



Face-Specific Activity in the Ventral Stream Visual Cortex Linked to Conscious Face Perception

Wenlu Li^{1,2} · Dan Cao¹ · Jin Li³ · Tianzi Jiang^{1,2,4,5}

Received: 6 October 2023 / Accepted: 25 November 2024

© Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences 2024

Abstract When presented with visual stimuli of face images, the ventral stream visual cortex of the human brain exhibits face-specific activity that is modulated by the physical properties of the input images. However, it is still unclear whether this activity relates to conscious face perception. We explored this issue by using the human intracranial electroencephalography technique. Our results showed that face-specific activity in the ventral stream visual cortex was significantly higher when the subjects subjectively saw faces than when they did not, even when face stimuli were presented in both conditions. In addition, the face-specific neural activity exhibited a more reliable neural response and increased posterior-anterior direction information transfer in the “seen” condition than the “unseen” condition. Furthermore, the face-specific neural activity was significantly correlated with performance. These findings support the view that face-specific activity in the ventral stream visual cortex is linked to conscious face perception.

Keywords Face-specific · Conscious face perception · Ventral stream visual cortex · Intracranial EEG

Introduction

Face perception is a fundamental cognitive process that allows humans to identify faces in their surroundings and engage in social interactions more effectively. Extensive research has identified a specific region in the ventral stream visual cortex of the brain that exhibits heightened activity in response to faces compared with other visual stimuli [1, 2]. These studies have shown that face-specific activity in the ventral stream visual cortex is related to face perception at the visual processing level. However, it remains unclear whether this activity is linked to face perception at the conscious level (e.g., the subjective perception of faces).

Comparing neural activity in different perceptual states at the perceptual threshold provides direct evidence for the involvement of these signals in conscious face perception [3–7]. Some studies have shown that when subjects are presented with face images, the face-specific activity in the ventral stream visual cortex increases when they report seeing a face (“seen” condition) compared with when they did not (“unseen” condition), suggesting that the face-specific activity in this area is involved in conscious face perception [4, 7, 8]. However, not all studies have supported this idea [5, 6]. The reasons for reaching different conclusions may be differences in experimental paradigms, signal acquisition methods, or analysis methods. These discrepancies highlight the need for further investigation, especially by using invasive techniques such as human intracranial electroencephalography (ECoG), which provides direct measurement of brain electrophysiological responses and offers high spatial and temporal resolution. In addition, at the level of

✉ Jin Li
7049@cnu.edu.cn

✉ Tianzi Jiang
jiangtz@nlpr.ia.ac.cn

¹ Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

² School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Psychology, Capital Normal University, Beijing 100048, China

⁴ Research Center for Augmented Intelligence, Zhejiang Lab, Hangzhou 311100, China

⁵ Xiaoxiang Institute for Brain Health and Yongzhou Central Hospital, Yongzhou 425000, China

conscious face perception, the interaction of face-specific activity at different locations in this area remains incompletely understood. Moreover, there is a lack of correlation studies between trial-wise face-specific neural activity and face perception performance (e.g., response latency and detection accuracy). Studies that address these could provide neurophysiological support for conscious face perception and contribute to our understanding of how neural activity influences perception performance.

To investigate whether the local face-specific activity in the ventral stream visual cortex is linked to conscious face perception, we compared the activity recorded by ECoG in the “seen” and “unseen” conditions. Our findings demonstrate that face-specific activity is significantly higher when subjects see faces. Also, based on single trials, the classification accuracy of predicting whether a face was seen or not surpassed chance. In addition, we recorded more reliable neural responses in the “seen” condition than in the “unseen” condition. We also found a significant increase in information transfer in the posterior-anterior direction in the “seen” condition. Furthermore, there was a significant correlation between face-specific neural activity in this area and human face perception performance.

Materials and Methods

Human Brain Data

We used publicly available ECoG recordings shared in a freely available library at <https://searchworks.stanford.edu/view/zk881ps0522> [9]. All the patients participated voluntarily after providing informed written consent under experimental protocols approved by the Institutional Review Board (IRB) of the University of Washington (#12193). All patient data was anonymized according to the IRB protocol, in accordance with the HIPAA (Health Insurance Portability and Accountability Act of 1996) mandates. The data used in this paper originally appeared in a manuscript published in PLoS Computational Biology in 2016 [10].

Recordings

From the above library, we included only those subjects who had completed behavioral and ECoG data from “faces_basic” and “faces_noisy” paradigms; 4 subjects met this criterion (2 males, average 35.3 years old). All of these were epileptic patients at Harborview Hospital in Seattle, Washington. Experiments were performed at the bedside, using SynAmps2 amplifiers (Neuroscan, El Paso, USA) in parallel with clinical recording. Stimuli were presented with a monitor using the general-purpose BCI2000 stimulus and acquisition program (interacting with proprietary Neuroscan

software), which also recorded the behavioral parameters and cortical data. Subdural grid and platinum arrays (4 mm diameter, 2.3 mm exposed, and 10 mm interelectrode distances; Ad-Tech, Racine, USA) were placed on their occipitotemporal cortex for clinical monitoring and localization of epileptic foci.

Experimental Design

The details of the experimental paradigm are described by Miller *et al.* (2017). In brief, participants completed a “faces_basic” task and a “faces_noisy” task. In the “faces_basic” task, they watched grayscale pictures of faces and houses, which were displayed in random order, one picture in a trial for 400 ms, with a 400 ms blank screen inter-stimulus interval. Subjects were asked to verbally report a simple target (an upside-down house) to maintain their fixation on the stimuli. Only the presentations of upright images were used in the analysis. A total of 300 stimuli were shown to each subject with 100 stimuli in each run. The stimuli were balanced for the number of face and house pictures (Fig. 1A left). In the “faces_noisy” task, the subjects were asked to perform a face detection task using images of faces and houses. The perceptibility of stimuli was manipulated using the “phase-scrambling” method, which involves systematically varying the spatial phase of the stimuli by adding noise (ranging from 0 to 100%, in 5% increments) [11]. This process guarantees that the Fourier amplitude spectra of these stimuli were consistent across conditions, thereby removing the possibility for differences in neural activity between conditions to be attributed to low-level image characteristics. Each image was shown for 1000 ms with no inter-stimulus interval. The subjects were instructed to press the “F” key if they believed the picture to be that of a face. A total of 630 stimuli were shown to each subject, 105 stimuli for each run. The stimuli were randomly interleaved and balanced for noise level and number of face and house images (Fig. 1A right).

