ORIGINAL ARTICLE

Electrophysiological correlates of face and object perception: A comparative analysis of 2D laboratory and virtual reality conditions

Merle Sagehorn[1] | Marike Johnsdorf[1] | Joanna Kisker[1] | Thomas Gruber[1] | Benjamin Schöne[1,2]

[1]Experimental Psychology I, Institute of Psychology, Osnabrück University, Osnabrück, Germany

[2]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence
Merle Sagehorn, Experimental Psychology I, Institute of Psychology, Osnabrück University, Lise-Meitner-Str. 3, Osnabrück 49076, Germany.
Email: merle.sagehorn@uni-osnabrueck.de

Abstract

Human face perception is a specialized visual process with inherent social significance. The neural mechanisms reflecting this intricate cognitive process have evolved in spatially complex and emotionally rich environments. Previous research using VR to transfer an established face perception paradigm to realistic conditions has shown that the functional properties of face-sensitive neural correlates typically observed in the laboratory are attenuated outside the original modality. The present study builds on these results by comparing the perception of persons and objects under conventional laboratory (PC) and realistic conditions in VR. Adhering to established paradigms, the PC- and VR modalities both featured images of persons and cars alongside standard control images. To investigate the individual stages of realistic face processing, response times, the typical face-sensitive N170 component, and relevant subsequent components (L1, L2; pre-, post-response) were analyzed within and between modalities. The between-modality comparison of response times and component latencies revealed generally faster processing under realistic conditions. However, the obtained N170 latency and amplitude differences showed reduced discriminative capacity under realistic conditions during this early stage. These findings suggest that the effects commonly observed in the lab are specific to monitor-based presentations. Analyses of later and response-locked components showed specific neural mechanisms for identification and evaluation are employed when perceiving the stimuli under realistic conditions, reflected in discernible amplitude differences in response to faces and objects beyond the basic perceptual features. Conversely, the results do not provide evidence for comparable stimulus-specific perceptual processing pathways when viewing pictures of the stimuli under conventional laboratory conditions.

KEYWORDS

EEG, ERP, face perception, late components, N170, realistic conditions

1 | INTRODUCTION

Faces are a constant presence in daily life and essential components of the visual experience. Despite their universal familiarity, faces are processed as intricate visual stimuli, reflected by complex neural mechanisms that evolved through real-life interactions, providing vital social cues for evaluation in various contexts.

In conventional laboratory research, this complex phenomenon is simplified by deconstructing it into accessible components. Researchers manipulate individual variables, maintaining experimental control, and often use two-dimensional screens to present facial stimuli, reducing them to their basic attributes (e.g., Blau et al., 2007; Civile et al., 2018; Dering et al., 2011; Jemel et al., 2003; Rossion, 2014a, 2014b). However, this reductionist approach removes the inherent complexity and natural context, focusing on isolated neuronal processes in a controlled laboratory setting, despite face perception's real-world integration. The same limitations extend to common control images used in face perception research, with cars being a prominent category (e.g., Boehm et al., 2011; Dering et al., 2009; Kloth et al., 2013; Kuefner et al., 2010; Ratcliff et al., 2009; Thierry et al., 2007). Whereas cars are chosen for their real-world relevance and familiarity (Rossion & Jacques, 2011), displaying them alongside facial stimuli of equal size on a 2D screen does not accurately represent their real-life proportions. In reality, perceiving a car or encountering a person walking by evokes a distinct response that differs from merely viewing images on a computer screen.

However, investigating the psychophysiological correlates underlying face perception, often assessed through measures like response times and electroencephalography (EEG), requires the strict experimental control provided by the reductionist approach, ensuring reproducibility and further development. Creating an equally controlled setup in natural settings resembling our brain's typical face processing environment is considerably more challenging. Consequently, there are limited studies exploring face perception with electrophysiological measures under more realistic conditions, including realistic virtual settings (e.g., Kirasirova et al., 2021; Sagehorn et al., 2023; Stolz et al., 2019), settings with real persons as stimulus material (Myllyneva & Hietanen, 2015; Pönkänen et al., 2011), and actual real-world settings (Johnston et al., 2015). Investigating face perception in these settings significantly enhances our understanding of the cognitive mechanisms and neural correlates involved in processing real faces within their natural contexts, as opposed to the two-dimensional images on screens. Yet, conducting research with live stimulus material or in real-world settings (Johnston et al., 2015; Myllyneva & Hietanen, 2015;

Pönkänen et al., 2011) presents methodological challenges, including the need for resources like time, funding, and personnel to meet strict requirements (e.g., trial sequencing and timing). For studies on real-life face perception, obtaining human stimuli and creating standardized emotional expressions for all participants requires additional efforts.

Virtual reality (VR) presents a valuable solution to bridge the gap between laboratory settings and real-life cognition to gain a comprehensive understanding of the functional properties of cognitive and emotional processes in real life. VR allows to present diverse stimuli in a conceptually plausible context and at the same time maintain high experimental control. It enhances the naturalism of conventional laboratory designs by presenting stimuli in real-world dimensions with depth and spatial proximity, yielding behavioral and neural responses akin to real-life scenarios (Blascovich et al., 2002; Gabana et al., 2017; Gromer et al., 2018; Kisker, Gruber, & Schöne, 2021; Kisker, Lange, et al., 2021; Newman et al., 2022; Rubo & Gamer, 2018; Schöne et al., 2019; Schöne et al., 2023; Schöne, Kisker, et al., 2021; Van Den Oever et al., 2022; Xu et al., 2021). Despite the fact that participants are aware of the virtual nature of their environment and stimuli (e.g., avatars), a potential bias similar to artificial environment awareness in laboratory studies (McCambridge et al., 2014), the immersive VR experience allows for individuals' faces to be perceived as close to lifelike.

In our prior study (Sagehorn et al., 2023), we used 3D-360° photographs of real persons and standard control images (i.e., blurred and scrambled images; see, e.g., Bombari et al., 2013; Civile et al., 2018; Herrmann et al., 2005; Kuefner et al., 2010; Rossion, 2014a, 2014b; Schwaninger et al., 2009) in VR, creating a photorealistic virtual environment in which participants encountered the stimuli in life-sized proportions and in close proximity to themselves, while simultaneously recording EEG. The study revealed distinct neural mechanisms for real-life face perception, challenging the generalizability of results from conventional laboratory paradigms. This trade-off between experimental control and ecological validity is a well-known issue in laboratory research (Parsons, 2015; Shamay-Tsoory & Mendelsohn, 2019).

Further evidence that different modalities yield altered cognitive-affective processes comes from previous studies comparing cognitive and emotional processes in conventional laboratory settings with realistic VR conditions, highlighting the limited ecological validity of laboratory methods (Johnsdorf et al., 2023; Kisker, Gruber, & Schöne, 2021; Pan & Hamilton, 2018; Parsons, 2015; Schöne, Sylvester, et al., 2021; Snow & Culham, 2021). Attentional, motivational, and memory processes operate differently in realistic compared to laboratory

conditions (Johnsdorf et al., 2023; Kisker, Gruber, & Schöne, 2021; Kisker, Lange, et al., 2021; Schöne, Kisker, et al., 2021; Schöne, Sylvester, et al., 2021). The transfer of established paradigms to realistic VR conditions not only changes the extent of the effects but also their morphology.

Electrophysiological research on face perception has primarily been focused on the event-related potential (ERP) component N170, a negative signal deflection at approx. 170 ms post-stimulus at occipito-temporal electrodes (i.e., Itier & Taylor, 2004; Kloth et al., 2013; Ratcliff et al., 2009; Rossion et al., 2003). In conventional laboratory studies, the N170 showed stronger amplitude deflections and shorter latencies in response to facial stimuli compared to objects and other non-facial stimuli (e.g., Churches et al., 2009; Goffaux et al., 2003; Itier & Taylor, 2004; Kuefner et al., 2010). However, previous research has pointed to inconsistencies in the results concerning the N170, with studies reporting contradictory effects or no significant amplitude difference at all (Dering et al., 2009; Kloth et al., 2013). Factors such as stimulus expertise (Bukach et al., 2006; Ip et al., 2017) and perceptual variability within the stimulus category (Boehm et al., 2011; Dering et al., 2009; Thierry et al., 2007) have been proposed to account for these discrepancies, emphasizing the need for cautious interpretation of the N170 as a neural correlate of comprehensive face perception. Furthermore, there is increasing recognition that complex processes like face perception cannot be fully captured by early exogenous ERPs. The N170 primarily captures stimulus-specific information (Berchicci et al., 2016), whereas later components such as the early posterior negativity (EPN: ≈200–400 ms at posterior electrodes) or the late positive potential (LPP: ≈400–1000 ms at contro-parietal electrodes) (e.g., Bublatzky et al., 2014; Herbert et al., 2013; Schindler et al., 2017; Wheatley et al., 2011) provide insights into more comprehensive stimulus processing (Nasr, 2010; Ratcliff et al., 2009; Zion-Golumbic & Bentin, 2007), influenced by social relevance, emotions, and decision-relevant factors (Bublatzky et al., 2014; Herbert et al., 2013; Sollfrank et al., 2021; Stolz et al., 2019). These insights can be further strengthened by locking the ERPs to participant responses (Nasr, 2010), as later components related to the processing of perceptual decisions are more strongly associated with the response than the stimulus onset (Berchicci et al., 2016). Thus, advanced face processing leading to personalized and contextually relevant representations is more accurately reflected in the later and response-locked components.

The findings from the EEG-VR study by Sagehorn et al. (2023) support the re-evaluation of the N170 as an insufficient neural marker for face perception and the potential contribution of later components, which also applies to realistic virtual conditions. Under laboratory conditions, the N170 discriminated faces and blurred faces, whereas it did not under realistic conditions in VR. The study further revealed greater discriminative power concerning faces and controls for potentials subsequent to the N170, reflected in distinct topographic distributions showing that unique neural mechanisms are applied for face compared to silhouette and object perception (Sagehorn et al., 2023). If minor changes to the experimental paradigm or the stimulus material cause substantial changes in the observed effects even under laboratory conditions, it remains to be addressed whether these effects are caused by the artificiality of the laboratory setting or whether they are also relevant for the processing of faces in reality.

The current study aims to expand on Sagehorn et al.'s (2023) research on the neural mechanisms employed for the perception of realistic faces as opposed to their monitor representation. Specifically, to investigate the initial perception and classification, conceptual identification and recognition, and post-decisional evaluation of realistic faces, we compared a conventional laboratory setting and a naturalistic setting in a blocked two-modality design (i.e., PC modality via monitor vs. VR modality via headset). In both modalities, we directly contrasted the perception of faces with standard objects and perceptual controls (i.e., cars and blurred images; see, e.g., Boehm et al., 2011; Bombari et al., 2013; Dering et al., 2009; Kuefner et al., 2010; Rossion, 2014a, 2014b; Schwaninger et al., 2009) within the same experimental context. Moreover, to enforce the face identification process and accurately differentiate it from the post-decisional stimulus evaluation process, participants were actively engaged in a standard face–car discrimination task (e.g., Boehm et al., 2011; Ratcliff et al., 2009).

In alignment with standard paradigms investigating face perception, the current study compares response times, N170 latencies, and N170 amplitudes in response to faces, cars, and their respective blurred images under conventional laboratory and realistic conditions. Following the approach of previous electrophysiological studies comparing differences in complex cognitive processing across varying levels of realism (Johnsdorf et al., 2023; Sagehorn et al., 2023), we furthermore aimed to investigate the processing mechanisms for both faces and cars on a more profound conceptual level. More precisely, to further determine how differences in face and object perception beyond basic physical features were reflected by electrophysiological responses within the N170 latency range, the blurred images served as a perceptual baseline before contrasting the stimulus categories within each modality, resulting in modality-specific face–object differences (FODs). As an index for face-specific processing, the

FODs were then compared between conventional laboratory and realistic conditions to determine the functioning of face processing depending on modality.

To investigate subsequent face-specific processing stages applied under realistic conditions, the FODs were also compared for later potentials and response-locked components. Since there are currently no established standards regarding the specific timing and location of these components, the analysis adopted a regional mean analysis approach, specifically examining the regional means of relevant electrode clusters that encompass the entire scalp surface. In line with previous studies that followed a similar exploratory approach (Johnsdorf et al., 2023; Sagehorn et al., 2023), the later components are defined as L1 (200–450 ms) and L2 (600–1350 ms) component complexes broadly corresponding to the established EPN and LPP in terms of their time windows but with a broader electrode distribution.

We expected the behavioral measures and neural correlates to exhibit discernible patterns that are specific to each modality. In particular, considering the heightened realism and thereby closer approximation to the natural environment in which the stimuli occur, we predicted that reaction times and N170 latencies would indicate faster stimulus processing and discrimination between stimulus categories (face vs. car) under realistic conditions. Furthermore, as the discrimination task increases stimulus attention and facilitates the perception and identification process, we hypothesized that the amplitude of the N170 would be noticeably influenced depending on modality. In addition to early perception, we also expected that distinct patterns of conceptual identification and recognition as well as post-decisional stimulus evaluation between modalities would emerge in modality-specific FODs of the late components and response-locked components.

2 | METHOD

The present article constitutes a follow-up study of a previously conducted experiment on face perception under realistic conditions (Sagehorn et al., 2023). The methodological approach is therefore partly equivalent to the preceding experiment.

2.1 | Participants

An a priori power analysis using Power Contour Estimation (Baker et al., 2020) based on the mean difference and within- and between-subject standard deviations from the previous study as well as the trial count of the present study yielded a total sample size of 55 participants to be necessary to find the hypothesized small to medium effects.

To compensate for possible technical issues during EEG acquisition or potential data loss due to poor signal quality, a total of 65 participants were recruited from the student population of Osnabrück University. The pre-experimental screening of all participants included questions about psychological and neurological disorders and regular drug use. Participants who did not meet the inclusion criteria did not participate in the experiment. If vision correction was necessary, participants were only allowed to participate when wearing contact lenses. Wearing glasses was not permitted because they may be uncomfortable beneath the EEG cap and the VR headset mounted over it for an extended period of time. Before recruiting, it was ensured that participants had not been photographed for the stimulus creation (see section 2.2), knew any of the people whose pictures were presented to them, or had already participated in the preceding study on face perception in VR using the same facial stimuli. All participants gave informed written consent and received either partial course credits or 18€ for their participation.

