Decoding the neural representations of digital humans' emotional faces in stereo- versus monoscopic viewing conditions – a study plan

Ammara Nasim[1,2], Jeroen de Mooij[3], Simon M. Hofmann[1,4], Vadim Nikulin[1], Werner Sommer[5,6,7], Arno Villringer[1,5], Michael Gaebler[1,5], Felix Klotzsche[1,5,6]

MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES

1. Max Planck Institute for Human Cognitive and Brain Sciences, Department of Neurology, Leipzig, Germany
2. Carl von Ossietzky Universität Oldenburg, Germany
3. Thefirstfloor.nl, Rotterdam, Netherlands
4. Fraunhofer Institute Heinrich Hertz, Department of Artificial Intelligence, Berlin, Germany
5. Humboldt-Universität zu Berlin, Berlin School of Mind and Brain, Germany
6. Humboldt-Universität zu Berlin, Department of Psychology, Germany
7. Zhejiang Normal University, Department of Psychology, Jinhua, China

nasim@cbs.mpg.de
klotzsche@cbs.mpg.de

Introduction

Emotional facial expressions are a cornerstone of social communication. **Virtual Reality** (VR) technology and **digital humans** offer new possibilities to study these processes in more lifelike yet highly controlled settings. One core feature of VR is that it allows for **stereoscopic** viewing conditions. Stereoscopic information can facilitate face recognition [1] and assists the construction of robust facial representations in memory [2]. Such enriched mental representations could be due to additional visual features in stereoscopic viewing conditions or because of increased relevance of more lifelike stimuli.

Combining EEG with multivariate decoding has been suggested as a sensitive approach to study the neurophysiological underpinnings of face processing [3].

If enriched mental representations manifest in better separable neurophysiological activation patterns (e.g., due to a higher number of represented features), this could be reflected in a higher decoding performance in the according experimental conditions.

Main Hypotheses

H1: The emotional expressions of **computer-generated human faces** can be **decoded** from the EEG of an observer in a virtual reality setup.

H2: The **decoding performance** is **significantly higher** under **stereoscopic** as compared to monoscopic viewing conditions.

References

Please visit this link for a list of the cited resources:

SCAN ME

https://tinyurl.com/VRstereofem-Poster-MBBS

Experimental Paradigm

- **Participants:**
  N = **30** (all females, 18-40 years)
  Tested for **stereopsis (stereo fly test)**
- **720 trials** in 6 blocks

Stereo Fly test

- **Stimuli:** Different renderings of digital humans' emotional facial expressions in mono- and stereoscopic viewing conditions (Figure 1)
- **Facial emotion recognition task** (Figure 2)
- Control measure: **Subjective rating** of the perceived **intensity** of the stimulus' facial emotion (Self-Assessment Manikins)