Statistical Analysis

Significant differences between the experimental conditions were analyzed by *t*-tests, covariance, and non-parametric cluster-based permutation tests [12]; differences were considered statistically significant at $P < 0.05$. Analysis of covariance was used to compare means between two or more groups while controlling for the effects of covariates. When using the non-parametric cluster-based permutation tests for significance, the time windows with statistical values above a threshold were selected and combined into clusters. Cluster-level statistics were computed by taking the sum of the *t* values within a cluster. To obtain a distribution of cluster-level statistics under the null hypothesis, 1000 surrogates

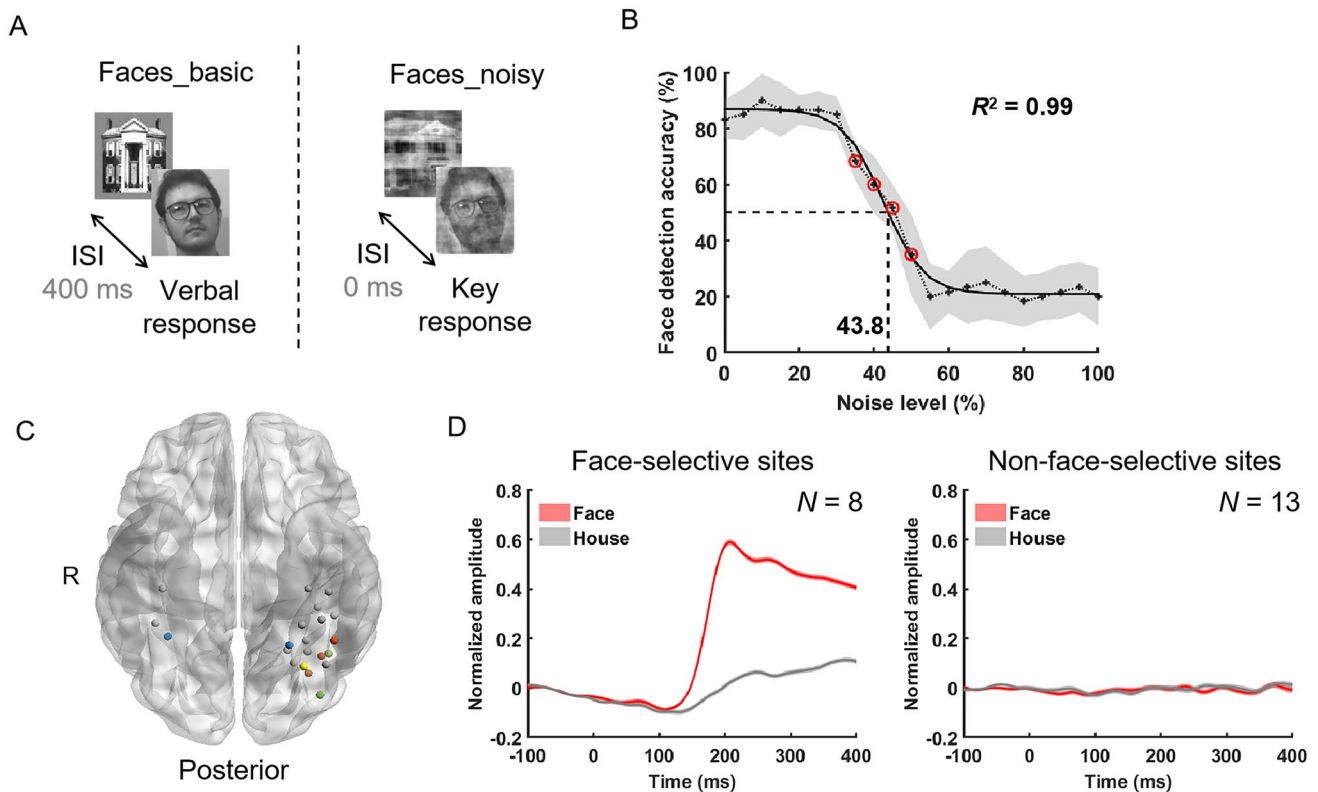


Fig. 1 Experimental paradigm, behavioral performance, and electrode localization. **A** Left: “faces_basic” task. Grayscale images of faces and houses are presented in a random order for 400 ms each, with a 400 ms inter-stimulus interval (ISI) of a blank screen. Right: “faces_noisy” task. Phase-scrambled pictures of faces and houses are displayed in random order for 1 s each, without an ISI. The subjects are instructed to press a key when they believe the picture depicts a face. For copyright reasons, the facial images presented are not the actual images viewed by the subjects. The face images shown here were sourced from a publicly available dataset, UMIST_Face (<http://images.ee.umist.ac.uk/>), and were solely used for illustrating the experimental paradigm, but not used in the experiments. **B** The detection accuracy demonstrates a sigmoidal pattern of change as the noise

level varies linearly. The crosses represent the real face detection accuracy, connected by black dotted lines, and the shadow represents the SEM of the detection accuracy. The solid black line represents a curve fit to real data using the sigmoid function and the gray shadow represents the SEM of detection accuracy across subjects. The red circle represents the noise level at the perception threshold (35%–50%). **C** Ventral view of the MNI cortex displaying the face-selective sites (colored dots) and non-face-selective sites (gray dots) of all four subjects. Each dot represents a site and the face-selective sites of different subjects are distinguished by different colors (S1: blue; S2: green; S3: yellow; S4: red). **D** BGA of face-selective sites (left, $N = 8$) and non-face-selective sites (right, $N = 13$) in the “faces_basic” task.

were created by randomly permuting the condition labels, and then the maximum cluster-level statistic in each permutation was extracted. The non-parametric statistical significance was determined by calculating the proportion of the surrogates within the permutation distribution that exceeded the observed cluster-level statistics.

