# Real or fake? Decoding realness levels of stylized face images with EEG

Yonghao Chen[1], Tilman Stephani[1], Milena Teresa Bagdasarian[2,3], Anna Hilsmann[3,4], Peter Eisert[3,4], Arno Villringer[1,5], Sebastian Bosse[2,3], Michael Gaebler[1,6], Vadim Nikulin[1].

1. Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany | 2. Department of Video Coding and Analytics, Fraunhofer HHI, Berlin | 3. Department of Vision and Imaging Technologies, Fraunhofer HHI, Berlin

4. Visual Computing Group, Humboldt-Universität zu Berlin, Berlin | 5. Clinic of Cognitive Neurology, University Hospital Leipzig, Leipzig, Germany, Berlin | 6. MindBrainBody Institute at the Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany

cheny@cbs.mpg.de

MAX PLANCK INSTITUTE FOR HUMAN COGNITIVE AND BRAIN SCIENCES

## Introduction

The **uncanny valley effect** may cause people to perceive highly, yet not perfectly, human-like faces as eerie, challenging techniques of generation of virtual faces. However, the neurocognitive understanding of the uncanny valley effect and **how human brain perceiving realness** remains elusive.

A study suggested that the amplitudes of N170 components were modulated by face-realism, and they found a **U-shaped modulation effect** between the realness and the amplitudes of N170 components, with the largest neural responses for most abstract and most realistic images [1].

Based on the same stimuli materials used in [1], a recent study chose **steady-state visual evoked potential (SSVEP)** as the neural marker of perceiving the realness of rendered faces [2]. To improve this study further, we reanalyzed this dataset using state-of-the-art EEG analysis techniques.

This study provides a basis for future research and benchmarking of real-time detection of face realness regarding three aspects: **SSVEP-based modulation effects of stylized images, efficient classification methods, and low-level stimulus confounders that need to be controlled.**

## Methods

**Stimulus materials:** In total, 36 face images were generated with six levels of stylization (R0-R5), two genders (male and female), and three emotions (neutral, happy, angry). Those images were presented through the SSVEP paradigm, which alternated images and backgrounds. (Fig 1).

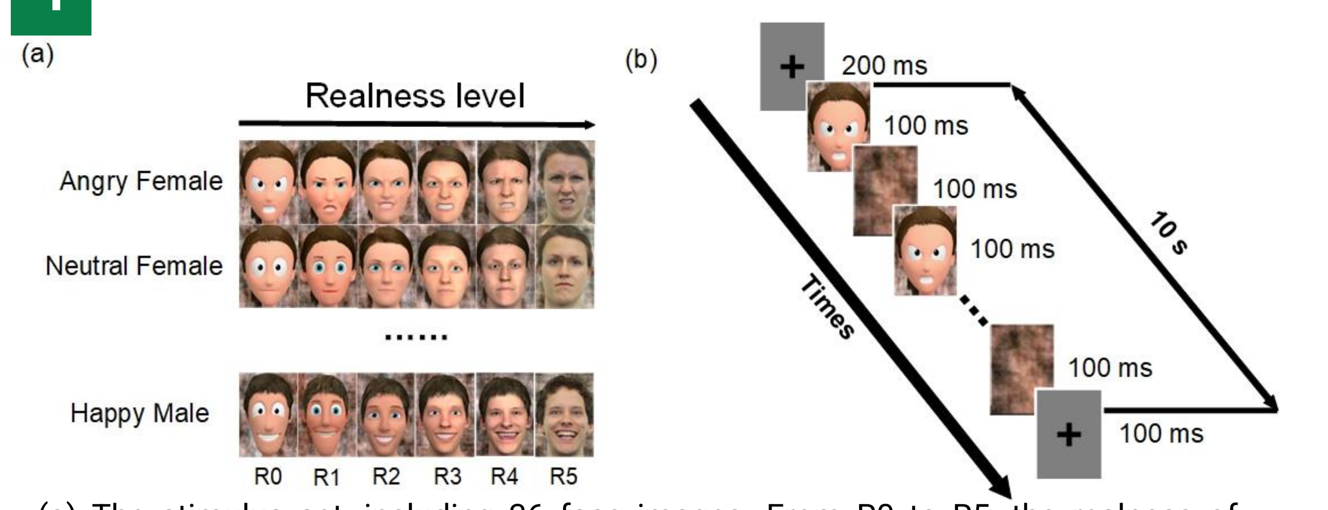
**EEG dataset:** Ten partcipants' EEG data were recorded, every partcipant's data consisted of eight sessions, each session lasted about seven minutes and comprised 36 trials. (Fig 1)