Due to unmet anamnesis criteria ($n=1$), insufficient data quality ($n=4$), or because they aborted the experiment ($n=2$), seven participants had to be excluded from participation. In the end, 58 data sets were selected for data analyses ($M_{age}=22.98$ years, $SD_{age}=4.3$ years, 39 female, 55 right-handed).

2.2 | Stimulus material

The stimulus material consisted of 120 pictures of regular cars and 120 pictures of persons sitting on stools, both with an empty garage yard as background and rendered as 2D and 3D-360° images. All persons had neutral facial expressions. For all cars, any brand logo as well as the license plate were removed. All images were shot with the Insta360Pro VR camera with an 8k resolution. The background image (i.e., the garage yard) was chosen as a neutral scene in which either a person or a car would not appear physically improbable.

Within each stimulus category (face and car), the images were randomized across participants and conditions (PC or VR), yielding 60 person images and 60 car images per condition, to ensure that participants would not see the same person or car twice. Additionally, cars and persons were presented as a blurred version as one of the standard perceptual control stimuli in face perception (Bombari et al., 2013; Kuefner et al., 2010; Rossion, 2014a, 2014b; Schwaninger et al., 2009). Thereby, only the high-frequency content of the stimuli was removed, and color information and perceptual frame were conserved, that is,

stimulus size and shape remained unchanged. The stimuli were thus still identifiable as either a person or a car, but the conceptual relevance was omitted (see Figure 1). By controlling for a perceptual baseline, a comparison of face-specific processing reflected by ERP amplitude differences among stimulus types and between the two modalities can be achieved. In total, each condition included 240 stimuli (60 normal persons, 60 blurred persons, 60 normal cars, and 60 blurred cars). The normal persons and cars and their matching blurred control images were always presented in the same modality, respectively.

In both conditions, participants were instructed to perform a simple face–car discrimination task. Participants were asked to respond by pressing the shoulder buttons on a basic USB gamepad with their index fingers as soon as they identified the image being presented as either a person or a car, irrespective of it being blurred or not. The button presses were alternated between participants (even participant numbers were pressing right for face and left for car, and odd participant numbers vice versa).

Each of the 240 trials of each condition followed a standardized sequencing structure, which was the same for both modalities (PC- and VR condition) (see Figure 1). Preceded by a fixation dot (0.5–0.8 s) that participants were asked to fixate at the beginning of each trial in both modalities, the pictures of all categories were presented for 1.5 s. They were followed by a short visual feedback on the correctness of the participants' response (0.2 s) and an interstimulus interval (ISI; background image without person or car; 1.5 s). To reduce ocular and movement artifacts, the participants were instructed to blink or move only during the ISI, that is, when the garage yard was empty. Each trial lasted between 3.7 and 4 s, yielding a total run time of approximately 18 min per condition.

2.3 | Procedure

All participants completed both, the PC and the VR condition, and the order of both conditions was alternated between participants. Both conditions were conducted in the same soundproof and electrically shielded room suitable for EEG measurements. Due to the sensitivity of EEG data to motion-induced artifacts, participants were asked to keep motion to a minimum and refrain from looking around, especially in the VR environment. Between conditions, they were given a 5-min break to relax while the EEG signal's quality was checked.

For the PC condition, participants were seated in front of a standard PC monitor (24″, 1920×1200 resolution)
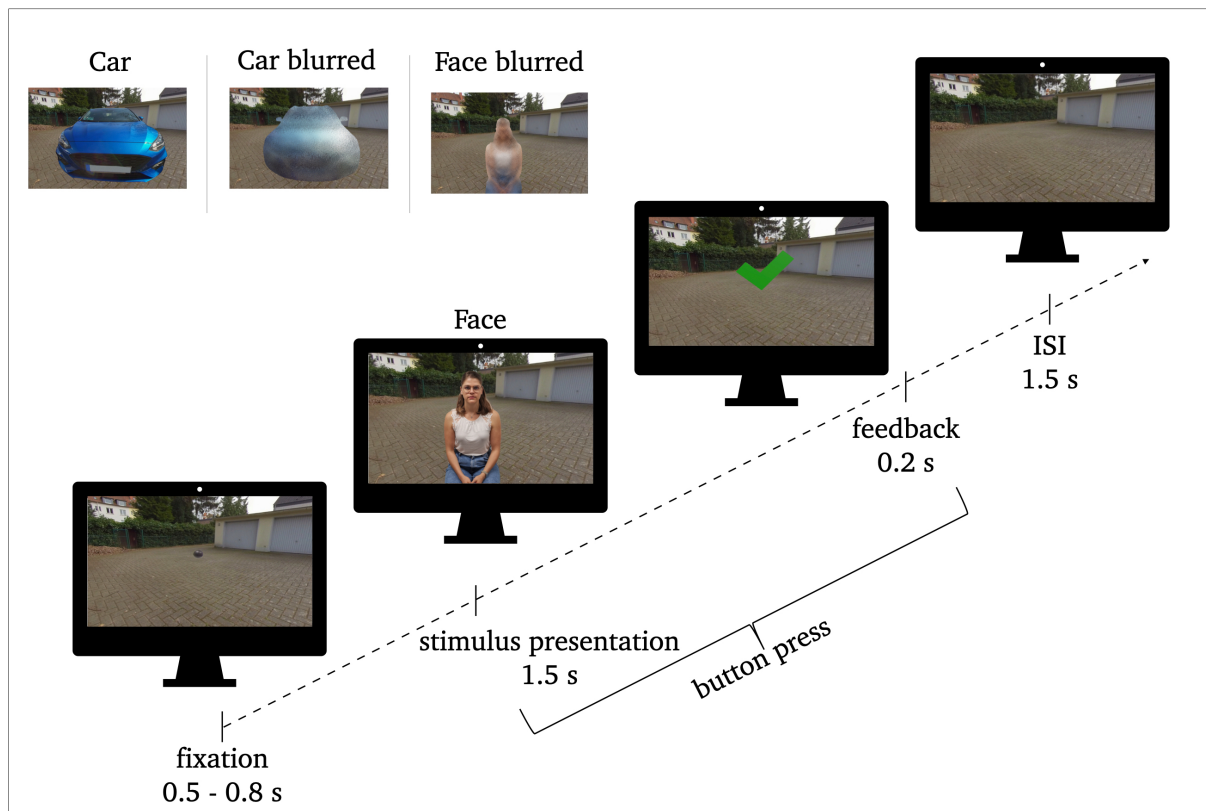

FIGURE 1 Procedure of stimulus presentation: 0.5–0.8 s fixation, 1.5 s stimulus presentation, 0.2 s feedback, 1.5 s interstimulus interval (ISI). Exemplary stimuli of face, car, and perceptual control (blurred) conditions are illustrated.

with a constant distance of 115 cm to the screen, resulting in a horizontal viewing angle of 7.5° and a vertical viewing angle of 5°. The pictures were presented in 2D in the center of the screen using the video-game engine Unity 5 (Version 2020) for stimulus presentation.

For the VR condition, participants remained seated and were equipped with a VR headset (HTC Vive Pro 2, 2448 × 2448 pixels per eye, up to 120° field of vision, 120 Hz refresh rate). The pictures were presented in 3D-360° in real-life size at a distance of 62 cm for the persons (horizontal viewing angle 42°; vertical viewing angle: 98°) and 92 cm for the cars (horizontal viewing angle 166°; vertical viewing angle: 168°) also via Unity 5. For both conditions, the trigger stream from Unity was transmitted to the Lab Streaming Layer (LSL by SCCN, https://github.com/sccn/labstreaminglayer) to synchronize the EEG data stream and Unity triggers. The precise timing of the stimulus onset trigger and the stimulus appearance on the monitor or VR headset were ensured by the use of a photodiode. The timepoint of the button press was also recorded and saved to a separate response time for each participant file by Unity for further analysis, and the respective triggers were later merged with the EEG data stream (see 2.4).

2.4 | Electrophysiological recordings and preprocessing

Electrophysiological recording (EEG) was derived with 128 electrodes, attached in accordance with the international 10–20 system. An Active-Two amplifier system from BioSemi (Amsterdam, Netherlands) was used with a sampling rate of 512 Hz and a bandwidth (3 dB) of 104 Hz. In addition, a horizontal electrooculogram (hEOG) and a vertical electrooculogram (vEOG) were recorded, and a common mode sense (CMS) and a driven right leg (DRL) electrode were applied as reference and ground electrodes (for details, see www.biosemi.com/fraq/cms&drl.htm).

All preprocessing steps were applied to the recordings of both modalities using MATLAB (version R2022a, MathWorks Inc.) and EEGLAB (version 2023, Delorme & Makeig, 2004). First, the EEG data stream and trigger stream had to be merged via the EEGLAB add-on MoBi-Lab (Ojeda et al., 2014). The data were re-referenced to the average reference, high-pass filtered at 0.25 Hz, and low-pass filtered at 25 Hz. Bad channel identification was achieved using the automatic channel removal add-on (ASR; Mullen et al., 2015), followed by respective channel interpolation. For the elimination of extended potential drifts, linear detrending was applied to all channels. Artifact rejection was applied for muscle, eye, heart, line noise, and channel noise artifacts using independent

component analysis (ICA; Delorme et al., 2007). For stimulus-locked epoching, the time window around the trigger onset was set from −500 to 1450 ms, with a baseline window of 300 ms before stimulus onset. For response-locked epoching, the time window around the trigger onset was set from −600 to 1400 ms, with a baseline from −600 to −450 ms pre-response.

For ERP analysis, grand means were computed per modality and within modality per stimulus category (face, face blurred, car, car blurred), resulting in eight individual stimulus-locked ERPs (i.e., VR-face, VR-face blurred, VR-car, VR-car blurred, PC-face, PC-face blurred, PC-car, PC-car blurred) and eight response-locked ERPs (i.e., see above).

2.5 | ERP components

2.5.1 | Stimulus-locked ERP components

Standard analysis: N170 Latency and amplitude
The time windows and electrode sites for the analysis of the N170 were selected based on prior literature (Boehm et al., 2011; Dering et al., 2011; Latinus & Taylor, 2006; Rossion & Jacques, 2008) combined with visual inspection of the root-mean-squared ERP (see Figure 2). The N170 was analyzed at temporal parieto-occipital electrodes (i.e., P7, P8, P9, P10, P07, and P08). Due to latency differences, the time windows were set individually for the two modalities, that is, from 145 to 191 ms for PC (see Figure 2a) and 112 to 150 ms for VR (see Figure 2b).

For the latency analysis of the N170, mean peak-amplitude latencies were collected for each stimulus type by locating the maximum negative amplitude deflection of each participant's mean ERP within the N170 time window. Peak latencies were validated by visual inspection at the individual subject level and manually corrected as needed.

Regional mean analysis: N170, L1, and L2
For the regional mean analysis, the time windows for the N170 component were the same as for the standard approach (see above). The time windows for the later components were chosen based on prior literature (Bublatzky et al., 2014; Johnsdorf et al., 2023; Stolz et al., 2019) combined with visual inspection of the root-mean-squared ERP (see Figure 2). The latency range for the L1 was defined from 200 to 450 ms and for the L2 component from 600 to 1350 ms post-stimulus onset, corresponding to the late components defined in Sagehorn et al. (2023).

All stimulus-locked components were analyzed at six regional means: frontal, temporal left, posterior left, temporal right, posterior right, and parieto-occipital (see
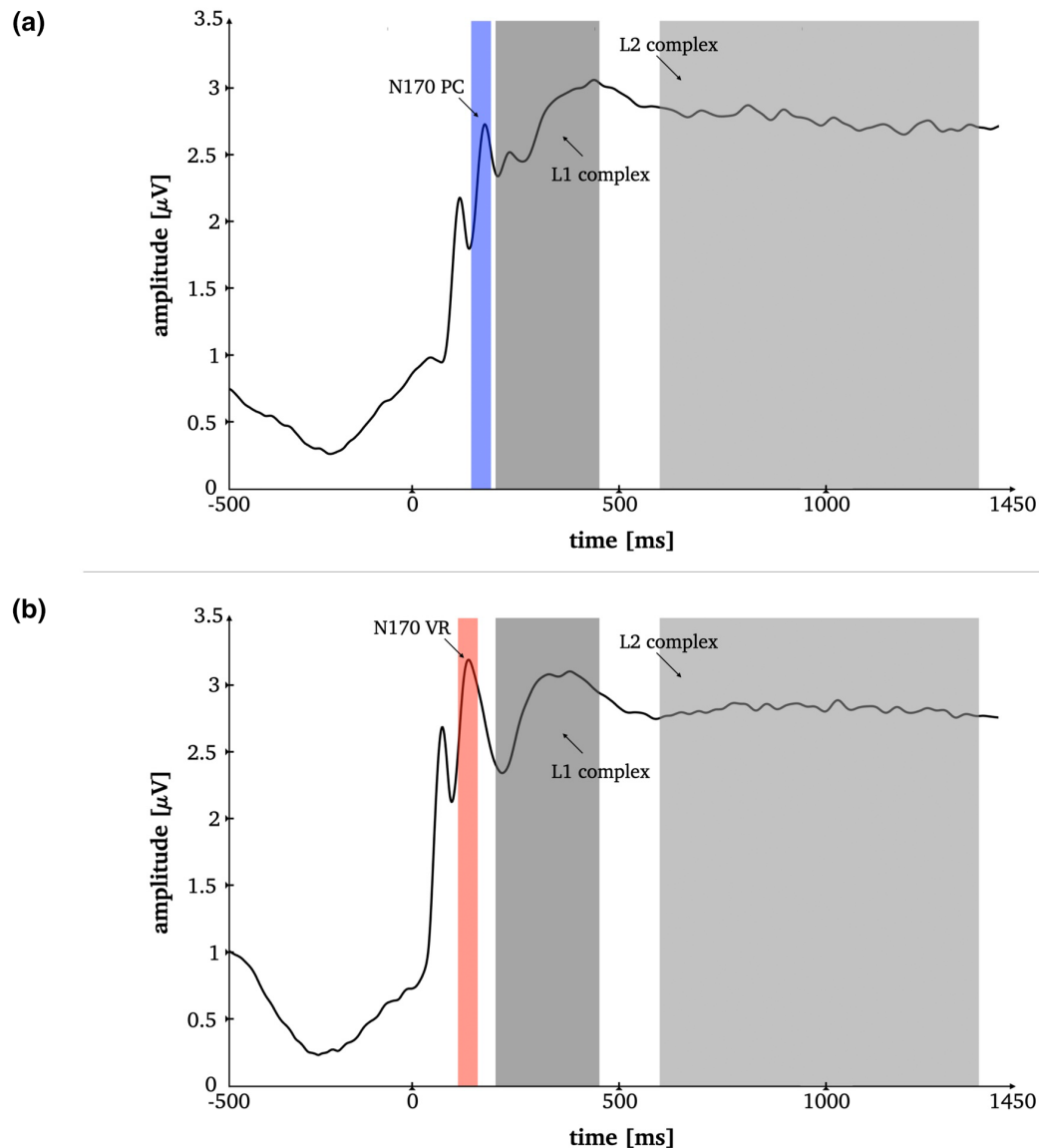
FIGURE 2 Time-by-amplitude plots of the root-mean-squared event-related potential (ERP) averaged over all electrodes per condition (panel a: PC, panel b: VR) for the selection of appropriate time windows for all stimulus-locked ERP components. Highlighted sections mark the time windows for N170 PC (145–191 ms), N170 VR (112–150 ms), L1 (200–450 ms), and L2 (600–1350 ms).