Preprocessing the ECoG Data

The ECoG data preprocessing process was described in detail in our previous work [13]. In brief, the ECoG data underwent several preprocessing steps, including filtering for line noise and spiky signals, re-referencing, baseline correction, and removal of signal-averaged event-related potentials. The potentials of each trial were analyzed through time-frequency decomposition using a Morlet wavelet with a

1 Hz step. Then, the power values within each frequency (1 Hz) were logarithmically transformed, and were Z-scored by subtracting the mean of the baseline values and dividing the SD of the baseline values. For the ECoG data collected during the “faces_basic” task, we extracted each epoch within the window from -200 to 600 ms relative to the stimulus onset. A baseline correction was applied between -200 and 0 ms surrounding the stimulus onset. For the ECoG data collected during the “faces_noisy” task, since there was no inter-stimulus interval, we extracted each epoch from 0 to 1000 ms following the stimulus onset. Then, because previous studies have shown that the response latency of face-selective broadband gamma activity (BGA) at 30 Hz– 150 Hz typically occurs after 100 ms, we applied a baseline correction within 0 to 100 ms [11, 13]. For each trial, the power spectral values were averaged across the broadband gamma

frequency of 30 Hz–150 Hz, and smoothed with a 50 ms width Gaussian window.

Electrode Localization and Face-selective Sites

The electrode positions and the brain regions in which they were located are described in the database. For simplicity, the word “site” is used to represent the electrode contact location. We analyzed each site in the temporal ventral visual cortex, which includes the temporal pole, para-hippocampal gyrus, inferior temporal gyrus, middle temporal gyrus, fusiform gyrus, lingual gyrus, and inferior occipital gyrus. In the “faces_basic” task, if the BGA of the face image within 100 to 300 ms after stimulus onset was significantly greater than the BGA of the baseline (–200 to 0 ms) and housed images based on a paired *t*-test ($P < 0.05$), then the site was defined as a “face-selective” site with 8 face-selective sites included in the subsequent analysis. The details of each face-selective site are shown in Table 1. Face-selective sites were mainly located in the fusiform gyrus. The remaining sites in the fusiform gyrus were defined as non-face-selective sites ($n = 13$). Using BrainNet Viewer, a toolbox in MatLab (The MathWorks, Inc., Natick, MA), we displayed the face-selective sites (colored dots) and the non-face-selective sites (gray dots) on the ventral view of a Montreal Neurological Institute (MNI) cortex in Fig. 1C, with each dot representing a site and the face-selective sites of different subjects distinguished by different colors (S1: blue; S2: green; S3: yellow; S4: red).

Classification Analysis

To determine the presence of information regarding “seen”/“unseen” conditions (participant reported seeing faces or not) *via* a single-trial BGA, we applied a binary nearest neighbor classification using a *k* value of 1 [14]. The single-trial neural responses of all face-selective sites

were divided into odd trials for training and even trials for testing. We calculated the Euclidean distance between the 100 ms time windowed data from each test trial and each training trial. The single-trial BGA from the test trial was assigned to the condition with the nearest neighbor classifier. This process was repeated for all possible pairs of trials, sliding the time windows in 5 ms increments between 100 and 900 ms after the onset. To statistically test the classification accuracy, we applied a permutation test. Specifically, we randomly permuted the labels of the conditions for all trials, and then applied the same classification procedure to these permuted trials. We extracted the maximum classification accuracy across the 100 to 900 ms time window. This permutation procedure was repeated 1000 times, and the mean value was defined as the permuted classification accuracy.

Granger Causality Analysis

To investigate the directionality of information transfer in the BGA, we applied spectral Granger causality (GC) analysis. This method quantifies the prediction error of the signal in the frequency domain by incorporating another time series, and is able to examine the direction and intensity of causal relationships between signals. Specifically, the GC analysis was performed between site pairs consisting of two face-selective sites from the same hemisphere of the same subject. Our analysis included a total of four site pairs, of which subject S2 had one site pair and subject S4 had three site pairs (Table 1). For each site pair, we subtracted the trial-wise mean from each trial to remove non-stationarity in the mean [15] and fitted the resulting data to an autoregressive model to compute the spectral GC. To determine the appropriate model order for each pair, we used the multivariate Granger causality MatLab toolbox [16]. The GC index was computed for both directions, i.e., from anterior to posterior and from posterior to anterior, during the time window from 200 to 700 ms for seen and unseen trials separately. To assess the statistical significance of the spectral GC, we applied an F-test and corrected for multiple null hypotheses using a significance level of $P = 0.05$.

Partial Correlation Analysis Between the BGA and Conscious Face Perception Performance

To investigate the correlation between the BGA and face perception performance in the “seen” condition, while accounting for the influence of noise levels, we conducted a partial correlation analysis. The partial correlation coefficient was calculated by the following formula:

Table 1 MNI coordinates and located brain region of face-selective sites

Site	Subject	MNI coordinates (x, y, z)	Located brain region
1	S1	–24, –52, –10	FG
2	S1	33, –47, 12	FG
3	S2	–42, –56, –8	FG
4	S2	–38, –78, 0	IOG
5	S3	–31, –63, –21	FG
6	S4	–33, –67, –14	FG
7	S4	–39, –58, –15	FG
8	S4	–45, –50, –17	FG

FG, fusiform gyrus; IOG, inferior occipital gyrus.

$$r_{xy|z} = \frac{r_{xy} - (r_{xz} * r_{yz})}{\sqrt{(1-r_{xz}^2) * (1-r_{yz}^2)}}$$

where $r_{xy|z}$ is the partial correlation coefficient between variable x and variable y after controlling for variable z . r_{xy} is the original correlation coefficient between x and y , r_{xz} is the correlation coefficient between x and z , and r_{yz} is the correlation coefficient between y and z .

The analysis focused on the noise level range from 0% to 55%, as the face detection accuracy remained consistently low beyond this threshold. When calculating the correlation between the peak latency of the BGA and the response latency of the subjects, for each trial of each face-selective site, we extracted the latency at which the maximum BGA occurred as well as the subject response latency in this trial. Since conscious perception usually occurs after 200 ms [17] and since the BGA enhancement in the “seen” condition occurred within the time window of 269 to 691 ms, trials with a subject response latency <200 ms and trials with a peak BGA latency <200 ms or >700 ms were excluded from the analysis. In calculating the correlation between the BGA amplitude and the detection accuracy, because it was not feasible to calculate detection accuracy at the individual trial level, we extracted the detection accuracy for each subject at each noise level. For each face-selective site, at each noise level, we averaged the BGA across trials in the “seen” condition and extracted the average power values within 200 to 700 ms of the average BGA as its amplitude.