**Signal processing:** This study reanalyzed the EEG data via cluster-analysis and spatio-spectral decomposition (SSD) [3].

**Classification:** Task-related component analysis (TRCA) [4] was employed in this study to classify stylization levels of stimuli through the EEG responses. (Fig 2)

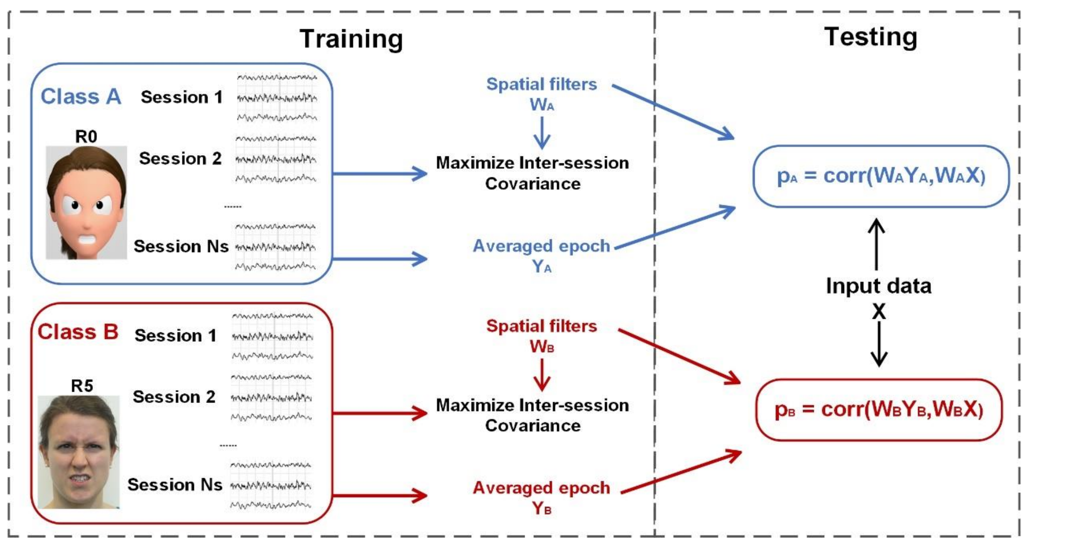
**Stastical analysis:** We used Linear mixed model (LMM) for model regression, and likelihood ratio test (LRT) for comparision of linear and quadratic model.

**1**




(a) The stimulus set, including 36 face images. From R0 to R5, the realness of images increased. (b) The framework of stimulation process.

**2**




Framework of Task-related component analysis (TRCA)-based classification algorithm (for two-class: R0&R5).

## Results

**SSVEP**: We found an **nonlinear relationship** between amplitudes of 5Hz SSVEP components and degrees of realness, the result suggested that the most stylized cartoon images and the real photographs evoked stronger responses than images with medium stylization. (Fig 3)

**ERP-like component**: ERP-like components were acquired by averaging every 200ms transient responses after the stimulus onset of every cycle. Some typical ERP-like components such as P100&N170 can be found. The amplitudes of N170-like components maintained an **nonlinear relationship** with degrees of realness. (Fig 4)

**Quadratic Vs Linear**: We use Linear mixed model (LMM) as the method for model regression:

SSVEP/N170 amplitudes ~ 1 + realness + (1 | sub)
SSVEP/N170 amplitudes ~ 1 + realness + I(realness^2) + (1| sub)

