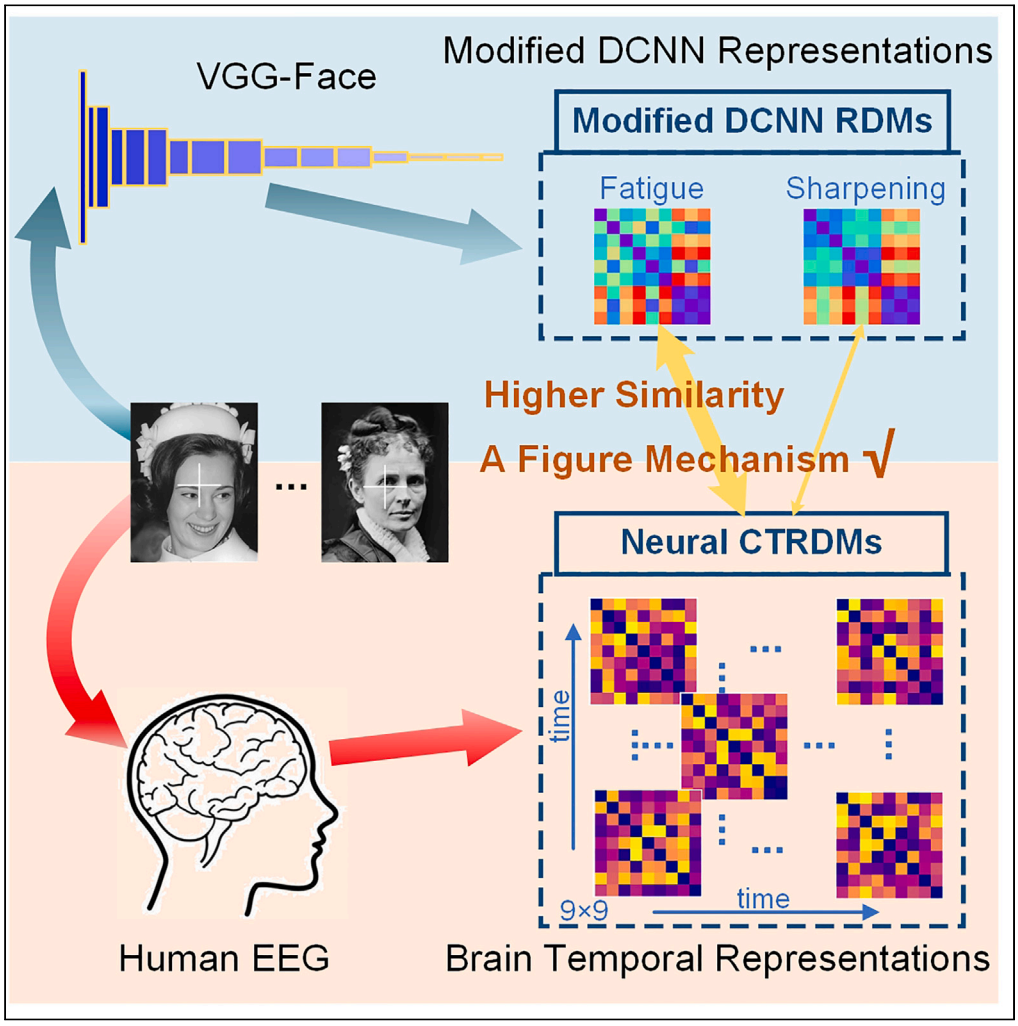
iScience

CellPress
OPEN ACCESS

Article

Bridging the gap between EEG and DCNNs reveals a fatigue mechanism of facial repetition suppression



Zitong Lu, Yixuan Ku

kuyixuan@mail.sysu.edu.cn

Highlights

EEG MVPA provides temporal evidence of facial repetition suppression

Comparisons between brains and DCNNs reveal a fatigue mechanism

Our reverse engineering framework offers a tool to investigate neural mechanisms

Lu & Ku, iScience 26, 108501
December 15, 2023
https://doi.org/10.1016/j.isci.2023.108501

iScience

CellPress
OPEN ACCESS

Article

Bridging the gap between EEG and DCNNs reveals a fatigue mechanism of facial repetition suppression

Zitong Lu[1] and Yixuan Ku[2,3,4,*]

SUMMARY

Facial repetition suppression, a well-studied phenomenon characterized by decreased neural responses to repeated faces in visual cortices, remains a subject of ongoing debate regarding its underlying neural mechanisms. Our research harnesses advanced multivariate analysis techniques and the prowess of deep convolutional neural networks (DCNNs) in face recognition to bridge the gap between human electroencephalogram (EEG) data and DCNNs, especially in the context of facial repetition suppression. Our innovative reverse engineering approach, manipulating the neuronal activity in DCNNs and conducted representational comparisons between brain activations derived from human EEG and manipulated DCNN activations, provided insights into the underlying facial repetition suppression. Significantly, our findings advocate the fatigue mechanism as the dominant force behind the facial repetition suppression effect. Broadly, this integrative framework, bridging the human brain and DCNNs, offers a promising tool for simulating brain activity and making inferences regarding the neural mechanisms underpinning complex human behaviors.

INTRODUCTION

In our daily experiences, we frequently encounter recurring or similar stimuli alongside new ones. When we repeatedly receive the same or similar information input, our brain's neural activity tends to diminish compared to the initial exposure, a phenomenon referred to as repetition suppression. Numerous electrophysiological studies have observed that neurons sensitive to visual information in the interior temporal cortex exhibit reduced responses when exposed to repetitive stimuli.[1–7] Additionally, research has shown that repeated stimuli can lead to a decrease in the blood oxygenation level-dependent (BOLD) response in functional magnetic resonance imaging (fMRI) studies.[8]

Within the field of face perception, numerous electroencephalogram (EEG) and magnetoencephalography (MEG) studies have reported various event-related potential (ERP) components associated with facial repetition suppression. These include the N170,[9–15] P200,[16–20] N250r,[21–26] and N400.[27–31] Despite these observations, the precise neuronal mechanism responsible for repetition suppression remains a topic of ongoing debate.

Previous research on facial repetition suppression has predominantly relied on univariate analysis methods, often overlooking the dynamic nature and variances in neural representations. However, in the past decade, cognitive neuroscience has seen a growing trend toward the adoption of multivariate analysis techniques. These include methods like correlation or classification-based Multivariate Pattern Analysis (MVPA)[32–36] and representational similarity analysis (RSA).[37,38] These multivariate analysis tools provide valuable insights into the neural mechanisms underlying complex cognitive processes by capturing the representational patterns through which our brains encode information. RSA, in particular, enables researchers to conduct representational comparisons across diverse modalities. For instance, it allows for the comparison of brain activity patterns with activations in computational models.[39–47] These advanced methods open up new avenues for a deeper understanding of neural processing and encoding in the brain.

Grill-Spector et al.[48] proposed three potential models (Fatigue, Sharpening, and Facilitation) to explain repetition suppression in neural coding patterns, drawing from a range of studies involving single-cell recordings, fMRI, and EEG/MEG. These models offer different hypotheses about how the brain processes repeated stimuli. The Fatigue model proposes that neurons with stronger initial responses to a stimulus exhibit higher repetition suppression. The Sharpening model suggests that neurons encoding irrelevant features of the stimulus show repetition suppression, leading to a more focused representation. The Facilitation model posits that repetition accelerates stimulus processing,

[1]Department of Psychology, The Ohio State University, Columbus, OH, USA
[2]Guangdong Provincial Key Laboratory of Brain Function and Disease, Center for Brain and Mental Well-being, Department of Psychology, Sun Yat-sen University, Guangzhou, China
[3]Peng Cheng Laboratory, Shenzhen, China
[4]Lead contact
*Correspondence: kuyixuan@mail.sysu.edu.cn
https://doi.org/10.1016/j.isci.2023.108501

reducing waiting time. To determine which of these models is more likely to underlie facial repetition suppression in human brains, multivariate analysis techniques can be employed.