Results

Human Behavioral Results

First, we examined the impact of noise on face detection accuracy. A total of four subjects (2 males, average 35.3 years old) participated in both the “faces_basic” task and the “faces_noisy” task and provided behavioral data. In the “faces_basic” task, the subjects viewed grayscale pictures of faces and houses without any added noise (Fig. 1A, left). During the “faces_noisy” experiment, the subjects were presented with noisy face and house images (Fig. 1A, right). Each noise level comprised 15 trials showing face images and 15 trials showing house images. The subjects indicated their perception of a face image by pressing the “F” key on the keyboard. Fig. 1B presents the average face detection accuracy (number of detected face images divided by the total number of face images) at different noise levels. The detection accuracy showed a sigmoidal change pattern as the noise level varied linearly. To define the perceptual threshold, we used a sigmoid function to fit the behavioral

measurement curve against the noise level of the face images ($R^2 = 0.99$). On the fitted curve, when the face detection accuracy reached 50%, the noise level was determined to be 43.8%. To include a more comprehensive dataset for analysis, we selected the two nearest noise levels below 43.8% (40% and 35%) and the two nearest noise levels >43.8% (45% and 50%) as the perceptual threshold (Fig. 1B, indicated by red circles). Within these four noise levels, the behavioral measurement curve exhibited a sharp transition, indicating the occurrence of a transition into perceptual awareness [18]. The positions of the face-selective sites for the four subjects are displayed on a ventral view of the MNI cortex, with each gray dot representing a non-face-selective site and each colored dot representing a face-selective site, and the face-selective sites of different subjects are represented by different colors (Fig. 1C; S1: blue; S2: green; S3: yellow; S4: red). Table 1 provides detailed information about these face-selective sites. Considering the BGA of 30 Hz–150 Hz is generally considered to be related to cognition, and neural activity related to conscious perception is usually reported on this broadband gamma frequency [4–6, 13], our analysis focused on the BGA in the human brain. The BGA of face-selective sites and non-face-selective sites in the “faces_basic” are shown in Fig. 1D. Face-selective sites exhibited significantly higher activation in response to face images compared to house images, whereas non-face-selective sites did not exhibit this characteristic.

Ventral Stream Visual Cortex BGA Is Linked to Human Conscious Face Perception

To investigate the relationship between face-specific neural activity in the ventral stream visual cortex and conscious face perception, we compared the BGA of face-selective sites in the “seen” (participants reported seeing faces) and “unseen” conditions (participants did not report seeing faces) at the perceptual threshold. Results from a cluster-based permutation test revealed a significantly higher BGA in the “seen” condition than in the “unseen” condition (256–691 ms, $P = 0.0003$) (Fig. 2A). To ensure this difference was not due to trial selection bias, we examined the BGA of non-face-selective sites in the same brain region, which did not exhibit any significant difference (Fig. 2B). To further test whether these differences were caused by the “seen” and “unseen” conditions, we compared the average BGA amplitude values within 200–700 ms between the two conditions at individual noise levels. Even at the same level of noise, some faces are likely to be more visible than others. Therefore, to eliminate the influence of visibility, we applied a covariance analysis (ANCOVA) with visibility as a control variable. The visibility of each face image was defined as the probability that four subjects detected a face in the image (e.g., when all four subjects detected a face in the image, the visibility of the image was assigned a value

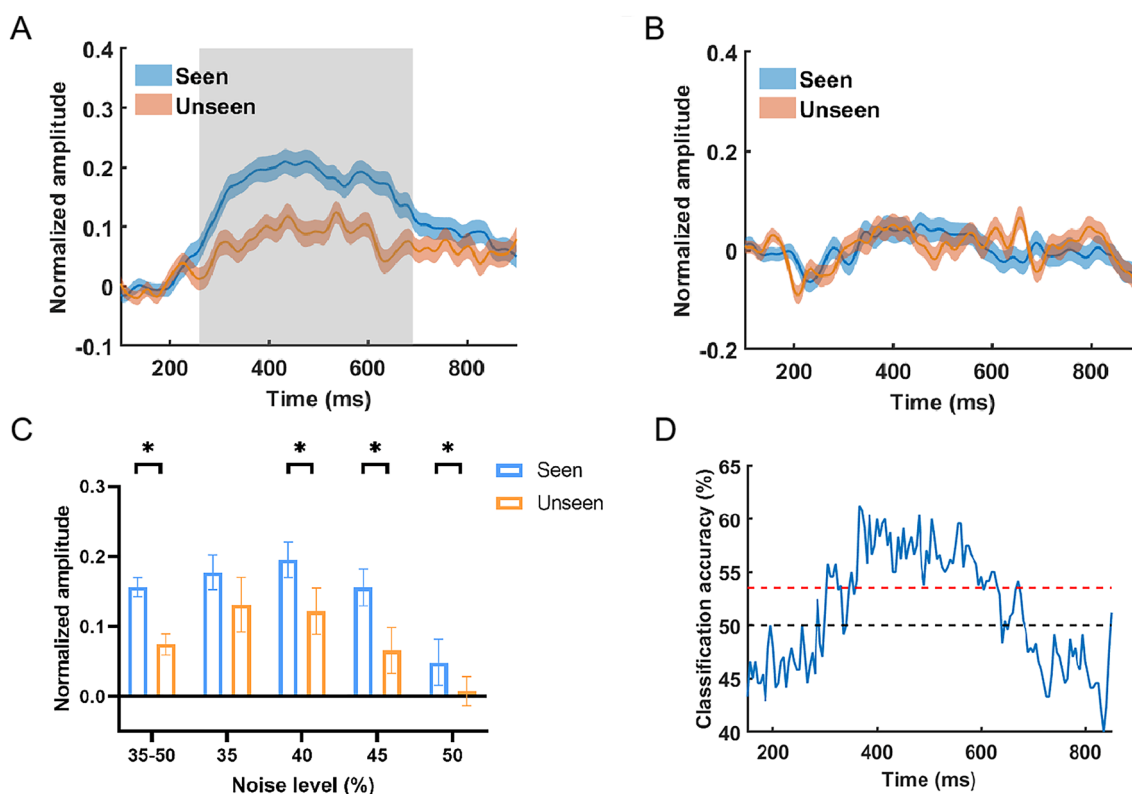


Fig. 2 Broadband gamma activity (BGA) of face-selective sites in the “seen” and “unseen” conditions. **A** BGA comparison between the “seen” and “unseen” conditions reveals a significantly higher BGA in the “seen” condition (256–691 ms, $P = 0.0003$, cluster-based permutation test). The blue line represents the trial-averaged BGA and the blue shadow represents the SEM of BGA across trials in the “seen” condition. The orange line represents the trial-averaged BGA and the orange shadow represents the SEM of BGA across trials in the “unseen” condition. The gray shadow represents the period with significant differences between “seen” and “unseen” conditions. **B** No significant difference in BGA is found between the “seen” and “unseen” conditions for the non-face-selective sites. The meaning of the legend is consistent with Fig. 2A. **C** Average BGA amplitude val-