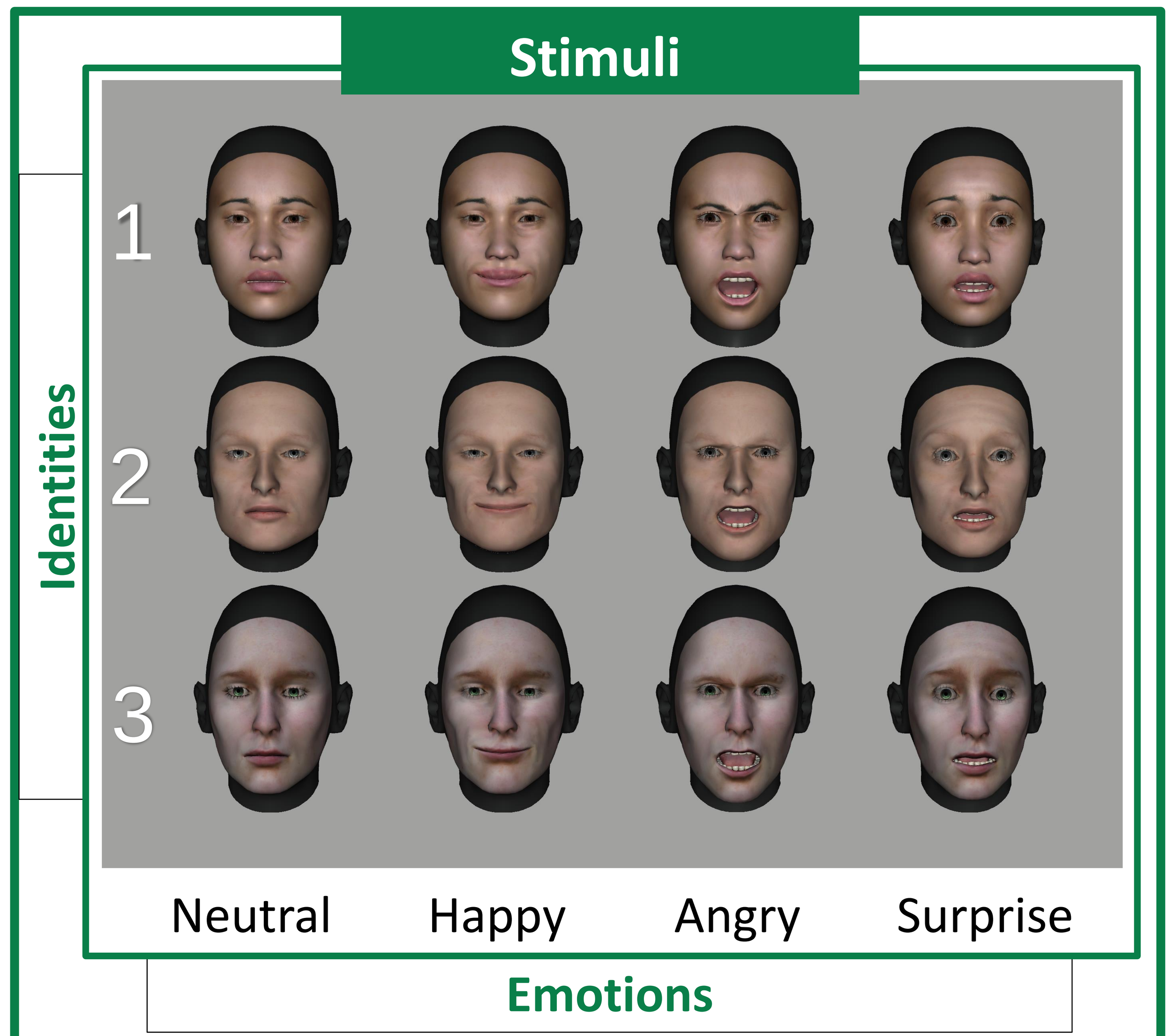


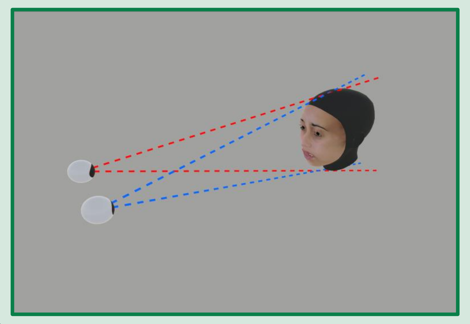
Figure 1: Manipulated features of the stimulus

Main Manipulation

Both viewing conditions are presented in a VR headset

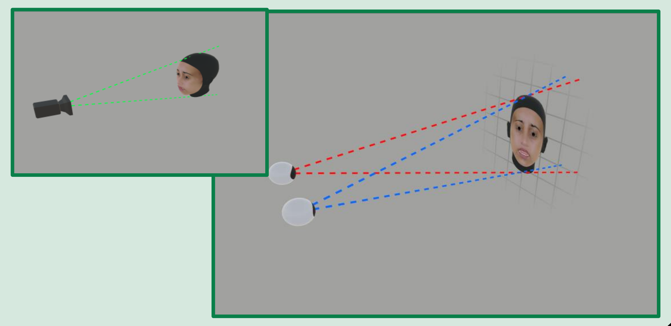
Stereoscopic

Two (virtual) cameras (one per eye) record a 3D object with slightly different perspectives. This creates the impression of spatial depth.



Monoscopic

The 3D object is captured with a single (virtual) camera and the resulting 2D image is presented as a plane in the virtual environment (comparable to showing a picture on a screen). The visual input to both eyes is therefore identical.



- Stimuli were generated using the FACSHuman [4] plugin for MakeHuman [5].
- Emotions are computer generated according to the FACS [6] with 100% intensity and have recently been validated [4].
- Total of **three identities** and **four** facial **expressions (neutral, happy, angry and surprise)**
- To ensure central presentation of the stimuli, the content is fixed to the participant's viewpoint.
- Stimuli are shown in a random sequence with all experimental conditions interleaved.