Recent advances in computer vision have led to the development of deep convolutional neural network (DCNN) models for face recognition[49–52] that have achieved human-level performance.[53] In parallel, researchers in cognitive neuroscience and computer science have begun exploring the similarities and differences between human brains and artificial intelligence (AI) models in information processing. Studies that combine brain activity measurements and DCNNs have found that the hierarchical structure of the ventral visual pathway and DCNNs share similar processing representations of visual information.[39,40,44,54] The idea of reverse engineering, where the representation of an AI model can be modified based on theoretical hypotheses to align with the representation of human brains, provides a promising solution to investigate the neural mechanism of repetition suppression. On the one hand, the current challenges in understanding the mechanisms of facial repetition suppression in the human brain is the inherent difficulty in recording single-neuron activity. However, DCNNs provide a unique platform where we can manipulate the activation of neurons, offering an avenue to probe these mechanisms more directly. On the other hand, while there has been a burgeoning interest in linking human brain activity with DCNNs to elucidate neural mechanisms underlying object and face perception, to our knowledge, no studies have extended this approach to explore the specific issue of facial repetition suppression. Therefore, we aimed to apply reverse engineering methods to bridge the gap between human brain activity and DCNNs. We can potentially manipulate neuronal activity in DCNNs and compare this at the population-level with measurable human EEG activity to shed light on the possible mechanism underlying facial repetition suppression.

Our study aimed to delve into the neural mechanism of facial repetition suppression using innovative computational approaches. Initially, we investigated the dynamic representations of facial information and confirmed the presence of the facial repetition suppression effect in human brains. We accomplished this using a classification-based MVPA method on human EEG data (Figures 1A–1C). Subsequently, we employed the concept of reverse engineering to develop two potential repetition suppression models, namely the Fatigue and Sharpening models. These models were used to manipulate the activation of a DCNN (Figures 1D and 1E). We then conducted cross-modal RSA between human brain activity and the activations in the modified DCNNs (Figure 1F). Our findings strongly suggest that the fatigue mechanism is the more plausible neural mechanism underlying facial repetition suppression in the human brain.

RESULTS

Facial repetition suppression in human brains

The classification-based EEG decoding results for the three presentation conditions are presented in Figure 2. Time-by-time decoding results are illustrated in Figure 2A. (1) New vs. Immediate Decoding: Decoding accuracies for both familiar and unfamiliar faces were consistently above chance levels from 200ms to 1500ms. Similarly, decoding accuracies for scrambled faces exceeded chance levels from 360ms to 1500ms. Notably, during specific time intervals, decoding accuracies for familiar faces surpassed those for unfamiliar faces: from 620ms to 800ms, 880ms–1100ms, and 1120ms–1220ms. Additionally, decoding accuracies for familiar faces were consistently superior to those for scrambled faces from 240ms to 1500ms. Furthermore, decoding accuracies for unfamiliar faces were significantly better than those for scrambled faces from 240ms to 960ms.

(2) New vs. Delayed Decoding: Significant decoding accuracies for familiar faces were detected within limited time intervals, specifically from 660ms to 780ms, 860ms–920ms, and 1100ms–1160ms. Moreover, during the period from 580ms to 800ms, decoding accuracies for familiar faces were significantly higher than those for scrambled faces.

(3) Immediate vs. Delayed Decoding: Decoding accuracies for familiar faces consistently outperformed chance levels from 280ms to 1180ms. Similarly, decoding accuracies for unfamiliar faces exceeded chance levels from 240ms to 980ms, and those for scrambled faces from 560ms to 900ms. Moreover, during the interval from 440ms to 840ms, decoding accuracies for familiar faces were significantly better than those for scrambled faces, while decoding accuracies for unfamiliar faces surpassed those for scrambled faces from 440ms to 780ms.

Cross-temporal decoding results are displayed in Figure 2B. The most pronounced differences in neural patterns were observed between the New and Immediate conditions. Conversely, there were minimal significant differences between the New and Delayed conditions. Additionally, the differences in neural patterns for familiar faces were more substantial compared to those for unfamiliar faces and scrambled faces.

These results offer valuable insights into the temporal dynamics of neural representations during different presentation conditions, emphasizing the robustness of decoding accuracies in distinguishing between these conditions, particularly in the New vs. Immediate decoding scenario. They shed light on the neural mechanism of facial repetition suppression in human brains. Despite the identical input face images, notable differences were observed in neural representations between new and immediate repetition conditions. This suppression effect diminished as the interval between repetitions increased. Consequently, when the same face image was repeatedly shown after several trials, the neural representation became more similar to that of the initial presentation. Furthermore, familiar faces elicited a stronger repetition suppression effect compared to unfamiliar and scrambled faces.

Modified representations based on different repetition suppression models in DCNNs

Regarding the DCNNs, we initially extracted features corresponding to the 450 face images from all even layers and computed 450 × 450 RDMs (Figures S1B and S2). To simulate the facial repetition suppression effect in DCNNs, we adjusted neural representations in both VGG-Face and untrained VGG using the Fatigue and Sharpening models, respectively. Figure 3 displays DCNN RDMs of layer 16 derived from DCNN's internal activations that were modified by the two repetition suppression models. Under the Fatigue model, only the activation of the node with a higher response to the face stimulus exhibited repetition suppression, and nodes with higher activation experienced more

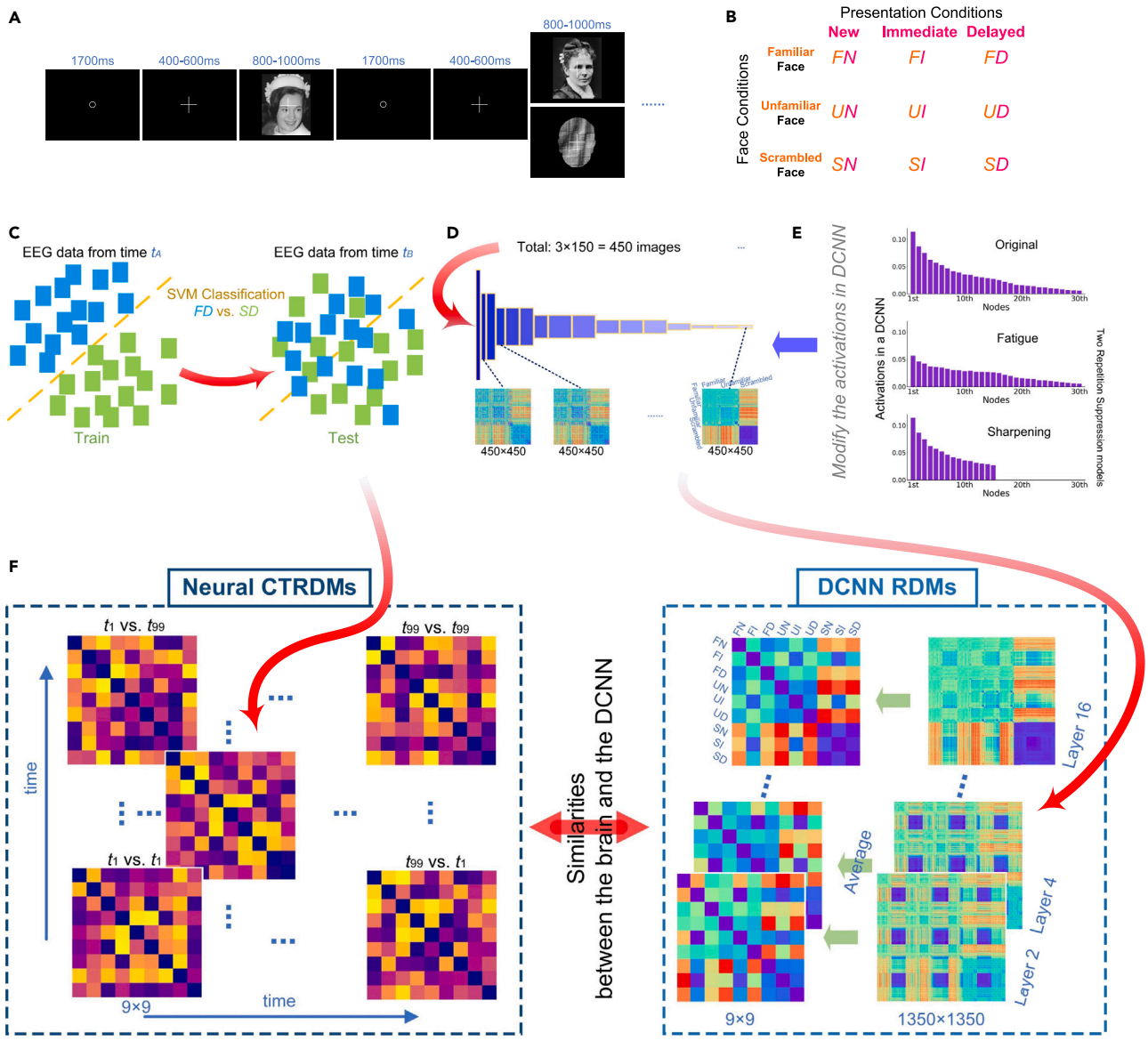


Figure 1. Experimental procedure and key analysis flow chart

(A) Overview of the experimental sequences.