Figure 3). The ERP amplitudes were computed by calculating the mean voltage across the respective time window and for each regional mean individually.

The investigation of electrophysiological responses during the N170 latency range in two separate analyses is performed to be consistent with previous research on the N170 component and simultaneously maintain consistency in analyzing all components of interest (early, late, and response-locked). The N170 component is conventionally analyzed by comparing the mean amplitude between stimulus types (faces vs. perceptual and object controls), which is performed for both modalities, PC and VR. In addition, analyzing the regional averages provides more comprehensive insights into the spatial distribution of neural processing during the N170 latency range,

allowing comparison with the later and response-locked components.

2.5.2 | Response-locked ERP components

Regional mean analysis: Pre- and post-response components

The explorative response-locked analysis extends the stimulus-locked analysis by providing precise timing information about the participants′ responses. Thus, the time windows for the response-locked ERP analysis were set liberally before (−400 to −50 ms) and after (50–450 ms) the response. Both were analyzed at the same six regional means as the stimulus-locked components:
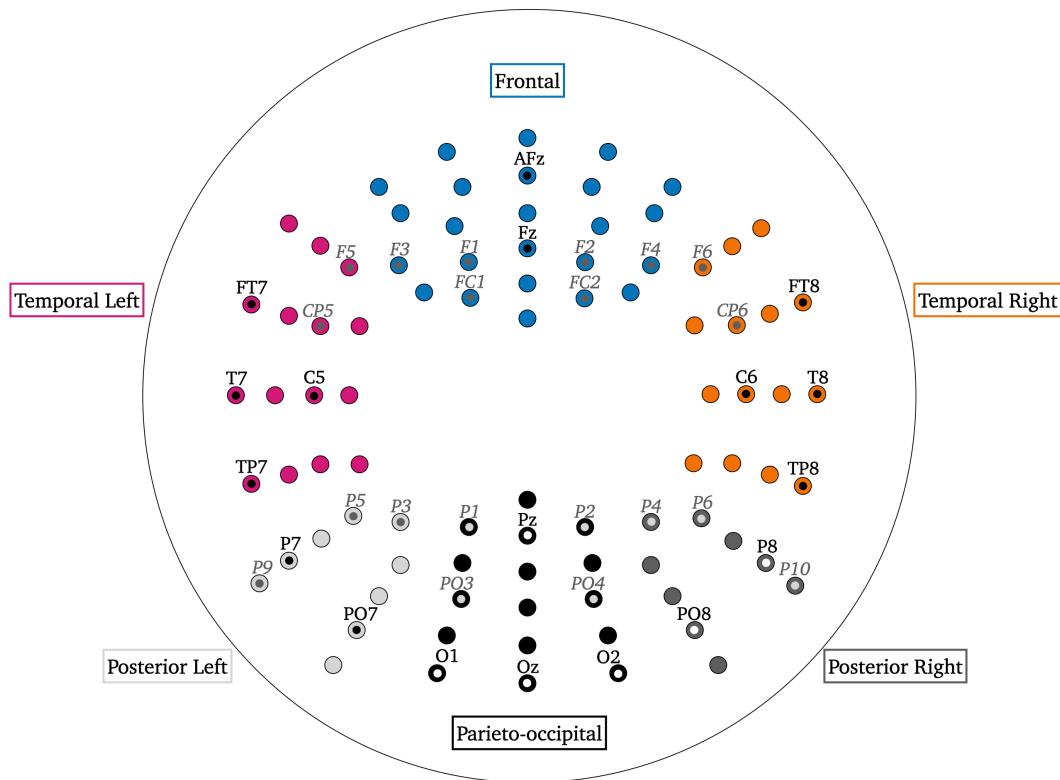
FIGURE 3 Electrode positions of the six regional means selected for the regional mean analysis approach for the stimulus-locked components (N170, L1, L2) and the response-locked components (pre- and post-response): frontal, temporal left, temporal right, posterior left, posterior right, and parieto-occipital. The 10–20 system positions are labeled when available (gray italic labels correspond to approx. electrode positions). For a complete BioSemi 128-channel ABC layout, see https://www.biosemi.com/headcap.htm.

frontal, temporal left, posterior left, temporal right, posterior right, and parieto-occipital (see Figure 3). The pre- and post-response component amplitudes were computed by calculating the mean voltage across the respective time window and for each regional mean individually.

2.6 | Statistical analysis

All statistical analyses were performed using SPSS Statistics (IBM, Version 28). All post hoc testing was corrected for multiple testing using Bonferroni correction, that is, the alpha-level of $\alpha = 0.05$ was adjusted according to the number of tests being performed (adjusted $\alpha = \alpha/$ no. of tests).

2.6.1 | Response times

Response times were averaged across trials for each participant and per stimulus type and the resulting means were analyzed using a 2×4 repeated-measurements ANOVA (rmANOVA) with the within-subject factors "modality" (VR vs. PC) and "stimulus type" (face vs. face blurred vs.

car vs. car blurred). Whenever necessary, Greenhouse–Geisser-corrected p values are reported. Significant effects of rmANOVA were complemented by post hoc t tests within each modality and between modalities.

2.6.2 | Standard analysis: N170 amplitude & latency

Both, the mean N170 amplitudes and latencies were analyzed using a 2×4 rmANOVA with the within-subject factors "modality" (VR vs. PC) and "stimulus type" (face vs. face blurred vs. car vs. car blurred). Whenever necessary, Greenhouse–Geisser-correction was applied. Significant effects of rmANOVA were complemented by post hoc t tests within each modality and between modalities.

2.6.3 | Regional mean analysis: N170, L1, L2, and response-locked components

The EEG data for all stimulus-locked ERP components (N170, L1, L2) and response-locked components (pre- and post-response) were analyzed using a $2 \times 4 \times 6$

rmANOVA with the within-subject factors "modality" (VR vs. PC), "stimulus type" (face vs. face blurred vs. car vs. car blurred), and "regional mean" (frontal vs. temporal left vs. temporal right vs. posterior left vs. posterior right vs. parieto-occipital). Whenever necessary, Greenhouse–Geisser-corrected *p* values are reported.

Planned comparisons

The goal of the regional mean analysis was to compare the processing mechanisms for faces and cars not only on a broader spatial frame but moreover on a deeper conceptual level to determine when, over which areas, and under what conditions potential differences occur. Specifically, the goal was to compare the differences in the processing of faces and objects (in this case, cars) beyond basic physical features between the two modalities.

Thus, to compare the perception of faces and objects beyond their basic perceptual features, the blurred image was first subtracted from the sharp image of the same stimulus category (face – face blurred; car – car blurred), before contrasting the two categories within each modality to determine neural processes inherently dedicated only to human face processing. The resulting FOD was then compared between modalities to determine whether and to what extent they differed between laboratory and realistic conditions.

Hence, significant effects of the rmANOVA were complemented by planned contrasts of the FODs between modalities per regional mean, that is, PC-FOD versus VR-FOD.

Topography similarity testing

To test whether the L1 and L2 components from the stimulus-locked analysis yield comparable topographies to the pre- and post-response components from the response-locked analysis, they were tested for similarity employing a previously applied approach that uses 2D correlation (Sagehorn et al., 2023).

3 | RESULTS

3.1 | Response times

The rmANOVA for the response times revealed significant main effects for the factor "modality" ($F_{modality}(1, 57) = 127.33$, $p < .001$, $\eta^2 = 0.69$) and "stimulus type" ($F_{stimulus}(2.4, 136.4) = 492.45$, $p < .001$, $\eta^2 = 0.97$), as well as the interaction effect ($F_{interaction}(1.3, 76) = 31.79$, $p < .001$, $\eta^2 = 0.44$). The respective descriptive statistics are given in Table S1.

Within the PC modality, response times for faces were significantly slower than all other stimulus categories. Cars and blurred cars both yielded faster response times than faces and blurred faces, but were not significantly different from each other. Within the VR modality, response times for faces were significantly slower than all other stimulus categories, whereas response times for cars were the fastest. Responses for blurred faces were slower than for blurred cars. For detailed statistics, see Table S2, and Figure 4.
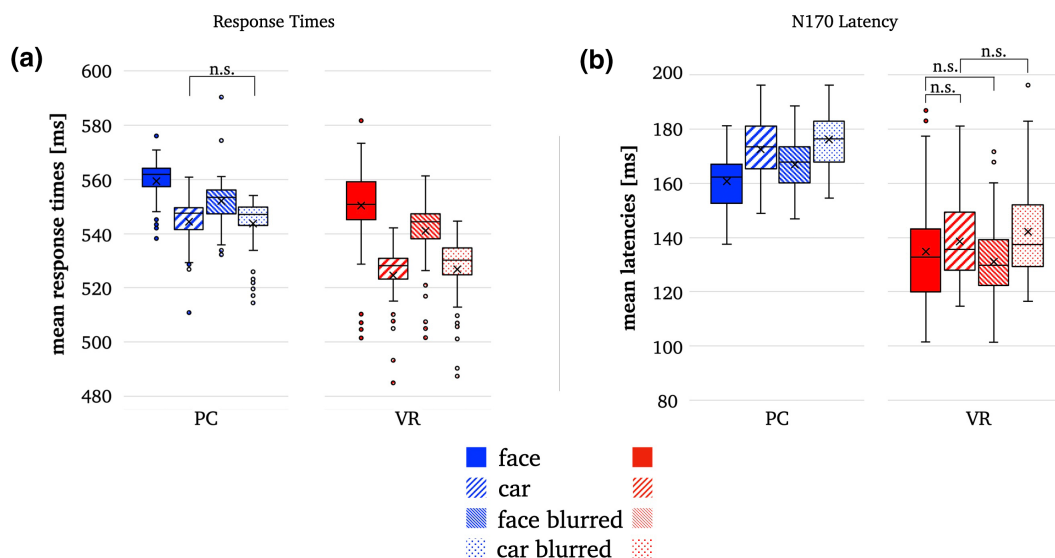

FIGURE 4 Boxplots illustrating the distributional characteristics of response times [panel a] and N170 latency [panel b] for all stimulus types in both modalities. The boxes depict the interquartile range (IQR), the exclusive median is depicted as a solid line, the mean as a cross, and outliers (>1.5×IQR) as dots. The whiskers comprise the range from Q1 to Q1.5×IQR to Q3+1.5×IQR. For the sake of clarity only the non-significant differences are marked (ns).

The between-modality comparisons revealed significantly faster response times in VR than PC for the same stimulus type (14 ms difference on average; all ps < .001; see Table S2).

3.2 | Stimulus-locked components

3.2.1 | Standard approach: N170 latency

The rmANOVA for the N170 latency revealed significant main effects for the factor "modality" ($F_{modality}(1, 57) = 222.43$, $p < .001$, $\eta^2 = 0.90$) and "stimulus type" ($F_{stimulus}(2.1, 120.1) = 41.28$, $p < .001$, $\eta^2 = 0.42$), as well as the interaction effect ($F_{interaction}(2.4, 136.2) = 9.10$, $p < .001$, $\eta^2 = 0.14$). The respective descriptive statistics are given in Table S1.

Within the PC modality, faces yielded the shortest N170 latencies compared to all other stimulus types. Cars and blurred cars both displayed longer latencies than faces and blurred faces, with shorter latencies for cars than blurred cars. Within the VR modality, blurred faces showed the shortest latency among all stimulus types, even though they were not significantly shorter than normal faces. The latencies for cars were not significantly different from normal faces or blurred cars. Blurred cars yielded significantly longer latencies than normal and blurred faces. For detailed statistics, see Table S2, and Figure 4.

The between-modality comparisons revealed significantly shorter latencies for VR than PC for the same stimulus type (approx. 30 ms difference on average; all ps < .001; see Table S2).

3.2.2 | Standard approach: N170 amplitude

The rmANOVA for the N170 amplitude revealed significant main effects for the factor "modality" ($F_{modality}(1, 57) = 30.75$, $p < .001$, $\eta^2 = 0.35$) and "stimulus type" ($F_{stimulus}(2, 113.4) = 67.81$, $p < .001$, $\eta^2 = 0.54$), as well as the interaction effect ($F_{interaction}(2.3, 131.5) = 19.94$, $p < .001$, $\eta^2 = 0.26$). The respective descriptive statistics are given in Table S1.

Classical pairwise comparisons between stimulus types within the PC modality revealed the weakest negative N170 mean amplitudes for faces and blurred faces, which were, however, not significantly different from each other. Cars elicited the strongest negative mean amplitudes. Within the VR modality, faces yielded the least negative mean amplitudes, followed by blurred faces. Cars and blurred cars elicited more negative mean amplitudes but were not significantly different from each other. For detailed statistics, see Table S2, and Figure 5.