ues within 200–700 ms in the “seen” condition are greater than those in the “unseen” condition at individual noise levels with visibility as a control variable (ANCOVA, 35%–50% noise level: $P = 0.041$, N of “seen” = 268, N of “unseen” = 212; 35%: $P = 0.11$, N of “seen” = 85, N of “unseen” = 35; 40%: $P = 0.038$, N of “seen” = 74, N of “unseen” = 46; 45%: $P = 0.021$, N of “seen” = 65, N of “unseen” = 55; 50%: $P = 0.047$, N of “seen” = 44, N of “unseen” = 76). $*P < 0.05$. The y-axis represents the mean activation values, and the error bar represents the SEM. **D** Time course of the “seen”/“unseen” classification accuracy based on single-trial BGA. The classification accuracy surpasses the chance level (black dashed line, 50%) and the permuted classification accuracy (red dashed line, 53.5%).

Table 2 Trial numbers of “seen” and “unseen” conditions of individual noise levels

Noise level (%)	Seen	Unseen
35	85	35
40	74	46
45	65	55
50	44	76
Total (35–50)	268	212

of 1). The trial numbers of “seen” and “unseen” conditions at individual noise levels are shown in Table 2. The results showed that the BGA amplitude in the “seen” condition was consistently greater than that in the “unseen” condition across 40%–50% individual noise levels (ANCOVA, 35% – 50% noise level: $P = 0.041$; 35%: $P = 0.11$; 40%: $P = 0.038$; 45%: $P = 0.021$; 50%: $P = 0.047$) (Fig. 2C). At the 35% noise level,

even though the BGA amplitude in the “seen” condition was greater than that in the “unseen” condition, there was no significant difference between the two conditions. This may be due to the relatively smaller sample size in the “unseen” condition. Furthermore, applying a nearest neighbor classification algorithm [14], we accurately predicted whether a face was seen or unseen based on single-trial BGA, surpassing chance level (black dashed line, 50%) and permuted classification accuracy (red dashed line, 53.5%) (Fig. 2D).

Ventral Stream Visual Cortex BGA Shows More Reliable Neural Response and Increased Posterior-anterior Information Transfer in Conscious Face Perception

To examine the reliability of neural response in conscious face perception, we applied a similarity analysis of neural activity for the two conditions separately. Specifically, the similarity was estimated by Spearman correlations between pairs of trials, with each trial containing BGA time series within 200–700 ms. For all trials in the same condition, we calculated pairwise correlation values between them (268 trials in the “seen” condition, resulting in $N = 268 \times 267 / 2 = 35,778$ similarity values; 212 trials in the “unseen” condition, resulting in $N = 212 \times 211 / 2 = 22,366$ similarity values; see Table 2 for details), and then we compared similarity values between conditions. The results showed that the similarity of face-specific neural activity in the “seen” condition was significantly greater than that in the “unseen” condition ($P < 0.0001$, two-tailed t -test). To rule out the possibility that this difference could be attributed to trial selection bias, we conducted similarity analyses on non-face-selective sites within the same brain region and did not find significant differences between the two conditions. This result suggests that face-specific neural activity has a more reliable response in the “seen” condition than in the “unseen” condition (Fig. 3A). Furthermore, to explore the directionality of information transfer in face perception, we conducted spectral GC analysis to examine the direction and intensity of causal relationships between signals. We found significant increases in GC values driven by BGA within 200–700 ms in the posterior-anterior direction in

the “seen” condition than in the “unseen” condition ($P = 0.019$, two-tailed t -test). Conversely, no significant difference in GC values was found in the anterior-posterior direction (Fig. 3B). These results demonstrated enhanced neural response reliability and neural interaction associated with conscious face perception, providing evidence for the idea that face-specific activity in the ventral stream visual cortex supports conscious face perception.

Ventral Stream Visual Cortex BGA Correlates with Performance in Human Face Perception

To investigate the relationship between BGA in the ventral stream visual cortex and face perception performance while eliminating the influence of noise levels, we examined the partial correlation between the BGA and face perception performance in the “seen” condition, setting noise level as the control variable. The analysis focused on the noise level ranging from 0% to 55%, as face detection accuracy remained consistently low beyond this threshold. First, we analyzed the correlation between the peak latency of the BGA and the response latency of the subjects at the individual trial level. Specifically, for each trial of each face-selective site, we extracted the latency at which the maximum BGA occurred and the subject response latency in this trial. A significant positive correlation was found between the peak latency of the BGA and the response latency of the subjects ($r = 0.44$, $P < 0.0001$, Spearman correlation) (Fig. 4A). Next, we analyzed the correlation between the amplitude of the BGA and detection accuracy of the subjects. Since the detection accuracy could not be calculated at the individual trial level, we extracted the detection accuracy