Experimental Setup

Hardware:
- **HTC Vive Pro Eye headset** (rf: 90Hz) incl. a
- **Tobii eye-tracker** (120Hz)
- 60 active **EEG electrodes** (10/20)
- BrainProducts LiveAmp (500Hz)

Software:
- Unity 3D (v2020.3.3)
- FACSHuman plugin [4]
- MakeHuman (v1.2.0) [5]
- Unity Experiment Framework (v2.3.4) [6]
- Lab Streaming Layer (LSL)

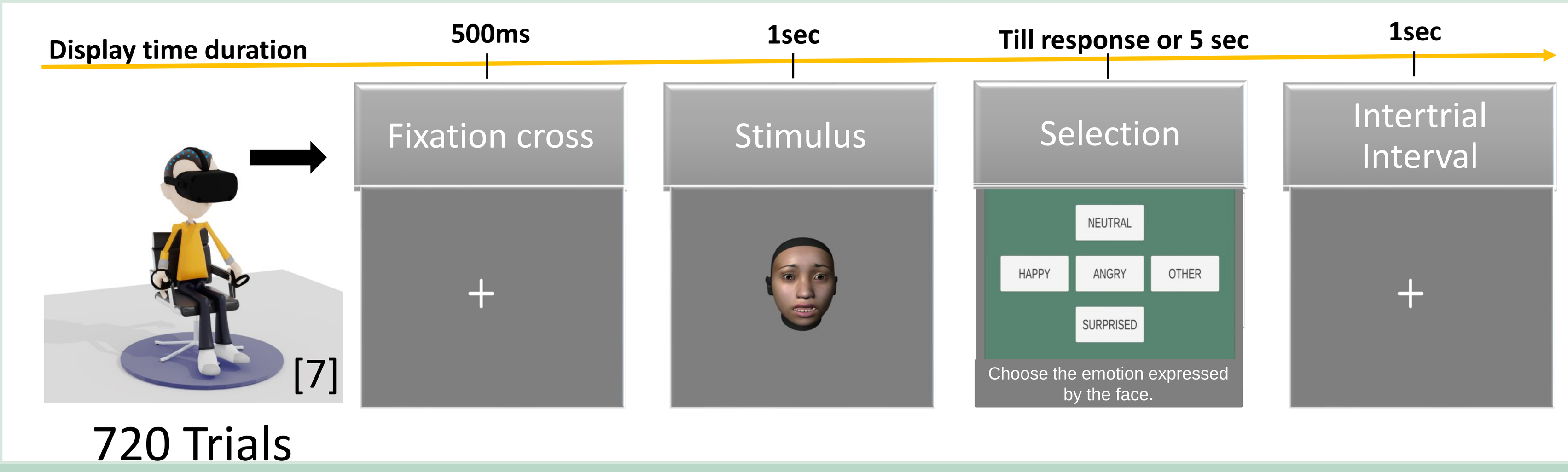


Figure 2: Schematic diagram of experimental paradigm

Analysis

Data analysis will be implemented using MNE-Python (v24.0) [8] and R (v4.0.0) [9].

Raw data

Filtering
- Band pass filter: 0.1-40Hz

Epoch
- Relative to stimulus onset
- -0.3-1.0s

ICA
- Filter (1-40Hz) + epoch a copy of the data
- Reject bad epochs (autoreject [10])
- Picard algorithm [11]
- Reject components correlated with EOG
- Apply ICA weights to regular epochs

Baseline Correction
- Mean activity of 0.2s pre-stimulus

Data cleaning
- Using autoreject (local) [10]
- Locally interpolate bad channels
- Reject bad epochs

Decoding (single subject level)
- Logistic regression (multiclass, one-vs-one)
- Decoding target: emotional expression
- L2 regularization
- Sliding window (20ms, no overlap)
- Features:
  - Mean activation per channel
  - Mini-batches: average of 3 trials
- 3-fold cross-validation (20 repetitions)

Statistical analysis (group level)
- Time windows of interest:
  - 0.08-0.12s (P1)
  - 0.14-0.20s (N170)
  - 0.15-0.35s (EPN)
  - 0.40-0.60s (LPC)
- Hypothesis 1:
  t-test mean classification performance against empirical chance level
  (null distribution: rep. permute labels)
- Hypothesis 2:
  Compare (*t*-test) mean classification performance in mono- vs. stereoscopic viewing conditions

Potential Challenge: Eye Movements

- Saccades lead to fixation-related EEG potentials [12]
- Systematic differences in saccade statistics between the exp. conditions might influence the classifier performance

- **Control measures:**
  - ❖ Compare saccade statistics
  - ❖ Run classifier on eye-tracking data & compare results
  - ❖ Isolate saccade-related components via deconvolution [13]