The between-modality comparisons revealed significantly stronger negative N170 amplitudes for VR than PC for the same stimulus type (all ps < .001), except for PC cars and VR cars (see Figure 5). Figure 6 depicts the lineplots comparing stimulus types within and between modalities.
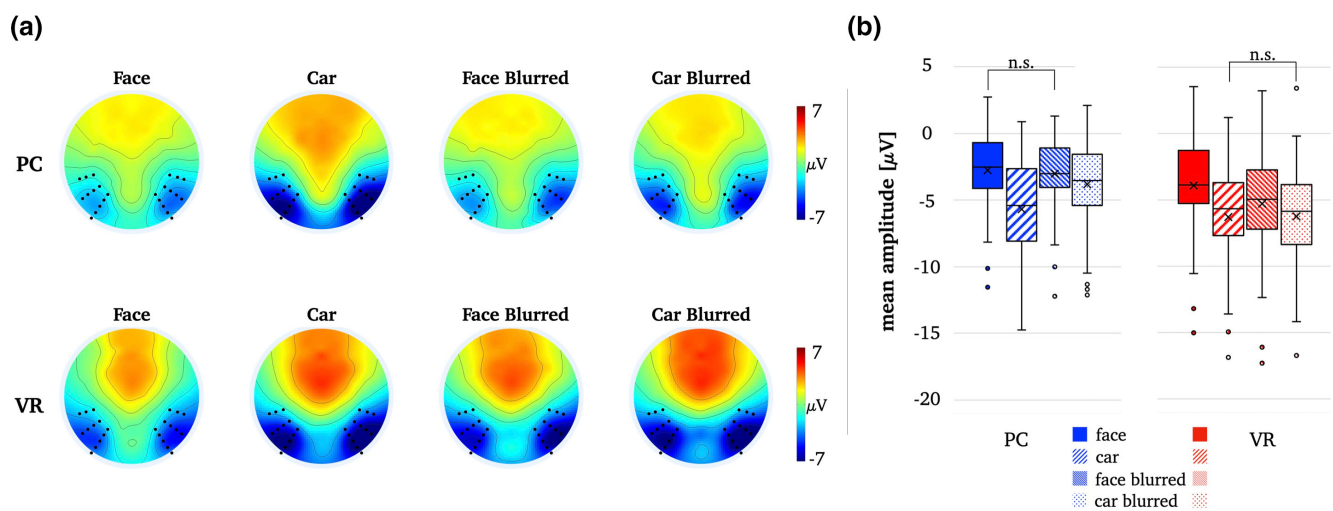

FIGURE 5 Panel a illustrates the N170 topographies for all stimulus types in both modalities. The electrodes used for analysis are highlighted (i.e., P7, P8, PO7, PO8, P10, P9, TP7, and TP8). Panel b illustrates the distributional characteristics of the N170 amplitude as boxplots for all stimulus types in both modalities. The boxes depict the interquartile range (IQR), the exclusive median is depicted as a solid line, the mean as a cross, and outliers (>1.5×IQR) as dots. The whiskers comprise the range from Q1 to Q1.5×IQR to Q3+1.5×IQR. For the sake of clarity, only the non-significant differences are marked (ns).
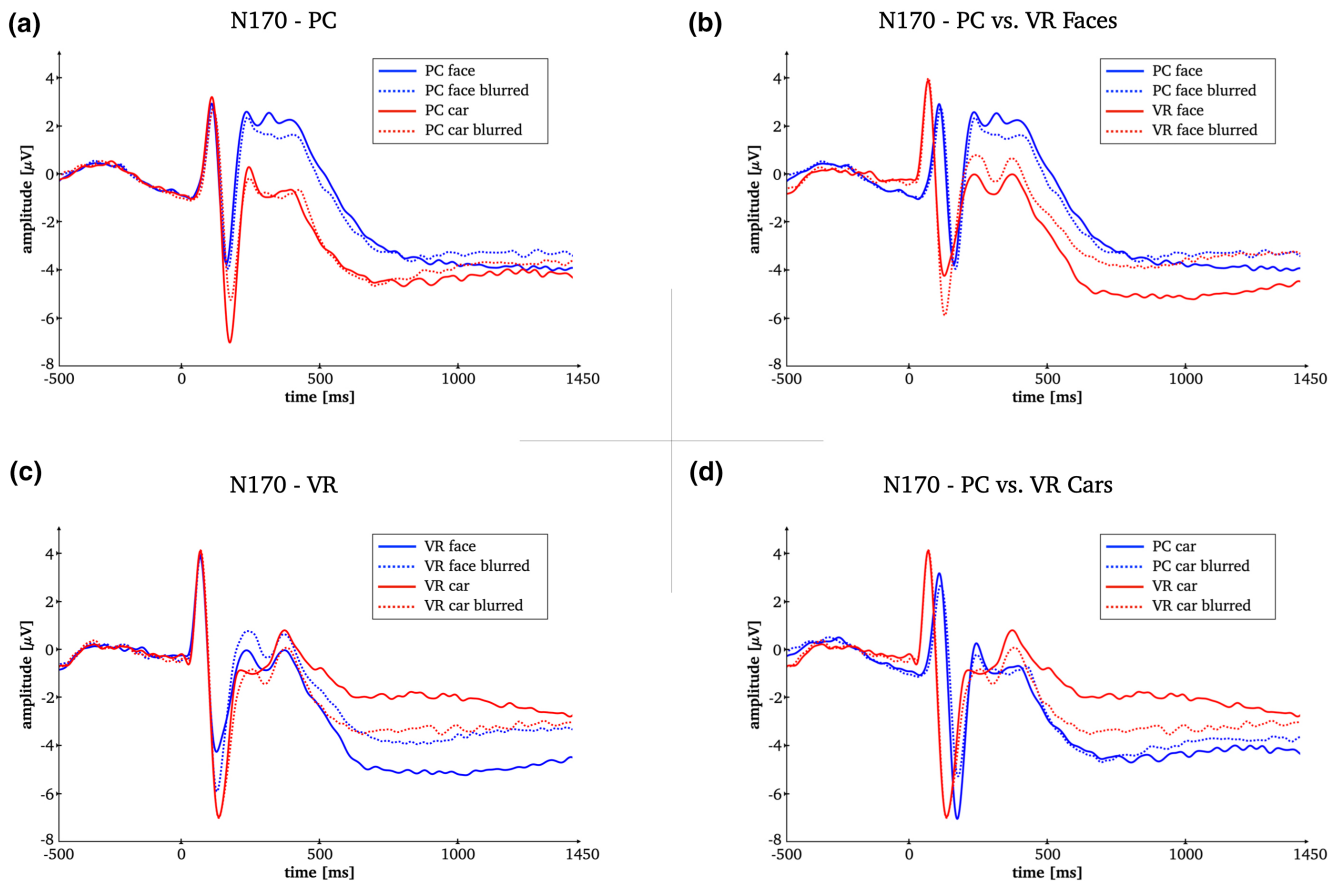
FIGURE 6 Time-by-amplitude plots of the mean N170 amplitudes for all stimulus types within the PC modality [panel a] and within the VR modality [panel c], for faces in both modalities [panel b], and for cars in both modalities [panel d].

3.2.3 | Regional mean analysis: N170, L1, & L2

N170

The rmANOVA for the regional mean analysis of the N170 revealed significant main effects for the factors "modality," "stimulus type," and "regional mean," as well as for all interactions of these factors (all $ps < .001$). The respective descriptive statistics are given in Table S3, and the detailed statistics in Table S8.

The planned contrasts between modalities for the FODs per regional mean revealed no significant effects for any of the regional means when corrected for multiple comparisons (all $ps \geq .01 >$ Bonferroni correction) (see Table 1). Figures S3 and S4, depict the regional mean lineplots and topographies for the FOD effects per modality for the N170 time window.

L1

The rmANOVA for the regional mean analysis of the L1 revealed significant main effects for the factors "modality," "stimulus type," and "regional mean," as well as for all interactions of these factors (all $ps < .001$). The respective descriptive statistics are given in Table S4, and the detailed statistics in Table S8. Figure 7 shows the mean topographies for the L1 component per modality and stimulus type.

The planned contrasts between modalities for the FODs per regional mean revealed significant effects for all regional means except the parieto-occipital ($p = .024 >$ Bonferroni correction) when corrected for multiple comparisons (see Table 1). For the VR modality, amplitudes were significantly more positive at the frontal regional sensors and more negative at bilateral temporal and temporo-posterior regional sensors than for the PC modality. Figure 8 depicts the topographies and regional mean lineplots for the FOD effects per modality for L1 and L2.

L2

The rmANOVA for the regional mean analysis of the L2 revealed no significant main effects for the factor "modality" ($F_{modality}(1, 57) = 0.28$, $p = .597$), but for the factors "stimulus type" and "regional mean" and for all interactions of these factors except "Modality × Regional mean" ($F_{ModalityxRegionalMean}(2.6, 150.6) = 2.05$, $p = .117$). The respective descriptive statistics are given in Table S5, and the detailed statistics in Table S8. Figure 7 shows the

TABLE 1 Planned contrasts between modalities (PC vs. VR): Face–object differences in amplitude for N170, L1, L2, and pre- and post-response components.

	df	t	p	Cohen's d
PC vs. VR				
N170				
Frontal	57	−1.73	.089	−0.23
Temporal left	57	.54	.590	0.07
Temporal right	57	1.46	.151	0.19
Posterior left	57	1.29	.202	0.17
Posterior right	57	2.62	.011*	0.34
Parieto-occipital	57	.71	.483	0.09
L1				
Frontal	57	−3.81	<.001**	−0.50
Temporal left	57	4.86	.004**	0.64
Temporal right	57	3.00	<.001**	0.39
Posterior left	57	5.42	<.001**	0.71
Posterior right	57	3.33	.002**	0.43
Parieto-occipital	57	−2.31	.024*	−0.30
L2				
Frontal	57	−3.91	<.001**	−0.51
Temporal left	57	4.35	<.001**	0.57
Temporal right	57	3.64	<.001**	0.48
Posterior left	57	4.50	<.001**	0.59
Posterior right	57	2.23	.002**	0.42
Parieto-occipital	57	−.57	.569	−0.08
Pre-response				
Frontal	57	−4.63	<.001**	−0.61
Temporal left	57	4.13	<.001**	0.54
Temporal right	57	3.85	<.001**	0.51
Posterior left	57	4.83	<.001**	0.64
Posterior right	57	3.42	.001**	0.45
Parieto-occipital	57	−.46	.646	−0.06
Post-response				
Frontal	57	−4.72	<.001**	−0.62
Temporal left	57	3.92	<.001**	0.52
Temporal right	57	3.61	<.001**	0.47
Posterior left	57	5.23	<.001**	0.69
Posterior right	57	3.41	.001**	0.45
Parieto-occipital	57	−.59	.559	−0.08

*Significant for $\alpha = .05$. **Significant after Bonferroni correction.

mean topographies for the L2 component per modality and stimulus type.

The planned contrasts between modalities for the FODs in amplitude per regional mean revealed significant effects for all regional means except the parieto-occipital ($p = .569$) (see Table 1). For the VR modality, amplitudes

were significantly more positive at the frontal regional sensors and more negative at bilateral temporal and temporo-posterior regional sensors than for the PC modality. Figure 8 depicts the topographies and regional mean lineplots for the FOD effects per modality for L1 and L2.

3.3 | Response-locked components

Pre-response

The rmANOVA for the regional mean analysis approach of the pre-response component revealed significant main effects for the factor "modality," "stimulus type," and "regional mean," as well as for all interactions of these factors (all ps < .001). The respective descriptive statistics are given in Table S6, and the detailed statistics in Table S8. Figure 9 shows the mean topographies for the pre-response component per modality and stimulus type.

The planned contrasts between modalities for the FODs per regional mean revealed significant effects for all regional means except the parieto-occipital ($p = .646$) (see Table 1). For the VR modality, amplitudes were significantly more positive at the frontal regional sensors and more negative at bilateral temporal and temporo-posterior regional sensors than for the PC modality (all ps ≤ .001). Figure 10 depicts the topographies and regional mean lineplots for the FOD effects per modality for the pre- and post-response components.

Post-response

The rmANOVA for the regional mean analysis of the post-response component revealed no significant main effects for the factor "modality" ($F_{modality}(1, 57) = 0.12$, $p = .735$), but for the factors "stimulus type" and "regional mean" and for all interactions of these factors except "Modality × Regional mean" ($F_{ModalityxRegionalMean}(2.6, 147.5) = 1,26$, $p = .290$). The respective descriptive statistics are given in Table S7, and the detailed statistics in Table S8. Figure 9 shows the mean topographies for the post-response component per modality and stimulus type.

The planned contrasts between modalities for the FODs per regional mean revealed significant effects for all regional means except the parieto-occipital ($p = .559$) (see Table 1). For the VR modality, amplitudes were significantly more positive at the frontal regional sensors and more negative at bilateral temporal and temporo-posterior regional sensors than for the PC modality. Figure 10 depicts the topographies and regional mean lineplots for the FOD effects per modality for the pre- and post-response components.
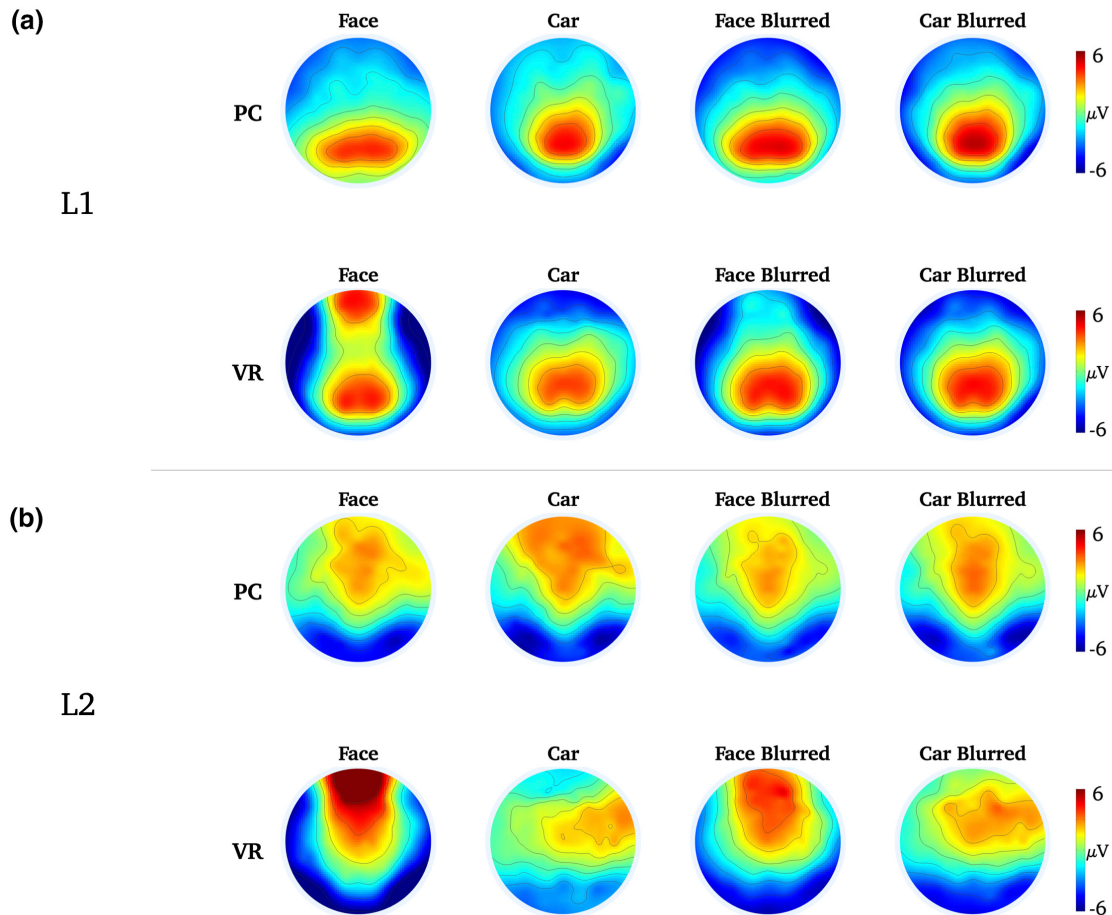
FIGURE 7 Topographies for all stimulus types and both modalities for L1 [panel a] and L2 [panel b].