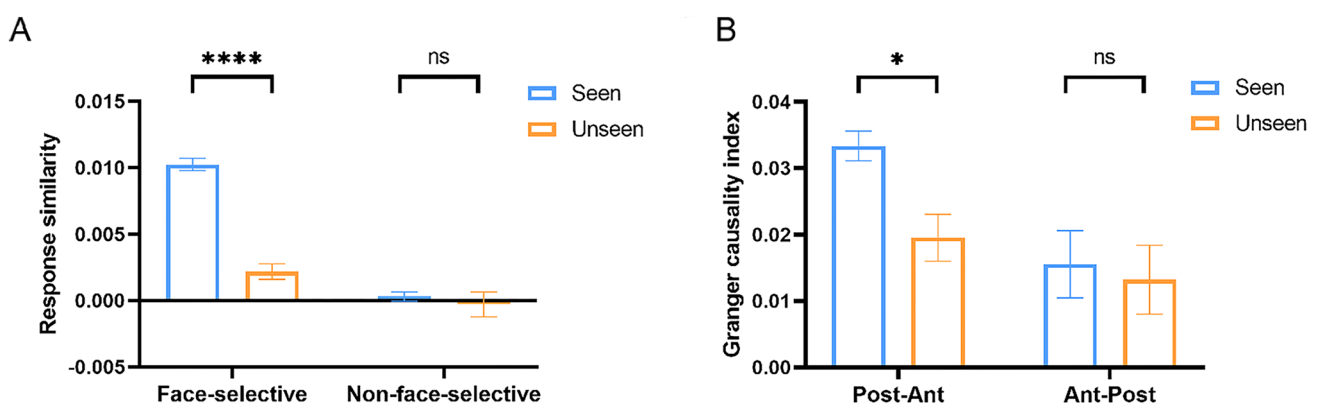


Fig. 3 Neural response reliability and information transfer in the “seen” and “unseen” conditions. **A** Response similarity analysis. The “seen” condition exhibits higher neural response similarity than the “unseen” condition (**** $P < 0.0001$, two-tailed t -test, N of “seen” = 35,778, N of “unseen” = 22,366), while no significant difference (ns) is found for the non-face-selective sites. The y-axis represents the similarity values, and the error bars represent the SEM. **B** Granger

causality (GC) index for two directions of information transfer. The “seen” condition exhibits stronger posterior-anterior GC than the “unseen” condition ($P = 0.019$, two-tailed t -test, N of “seen” = 4, N of “unseen” = 4); whereas no significant difference is found in the anterior-posterior direction (Post: posterior; Ant: anterior). * $P < 0.05$. The y-axis represents the mean activation values, and the error bars represent the SEM.

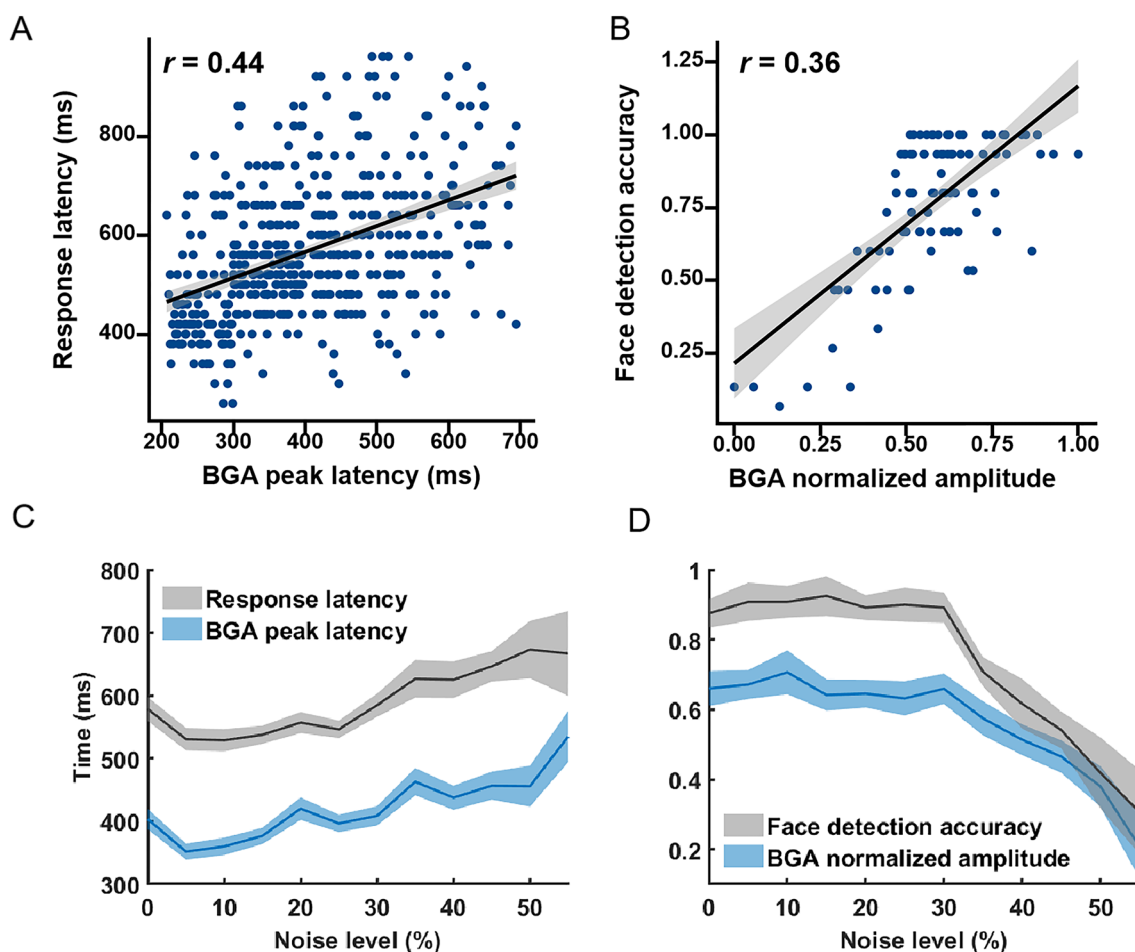


Fig. 4 Correlation between the BGA of the face-selective sites and face perception performance of the subjects. **A** Significant positive partial correlation between the peak latency of the BGA and the response latency of the subjects ($r = 0.44$, $P < 0.0001$, Spearman correlation). Each dot represents a trial of a face-selective site. The black line represents a linear fitting line for the x -axis and y -axis values and the gray shadow represents the 95% confidence interval. **B** Significant positive partial correlation between the amplitude of the BGA and the face detection accuracy of the subjects ($r = 0.36$, P

< 0.0001 , Spearman correlation). Each dot represents the noise level of a face-selective site. The black line represents a linear fitting line for the x -axis and y -axis values and the gray shadow represents the 95% confidence interval. **C** BGA peak latency and subject response latency exhibit similar nonlinear dependence on noise levels, with the response latency consistently lagging behind the peak latency. Gray and blue shadows represent the SEM. **D** BGA amplitude and face detection accuracy show similar nonlinear dependence on noise levels. Gray and blue shadows represent the SEM.

for each subject at the single noise level. For each face-selective site at each noise level, we averaged the BGA results across trials in the “seen” condition and extracted the average power values within 200–700 ms of the average BGA as its amplitude. A significant positive correlation was found between the amplitude of the BGA and the face detection accuracy of the subjects ($r = 0.36$, $P < 0.0001$, Spearman correlation) (Fig. 4B). Furthermore, BGA peak latency and subject response latency exhibited similar nonlinear dependence on noise levels, with the response latency consistently lagging behind the peak latency (Fig. 4C). The BGA amplitude and face detection accuracy showed similar patterns of dependence on noise levels, with both sharply decreasing beyond the 30% noise level (Fig. 4D). These findings

provide additional evidence supporting the link between the ventral stream visual cortex BGA and conscious face perception.