(B) Illustration of the 9 experimental conditions, consisting of 3 face conditions by 3 presentation conditions.

(C) Schematic representation of cross-temporal EEG decoding.

(D) Diagram depicting the process of calculating DCNN RDMs (Representational Dissimilarity Matrices).

(E) Schematic illustration of the Fatigue and Sharpening repetition suppression models.

(F) Flowchart outlining the procedure for cross-modal RSA comparisons between EEG and DCNNs. Note: The two face images displayed here are from the public domain and are available at https://commons.wikimedia.org for illustrative purposes only. The actual images used during the experiment were described in Wakeman & Henson, 2015.

suppression. Under the Sharpening model, only the activation of the node with a lower response to the face stimulus exhibited repetition suppression, and nodes with low responses were not activated under repetition conditions. Modified 1350 × 1350 RDMs for all even layers are presented in Figure S3.

Comparisons between brains and DCNNs revealing a fatigue mechanism

After compressing the DCNN RDMs, we computed the similarity between EEG RDMs and the modified DCNN models, then calculated the difference in cross-modal similarity between VGG-Face and the untrained VGG as the valid similarity. Figure 4A illustrates the valid

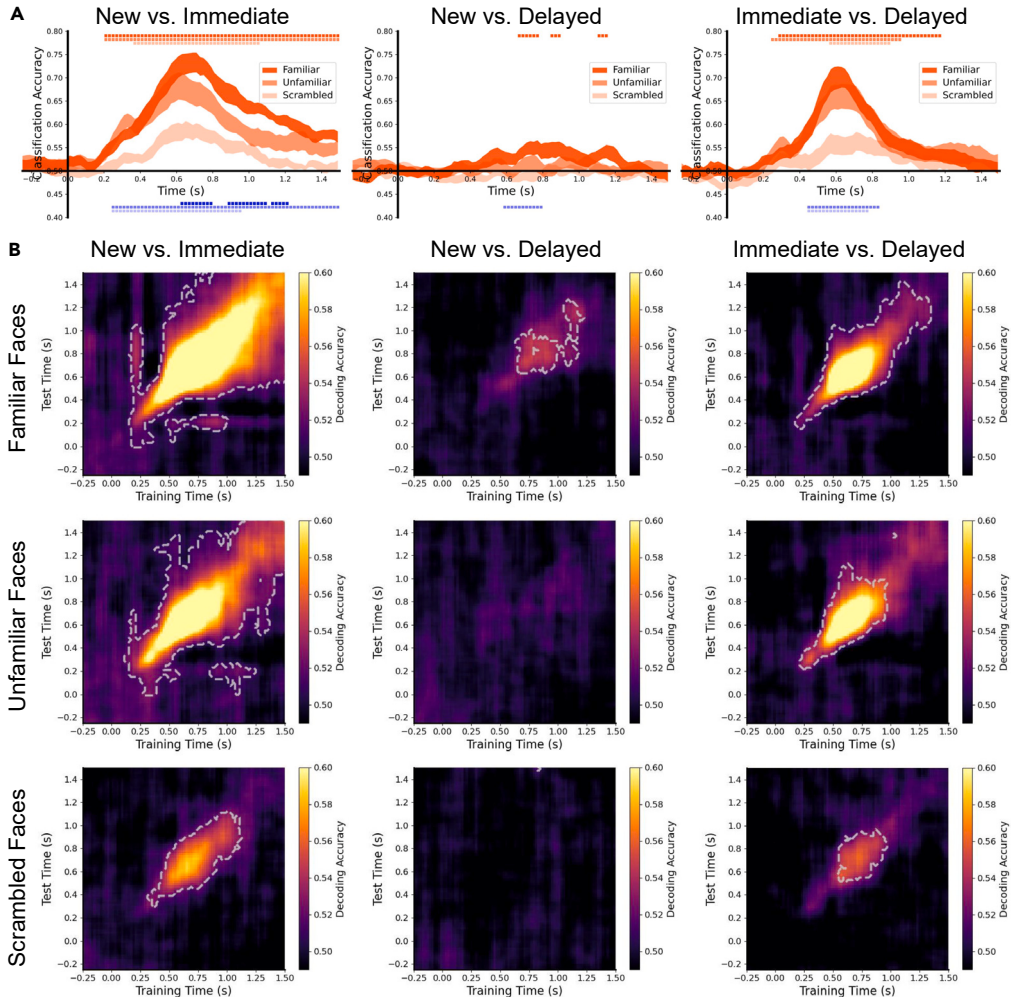


Figure 2. Temporal facial repetition suppression in human brains

(A) Time-by-time decoding results. The top of each plot is adorned with color-coded small squares indicating $p < 0.01$ (cluster-based permutation test) of decoding accuracy significantly greater than chance (ranging from dark to light orange for familiar, unfamiliar, and scrambled faces). The bottom of each plot features color-coded small squares indicating $p < 0.01$ (cluster-based permutation test) of significant differences in decoding accuracy between two face conditions (ranging from dark to light blue for familiar vs. unfamiliar faces, familiar vs. scrambled faces, and unfamiliar vs. scrambled faces). Line width reflects $\pm$SEM.

(B) Cross-temporal decoding results. The baseline for classification-based decoding accuracy is 50%. Regions where average accuracy significantly exceeds chance are highlighted with a light gray outline (cluster-based permutation test, $p < 0.01$).

representational similarity between activations modified by the Fatigue model for all even layers in DCNNs and EEG signals. DCNN representations modified by the Fatigue model exhibited significant representational similarity with human brains. We observed significant valid similarities in many layers (layer 2: 120ms–640ms, 700ms–1420ms; layer 4: 120ms–460ms, 1040ms–1220ms; layer 10: 480ms–560ms, 1380ms–1500ms; layer 12: 260ms–380ms, 420ms–560ms, 1380ms–1500ms; layer 14: 160ms–1000ms; layer 16: 200ms–1320ms). However, DCNN representations modified by the Sharpening model showed almost no significant similarity with human brains. We only found significant valid similarities from 640ms to 680ms in layer 8.

In detail, Figure 4B presents the cross-temporal valid similarity of layer 16, which is the last layer in the VGG structure and contains the most relevant information for face recognition. When modified by the Fatigue model, the DCNN's representations exhibited strong and extensive valid similarity with human brains. However, representations of the DCNN modified by the Sharpening model only displayed weak similarities with brain activity.

Cross-modal comparisons suggest that simulating the activation in DCNNs based on a fatigue mechanism, rather than a sharpening mechanism, could induce more similar representations with human brain activities. Therefore, the facial repetition suppression effect in face perception is more likely caused by the fatigue mechanism.

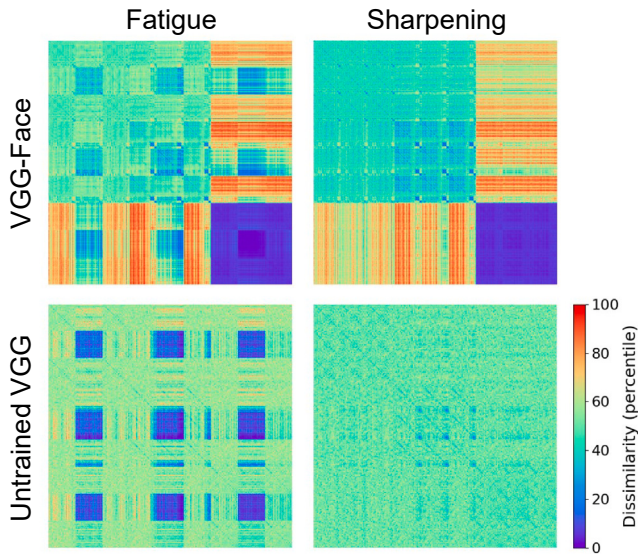


Figure 3. Modified DCNN RDMs of layer 16 based on two different repetition suppression mechanisms

DISCUSSION

In this study, we employed a unique approach that combines human EEG and DCNNs to delve into the neural mechanism of facial repetition suppression in face perception. We accomplished this through classification-based EEG decoding and cross-modal RSA and revealed a fatigue mechanism of human face repetition suppression.