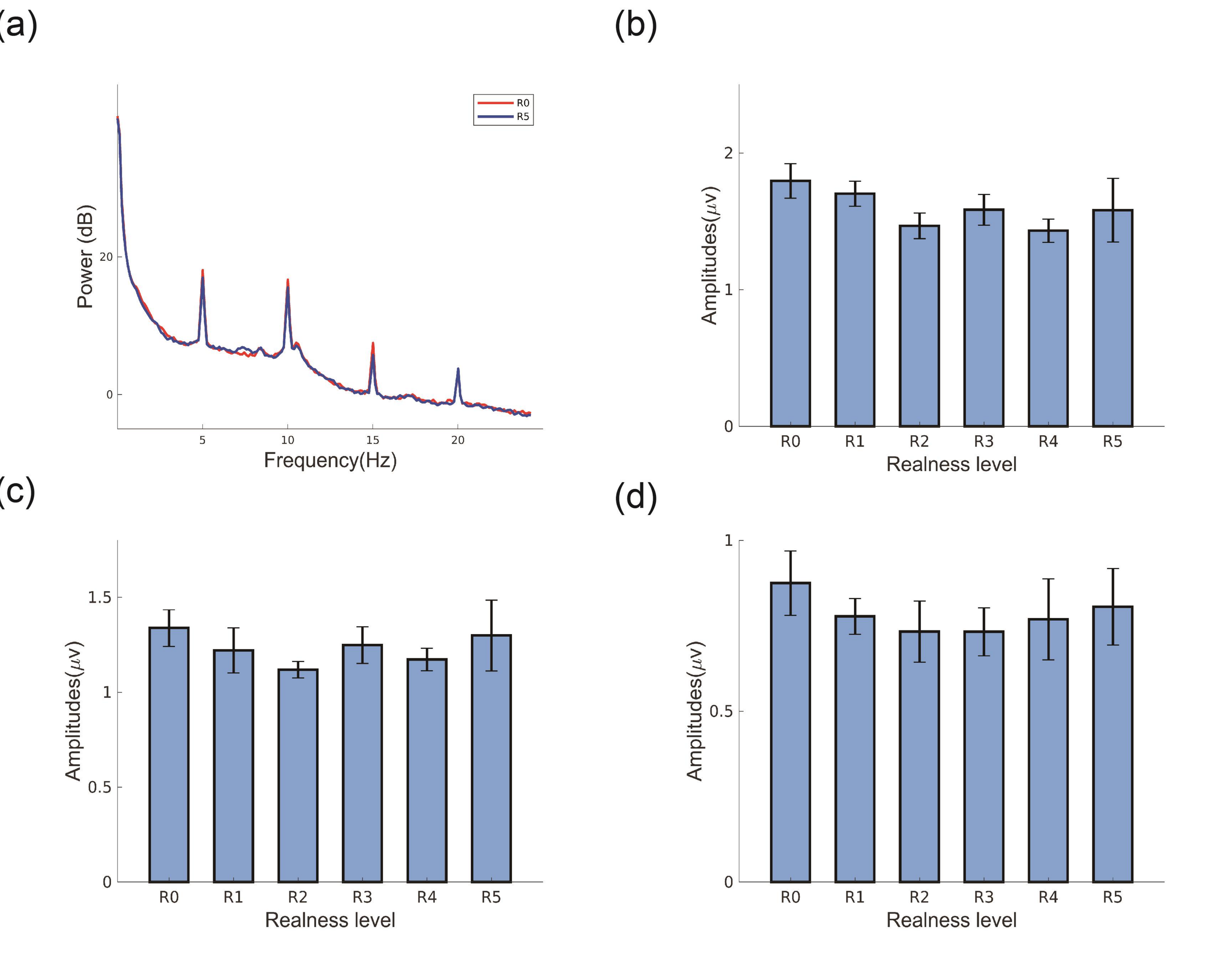
The results suggested that the likelihoods of quadratic models were higher than linear model. ($p$<0.01 for SSVEP amplitudes, $p$<0.001 for N170 amplitudes)