3.4 | Topography similarity testing

The 2D-correlation similarity testing between the topographies of the L1 and pre-response components revealed moderate similarity for the PC condition ($r_{2D} = .67$) and high similarity for the VR condition ($r_{2D} = .90$). The 2D-correlation between the topographies of the L2 and post-response components showed high similarity for both the PC ($r_{2D} = .89$) and VR ($r_{2D} = .92$) conditions.

4 | DISCUSSION

This study aimed to expand on previous insights into neural mechanisms for processing realistic human faces by comparing a 2D laboratory approach to a 3D VR setting (Sagehorn et al., 2023). We maintained laboratory conventions in the setup and methods, presenting images of people and cars in both 2D and VR conditions. In the VR setting, participants encountered life-sized 3D representations, requiring the processing of contextual and self-relevant information. Participants actively discriminated stimulus categories via button presses to distinguish the identification process from post-decisional evaluation.

To control for stimulus- or modality-dependent differences in perceptual characteristics and access the neural mechanisms of realistic face processing, we contrasted modality-specific FODs, comparing component amplitude differences of the two stimulus categories between the PC and realistic conditions, beyond basic perceptual processes.

We conducted a comparative analysis of behavioral data (response times) and ERP components associated with face processing, that is, the N170 component, L1, L2, and response-locked components, to comprehensively investigate realistic face processing, including initial perception and classification, conceptual identification and recognition, and post-decisional evaluation. In line with conventional laboratory studies, the N170 component was analyzed for latency and amplitude. To further compare neural mechanisms between monitor and realistic face processing in a broader spatial and temporal frame, we analyzed the FODs of all components using a regional mean analysis.
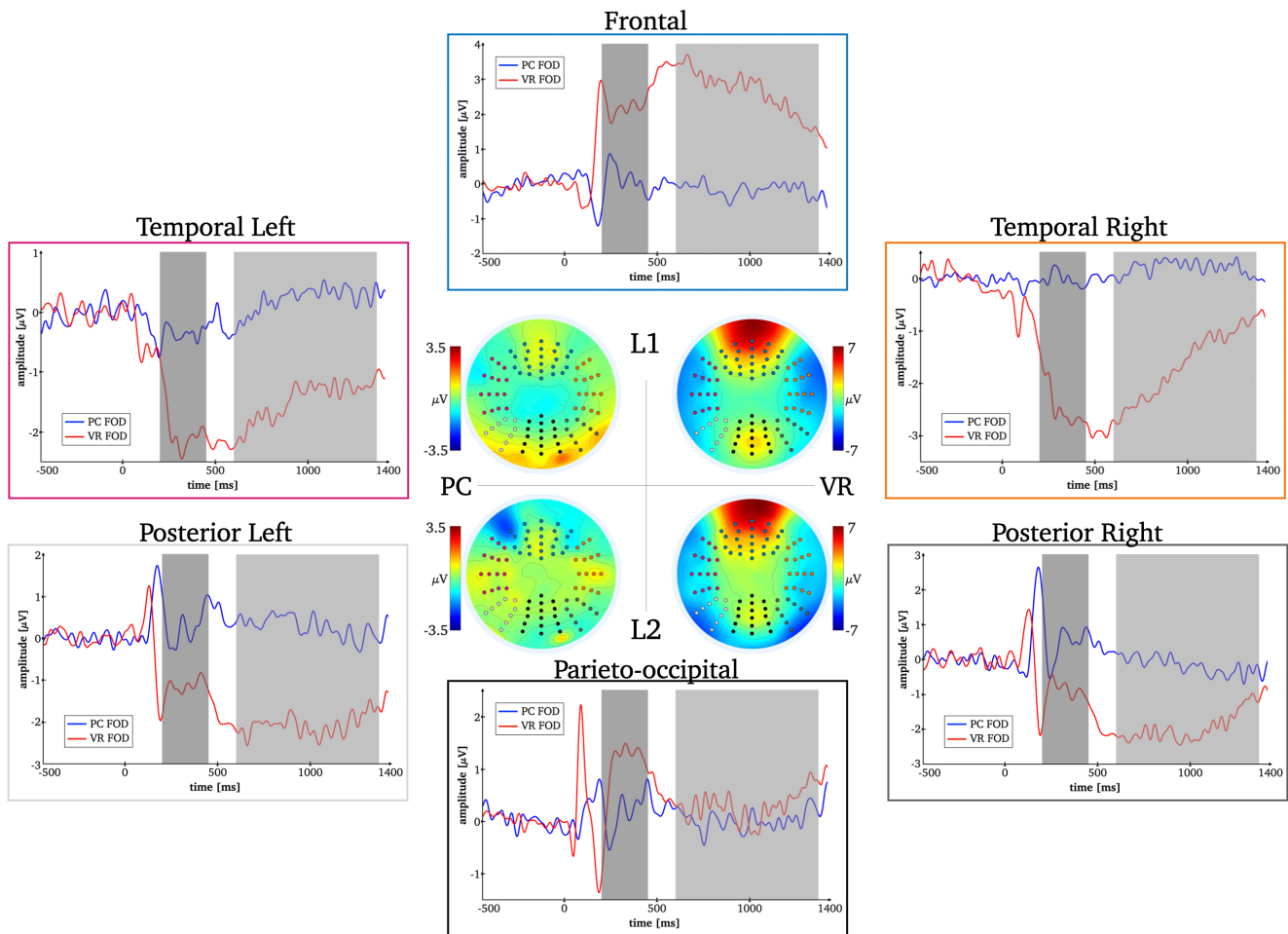
FIGURE 8 Time-by-amplitude plots of the mean face–object difference (FOD) for both modalities for all regional means (frontal, temporal left, temporal right, posterior left, posterior right, and parieto-occipital). Time windows of interest (L1 and L2) are highlighted in gray. Topographies for the modality-specific FODs are depicted for the L1 and L2 components.

We elicited N170 responses in laboratory conditions, demonstrating distinctions in latency and amplitude between face and car stimuli, consistent with previous laboratory findings demonstrating distinctions in latency and amplitude between face and car stimuli in contrast to the established face effect (Dering et al., 2009; Kloth et al., 2013; Kuefner et al., 2010), providing evidence for the general feasibility of the approach used in this study. The between-modality comparison showed faster stimulus processing under realistic conditions. However, the N170 component did not exhibit conclusive face-sensitive characteristics relevant for realistic face processing. Analyses of later and response-locked components revealed more in-depth face processing under realistic conditions based on specific FOD effects not evident under laboratory conditions.

4.1 | Response times

The analysis of response times obtained from the active discrimination task performed by the participants

showed faster stimulus discrimination under realistic conditions. For each stimulus type, participants were able to decide faster whether they perceived a (blurred) face or a (blurred) car under realistic conditions than under conventional laboratory conditions, and the difference pattern across stimulus types was comparable between modalities.

For both categories in both modalities, reaction times for the blurred images were faster than their sharp equivalents except for cars and blurred cars in PC, which is consistent with previous results (e.g., Kuefner et al., 2010). The expedited processing of global features or low-frequency information in stimuli, in contrast to the slower processing of detailed features or high-frequency information, has long been acknowledged (e.g., Bar, 2003; Hoeger, 1997; Navon, 1977). In addition, the amount of spatial frequency information required for stimulus processing was found to depend on the task itself (e.g., identification or recognition) and the task demands (e.g., degree of feature similarity of the stimuli to be discriminated) (Ruiz-Soler &
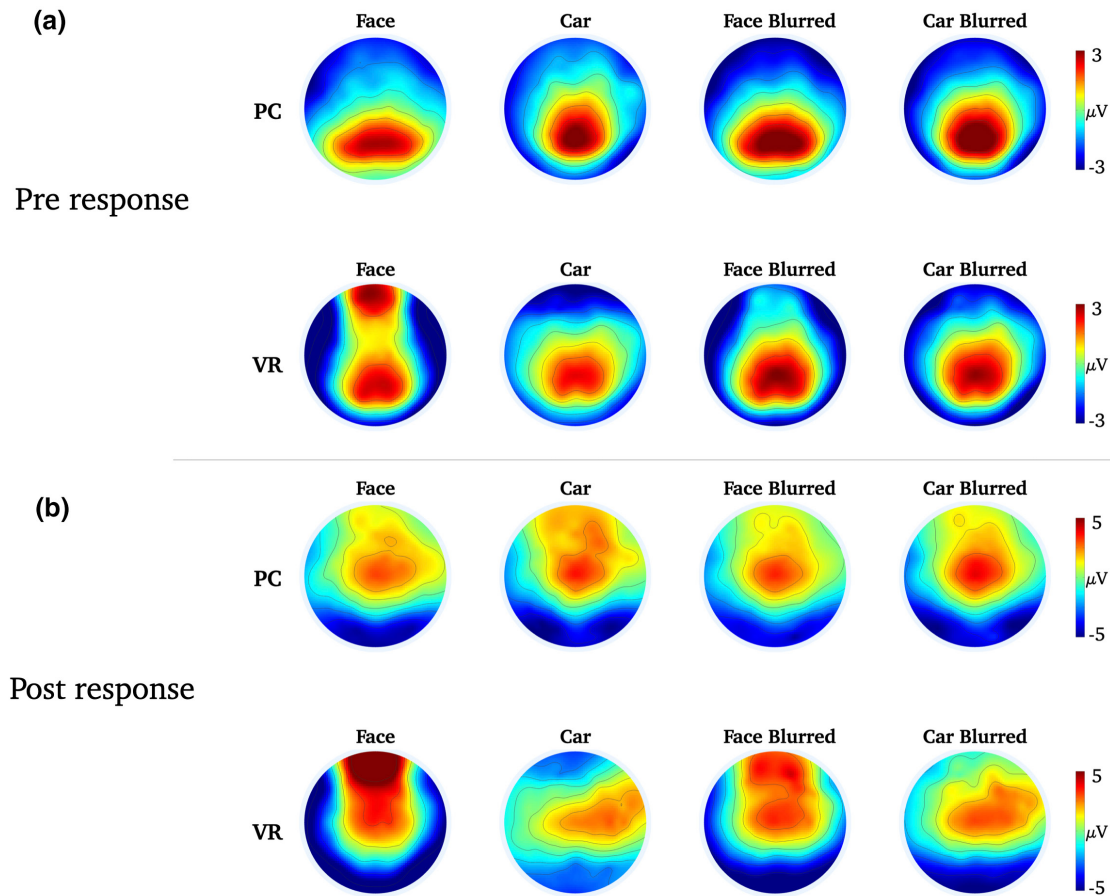
FIGURE 9 Topographies for all stimulus types and both modalities for the pre-response component [panel a] and the post-response component [panel b].

Beltran, 2006). Assuming that the silhouette of a person or car as a prominent global feature is sufficient to identify the stimulus category correctly in the rather simple discrimination task at hand, the faster identification of perceptual controls can be attributed to the absence of high-frequency information. Only the sharp stimuli contain high-frequency information, yet its presence is not essential for identification of the stimulus category but rather prolongs the process by requiring additional processing resources.

Regarding the differences in response time between the two stimulus categories (faces vs. cars), previous studies found that faces are identified faster than (Kuefner et al., 2010) or at the same rate as cars (Dering et al., 2009), whereas we found that cars are consistently identified faster than faces across modalities. The faster reaction times for cars compared to people cannot be attributed solely to the difference in relative stimulus size. For example, reaction times for cars in the PC condition were faster than for faces in the VR condition, even though the latter were considerably larger (see Figure 4). The difference in reaction time between faces and cars on both modalities can be better explained by

a difference in the amount of high-frequency information, with faces being the most complex and detailed stimuli. The identification of an inanimate non-face object is thus easier, causing faster response times for cars and perceptual controls.