Initially, our EEG-based decoding results provided insights into the temporal dynamics of neural representations across different face presentation conditions. We observed a facial repetition suppression effect that was more pronounced for familiar faces and less so for scrambled faces. This effect decreased as the interval between repeated viewings increased. Notably, our findings revealed distinct neural representations between new and immediately repeated conditions, demonstrating the influence of repetition suppression. Furthermore, the repetition suppression effect was more robust for familiar faces than for unfamiliar or scrambled ones.

Subsequently, we embarked on a reverse engineering endeavor to delve into the mechanisms underlying facial repetition suppression in the human brain. Our approach involved the modification of neural activations in AI models capable of achieving human-level performance in face recognition. Specifically, we honed in on the repetition suppression mechanisms pertaining to face-specific information using the VGG-Face model, a neural network trained on an extensive dataset of facial images for the purpose of face identification. To distill the information relevant to facial features within the DCNN, we took a unique approach. We contrasted the results obtained from the VGG-Face model with those of an untrained VGG model. This juxtaposition allowed us to extract the valid representational similarity between the trained DCNN and the human brain, focusing on their shared representations of facial stimuli. The outcome of this analysis revealed a significant convergence between the representation of the DCNN model and the neural patterns observed in the human brain when employing the Fatigue model for modification. This alignment was not only observed in the later layers of the DCNN but also extended to the early layers, mirroring the hierarchical structure of the human brain's visual processing pathway.

Moreover, our cross-temporal analysis illuminated that the similarities between the DCNN model's representations and the neural patterns observed in the human brain encompassed a broader temporal range. Intriguingly, these similarities were not confined solely to the later layers of the DCNN; they extended to the early layers as well. This finding aligns with previous research that has drawn parallels between the hierarchical structure of DCNNs and the visual processing pathway in the human brain.[39,40] Collectively, these results provide compelling evidence that the attenuation of neural activations in the process of repetition suppression, driven by the fatigue mechanism, occurs not only in neurons processing high-level face features but also in those handling low-level facial information. This underscores the robustness and comprehensiveness of the fatigue-based repetition suppression phenomenon across different neural processing stages, mirroring the multifaceted nature of facial perception in the human brain.

It is worth noting that while numerous human fMRI studies[55–62] have given support to the idea that repetition suppression is indicative of a reduction in prediction error within the predictive coding framework,[63,64] recent electrophysiological investigations have put forth an alternative perspective. These electrophysiological studies have provided evidence in favor of the fatigue mechanism, positing that it involves bottom-up or local adaptation, as opposed to sharpening or sparseness representations and the predictive coding hypothesis.

However, the collection of human electrophysiological data presents substantial challenges, leading to a scarcity of human neuroimaging studies that could corroborate these findings.[65] Our present study is innovative in that it introduces state-of-the-art computational methods, combining noninvasive human EEG with DCNNs, to delve into the neural underpinnings of facial repetition suppression. In doing so, we have provided compelling and robust evidence to support the fatigue mechanism as the driving force behind repetition suppression in the context of facial perception.

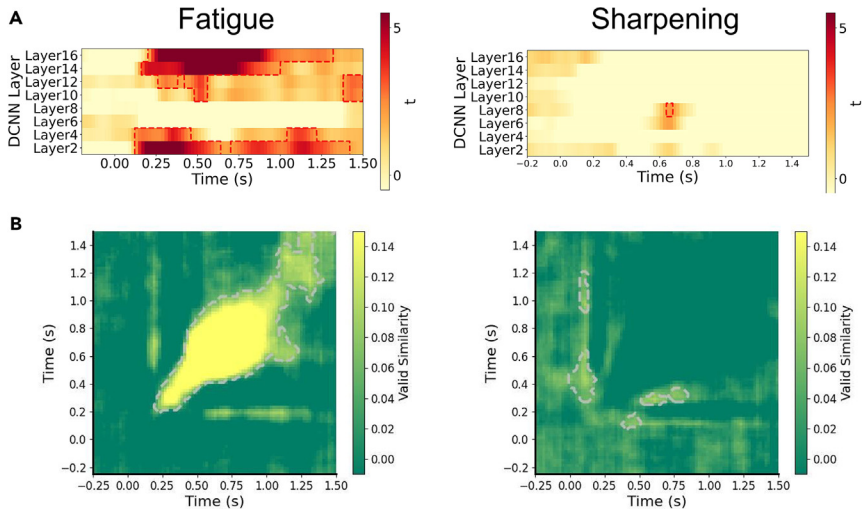


Figure 4. Representational comparison between brain activity and modified DCNN activations

(A) Layer-by-layer temporal valid similarity between EEG and modified DCNNs.

(B) Cross-temporal valid similarity between EEG and modified DCNNs on layer 16. The baseline of valid similarity is zero. Outlines indicate significant clusters (cluster-based permutation test, $p < 0.05$).

Nevertheless, whether DCNNs represent the best brain-like models still needs investigation. While DCNNs are widely employed in neuroscience, the field of computer vision has witnessed rapid advancements, resulting in the development of new ANN models that demonstrate exceptional task performance and feature space effectiveness. Some novel models with architectures distinct from DCNNs, such as Generative Adversarial Networks (GANs),[66] Vision Transformers (ViTs),[67] Contrastive Language-Image Pretraining (CLIP),[68] and Spiking Neural Networks (SNNs),[69] may process visual embeddings differently than DCNNs. For instance, GANs perform end-to-end compression of image information, ViTs use attention mechanisms for a more coarse-grained extraction of image features, CLIP incorporates contrastive learning with linguistic information to attain multi-modal image representations, and SNNs simulate the temporal dynamics of neural firing, offering a different paradigm for information processing. Recent studies provided unsupervised models[70] and some other brain-inspired models[71–73] may show more similar representations to human brains. However, we chose to use DCNNs for several reasons: (1) It is a model that shows similar hierarchical processing to human visual cortex, extracting from lower-to higher-level visual information; (2) It allows us to easily manipulate at the neuron-level; (3) Our primary focus is on the processing changes in visual information during the facial repetition suppression changes, and we do not need to overly concern ourselves with the subtle differences between different visual extraction models. Evaluating which model aligns most closely with the human brain is a complex challenge, and it remains a cutting-edge area of research that garners attention from both computer scientists and neuroscientists. From the perspective of extracting representations from human EEG signals, various EEG-based models directly trained on EEG signals[74–76] might offer various schemes for feature extraction, rather than directly using electrodes as features for subsequent RSA, as done in our current study. Whether and how these different models could provide more insights in not only repetition suppression but also more generally visual perception area is worth evaluating in the future. Our present study provides a valuable framework for investigating neural mechanisms within the human brain through the use of reverse engineering and cross-modal RSA.

In conclusion, this study represents an innovative foray into the realm of neuroscience using state-of-the-art methodologies that fuse AI and neuroimaging techniques. Through the strategic modification of AI model representations, we have endeavored to identify the condition that most closely approximates the neural representations within the human brain. Our exploration of cross-modal representations not only facilitates the unraveling of intricate neuroscience questions that are challenging to address through conventional noninvasive methods such as fMRI or EEG experiments but also holds the potential to inspire advancements in AI models from a neuroscientific perspective. The insights gleaned from this research are poised to be of significant importance for the future development of brain-inspired artificial intelligence. By bridging the gap between AI and neuroscience, this work contributes to a deeper understanding of neural mechanisms and offers invaluable contributions to the ongoing quest for brain-like intelligence.