**Topography**: Similar to other face-related studies, neural responses were lateralized at **right hemisphere**, SSVEP components (fundamental and harmonics) mostly localized in primary visual cortex, and ERP-like components mostly localized in the anterior parts of the occipital lobe . (Fig 5)

**Classification**: For the six-class (R0-R5) classification task, when the interested window was 2s, the average classification accuracy across all subjects and all emotion states was 39.48±9.58%. Even for two-class R4&R5 that had close-by stylization levels of high similarity, the classifier can reach 59.29±7.19%. (Fig 6)

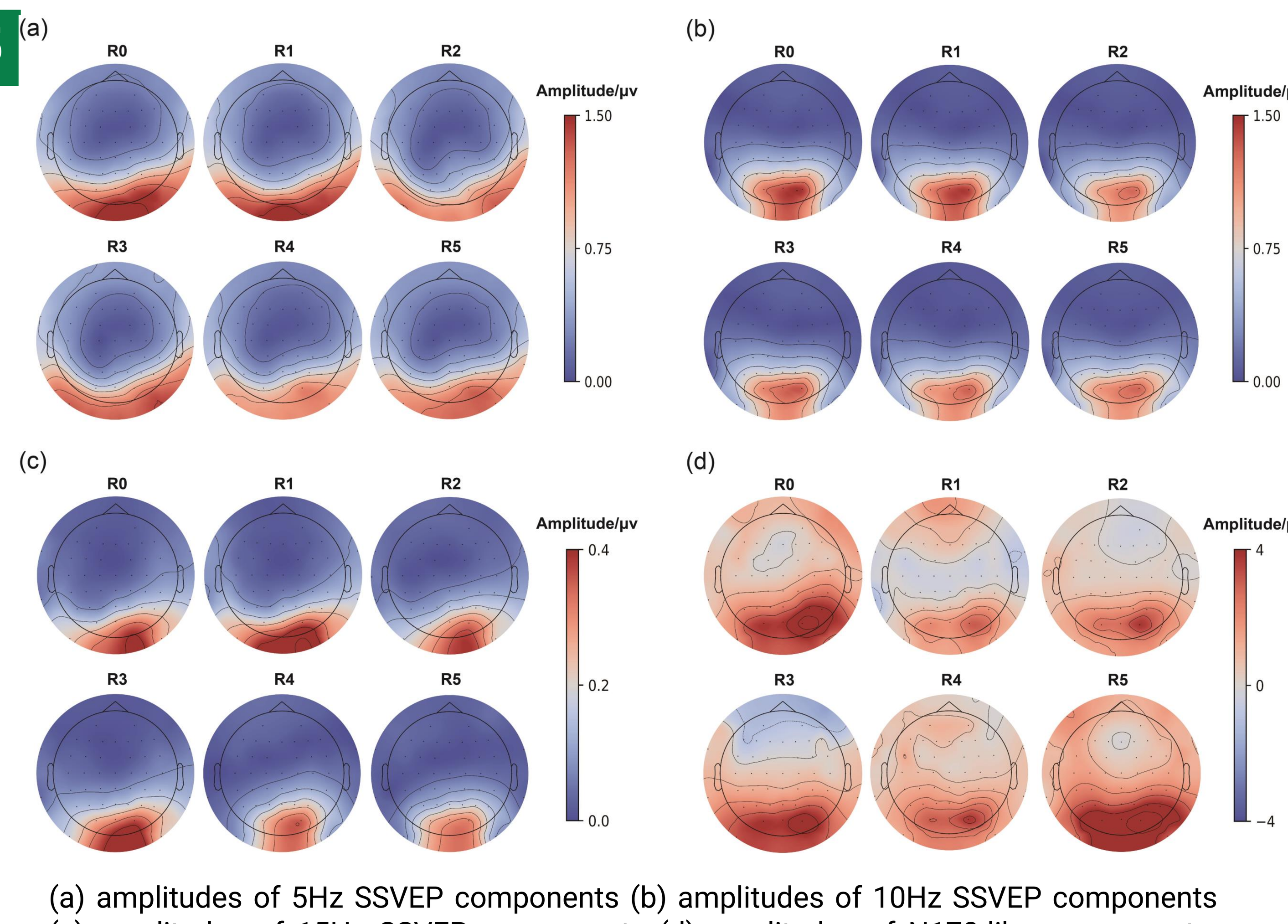
**Confouding varaibles:** We found that eye size have a negative correlation with realness level, thus the eye sizes should be controlled in the future. (Fig 7)