Discussion

We aimed to investigate whether the face-specific activity in the ventral stream visual cortex is linked to conscious face perception. To address the question, we compared the face-specific BGA in the ventral stream visual cortex when subjects saw faces *versus* when they did not. Our results showed that face-specific activity was significantly higher when seeing faces. The classification accuracy of predicting whether a

face was seen or not, based on the BGA of a single trial was higher than the chance level and than the permuted classification accuracy. In addition, the face-specific neural activity in the “seen” condition exhibited a more reliable neural response than in the “unseen” condition. Also, the intensity of information transfer in the posterior-anterior direction was significantly higher in the “seen” condition. Furthermore, we found a significant positive correlation between the peak latency of the BGA and the response latency of the subjects as well as between the amplitude of the BGA and the face detection accuracy of the subjects. Taken together, our results support the idea that face-specific activity in the ventral stream visual cortex is associated with conscious face perception.

Activation of the ventral stream visual cortex in response to faces was primarily recorded in the high-gamma frequency (30 Hz–150 Hz) of ECoG signals in the human brain. Broadband gamma activity has gained attention in cognitive neuroscience due to its proposed role in cortical object representations and in the perception and encoding of objects [4–6, 13]. Studies have suggested that the induced gamma response to complex stimuli, such as faces, supports higher-order processing rather than a mere response to simple features [19]. Based on these reports, we speculate that the heightened BGA during the conscious perception of faces may be attributed to the extraction of face-specific features. It should be noted that some faces may be more visible than others, even under the same noise level. It makes sense to consider the effect of face visibility. However, since our data came from a public repository, there was a lack of detailed information about the stimuli seen by the subjects. Consequently, we were unable to fully explore the potential effects of face visibility. We defined the face visibility of an image as the probability that a face was perceived by subjects, and used ANCOVA to control the influence of this factor. Although this method controls the influence of face visibility to a certain extent, this estimate of visibility may be somewhat subjective due to possible differences in how subjects judge the face visibility of each image. We believe that future studies will benefit from obtaining more comprehensive data, especially examining different trials with identical stimuli. These details will be critical to further validate the robustness of our conclusions and gain a deeper understanding of the properties of face-specific neural activity in the ventral visual cortex.

Our research revealed that face-specific neural activity in the “seen” condition exhibited more similar neural responses than in the “unseen” condition. When face images are “seen”, it is possible that higher-level abstract face representations are formed in the ventral stream visual cortex and therefore lead to more reliable neural activity. It is worth noting that the similarity values of neural activity under the “seen” condition were relatively low, although significantly

higher than those obtained under the “unseen” condition. This seemingly contradictory finding is reasonable. The low similarity in the “seen” condition may reflect the ability of the neural activity to distinguish between different stimuli in the “seen” condition. Even though all belong to the “seen” condition, different trials in this condition have distinct stimuli features (e.g., different face identities), which the brain appears to finely tune to process, resulting in the low similarity. On the other hand, subjects in the “seen” condition perceived the same stimuli features (overall face structure) in all trials [20]. The similarity values between them were therefore significantly higher than the similarity values obtained in the “unseen” condition. Thus, this is an intriguing finding that invites a deeper exploration of the neural mechanisms underlying visual perception, recognition, and fine-grained categorization. Furthermore, we found that information transfer in the posterior-anterior direction was significantly increased in the “seen” condition compared with the “unseen” condition, but not in the anterior-posterior direction. This posterior-to-anterior interaction supports the role of the posterior ventral stream visual cortex in face perception, as was found in previous human studies [21, 22]. The ventral visual stream is known to exhibit hierarchical representations, with visual information flowing from the posterior to the anterior regions [23]. The posterior regions provide coarse representations, while the anterior regions offer finer representations. Given that our participants were engaged in forming perceptions of faces, we propose that this process necessitates the activation of finer-grained representations, resulting in increased posterior-to-anterior information transfer.

We found significant positive correlations and consistent non-linear patterns of change between the BGA and face perception performance. This change pattern is consistent with a fundamental characteristic of human perception [6, 13, 24], wherein small stimulus changes within the threshold range can lead to abrupt changes in the perceptual state [25]. The relationship between gamma power and the response latency of the subjects has been described for individual subjects before [14]. Our results further demonstrated the relationship between the peak latency of gamma activity and the response latency of the subjects for each site, providing more refined and comprehensive evidence for the relationship between neural activity and behavioral performance. In addition, our results demonstrated consistency in two areas: the noise level at which face perception performance begins to change (evidenced in increased response latency and decreased face detection accuracy at the 30% noise level) and the congruence between the change patterns of the BGA and the face perception performance (between the peak latency of the BGA and the response latency of the subjects and between the amplitude of the BGA and the face detection accuracy). The consistency of the trends found

in the BGA and the behavioral performance further support the idea that face-specific activity in the ventral stream visual cortex serves as a neural marker of conscious face perception.

While prior research explored how the neural response changes with varying levels of noise, they only analyzed the response of neuronal populations and lacked a direct correlation analysis between neural activity and behavior performance [11]. Our correlation analysis was performed at the individual site and the single trial level, revealing a more direct and fine-grained correlation between neural activity and behavioral performance. In addition, it is necessary to compare neural activity in different perceptual conditions when analyzing the relationship between neural activity and perception. Miller *et al.* compared the neural activity of all trials (including the “seen” and the “unseen” conditions in our analysis) with the neural activity of correct trials only (including the “seen” condition in our analysis). However, this method cannot separate the neural activity in different perceptual conditions, especially comparing the neural activity in the “seen” condition with that in the “unseen” condition. The paradigm of Fisch *et al.* tested sensitivity to the presence of stimuli [4]. In this study, participants consistently remained aware of the presence of a stimulus in every trial, and the investigation instead focused on sensitivity to meaningful structures (e.g., a face) within the stimuli.