Limitations of the study

Our study brings to light several areas that warrant further investigation and potential avenues for improvement. First, it is important to consider whether there are alternative mechanisms underlying facial repetition suppression. While our study examined two primary models, the Fatigue and Sharpening models, it is conceivable that other mechanisms may be at play. One limitation of pure DCNN models is their lack of inherent temporal processing capability. Future research could explore the inclusion of timing process components, such as recurrent structures,[54,77] within DCNN models. This could lead to the discovery of additional mechanisms, including the possibility of a Facilitation

mechanism[48] contributing to repetition suppression. Additionally, investigating repetition suppression mechanisms under different conditions and for various types of facial information or tasks could provide valuable insights.

Second, our study has certain limitations stemming from the original experimental design of the open dataset we utilized. There exist numerous nuanced facets of facial information that have not been thoroughly explored and deeply analyzed due to these limitations. These unexplored dimensions include variations in hairstyle, skin color, viewpoint (e.g., upright or inverted faces), gender, race, and facial expressions. These dimensions are critical as they can influence how faces are perceived and can potentially alter repetition suppression dynamics. Furthermore, expanding our investigations to encompass the spatial aspects of neural information processing through fMRI and MEG data could offer a more comprehensive perspective on the neural mechanisms underlying facial repetition suppression.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data and experimental information
  - Classification-based EEG decoding
  - DCNN models
  - RSA
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108501.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32171082), the National Social Science Foundation of China (17ZDA323), the Neuroeconomics Laboratory of Guangzhou Huashang College (2021WSYS002) and the Leading talent program (31620016) at Sun Yat-sen University.

AUTHOR CONTRIBUTIONS

Z.L. and Y.K. designed research; Z.L. performed research; Y.K. supervised research; Z.L. and Y.K. wrote the paper.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: September 22, 2023
Revised: October 26, 2023
Accepted: November 17, 2023
Published: November 22, 2023

REFERENCES

1. Baylis, G.C., and Rolls, E.T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp. Brain Res. *65*, 614–622.
2. Kaliukhovich, D.A., and Vogels, R. (2011). Stimulus Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferior Temporal Cortex. Cereb. Cortex *21*, 1547–1558.
3. Kaliukhovich, D.A., and Vogels, R. (2012). Stimulus repetition affects both strength and synchrony of macaque inferior temporal cortical activity. J. Neurophysiol. *107*, 3509–3527.
4. Miller, E.K., Li, L., and Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. Science *254*, 1377–1379.
5. Ringo, J.L. (1996). Stimulus specific adaptation in inferior temporal and medial temporal cortex of the monkey. Behav. Brain Res. *76*, 191–197.
6. Sawamura, H., Orban, G.A., and Vogels, R. (2006). Selectivity of Neuronal Adaptation Does Not Match Response Selectivity: A Single-Cell Study of the fMRI Adaptation Paradigm. Neuron *49*, 307–318.
7. Sobotka, S., and Ringo, J.L. (1994). Stimulus specific adaptation in excited but not in inhibited cells in inferotemporal cortex of Macaque. Brain Res. *646*, 95–99.
8. Henson, R.N.A., and Rugg, M.D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. Neuropsychologia *41*, 263–270.

9. Kloth, N., and Schweinberger, S.R. (2010). Electrophysiological correlates of eye gaze adaptation. J. Vis. *10*, 17.
10. Kloth, N., Schweinberger, S.R., and Kovács, G. (2010). Neural Correlates of Generic versus Gender-specific Face Adaptation. J. Cogn. Neurosci. *22*, 2345–2356.
11. Kovács, G., Zimmer, M., Bankó, E., Harza, I., Antal, A., and Vidnyánszky, Z. (2006). Electrophysiological Correlates of Visual Adaptation to Faces and Body Parts in Humans. Cereb. Cortex *16*, 742–753.
12. Maurer, U., Rossion, B., and McCandliss, B.D. (2008). Category specificity in early perception: Face and word N170 responses differ in both lateralization and habituation properties. Front. Hum. Neurosci. *2*, 18.
13. Mercure, E., Kadosh, K.C., and Johnson, M.H. (2011). The N170 shows differential repetition effects for faces, objects, and orthographic stimuli. Front. Hum. Neurosci. *5*, 1–10.
14. Schweinberger, S.R., Kaufmann, J.M., Moratti, S., Keil, A., and Burton, A.M. (2007). Brain responses to repetitions of human and animal faces, inverted faces, and objects — An MEG study. Brain Res. *1184*, 226–233.
15. Walther, C., Schweinberger, S.R., Kaiser, D., and Kovács, G. (2013). Neural correlates of priming and adaptation in familiar face perception. Cortex *49*, 1963–1977.
16. Burkhardt, A., Blaha, L.M., Jurs, B.S., Rhodes, G., Jeffery, L., Wyatte, D., Delong, J., and Busey, T. (2010). Adaptation modulates the electrophysiological substrates of perceived facial distortion: Support for opponent coding. Neuropsychologia *48*, 3743–3756.
17. Kaufmann, J.M., and Schweinberger, S.R. (2012). The faces you remember: Caricaturing shape facilitates brain processes reflecting the acquisition of new face representations. Biol. Psychol. *89*, 21–33.
18. Latinus, M., and Taylor, M.J. (2006). Face processing stages: Impact of difficulty and the separation of effects. Brain Res. *1123*, 179–187.
19. Schulz, C., Kaufmann, J.M., Kurt, A., and Schweinberger, S.R. (2012). Faces forming traces: Neurophysiological correlates of learning naturally distinctive and caricatured faces. Neuroimage *63*, 491–500.
20. Zheng, X., Mondloch, C.J., and Segalowitz, S.J. (2012). The timing of individual face recognition in the brain. Neuropsychologia *50*, 1451–1461.
21. Dörr, P., Herzmann, G., and Sommer, W. (2011). Multiple contributions to priming effects for familiar faces: Analyses with backward masking and event-related potentials. Br. J. Psychol. *102*, 765–782.
22. Herzmann, G., Schweinberger, S.R., Sommer, W., and Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. Psychophysiology *41*, 688–701.
23. Pfütze, E.M., Sommer, W., and Schweinberger, S.R. (2002). Age-related slowing in face and name recognition: Evidence from event-related brain potentials. Psychol. Aging *17*, 140–160.
24. Schweinberger, S.R., Pfütze, E.M., and Sommer, W. (1995). Repetition Priming and Associative Priming of Face Recognition: Evidence From Event-Related Potentials. J. Exp. Psychol. Learn. Mem. Cogn. *21*, 722–736.
25. Schweinberger, S.R., and Burton, A.M. (2003). Covert Recognition and the Neural System for Face Processing. Cortex *39*, 9–30.
26. Wiese, H., Kachel, U., and Schweinberger, S.R. (2013). Holistic face processing of own- and other-age faces in young and older adults: ERP evidence from the composite face task. Neuroimage *74*, 306–317.
27. Barrett, S.E., and Rugg, M.D. (1989). Event-related potentials and the semantic matching of faces. Neuropsychologia *27*, 913–922.
28. Bentin, S., McCarthy, G., and Wood, C.C. (1985). Event-related potentials, lexical decision and semantic priming. Electroencephalogr. Clin. Neurophysiol. *60*, 343–355.
29. Rugg, M.D. (1985). The Effects of Semantic Priming and Word Repetition on Event-Related Potentials. Psychophysiology *22*, 642–647.
30. Schweinberger, S.R. (1996). How gorbachev primed yeltsin: Analyses of associative priming in person recognition by means of reaction times and event-related brain potentials. J. Exp. Psychol. Learn. Mem. Cogn. *22*, 1383–1407.
31. Stevenage, S.V., Hale, S., Morgan, Y., and Neil, G.J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. Br. J. Psychol. *105*, 1–16.
32. Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.
33. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
34. Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. *10*, 424–430.
35. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.
36. Golomb, J.D., and Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic, Not Spatiotopic, Object Location. Cereb. Cortex *22*, 2794–2810.
37. Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *4*.
38. Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron *60*, 1126–1141.
39. Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., and Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. *6*, 27755–27813.
40. Güçlü, U., and van Gerven, M.A.J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J. Neurosci. *35*, 10005–10014.
41. Kuzovkin, I., Vicente, R., Petton, M., Lachaux, J.P., Baciu, M., Kahane, P., Rheims, S., Vidal, J.R., and Aru, J. (2018). Activations of deep convolutional neural networks are aligned with gamma band activity of human visual cortex. Commun. Biol. *1*, 107–112.
42. Urgen, B.A., Pehlivan, S., and Saygin, A.P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. Neuropsychologia *127*, 35–47.
43. Xie, S., Kaiser, D., and Cichy, R.M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. Curr. Biol. *30*, 2621–2627.e5.
44. Yamins, D.L.K., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA *111*, 8619–8624.
45. Xu, Y., and Vaziri-Pashkam, M. (2021). Examining the Coding Strength of Object Identity and Nonidentity Features in Human Occipito-Temporal Cortex and Convolutional Neural Networks. J. Neurosci. *41*, 4234–4252.
46. Dobs, K., Isik, L., Pantazis, D., and Kanwisher, N. (2019). How face perception unfolds over time. Nat. Commun. *10*, 1258–1310.
47. Lu, Z., and Golomb, J.D. (2023). Human EEG and artificial neural networks reveal disentangled representations of object real-world size in natural images. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.26.538469.
48. Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. *10*, 14–23.
49. Taigman, Y., Marc', M.Y., Ranzato, A., and Wolf, L. (2014). DeepFace: closing the gap to human-level performance in face verification. Proceedings of the IEEE conference on computer vision and pattern recognition (IEEE), pp. 1701–1708.
50. Parkhi, O.M., Vedaldi, A., and Zisserman, A. (2015). Deep Face Recognition. In BMVC 2015 - Proceedings of the British Machine Vision Conference 2015 (British Machine Vision Association), pp. 41.1–41.41.
51. Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. Preprint at arXiv. https://doi.org/10.48550/arXiv.1503.03832.
52. Ranjan, R., Patel, V.M., and Chellappa, R. (2019). HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition. IEEE Trans. Pattern Anal. Mach. Intell. *41*, 121–135.
53. Phillips, P.J., Yates, A.N., Hu, Y., Hahn, C.A., Noyes, E., Jackson, K., Cavazos, J.G., Jeckeln, G., Ranjan, R., Sankaranarayanan, S., et al. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. Proc. Natl. Acad. Sci. USA *115*, 6171–6176.
54. Kietzmann, T.C., Spoerer, C.J., Sörensen, L.K.A., Cichy, R.M., Hauk, O., and Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. Proc. Natl. Acad. Sci. USA *116*, 21854–21863.
55. Kovács, G., Iffland, L., Vidnyánszky, Z., and Greenlee, M.W. (2012). Stimulus repetition probability effects on repetition suppression are position invariant for faces. Neuroimage *60*, 2128–2135.
56. Kovács, G., Kaiser, D., Kaliukhovich, D.A., Vidnyánszky, Z., and Vogels, R. (2013). Repetition Probability Does Not Affect fMRI Repetition Suppression for Objects. J. Neurosci. *33*, 9805–9812.
57. Grotheer, M., Hermann, P., Vidnyánszky, Z., and Kovács, G. (2014). Repetition probability