**3**



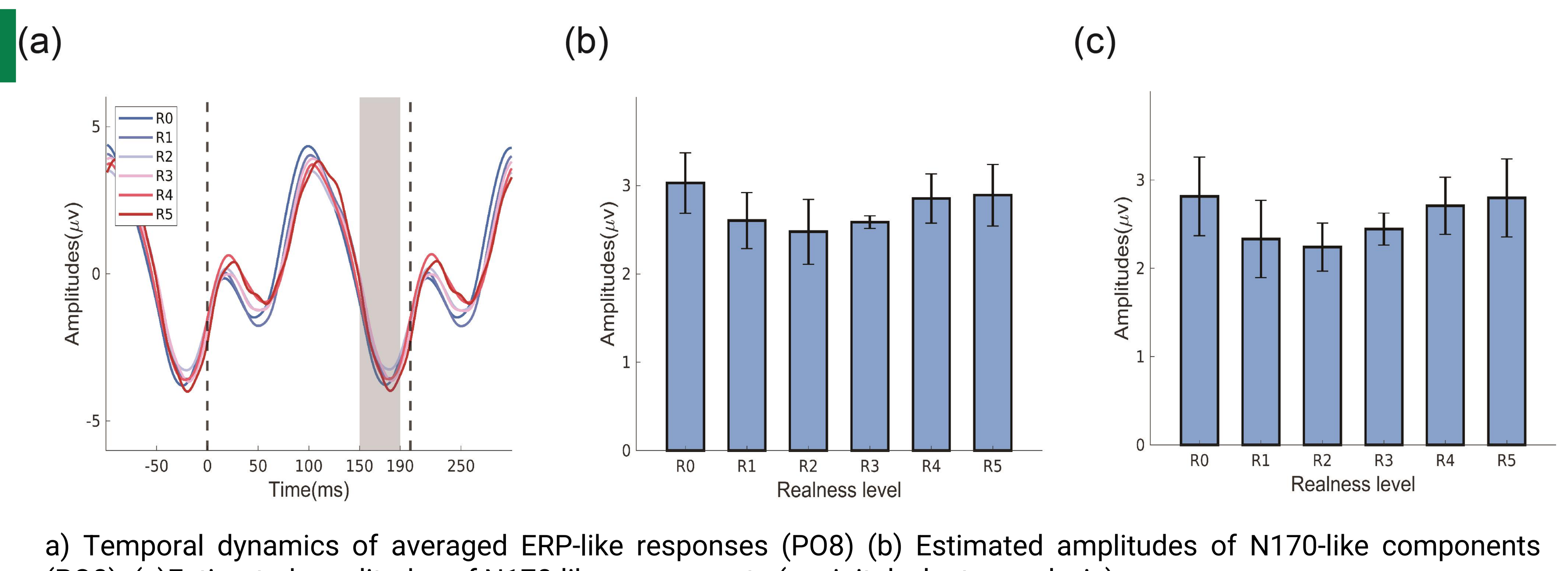

(a) Power spectral of concatenated responses (Oz) (b) FFT amplitudes of 5Hz (Oz electrode) (c) FFT amplitudes of 5Hz (Occipital cluster analysis) (c) FFT amplitudes of 5Hz (after SSD)