In general, the response times are a rather relative measure that comprises several processing steps. However, the overall faster response times in VR for stimuli in the same category can be attributed to the more immersive environment in which the stimuli were presented under realistic conditions. The resulting increased stimulus salience (Schöne, Sylvester, et al., 2021) and subjective presence and self-relevance (Dan & Reiner, 2017) induce heightened stimulus attention to be attributed to a life-sized person or car positioned directly in front of oneself. This observation accurately reflects real-life scenarios in which perceiving any stimulus from a personal distance requires a more immediate decision than seeing a picture of the same stimulus on a monitor. The accurate stimulus sizes contributed to faster decision-making under realistic conditions, at least on a behavioral level, corresponding to the behavior expected in reality.
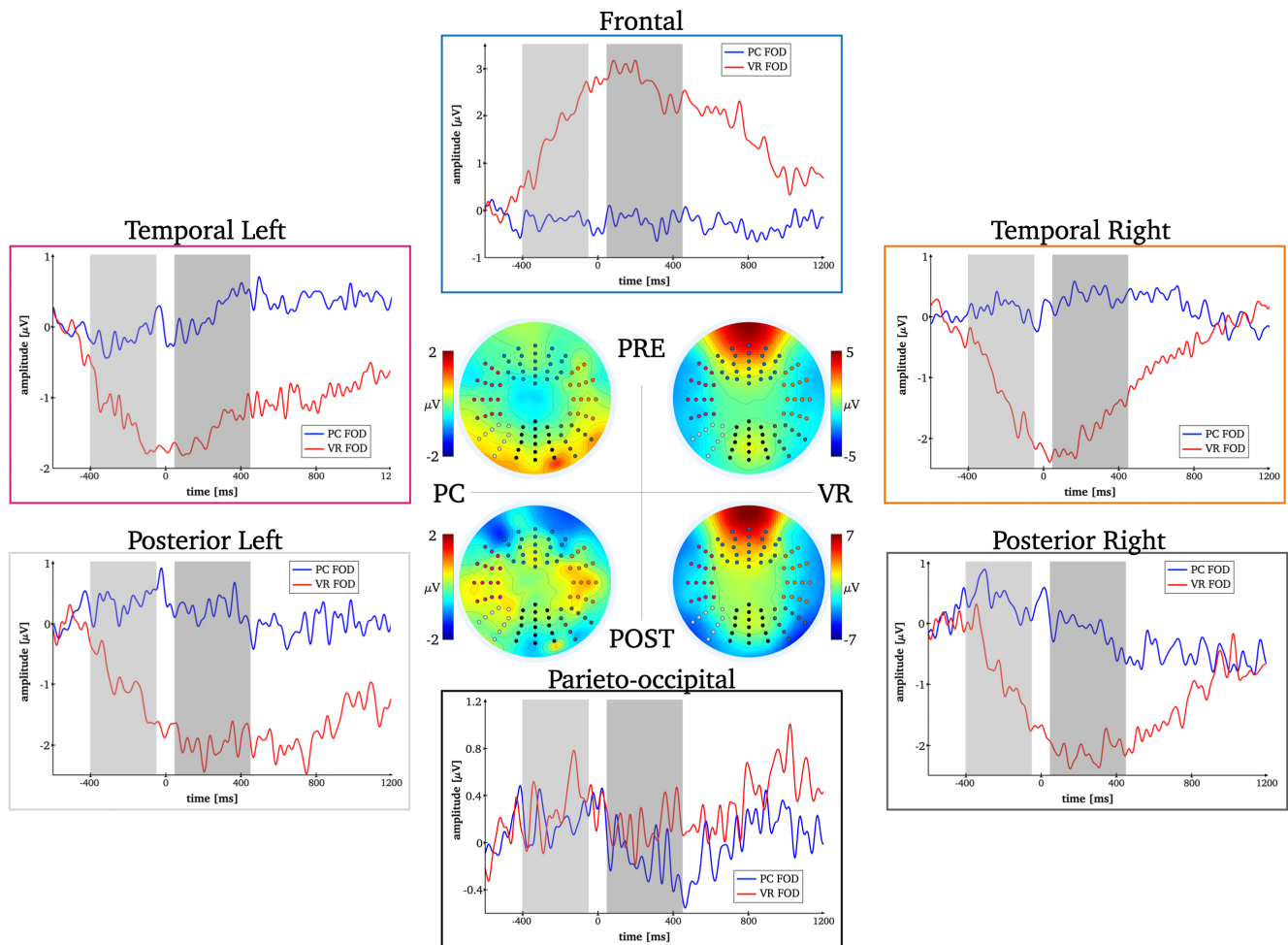
FIGURE 10 Time-by-amplitude plots of the mean face–object difference (FOD) for both modalities for all regional means (frontal, temporal left, temporal right, posterior left, posterior right, and parieto-occipital). Time windows of interest (pre- and post-response) are highlighted in gray. Topographies for the modality-specific FODs are depicted for the pre- and post-response components.

It should be noted that the face–car discrimination task can, in theory, be executed without the necessity of perceiving the stimuli in their entirety, as the distinction in size alone offers sufficient information to determine the stimulus type. Nevertheless, due to significant disparities in reaction times and physiological responses when contrasting normal and blurred stimuli of the same size, it can be assumed that participants perceived the normal faces and cars in their entirety rather than merely their outlines, even though the discrimination task did not explicitly demand a full perception.

4.2 | N170 latency

As expected, an overall faster stimulus processing under realistic conditions is also reflected in the respective differences in the N170 latencies between modalities. Under realistic conditions, the N170 appeared considerably earlier for all stimuli compared to their counterparts under conventional laboratory conditions, although displaying distinct modality-specific patterns of latency variations across different types of stimuli.

Our results from the PC condition align with previous laboratory studies that found faster processing of faces indexed by shorter N170 latencies for faces than for cars (Dering et al., 2009; Kloth et al., 2013; Kuefner et al., 2010). Thus, even though smaller stimuli usually elicit longer latencies (De Cesarei & Codispoti, 2006), faces still elicited shorter N170 latencies than cars, which were larger in stimulus size. The earlier occurrence of the N170 in response to faces does, however, not correspond to faster response times on a behavioral level (see section 4.1). In contrast, we observe only limited stimulus discrimination based on N170 latency differences under realistic conditions that is also not corresponding to the response time pattern.

Hence, the faster yet less distinct processing of realistic faces provides no evidence that the observed latency effect, commonly observed in conventional laboratory studies, can be extended to more realistic scenarios.

4.3 | N170 amplitude

4.3.1 | Standard analysis

The standard analysis approach of the N170 amplitude revealed complex differences within and between modalities. In general, all stimulus types except normal cars elicited stronger negative N170 responses under realistic conditions than under conventional laboratory conditions when modalities were directly compared. Taking into account the larger size and the associated increased amount of visual resources required for real-life-sized stimuli, these amplitude differences are likely to be considerably influenced by the physical properties of the stimuli (Busch et al., 2004; De Cesarei & Codispoti, 2006; Pfabigan et al., 2015). However, the modality-dependent amplitude differences are descriptively smaller than previously observed (Sagehorn et al., 2023), suggesting that the overall increased stimulus attention achieved by the active discrimination task reduces the influence of the basic physical features (i.e., stimulus size) on the N170 amplitude. Moreover, in both the PC and VR conditions, stimuli of the same size elicited significantly different N170 amplitudes, that is, PC-cars versus PC-blurred cars and VR-faces versus VR-blurred faces, and PC cars elicited stronger negative amplitude deflections than VR faces, although they were smaller (see Figure 5). Hence, though stimulus size remains an important factor that should not be disregarded, it cannot fully account for the observed amplitude differences. Nevertheless, direct comparison between modalities is of limited value when investigating the face and object perception and the underlying neural mechanisms on a more profound conceptual level.

The intra-modality amplitude patterns of the N170 reveal the extent to which the N170 differentiates stimulus types under conventional laboratory and realistic conditions. The typical N170 face effect found repeatedly in conventional laboratory studies describes a stronger negative signal deflection in response to the presentation of a face compared to perceptual and object controls (e.g., Eimer, 2011; Rossion & Jacques, 2011). The majority of studies report stronger negative amplitudes for faces compared to cars (Boehm et al., 2011; Flevaris et al., 2008; Kuefner et al., 2010; Rossion, 2014a, 2014b), although there are some studies that align with our findings and observe the opposite (Dering et al., 2009) or even no significant difference at all (Kloth et al., 2013). In general, we did not replicate this established face effect, that is, stronger N170 responses to face stimuli as to non-face or object stimuli, in either modality. Despite observing significant amplitude differences among the different stimulus types, both monitor and realistic faces consistently elicited the least pronounced N170 amplitudes, which stands in direct contrast to the typical pattern of findings.

Within our PC modality, the N170 amplitude did not differentiate between faces and blurred faces. It would have been expected for this outcome to manifest under conventional laboratory conditions, given the robustness of this effect that was also observed in the previous study (Sagehorn et al., 2023). Surprisingly, this distinction between normal and blurred faces was observed under realistic conditions, which was not apparent previously (Sagehorn et al., 2023). The implementation of an active discrimination task likely increased general stimulus attention, which might have caused the change in the N170 result pattern. The component is prevalent to top-down modulation as its face effect can only be observed under very specific monitor conditions. Under realistic conditions, the increased but also more natural attentional and contextual demands show that the effect is only of limited relevance for real-life face processing.

4.3.2 | Regional mean analysis

Regional mean analysis of the N170 revealed only a significant FOD effect at the posterior right regional mean, which, however, did not survive Bonferroni correction. Thus, when examining the disparities in face and object perception between conventional laboratory conditions and realistic conditions during the N170 time window, we do not yet observe specific face processing for realistic faces. Although there is a discernible stimulus discrimination when directly comparing N170 amplitudes (standard analysis), the perception of faces and objects up until this processing stage remains fundamentally similar between conventional laboratory settings and realistic conditions. This lends support to the idea that the N170 serves a distinct role within the face perception process, deviating from conventional assumptions. Although it appears to reflect the initial perceptual awareness of whether a visual stimulus is a face or a non-face, acting as a sort of gating mechanism for basic-level visual categorization based on overall configuration and specific facial features (Sollfrank et al., 2021), this alone is insufficient to serve as a neural marker for real-life relevant face perception. Besides its involvement in perceptual awareness, the N170 has also been considered a prediction error signal (Baker et al., 2021; Johnston et al., 2017). Manipulating attention and expectancy violations have independently amplified the N170 response for both cued and unexpected events, suggesting an alternative role for the N170 in signaling prediction errors (Baker et al., 2021; Johnston et al., 2017).

These observations highlight the need to examine established neural correlates like the N170 component

regarding their generalizability to realistic settings, as they are also frequently used in diagnostic and clinical contexts, for example, the N170 as a biomarker in disorders involving social dysfunction (Key & Corbett, 2020).

4.4 | Later components & response-locked analysis

The regional mean analysis of subsequent components revealed distinct patterns of face-specific processing depending on modality on a broader timescale, as indexed by significant FOD effects at the frontal, bilateral temporal, and bilateral posterior electrode regional means for the later components (L1 and L2) and response-locked components (pre- and post-response).

The L1 and pre-response components, as well as the L2 and post-response components, do not completely overlap in time but represent the same neuronal mechanisms to which they are linked to varying degrees. Looking at the mean response times and their 95% confidence intervals (see Table S1), the response times for each stimulus type in both conditions lie well between the L1 and L2 components, allocating them similarly to the pre- and post-response components before and after the participants' response. However, employing response-locking further refines this distinction, allowing for the exclusion of periods predominantly affected by motor activity (such as button pressing). This level of specificity is not attainable with stimulus locking unless variations in reaction times are meticulously considered. In addition, response-locking of ERPs can attenuate the effects of later potentials related to the processing of perceptual decisions (Berchicci et al., 2016; Nasr, 2010), which is why they provide more meaningful results than the late stimulus-locked components. As response-locking is not as commonly used and is not possible when the experiment does not require participants to respond, we interpret the components of both types of locking in parallel. Moreover, the similarity testing of the topographies of the L1 and pre-response as well as the L2 and post-response components showed a clear similarity between the stimulus-locked and response-locked domains. Although similar topographies do not necessarily reflect similar underlying neural generators, we take this as a strong indication that the L1 and pre-response components, as well as the L2 and post-response components, reflect similar or at least strongly overlapping neural processes.

The comparison of face and object perception, excluding stimulus- and modality-dependent perceptual features, revealed distinct neural mechanisms being recruited under realistic as opposed to conventional laboratory conditions. The analysis of later components therefore extends the insights on the face perception process, both in spatial and temporal dimensions.

Under conventional laboratory conditions, there seems to be no pronounced differentiation in the processing of screen faces and cars, as reflected by later potentials. This is consistent with the results of the previous study suggesting that similar processing mechanisms are employed for the perception of face and non-face stimuli presented in 2D (Sagehorn et al., 2023). In contrast, the processing of faces and objects under realistic conditions in VR involves distinct neural mechanisms depending on the stimulus category. Thus, encountering persons and cars in real size requires adaptation of processing depending on the stimulus category, whereas viewing images on a monitor with only the laboratory as context does not.

As several studies investigating emotional and cognitive processing properties under more realistic conditions have already shown, providing a realistic and immersive experience where different types of stimuli are presented in real-world dimensions reveals unique neural mechanisms that are not evident under conventional laboratory conditions (Blascovich et al., 2002; Gabana et al., 2017; Gromer et al., 2018; Kisker, Gruber, & Schöne, 2021; Kisker, Lange, et al., 2021; Newman et al., 2022; Rubo & Gamer, 2018; Schöne et al., 2019; Schöne et al., 2023; Schöne, Kisker, et al., 2021; Van Den Oever et al., 2022; Xu et al., 2021).

In Sagehorn et al. (2023), analysis of later components already indicated different neural mechanisms involved in the perception of faces compared to silhouettes and objects under realistic conditions. In conjunction with results from conventional laboratory studies suggesting that later potentials (i.e., LPP) distinctly index human face processing (e.g., Schindler et al., 2017; Wheatley et al., 2011), it was expected that these potentials would also be influenced by the transfer to more realistic conditions. Already in the conventional laboratory, the influence of social relevance, emotional and motivational salience, and perceived realism was found to be reflected in specific late potentials (Bublatzky et al., 2014; Schindler et al., 2017; Wheatley et al., 2011). Perception of human faces under more realistic conditions is associated with an increase in all of these dimensions, which resulted in the sustained and much stronger positive activation over frontal electrodes in the present study, providing the first evidence for more elaborate face processing.

Perceiving human faces in a natural context, proportionate to object control images (e.g., cars), further strengthens the notion that actual stimulus identification and recognition on a conceptual level occurs over a longer timeframe, extending beyond 200 milliseconds after the stimulus presentation. During the late component L1 and pre-response component, we observed a

pronounced face-specific processing of realistic faces on a conceptual level that continues into the post-decisional evaluation phase, reflected in the late component L2 and post-response component. Under realistic conditions, encountering a person or a car necessitates the use of specific neural mechanisms for identifying and evaluating the perceived stimuli. Conversely, viewing a picture of these stimuli in the PC condition does not appear to engage these stimulus-dependent perceptual processing mechanisms.