effects for inverted faces. Neuroimage *102 Pt 2*, 416–423.
58. Grotheer, M., and Kovács, G. (2014). Repetition Probability Effects Depend on Prior Experiences. J. Neurosci. *34*, 6640–6646.
59. Mayrhauser, L., Bergmann, J., Crone, J., and Kronbichler, M. (2014). Neural repetition suppression: Evidence for perceptual expectation in object-selective regions. Front. Hum. Neurosci. *8*, 225.
60. Ewbank, M.P., von dem Hagen, E.A.H., Powell, T.E., Henson, R.N., and Calder, A.J. (2016). The effect of perceptual expectation on repetition suppression to faces is not modulated by variation in autistic traits. Cortex *80*, 51–60.
61. Larsson, J., and Smith, A.T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or Stimulus Expectation? Cereb. Cortex *22*, 567–576.
62. Andics, A., Gál, V., Vicsi, K., Rudas, G., and Vidnyánszky, Z. (2013). FMRI repetition suppression for voices is modulated by stimulus expectations. Neuroimage *69*, 277–283.
63. Friston, K. (2005). A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. *360*, 815–836.
64. Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.M., and Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. Nat. Neurosci. *11*, 1004–1006.
65. Stam, D., Huang, Y.A., Vansteelandt, K., Sunaert, S., Peeters, R., Sleurs, C., Vrancken, L., Emsell, L., Vogels, R., Vandenbulcke, M., and Van den Stock, J. (2021). Long term fMRI adaptation depends on adapter response in face-selective cortex. Commun. Biol. *4*, 712–719.
66. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. Sci. Robot. *3*, 2672–2680.
67. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. Preprint at arxiv. https://doi.org/10.48550/arxiv.2010.11929.
68. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. In Proceedings of the International Conference on Machine Learning (ICML).
69. Ghosh-Dastidar, S., and Adeli, H. (2009). Spiking neural networks. Int. J. Neural Syst. *19*, 295–308.
70. He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R. (2019). Momentum Contrast for Unsupervised Visual Representation Learning. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 9726–9735.
71. Zeng, T., and Si, B. (2021). A brain-inspired compact cognitive mapping system. Cogn. Neurodyn. *15*, 91–101.
72. Li, J., Huang, Q., Han, Q., Mi, Y., and Luo, H. (2021). Temporally coherent perturbation of neural dynamics during retention alters human multi-item working memory. Prog. Neurobiol. *201*, 102023.
73. Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N.J., Issa, E.B., Bashivan, P., Prescott-Roy, J., Schmidt, K., et al. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. Adv. Neural Inf. Process. Syst. *32*.
74. Zhang, G., Yu, M., Liu, Y.J., Zhao, G., Zhang, D., and Zheng, W. (2023). SparseDGCNN: Recognizing Emotion from Multichannel EEG Signals. IEEE Trans. Affect. Comput. *14*, 537–548.
75. Liang, Z., Zhang, X., Zhou, R., Zhang, L., Li, L., Huang, G., and Zhang, Z. (2022). Cross-individual affective detection using EEG signals with audio-visual embedding. Neurocomputing *510*, 107–121.
76. Lu, Z., and Golomb, J.D. (2023). Generate your neural signals from mine: individual-to-individual EEG converters. Preprint at arXiv. https://doi.org/10.48550/arXiv.2304.10736.
77. Spoerer, C.J., McClure, P., and Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. Front. Psychol. *8*, 1551.
78. Wakeman, D.G., and Henson, R.N. (2015). A multi-subject, multi-modal human neuroimaging dataset. Sci. Data *2*, 150001–150010.
79. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21.
80. Lu, Z., and Ku, Y. (2020). NeuroRA: A Python Toolbox of Representational Analysis From Multi-Modal Neural Data. Front. Neuroinform. *14*, 61.
81. Lu, Z. (2020). PyCTRSA: A Python Package for Cross-Temporal Representational Similarity Analysis-Based E/MEG Decoding. Zenodo.
82. Drisdelle, B.L., Aubin, S., and Jolicoeur, P. (2017). Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory: ICA correction versus epoch rejection. Psychophysiology *54*, 83–99.
83. Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T.J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin. Neurophysiol. *111*, 1745–1758.
84. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
EEG data of 18 human subjects	Wakeman & Henson[78]	https://openneuro.org/datasets/ds002718/
Software and algorithms		
Python 3.8	Python Software Foundation	https://www.python.org/
EEGLAB toolbox	Delorme & Makeig[79]	https://doi.org/10.1016/j.jneumeth.2003.10.009
NeuroRA toolbox	Lu & Ku[80]	https://doi.org/10.3389/fninf.2020.563669
PyCTRSA toolbox	Lu[81]	https://doi.org/10.5281/ZENODO.4273674
Other		
All code used in this paper	This paper	https://osf.io/unhzm

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yixuan Ku (kuyixuan@mail.sysu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- All original code has been deposited at https://osf.io/unhzm and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The EEG data used in our study are from a comprehensive multi-subject, multi-modal human neuroimaging dataset.[78] Nineteen participants were recruited from the MRC Cognition & Brain Sciences Unit, of which 8 were female, and 11 were male, with an age range 23-37 years. All were Caucasian except for one Asian participant, who had spent many years in the UK. The study was approved by Cambridge University Psychological Ethics Committee. Written informed consent was obtained from each participant prior to and following each phase of the experiment. For the preprocessed EEG data, they excluded the EEG data from "sub-01" with unknown reasons (OpenNeuro Database: https://openneuro.org/datasets/ds002718/). Thus, we applied the final eighteen participants' EEG data they provide for our current study.