**5**



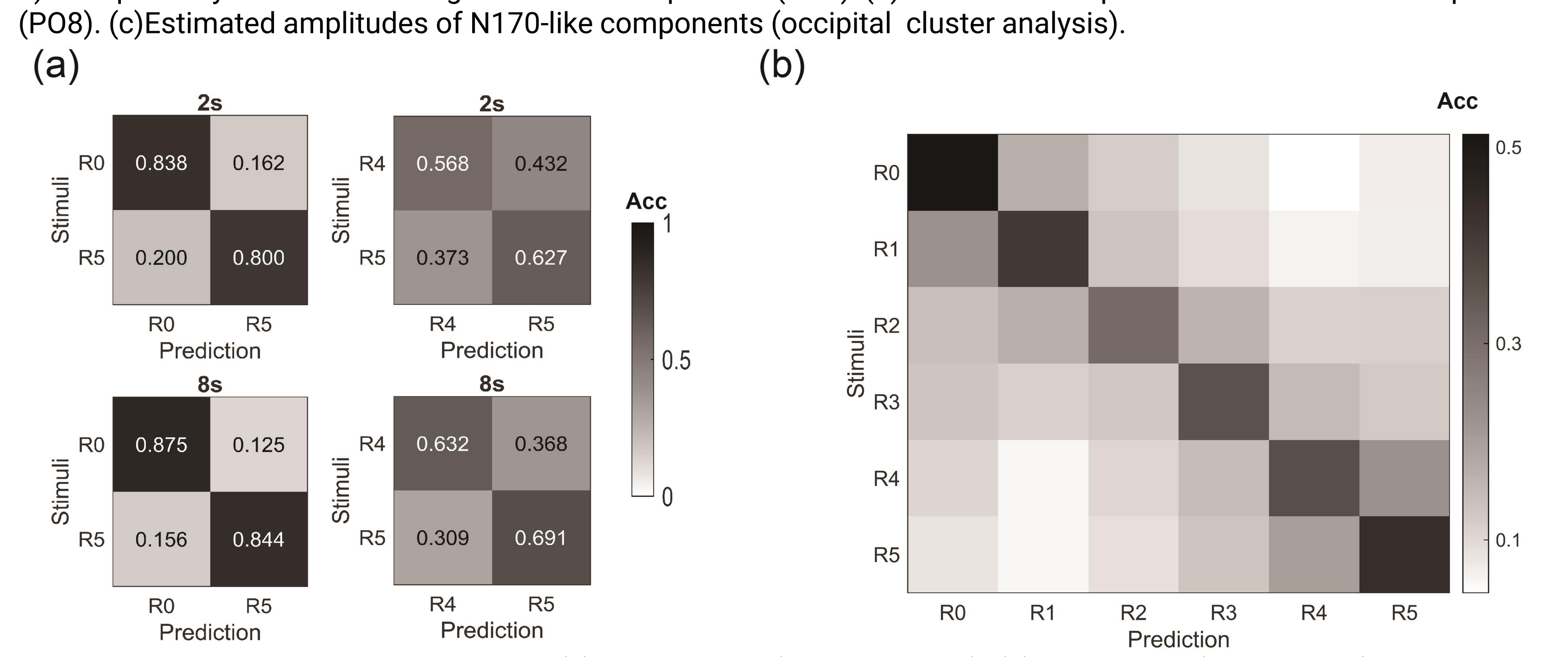

(a) amplitudes of 5Hz SSVEP components (b) amplitudes of 10Hz SSVEP components (c) amplitudes of 15Hz SSVEP components (d) amplitudes of N170-like components. The amplitudes were the averaged result across all participants and all sessions.

