrist
22	Jaw
23	Left Eye
24	Right Eye
25-27	Left Index1-3
28-30	Left Middle1-3
31-33	Left Pinky1-3
34-36	Left Ring1-3
37-39	Left Thumb1-3
40-42	Right Index1-3
43-45	Right Middle1-3
46-48	Right Pinky1-3
49-51	Right Ring1-3
52-54	Right Thumb1-3