4.5 | Limitations and outlook

The transfer of a classical face perception paradigm to realistic conditions using VR yielded broadly similar ERPs in the early processing phases (i.e., P1, N170), demonstrating the general feasibility of this approach. However, the processing of realistic stimuli also suggests that different neural mechanisms are recruited than under conventional laboratory conditions.

At this point, it is important to note that the experimental paradigm used in the current study, although more realistic than viewing images on a computer screen, is still unrealistic in terms of the physical probability of the stimulus occurrence. It is highly unlikely that immobile people and cars will randomly appear out of nowhere in front of someone. Nevertheless, this approach is essential to maintain consistency with previous laboratory designs and enable a reasonable comparison of physiological correlates. To take the next step toward investigating realistic face processing and to further test the neural correlates that seem to be relevant in real-life face perception, the experimental setup needs to become more realistic and dynamic without losing too much experimental control. Prior studies have employed dynamic 2D presentations of faces and bodies, revealing several important findings (Kilts et al., 2003; Puce et al., 2000; Wheaton et al., 2004). For instance, they demonstrated that facial movements engage distinct processing mechanisms not implicated in the perception of static images (Puce et al., 2000). Additionally, these studies highlighted differences in the processing of dynamic emotional facial expressions compared to static ones (Kilts et al., 2003). They also found that observing the movements of body parts activates a more extensive array of brain regions in comparison to merely viewing their static counterparts (Wheaton et al., 2004). With regard to the investigation of realistic face and person perception, the ultimate goal should therefore be to use dynamic 3D presentations of faces and the corresponding bodies in order to obtain a comprehensive picture of the underlying neural mechanisms.

Other approaches that focus on conducting their experiments in a natural environment brought the laboratory directly into the real world (e.g., Gert et al., 2022). Although the results obtained in a real-life environment provide valuable insights into new aspects of face processing, they are also very uncontrollable and make it difficult to reproducibly access physiological correlates. The context and task demands within an experimental paradigm are undoubtedly of great importance for the cognitive processes under investigation, and it should therefore be possible to control them systematically.

As the complexity of experimental setups and paradigms continues to evolve to align with the intricate neural functions under investigation, there arises a concurrent demand for the development of more sophisticated analytical approaches. Within this context, EEG data, characterized by its temporal variability, represent an ideal candidate for functional data analysis—an approach of growing significance in EEG research (Ullah & Finch, 2013; Wang et al., 2016). This methodology allows for a holistic interpretation of data in its continuous, not necessarily linear nature, facilitating the application of concepts like component analysis (Shangguan et al., 2020; Zhang et al., 2019), functional clustering and classification (Yi et al., 2022), and predictive modeling (Zhang et al., 2020).

Although mean-based analyses have traditionally played a pivotal role in EEG research, it is essential to recognize the potential of more robust statistical measures. Given the frequent interest in identifying differences in locations and scales in EEG signals, the utilization of median analysis, such as the determination of median peak-amplitude latencies (see Figure S1) or the calculation of coefficients of variance (Ospina & Marmolejo-Ramos, 2019), proves highly advantageous. Furthermore, the exploration of alternative techniques, including robust statistics and distributional regression (Klein, 2024), warrants consideration as alternatives to conventional parametric statistical methods like ANOVAs. Additionally, representing data in a manner congruent with its underlying distributional characteristics enables more comprehensive data exploration (e.g., cumulative distribution function plots; Marmolejo-Ramos et al., 2023) (see Figure S2).

At the same time, another challenge arises. Representing realistic and, at the same time, dynamic and responsive human avatars in virtual environments remains a demanding task (Di Natale et al., 2023; Hepperle et al., 2022). Virtual faces created with today's advanced digital graphics technologies, although appearing remarkably realistic, are still perceived as not fully human, which is reflected at both the behavioral and neural levels (*uncanny valley effect*; Di Natale et al., 2023; Hepperle

et al., 2022; Miller et al., 2023; Schindler et al., 2017; Sollfrank et al., 2021).

These developments also have an impact on practical research areas that focus on recognizing, interpreting, and responding to human emotions, such as emotional artificial intelligence (AI), where AI systems are used to improve the interaction between machines and humans in this regard. Here, EEG data is being utilized to gain deeper insights into emotion recognition processes, with the aim of modeling them (Kamble & Sengupta, 2023; J. Wang & Wang, 2021). Gaining a more nuanced understanding of the neural processes underpinning real-world face perception and recognition and how these are expressed in differentiated EEG signals has implications for research in the field of emotional AI.

Consequently, future studies face the task of achieving three objectives: creating realistic and physically probable environments, implementing realistic dynamic human avatars, and maintaining an appropriate level of experimental control to systematically manipulate context and task demands.

4.6 | Conclusion

Implementing established laboratory conditions in a more authentic VR environment represents a crucial stage of progress in the study of real-life relevant face perception. To our knowledge, together with Sagehorn et al. (2023), the direct comparison of face perception between a conventional 2D monitor and realistic VR conditions is still a novel approach that addresses the challenge of transferability between the laboratory and reality in this domain.

In line with conventional laboratory results, our standard analysis of the N170 latency and amplitude confirms the involvement of the component in the face perception process, however, only to a certain degree. As the N170 does not exhibit conclusive face-sensitive characteristics and is prone to contextual and attentional factors, it might rather represent a pre-selective gating mechanism involved in the initial awareness of faces.

The processing of faces beyond the initial perceptual level takes place across a broader spatial and temporal distribution, which can be observed through later components linked to the response. Notably, both the pre- and post-response components exhibit face-specific characteristics in the conceptual identification and post-decisional evaluation of faces only under realistic conditions.

Our study aligns with previous research that compares electrophysiological markers obtained in both 2D and VR conditions, offering evidence that these markers and the underlying neural properties are domain-specific. Conventional paradigms developed according to the reductionist approach are limited in their ecological validity as they lack specific characteristics of the real environment, that is, depth and spatial proximity as well as physical and conceptual plausibility, from which the cognitive process of interest cannot be isolated.

To summarize, face perception involves intricate neural mechanisms that operate over a broader timeframe, extending beyond approximately 200 ms after stimulus presentation. The fundamental perceptual processing phase of perceiving faces exhibits substantial overlap when viewing them in 2D on a screen (conventional laboratory conditions) or encountering them in their real size and 3D (realistic VR conditions). However, when it comes to in-depth processing at the conceptual level and evaluation of the perceived stimulus after initial categorical perception, distinct face-specific processing mechanisms are engaged for realistic faces that are not captured by conventional laboratory conditions.

AUTHOR CONTRIBUTIONS
Merle Sagehorn: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; visualization; writing – original draft. **Marike Johnsdorf:** Conceptualization; software; writing – review and editing. **Joanna Kisker:** Conceptualization; investigation; methodology; project administration; software; writing – review and editing. **Thomas Gruber:** Conceptualization; resources; writing – review and editing. **Benjamin Schöne:** Conceptualization; methodology; project administration; supervision; writing – review and editing.

CONFLICT OF INTEREST STATEMENT
All authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT
The data sets and the code to extract the components analyzed in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/w63hp/?view_only=e337d28889ff40dc99fcb71c56e5f9bc. The stimulus material used in this study is available upon reasonable request and may be used for scientific purposes only. Access under: https://www.psycho.uni-osnabrueck.de/fachgebiete/allgemeine_psychologie_i/luvre.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local ethic committee of Osnabrück University, Germany. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

ORCID

Merle Sagehorn https://orcid.org/0000-0002-5295-9577
Marike Johnsdorf https://orcid.org/0000-0002-1231-7158
Joanna Kisker https://orcid.org/0000-0002-5826-264X
Benjamin Schöne https://orcid.org/0000-0001-7926-7426

REFERENCES

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., & Andrews, T. J. (2020). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*(3), 295–314. https://doi.org/10.1037/met0000337

Baker, K. S., Pegna, A. J., Yamamoto, N., & Johnston, P. (2021). Attention and prediction modulations in expected and unexpected visuospatial trajectories. *PLoS One*, *16*(10), e0242753. https://doi.org/10.1371/journal.pone.0242753

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. https://doi.org/10.1162/089892903321662976

Berchicci, M., Spinelli, D., & Di Russo, F. (2016). New insights into old waves. Matching stimulus- and response-locked ERPs on the same time-window. *Biological Psychology*, *117*, 202–215. https://doi.org/10.1016/j.biopsycho.2016.04.007

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Crystal, L., Inquiry, S. P., Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blau, V. C., Maurer, U., Tottenham, N., & McCandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, *3*, 7. https://doi.org/10.1186/1744-9081-3-7

Boehm, S. G., Dering, B., & Thierry, G. (2011). Category-sensitivity in the N170 range: A question of topography and inversion, not one of amplitude. *Neuropsychologia*, *49*(7), 2082–2089. https://doi.org/10.1016/j.neuropsychologia.2011.03.039

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *Quarterly Journal of Experimental Psychology*, *66*(12), 2426–2442. https://doi.org/10.1080/17470218.2013.789065

Bublatzky, F., Gerdes, A. B. M., White, A. J., Riemer, M., & Alpers, G. W. (2014). Social and emotional relevance in face processing: Happy faces of future interaction partners enhance the late positive potential. *Frontiers in Human Neuroscience*, *8*, 493. https://doi.org/10.3389/fnhum.2014.00493

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159–166. https://doi.org/10.1016/j.tics.2006.02.004

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004). Size matters: Effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clinical Neurophysiology*, *115*(8), 1810–1820. https://doi.org/10.1016/j.clinph.2004.03.015

Churches, O., Baron-Cohen, S., & Ring, H. (2009). Seeing face-like objects: An event-related potential study. *Neuroreport*, *20*(14), 1290–1294. https://doi.org/10.1097/WNR.0b013e3283305a65

Civile, C., Elchlepp, H., McLaren, R., Galang, C. M., Lavric, A., & McLaren, I. P. L. (2018). The effect of scrambling upright and inverted faces on the N170. *Quarterly Journal of Experimental Psychology*, *71*(11), 2464–2476. https://doi.org/10.1177/1747021817744455

Dan, A., & Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International Journal of Psychophysiology*, *122*, 75–84. https://doi.org/10.1016/j.ijpsycho.2016.08.013

De Cesarei, A., & Codispoti, M. (2006). When does size not matter? Effects of stimulus size on affective modulation. *Psychophysiology*, *43*(2), 207–215. https://doi.org/10.1111/j.1469-8986.2006.00392.x

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*(4), 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Dering, B., Martin, C. D., Moro, S., Pegna, A. J., & Thierry, G. (2011). Face-sensitive processes one hundred milliseconds after picture onset. *Frontiers in Human Neuroscience*, *5*, 93. https://doi.org/10.3389/fnhum.2011.00093

Dering, B., Martin, C. D., & Thierry, G. (2009). Is the N170 peak of visual event-related brain potentials car-selective? *Neuroreport*, *20*(10), 902–906. https://doi.org/10.1097/WNR.0b013e328327201d

Di Natale, A. F., Simonetti, M. E., La Rocca, S., & Bricolo, E. (2023). Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Computers in Human Behavior Reports*, *10*, 100288. https://doi.org/10.1016/j.chbr.2023.100288

Eimer, M. (2011). The face-sensitivity of the N170 component. *Frontiers in Human Neuroscience*, *5*, 119. https://doi.org/10.3389/fnhum.2011.00119

Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: An ERP analysis. *Brain Research*, *1194*, 100–109. https://doi.org/10.1016/j.brainres.2007.11.071

Gabana, D., Tokarchuk, L., Hannon, E., & Gunes, H. (2017). Effects of valence and arousal on working memory performance in virtual reality gaming. 2017 7th international conference on affective computing and intelligent interaction, ACII 2017, 36–41. https://doi.org/10.1109/ACII.2017.8273576

Gert, A. L., Ehinger, B. V., Timm, S., Kietzmann, T. C., & König, P. (2022). WildLab: A naturalistic free viewing experiment reveals previously unknown electroencephalography signatures of face processing. *European Journal of Neuroscience*, *56*(11), 6022–6038. https://doi.org/10.1111/ejn.15824

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*(3), 416–424. https://doi.org/10.1016/S0926-6410(03)00056-9

Gromer, D., Madeira, O., Gast, P., Nehfischer, M., Jost, M., Müller, M., Mühlberger, A., & Pauli, P. (2018). Height simulation in a virtual reality cave system: Validity of fear responses and effects of an immersion manipulation. *Frontiers in Human Neuroscience*, *12*, 372. https://doi.org/10.3389/fnhum.2018.00372

Hepperle, D., Purps, C. F., Deuchler, J., & Wölfel, M. (2022). Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley. *Visual Computer*, *38*(4), 1227–1244. https://doi.org/10.1007/s00371-021-02151-0

Herbert, C., Sfaerlea, A., & Blumenthal, T. (2013). Your emotion or mine: Labeling feelings alters emotional face perception-an ERP study on automatic and intentional affect labeling. *Frontiers in Human Neuroscience*, *7*, 378. https://doi.org/10.3389/fnhum.2013.00378

Herrmann, M. J., Ehlis, A. C., Ellgring, H., & Fallgatter, A. J. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of Neural Transmission*, *112*(8), 1073–1081. https://doi.org/10.1007/s00702-004-0250-8

Hoeger, R. (1997). Speed of processing and stimulus complexity in low-frequency and high-frequency channels. *Perception*, *26*(8), 1039–1045. https://doi.org/10.1068/p261039

Ip, C., Wang, H., & Fu, S. (2017). Relative expertise affects N170 during selective attention to superimposed face-character images. *Psychophysiology*, *54*(7), 955–968. https://doi.org/10.1111/psyp.12862

Itier, R. J., & Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, *15*(8), 1261–1265. https://doi.org/10.1097/01.wnr.0000127827.73576.d8

Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuroreport*, *14*(16), 2035–2039. https://doi.org/10.1097/00001756-200311140-00006

Johnsdorf, M., Kisker, J., Gruber, T., & Schöne, B. (2023). Comparing encoding mechanisms in realistic virtual reality and conventional 2D laboratory settings: Event-related potentials in a repetition suppression paradigm. *Frontiers in Psychology*, *14*, 1051938. https://doi.org/10.3389/fpsyg.2023.1051938

Johnston, P., Molyneux, R., & Young, A. W. (2015). The N170 observed "in the wild": Robust event-related potentials to faces in cluttered dynamic visual scenes. *Social Cognitive and Affective Neuroscience*, *10*(7), 938–944. https://doi.org/10.1093/scan/nsu136

Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., Kaufman, J., & Young, A. W. (2017). Temporal and spatial localization of prediction-error signals in the visual brain. *Biological Psychology*, *125*, 45–57. https://doi.org/10.1016/j.biopsycho.2017.02.004

Kamble, K., & Sengupta, J. (2023). A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals. *Multimedia Tools and Applications*, *82*(18), 27269–27304. https://doi.org/10.1007/s11042-023-14489-9

Key, A. P., & Corbett, B. A. (2020). The unfulfilled promise of the N170 as a social biomarker. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(3), 342–353. https://doi.org/10.1016/j.bpsc.2019.08.011

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, *18*(1), 156–168. https://doi.org/10.1006/nimg.2002.1323

Kirasirova, L. A., Zakharov, A. V., Morozova, M. V., Kaplan, A. Y., & Pyatin, V. P. (2021). Erp correlates of emotional face processing in virtual reality. *Opera Medica et Physiologica*, *8*(3), 12–19. https://doi.org/10.24412/2500-2295-2021-3-12-19

Kisker, J., Gruber, T., & Schöne, B. (2021). Virtual reality experiences promote autobiographical retrieval mechanisms: Electrophysiological correlates of laboratory and virtual experiences. *Psychological Research*, *85*(7), 2485–2501. https://doi.org/10.1007/s00426-020-01417-x

Kisker, J., Lange, L., Flinkenflügel, K., Kaup, M., Labersweiler, N., Tetenborg, F., Ott, P., Gundler, C., Gruber, T., Osinsky, R., & Schöne, B. (2021). Authentic fear responses in virtual reality: A mobile EEG study on affective, behavioral and electrophysiological correlates of fear. *Frontiers in Virtual Reality*, *2*, 716318. https://doi.org/10.3389/frvir.2021.716318

Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application*, *11*. https://doi.org/10.1146/annurev-statistics-040722-053607

Kloth, N., Itier, R. J., & Schweinberger, S. R. (2013). Combined effects of inversion and feature removal on N170 responses elicited by faces and car fronts. *Brain and Cognition*, *81*(3), 321–328. https://doi.org/10.1016/j.bandc.2013.01.002

Kuefner, D., de Heering, A., Jacques, C., Palmero-Soler, E., & Rossion, B. (2010). Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood. *Frontiers in Human Neuroscience*, *3*, 67. https://doi.org/10.3389/neuro.09.067.2009

Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research*, *1123*(1), 179–187. https://doi.org/10.1016/j.brainres.2006.09.031

Marmolejo-Ramos, F., Barrera-Causil, C., Kuang, S., Fazlali, Z., Wegener, D., Kneib, T., De Bastiani, F., & Martinez-Flórez, G. (2023). Generalised exponential-gaussian distribution: A method for neural reaction time analysis. *Cognitive Neurodynamics*, *17*(1), 221–237. https://doi.org/10.1007/s11571-022-09813-2

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Miller, E. J., Foo, Y. Z., Mewton, P., & Dawel, A. (2023). How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. *Computers in Human

Behavior Reports, *10*, 100283. https://doi.org/10.1016/j.chbr.2023.100283

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., & Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Bio-Medical Engineering*, *62*(11), 2553–2567. https://doi.org/10.1109/TBME.2015.2481482

Myllyneva, A., & Hietanen, J. K. (2015). There is more to eye contact than meets the eye. *Cognition*, *134*, 100–109. https://doi.org/10.1016/j.cognition.2014.09.011

Nasr, S. (2010). Differential impact of attention on the early and late categorization related human brain potentials. *Journal of Vision*, *10*(11), 18. https://doi.org/10.1167/10.11.18

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Newman, M., Gatersleben, B., Wyles, K. J., & Ratcliffe, E. (2022). The use of virtual reality in environment experiences and the importance of realism. *Journal of Environmental Psychology*, *79*, 101733. https://doi.org/10.1016/j.jenvp.2021.101733

Ojeda, A., Bigdely-Shamlo, N., & Makeig, S. (2014). MoBILAB: An open source toolbox for analysis and visualization of mobile brain/body imaging data. *Frontiers in Human Neuroscience*, *8*, 121. https://doi.org/10.3389/fnhum.2014.00121

Ospina, R., & Marmolejo-Ramos, F. (2019). Performance of some estimators of relative variability. *Frontiers in Applied Mathematics and Statistics*, *5*(43). https://doi.org/10.3389/fams.2019.00043

Pan, X., & Hamilton, A. F. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, *9*, 660. https://doi.org/10.3389/fnhum.2015.00660

Pfabigan, D. M., Sailer, U., & Lamm, C. (2015). Size does matter! Perceptual stimulus properties affect event-related potentials during feedback processing. *Psychophysiology*, *52*(9), 1238–1247. https://doi.org/10.1111/psyp.12458

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, *6*(4), 486–494. https://doi.org/10.1093/scan/nsq068

Puce, A., Smith, A., & Allison, T. (2000). ERPs evoked by viewing facial movements. *Cognitive Neuropsychology*, *17*(1–3), 221–239. https://doi.org/10.1080/026432900380580

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6539–6544. https://doi.org/10.1073/pnas.0812589106

Rossion, B. (2014a). Understanding face perception by means of human electrophysiology. *Trends in Cognitive Sciences*, *18*(6), 310–318. https://doi.org/10.1016/j.tics.2014.02.013

Rossion, B. (2014b). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience (Elite Edition)*, *6*(2), 258–307. https://doi.org/10.2741/e706

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*(4), 1959–1979. https://doi.org/10.1016/j.neuroimage.2007.10.011

Rossion, B., & Jacques, C. (2011). The N170: understanding the time course of face perception in the human brain. In E. Kappenman & S. Luck (Eds.), *The oxford handbook of event-related potential components* (pp. 115–142). Oxford University Press.

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

Rubo, M., & Gamer, M. (2018). Virtual reality as a proxy for real-life social attention? Eye Tracking Research and Applications Symposium (ETRA) https://doi.org/10.1145/3204493.3207411

Ruiz-Soler, M., & Beltran, F. S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research*, *70*(4), 273–292. https://doi.org/10.1007/s00426-005-0215-z

Sagehorn, M., Johnsdorf, M., Kisker, J., Sylvester, S., Gruber, T., & Schöne, B. (2023). Real-life relevant face perception is not captured by the N170 but reflected in later potentials: A comparison of 2D and virtual reality stimuli. *Frontiers in Psychology*, *14*, 1050892. https://doi.org/10.3389/fpsyg.2023.1050892

Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, *7*, 45003. https://doi.org/10.1038/srep45003

Schöne, B., Kisker, J., Lange, L., Gruber, T., Sylvester, S., & Osinsky, R. (2023). The reality of virtual reality. *Frontiers in Psychology*, *14*, 1093014. https://doi.org/10.3389/fpsyg.2023.1093014

Schöne, B., Kisker, J., Sylvester, R. S., Radtke, E. L., & Gruber, T. (2021). Library for universal virtual reality experiments (luVRe): A standardized immersive 3D/360° picture and video database for VR based research. *Current Psychology*, *42*(7), 5366–5384. https://doi.org/10.1007/s12144-021-01841-1

Schöne, B., Sylvester, R. S., Radtke, E. L., & Gruber, T. (2021). Sustained inattentional blindness in virtual reality and under conventional laboratory conditions. *Virtual Reality*, *25*(1), 209–216. https://doi.org/10.1007/s10055-020-00450-w

Schöne, B., Wessels, M., & Gruber, T. (2019). Experiences in virtual reality: A window to autobiographical memory. *Current Psychology*, *38*(3), 715–719. https://doi.org/10.1007/s12144-017-9648-y

Schwaninger, A., Lobmaier, J. S., Wallraven, C., & Collishaw, S. (2009). Two routes to face perception: Evidence from psychophysics and computational modeling. *Cognitive Science*, *33*(8), 1413–1440. https://doi.org/10.1111/j.1551-6709.2009.01059.x

Shamay-Tsoory, S. G., & Mendelsohn, A. (2019). Real-life neuroscience: An ecological approach to brain and behavior research. *Perspectives on Psychological Science*, *14*(5), 841–859. https://doi.org/10.1177/1745691619856350

Shangguan, P., Qiu, T., Liu, T., Zou, S., Liu, Z., & Zhang, S. (2020). Feature extraction of EEG signals based on functional data analysis and its application to recognition of driver fatigue

state. *Physiological Measurement*, *41*(12), 125004. https://doi.org/10.1088/1361-6579/abc66e

Snow, J. C., & Culham, J. C. (2021). The treachery of images: How realism influences brain and behavior. *Trends in Cognitive Sciences*, *25*(6), 506–519. https://doi.org/10.1016/j.tics.2021.02.008

Sollfrank, T., Kohnen, O., Hilfiker, P., Kegel, L. C., Jokeit, H., Brugger, P., Loertscher, M. L., Rey, A., Mersch, D., Sternagel, J., Weber, M., & Grunwald, T. (2021). The effects of dynamic and static emotional facial expressions of humans and their avatars on the EEG: An ERP and ERD/ERS study. *Frontiers in Neuroscience*, *15*, 651044. https://doi.org/10.3389/fnins.2021.651044

Stolz, C., Endres, D., & Mueller, E. M. (2019). Threat-conditioned contexts modulate the late positive potential to faces—A mobile EEG/virtual reality study. *Psychophysiology*, *56*(4), e13308. https://doi.org/10.1111/psyp.13308

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*(4), 505–511. https://doi.org/10.1038/nn1864

Ullah, S., & Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, *13*(1), 43. https://doi.org/10.1186/1471-2288-13-43

Van Den Oever, F., Gorobets, V., Saetrevik, B., Fjeld, M., & Kunz, A. (2022). Comparing visual search between physical environments and VR. Proceedings - 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct, ISMAR-Adjunct 2022, 411–416. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00089

Wang, J., & Wang, M. (2021). Review of the emotional feature extraction and classification using EEG signals. *Cognitive Robotics*, *1*(1), 29–40. https://doi.org/10.1016/j.cogr.2021.04.001

Wang, J. L., Chiou, J. M., & Müller, H. G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*, 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

Wheatley, T., Weinberg, A., Looser, C., Moran, T., & Hajcak, G. (2011). Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP. *PLoS One*, *6*(3), e17960. https://doi.org/10.1371/journal.pone.0017960

Wheaton, K. J., Thompson, J. C., Syngeniotis, A., Abbott, D. F., & Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *NeuroImage*, *22*(1), 277–288. https://doi.org/10.1016/j.neuroimage.2003.12.043

Xu, C., Demir-Kaymaz, Y., Hartmann, C., Menozzi, M., & Siegrist, M. (2021). The comparability of consumers' behavior in virtual reality and real life: A validation study of virtual reality based on a ranking task. *Food Quality and Preference*, *87*, 104071. https://doi.org/10.1016/j.foodqual.2020.104071

Yi, Y., Billor, N., Liang, M., Cao, X., Ekstrom, A., & Zheng, J. (2022). Classification of EEG signals: An interpretable approach using functional data analysis. *Journal of Neuroscience Methods*, *376*, 109609. https://doi.org/10.1016/j.jneumeth.2022.109609

Zhang, J., Siegle, G. J., D'Andrea, W., and Krafty, R. T. (2019). Interpretable principal components analysis for multilevel multivariate functional data, with application to EEG experiments.

Arvix *15261*. 1–33. http://arxiv.org/abs/1909.08024 (accessed January 8, 2024).

Zhang, Y., Wang, C., Wu, F., Huang, K., Yang, L., & Ji, L. (2020). Prediction of working memory ability based on EEG by functional data analysis. *Journal of Neuroscience Methods*, *333*, 108552. https://doi.org/10.1016/j.jneumeth.2019.108552

Zion-Golumbic, E., & Bentin, S. (2007). Dissociated neural mechanisms for face detection and configural encoding: Evidence from N170 and induced gamma-band oscillation effects. *Cerebral Cortex*, *17*(8), 1741–1749. https://doi.org/10.1093/cercor/bhl100

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Time-by-amplitude plots of the root-mean-squared ERP averaged over all electrodes per condition (mean PAL) (panel a: PC, panel b: VR) for the selection of appropriate time windows for all stimulus-locked ERP components. Highlighted sections mark the time windows for N170 PC (145–191 ms), N170 VR (112–150 ms), L1 (200–450 ms), and L2 (600–1350 ms). The gray line illustrates the participant-based median ERP (median PAL).

Figure S2. Cumulative distribution functions (CDF) illustrating the distributions of response times [panel a], N170 latencies [panel b], and N170 amplitudes [panel c] for all stimulus types in both modalities.

Figure S3. Time-by-amplitude plots of the mean face–object difference (FOD) for both modalities for all regional means (frontal, temporal left, temporal right, posterior left, posterior right, and parieto-occipital). Time windows of interest are highlighted in red (N170 VR) and blue (N170 PC).

Figure S4. Topographies of the mean face–object difference (FOD) for the N170 time windows for both modalities (PC: 145–191 ms; VR: 112–150 ms).

Table S1. Descriptive statistics (mean, standard deviation, confidence interval) for response times, N170 latency and N170 amplitude (standard approach) for both modalities and all stimulus types.

Table S2. Pairwise comparison of response times, N170 latency and N170 amplitude (standard approach) within and between modalities.

Table S3. Mean amplitudes, standard deviations and confidence intervals for N170 component per modality, stimulus type and regional mean.

Table S4. Mean amplitudes, standard deviations and confidence intervals for L1 component per modality, stimulus type and regional mean.

Table S5. Mean amplitudes, standard deviations and confidence intervals for L2 component per modality, stimulus type and regional mean.

Table S6. Mean amplitudes, standard deviations and confidence intervals for pre-response component per modality, stimulus type and regional mean.

Table S7. Mean amplitudes, standard deviations and confidence intervals for post-response component per modality, stimulus type and regional mean.

Table S8. Results of 2×4×6 repeated-measurements ANOVA (rmANOVA) with the within-subject factors "modality" (VR vs. PC), "stimulus type" (face vs. car vs. face blurred vs. car blurred) and "regional mean" (frontal, temporal left, temporo-posterior left, temporal right, temporo-posterior right, parieto-occipital) for N170, L1, L2, pre-response and post-response components.