METHOD DETAILS

Data and experimental information

The data utilized in this study were sourced from an EEG dataset focusing on face perception, accessible on OpenNeuro (OpenNeuro Database: https://openneuro.org/datasets/ds002718/). This dataset is a comprehensive multi-subject, multi-modal human neuroimaging dataset.[78] In our analysis, we specifically utilized EEG data from 18 participants, each consisting of 70 valid channels. All participants participated in a face perception task (Figure 1A) that involved the presentation of 450 grayscale face stimuli. This set comprised 150 familiar faces, 150 unfamiliar faces, and 150 scrambled faces. Each scrambled face was generated from either the famous face or the non-famous face of the same stimulus number. These were scrambled by taking the 2D-Fourier transform of the faces, permuting the phase information, and then inverse-transforming back into the image space. To match the overall approximate shape and size of the original faces, the scrambled images were finally cropped to a mask created by a combination of one familiar and one nonfamiliar face. Stimuli were projected onto a screen approximately 1.3 m in front of the participant, subtending horizontal and vertical visual angles of approximately 3.66° and 5.38° respectively. The photographs were presented against a black background, with a white fixation cross in the center. Participants were asked to fix on the center of the screen and conducted a 'more' or 'less symmetric' judgement task (compare to the average based on a pre-task

consisting of 23 separate photos) to ensure their attention. Each face stimulus was presented twice, with a 50% probability of seeing the same face either immediately or after a delay of 5 to 15 trials. In total, it included 900 trials.

In our analysis, we labeled the first time a participant saw a specific face as 'new,' the second time they saw the same face immediately as 'immediate (repetition),' and the second time they saw the face after several trials as 'delayed (repetition).' This resulted in a total of 9 conditions (FN, 150 trials; FI, 75 trials; FD, 75 trials; UN, 150 trials; UI, 75 trials; UD, 75 trials; SN, 150 trials; SI, 75 trials; SD, 75 trials) derived from the combination of 3 face conditions (F: familiar; U: unfamiliar; S: scrambled; 300 trials per condition) and 3 presentation conditions (N: new, 450 trials; I: immediate, 225 trials; D: delayed, 225 trials) (Figure 1B).

The EEG data employed in this study had already undergone partial preprocessing and had been resampled to 250 Hz as part of the dataset. We applied band-pass filtering to retain data within the 0.1 to 30 Hz range. To identify and eliminate blinks and eye movements, we utilized independent components analysis (ICA)[82,83] using EEGLAB.[79] Epochs of -500 to 1500ms from stimulus onset were created. No baseline correction was applied.

Classification-based EEG decoding

Time-by-time EEG decoding

Each EEG trial corresponded to two distinct labels: one indicating the face condition (F, U, or S) and the other indicating the presentation condition (N, I, or D). To investigate neural representations across different presentation conditions, we performed nine separate classification-based decoding analyses, aiming to distinguish between them. These analyses encompassed the following condition pairs: FN vs. FI, UN vs. UI, SN vs. SI, FN vs. FD, UN vs. UD, SN vs. SD, FI vs. FD, UI vs. UD, SI vs. SD.

For each of these classification-based decoding analyses, we employed Support Vector Machines (SVM) with a 'linear' kernel, a 1e-3 tolerance for stopping criterion, and the default regularization parameter (value=1). Our approach involved binarizing the trial labels for the two conditions being compared. To reduce the dimensionality of the EEG data, we downsampled it by averaging every five time-points. The original 500 time-points spanning from -500 to 1500ms were compressed into 100 time-points. This yielded a label vector for each participant, containing labels for all trials, as well as a three-dimensional matrix with dimensions for time, trial, and channels, facilitating time-by-time classification.

In our process, we randomized the order of trials and then averaged the data every five trials. The classifier was trained and tested separately for each time-point. Specifically, we randomly selected 2/3 of the trials for training and used the remaining 1/3 for testing in each iteration. This entire sequence of random shuffling, averaging, classification training, and testing was repeated 100 times for each time-point. If there was an imbalance in the number of trials under two conditions, we applied the under-sampling approach. In each iteration, before training the SVM, we identified the condition with the fewer samples and randomly selected the same samples from the condition with the greater number of samples to ensure that both conditions have an equal number of samples to train the classifier. By adopting this under-sampling strategy, we made sure that the SVM was trained on a balanced dataset for each time, mitigating the bias that could arise due to the initial sample imbalance. We repeated these steps for each participant and for all nine classification condition pairs. To obtain more reliable time-by-time decoding accuracies, we averaged the classification accuracies across all iterations. This entire process was replicated for the 18 participants in our study.

Cross-temporal EEG decoding

Additionally, we carried out cross-temporal EEG decoding, which represents an extension of the time-by-time decoding approach. The fundamental idea behind cross-temporal decoding is to train the classifier on data at one specific time-point and then test it on data from other time-points to assess whether the encoding patterns of the information of interest remain consistent across different times.

Similar to building upon the time-by-time decoding methodology described earlier, we trained the classifier on data for each time-point and test this pre-trained classifier on data for all 100 time-points respectively. Thus, compared to time-by-time decoding, the only difference of the cross-temporal EEG decoding we conduct here was to test the classifier on data for not only the identical time-point with the training time-point but also every other time-point in the timecourse to see the temporal generalization of each time-point-based EEG classifier (see Figure 1C for a graphical representation). Similar to the previous approach, we obtained the final decoding accuracies by averaging results across 100 iterations. Consequently, each participant generated nine temporal generalization matrices for cross-temporal decoding, each corresponding to one of the nine classification condition pairs. All the EEG decoding processes described above were executed using the NeuroRA toolbox.[80]

DCNN models

In this study, we employed a DCNN model commonly utilized in the field of face recognition known as VGG-Face[50] (Figure S1A). VGG-Face was pretrained on a dataset consisting of 2622 unique identities, each with 1000 face images per person. The model exhibited impressive test accuracies, achieving 97.27% on the IFW dataset and 92.8% on the YouTube Faces dataset. VGG-Face essentially followed the structure of a VGG-16 model, comprising 13 convolutional layers and 3 fully connected layers. In our study, we utilized the VGG-Face model as a DCNN for extracting facial features. For the sake of comparison, we also incorporated an additional VGG-16 model that was left untrained and initialized with random weights. This untrained VGG model served as a DCNN model that did not possess any learned facial features. In RSA section

below, we introduced more details about how we extract features from these two VGG models, how we measured the representations of face perception in these two models, and how we compared the representational similarity between human EEG and these two models.

RSA

EEG RDMs

Given that we had 3 face conditions and 3 presentation conditions, we utilized EEG classification-based decoding accuracy as the dissimilarity metric to build a set of 9x9 neural Representational Dissimilarity Matrices (RDMs). For example, at a certain timepoint *t*, we assigned the decoding accuracy for the FD vs. SD conditions at timepoint *t* as the dissimilarity index between the FD condition and the SD condition in RDM as timepoint *t*.