Here, we only examined the results of face perception due to the limitations of the experimental paradigm. Therefore, it remains unclear whether these findings are specific to face perception or could be attributed to “seen” or “unseen” images in general. Although our analysis of non-face-selective sites did not yield similar effects, further examination of these sites showed that they also exhibited non-face response properties, with post-stimulus neural activity not being significantly greater than baseline activity. This attribute, on one hand, accounts for the absence of significant results in these sites. On the other hand, it does not exclude the possibility that the current findings are not unique to face perception, suggesting a potential universality in the recorded neural characteristics. This limitation provides an opportunity for future research to explore the mechanisms underlying the processing of “seen” and “unseen” images across different object categories, building upon our current findings. In addition, broader investigations may contribute to a deeper understanding of the role of the high-order visual cortex in processing various types of visual information, extending beyond just faces. Due to data limitations, we were unable to conduct an analysis of the dorsal face processing pathway, such as the superior temporal sulcus. However, it is well-recognized that the ventral and dorsal processing pathways typically serve distinct functions in face processing [26]. The ventral stream is associated with face recognition and conscious perception, whereas the dorsal stream is commonly

linked to motion-related aspects of face processing. This is why we primarily focused on the ventral stream. We suggest that future research could explore the role of the dorsal pathway in conscious face perception. This would require a comprehensive understanding of the functions of both the ventral and dorsal pathways in face processing, as well as an examination of their interactions. Such investigations could significantly contribute to a more comprehensive understanding of conscious face perception.

In conclusion, we demonstrated a link between the face-specific activity in the human ventral stream visual cortex and conscious face perception, providing insights into the neurophysiological mechanisms behind face perception.

Acknowledgements This work was supported by the Science and Technology Innovation 2030 - Brain Science and Brain-Inspired Intelligence Project (2021ZD0200200), the National Natural Science Foundation of China (62327805, 82151307, and 32271085), and the Beijing Natural Science Foundation (5244049). We are grateful to R. E. Perozzi and E. F. Perozzi for their assistance in language editing.

Data Availability The datasets analyzed during the current study are available in a freely available library at <https://searchworks.stanford.edu/view/zk881ps0522>.

Conflict of interest The authors declare that there are no conflicts of interest.

References

1. Hoehl S, Peykarjou S. The early development of face processing—What makes faces special? *Neurosci Bull* 2012, 28: 765–788.
2. Grill-Spector K, Weiner KS, Kay K, Gomez J. The functional neuroanatomy of human face perception. *Annu Rev Vis Sci* 2017, 3: 167–196.
3. Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: Progress and problems. *Nat Rev Neurosci* 2016, 17: 307–321.
4. Fisch L, Privman E, Ramot M, Harel M, Nir Y, Kipervasser S. Neural “ignition”: Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 2009, 64: 562–574.
5. Aru J, Axmacher N, Do Lam ATA, Fell J, Elger CE, Singer W, *et al.* Local category-specific gamma band responses in the visual cortex do not reflect conscious perception. *J Neurosci* 2012, 32: 14909–14914.
6. Perry G. The visual gamma response to faces reflects the presence of sensory evidence and not awareness of the stimulus. *R Soc Open Sci* 2016, 3: 150593.
7. Jiang Y, He S. Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Curr Biol* 2006, 16: 2023–2029.
8. Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, *et al.* Cortical mechanisms specific to explicit visual object recognition. *Neuron* 2001, 29: 529–535.
9. Miller KJ. A library of human electrocorticographic data and analyses. *Nat Hum Behav* 2019, 3: 1225–1235.
10. Miller KJ, Schalk G, Hermes D, Ojemann JG, Rao RP. Spontaneous decoding of the timing and content of human object perception from cortical surface recordings reveals complementary

- information in the event-related potential and broadband spectral change. *PLoS Comput Biol* 2016, 12: e1004660.
11. Miller KJ, Hermes D, Pestilli F, Wig GS, Ojemann JG. Face percept formation in human ventral temporal cortex. *J Neurophysiol* 2017, 118: 2614–2627.
 12. Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 2007, 164: 177–190.
 13. Li W, Li J, Cao D, Luo N, Jiang T. Neural mechanism of noise affecting face recognition. *Neuroscience* 2021, 468: 211–219.
 14. Ghuman AS, Brunet NM, Li Y, Konecky RO, Pyles JA, Walls SA, *et al.* Dynamic encoding of face information in the human fusiform gyrus. *Nat Commun* 2014, 5: 5672.
 15. Ding M, Bressler SL, Yang W, Liang H. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: Data preprocessing, model validation, and variability assessment. *Biol Cybern* 2000, 83: 35–45.
 16. Barnett L, Seth AK. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J Neurosci Methods* 2014, 223: 50–68.
 17. Dehaene S, Changeux JP. Experimental and theoretical approaches to conscious processing. *Neuron* 2011, 70: 200–227.
 18. Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I. Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci U S A* 2008, 105: 3599–3604.
 19. Engel AK, Singer W. Temporal binding and the neural correlates of sensory awareness. *Trends Cogn Sci* 2001, 5: 16–25.
 20. Ren S, Shao H, He S. Interaction between conscious and unconscious information-processing of faces and words. *Neurosci Bull* 2021, 37: 1583–1594.
 21. Gaillard R, Dehaene S, Adam C, Clémenceau S, Hasboun D, Baulac M, *et al.* Converging intracranial markers of conscious access. *PLoS Biol* 2009, 7: e61.
 22. Schrouff J, Raccach O, Baek S, Rangarajan V, Salehi S, Mourão-Miranda J, *et al.* Fast temporal dynamics and causal relevance of face processing in the human temporal cortex. *Nat Commun* 2020, 11: 656.
 23. Hong H, Yamins DL, Majaj NJ, DiCarlo JJ. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 2016, 19: 613–622.
 24. Grill-Spector K, Kushnir T, Hendler T, Malach R. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 2000, 3: 837–843.
 25. Lloyd MA, Appel JB. Signal detection theory and the psychophysics of pain: An introduction and review. *Psychosom Med* 1976, 38: 79–94.
 26. Duchaine B, Yovel G. A revised neural framework for face processing. *Annu Rev Vis Sci* 2015, 1: 393–416.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.