**4**



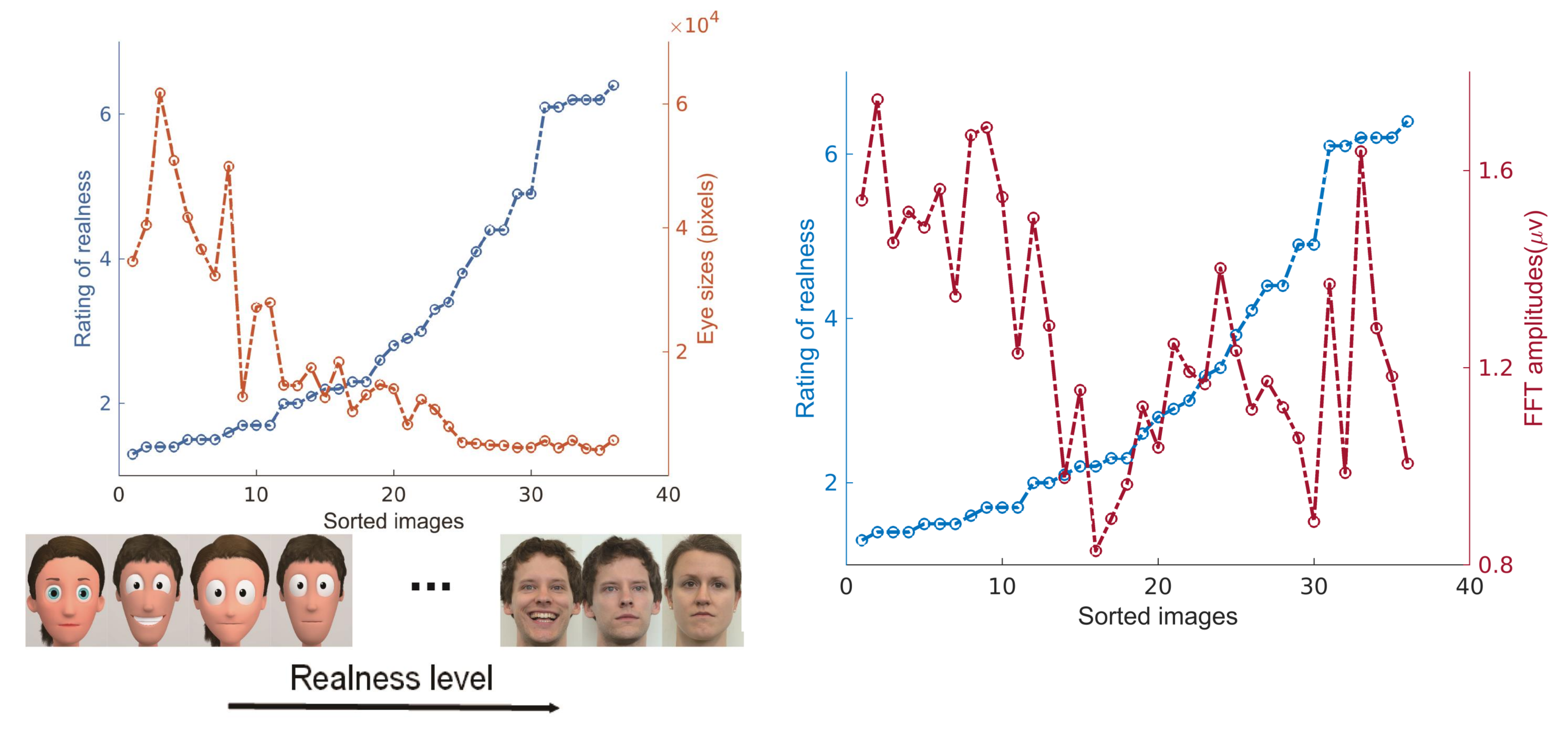

a) Temporal dynamics of averaged ERP-like responses (PO8) (b) Estimated amplitudes of N170-like components (PO8). (c)Estimated amplitudes of N170-like components (occipital cluster analysis).

**6**




Confusion matrix of classification results. (a) Two classes (R0&R5, R4&R5). (b) Six classes (R0 to R5, 2s). The number indicates the classification accuracy.

**7**




Relation between subjective rating of realness/5Hz FFT amplitudes and measured eye size.

## Discussion

The non-linear relation between amplitudes of neural responses (SSVEP&N170) and level of realness might be modulated by the uncanny valley effect and other factors. In our future study, it is worthwhile to disentangle how human brain perceiving realness in a more specific way.

## Reference

[1] S Schindler et al. (2017). *Scientific Reports*
[2]MT Bagdasarian et al. (2020). *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*
[3] VV Nikulin et al. (2011). *Neuroimage*
[4] Nakanishi et al. (2017). *IEEE Trans Biomed Eng (TBME)*