Additionally, we generated Cross-Temporal RDMs (CTRDMs) instead of traditional RDMs to conduct Cross-Temporal Representational Similarity Analysis (CTRSA).[81] In CTRSA, we incorporated cross-temporal decoding accuracy instead of traditional RSA. Here's how it worked:

As illustrated in Figure 1C, we trained a Support Vector Machine (SVM) classifier on EEG data for the FD vs. SD conditions at timepoint t_A and then tested it on data for the FD vs. SD conditions at timepoint t_B. The resulting test accuracy was employed as the dissimilarity index between the FD condition at t_A and the SD condition at t_B in a t_A -> t_B CTRDM for a particular participant. Due to the directionality from the training time to the testing time, the CTRDM for t_A -> t_B was distinct from the CTRDM for t_B -> t_A. The former was constructed using data from t_A for training and t_B for testing, while the latter was constructed using data from t_B for training and t_A for testing. As we couldn't perform classification between two identical conditions, the diagonal values in the decoding-based CTRDMs were consistently set to 0. By following this approach, we established a CTRDM for each pair of directed timepoints (from one timepoint to another) as described above. Consequently, we generated a total of 100 (timepoints) x 100 (timepoints) CTRDMs based on cross-temporal EEG decoding. The EEG RDMs section was executed using the NeuroRA toolbox.[80]

DCNN RDMs

To handle the substantial number of nodes in each layer of the VGG-16 model, we initially applied Principal Component Analysis (PCA) to reduce the feature dimension. For instance, the second layer encompassed 64 112x112 feature maps, equating to a total of 802,816 nodes. For each image fed into the VGG-16 model, the layer 2 activation could be represented as a 1x802,816 vector. We conducted PCA on these vectors and organized the principal components in descending order of their contribution rates. We retained principal components that collectively contributed to over 95% of the variance, discarding the rest. This process effectively reduced the feature dimension of each layer. As depicted in Figure S1B, the dimensionality of layer 2 was trimmed to 307 after PCA. In a similar manner, we performed dimension reduction for all layers in both the VGG-Face and untrained VGG models. This dimension reduction step was implemented using the Scikit-learn toolkit.[84]

To process the images, we fed each one into both the VGG-Face and untrained VGG models, extracting activation vectors for every even layer (e.g., layer 2, 4, …, and layer 16) after dimension reduction. This resulted in 450 activation vectors for each layer, corresponding to the 450 images in our dataset. To manage the computational load, we calculated the Pearson correlation coefficient (r) between the activation vectors for any two images within each layer. We then used 1 minus the correlation coefficient (1 - r) as the dissimilarity index. For a given layer, we constructed a 450x450 Representational Dissimilarity Matrix (RDM) in the order of 150 familiar faces, 150 unfamiliar faces, and 150 scrambled faces. This procedure allowed us to obtain DCNN RDMs for all even layers in both the VGG-Face and untrained VGG models. The calculation of DCNN RDMs was carried out using the NeuroRA toolbox.[80]

Repetition suppression simulations in DCNN

To investigate the neural mechanism of facial repetition suppression using "reverse engineering," we developed two possible neuronal-level models: the Fatigue model and the Sharpening model. These models were designed to simulate how the neural responses in a deep convolutional neural network (DCNN) change under facial repetition suppression conditions. We set the activation vector of a face image *p* at layer *i* of a DCNN (before PCA) as $A = (a_1, a_2, \ldots\ldots, a_n)$, where the activation values were ordered in descending order ($a_1 > a_2 > a_3 > \ldots > a_{m-1} > a_m = a_{m+1} = \ldots = a_n = 0$), which meant that there were *n-m* nodes with nonzero activation value and m nodes with activation value of zero).

For Fatigue model, we assumed that the activation of the node with higher response to face stimulus was weaken under repetition suppression condition. The activation of the node with low response remained unchanged, but the node with higher activation had more attenuations. Thus, if the face image *p* was viewed repeatedly, the new activation vector obtained based on Fatigue model would be $A_F = (a_{F1}, a_{F2}, \ldots\ldots, a_{Fn})$, and its internal activations were:

$$a_{Fi} = \begin{cases} \left(1 - \alpha + \frac{(i-1)\alpha}{\beta n}\right) \cdot a_i & (1 \le i \le \beta n) \\ a_i & (\beta n < i \le n) \end{cases} \quad \text{(Equation 1)}$$

where α was the maximum fatigue coefficient, and β was the proportion of the nodes that would be attenuated. Thus, the first βn nodes would be attenuated when the same face image was viewed repeatedly, and the nodes from the strongest one to the βn th one would be weakened in proportion from α to $\alpha/(\beta n)$.

For Sharpening model, we assumed that the nodes with lower response to face stimulus were no longer activated under repetition suppression condition, and the nodes with higher response kept same activations. Thus, the new activation vector obtained based on Sharpening model would be $A_S = (a_{S1}, a_{S2}, \ldots\ldots, a_{Sn})$, and its internal activations were:

$$a_{Si} = \begin{cases} a_i & (1 \leq i < (1-\theta)n) \\ 0 & ((1-\theta)n \leq i \leq n) \end{cases} \quad \text{(Equation 2)}$$

where θ was the proportion of the nodes which would be not activated. Thus, the last θn nodes' activations would become zero when the same face image was viewed repeatedly.

Figure 1E is a schematic diagram of how these two repetition suppression models simulated in DCNNs. Here, Figure 1E shows the original activations of 30 active nodes and the activations under repetition suppression condition based on Fatigue model ($\alpha = 0.5$, $\beta = 0.5$) and Sharpening model ($\theta = 0.5$) respectively. For all 450 images, we input them into both VGG-Face and untrained VGG respectively, and then we calculated activation vectors of different layers based on different simulation models of repetition suppression.

Modified DCNN RDMs

Based on the above two facial repetition suppression models, we set model parameters corresponding to three types of face stimuli here: For Fatigue model, α was set to 0.9, which meant that the first 90% nodes with nonzero activation value were set as the nodes with high activation. And we set the maximum fatigue coefficient β to 0.5 for immediate repetition condition and 0.05 for delayed repetition condition. For Sharpening model, θ was set to 0.5 for immediate repetition condition and 0.05 for delayed repetition condition. Therefore, three activation vectors corresponding to new, immediate, and delayed conditions were calculated from the activation in each layer in DCNNs of each image based on each repetition suppression model. Then, we input these vectors into PCA to get feature vectors after dimension reduction. For each even layer, face image and repetition suppression model, we calculated the dissimilarity (1-Pearson correlation coefficient) between each pair of two feature vectors and got two 1350×1350 RDMs for VGG-Face and untrained VGG.

RSA between EEG and DCNNs

For VGG-Face and untrained VGG, 8 1350×1350 RDMs corresponding to 8 even layers were obtained. For EEG, each participant corresponded 100 time-by-time 9×9 RDMs and 100×100 cross-temporal 9×9 CTRDMs. To establish the connection between DCNNs and human brains, we averaged the cells under 9 conditions respectively (FN, FI, FD, UN, UI, UD, SN, SI, SD) in 1350×1350 RDMs to get compressed 9×9 DCNN RDMs. To measure the representational similarity between neural (CT)RDMs and DCNN RDMs, we first extract all cells of the top half of the diagonal in each RDM to get a corresponding vector which including 36 values, then we calculated the Spearman correlation coefficient as the similarity between every CT(RDM) and every DCNN layer RDMs. To obtain the similarity between the representation of face information that a DCNN learned for face recognition and human brains, we calculated the valid representational similarity S_{valid} below:

$$S_{valid} = S_{VGG-Face} - S_{Untrained\ VGG} \quad \text{(Equation 3)}$$

where $S_{VGG-Face}$ was the representational similarity between VGG-Face and neural activity, and $S_{Untrained\ VGG}$ was the representational similarity between untrained VGG and neural activity. Here, we applied untrained VGG's representations as a baseline that there was not enough face-specific information and calculated valid similarities for two repetition suppression models, respectively. To get face-specific repetition suppression mechanism, the pre-train VGG-Face trained on face recognition task would have stronger similarity with human brains than untrained VGG.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the classification-based decoding results, we assessed whether neural representations in the brain encoded information at specific time-points. We assumed that if this was the case, it would be possible to linearly classify between two conditions, resulting in decoding accuracy greater than chance, which is 50%.

Regarding the RSA results, we aimed to determine whether the valid similarity between conditions was significantly greater than zero. If the valid similarity values were either zero or less than zero, it would suggest that the corresponding repetition suppression mechanism was not specific to face information.

To compare the decoding accuracy to chance and assess whether valid similarity differed significantly from zero at each time-point, while also controlling for multiple comparisons, we employed cluster-based permutation tests. Here's a step-by-step outline of the procedure: (1) Calculate t-values for each time-point and identify significant clusters; (2) Compute the clustering statistic as the sum of t-values within each cluster; (3) Perform 5000 permutations to establish the maximum permutation cluster statistic; (4)Assign p-values to each cluster in the actual decoding accuracies or similarities dataset by comparing their cluster statistic to the permutation distribution. This approach helps us determine the statistical significance of the observed decoding accuracy and similarity values while accounting for multiple comparisons.