Detection of AI-Generated Contents Based on Dyadic-Brain Neural Synchronization

Shiang Hu[1], Lihao Fu[1], Xiao Gong[1], Piqiang Zhang[1], Yuhan Lin[2], Juan Hou[3], Zhao Lv[1]

[1]Key Lab of Intelligent Computing and Signal Processing of Ministry of Education, Anhui Provincial Key Lab of Multimodal Cognitive Computation, School of Computer Science and Technology, Anhui University, Hefei 230601, China

[2]Stony Brook Institute at Anhui University, Hefei 230039, China

[3]Department of Psychology, School of Philosophy, Anhui University, Hefei 230039, China

(Corresponding author: Shiang Hu, shu@ahu.edu.cn)

Abstract

The proliferation of DeepFake has engendered widespread societal concerns, positioning its detection as a pressing imperative. Although existing studies have utilized single-subject EEG to distinguish between real and AI-generated content (AIGC), there is still a lack of research exploring dual-brain EEG and multimodal experimental paradigms. This study introduced a novel experimental paradigm, employing EEG hyperscanning to construct a dyadic EEG dataset for AIGC detection. This study employed inter-subject correlation (ISC) analysis to investigate the differences of interpersonal neural synchronization (INS). Additionally, this study proposed a novel neural network model named Squeeze-and-Excitation Depthwise Separable Convolution (SEDSC) for predicting the authenticity of real vs. AIGC. ISC analysis revealed apparent differences in INS under different modalities, valences, and animacy. Specifically, across the four frequency bands, both text and audio modalities elicited higher inter-brain synchronization under real materials than under AIGC materials. SEDSC utilized the phase locking value to assess inter-brain functional connectivity and weighted the inputs from four frequency bands before feeding them into the network for classification. This approach achieved a classification accuracy of 92.42% in distinguishing real from fake content. This study designed a new experimental paradigm and constructed a dataset, confirming that there were evident differences in INS during tasks involving real and AIGC materials. Furthermore, SEDSC successfully predicted the authenticity of the content.

Keywords: Hyperscanning, EEG, AIGC, Deep Learning, Neural Synchrony

Introduction

With the rapid development of generative AI technology, it has demonstrated excellence in AI generated content (AIGC) such as image generation, text creation, and speech synthesis. The widespread application of AIGC has also introduced various potential social risks and harms (Busacca & Monaca, 2023; Tidler & Catrambone, 2021). Existing researches have shown that traditional visual perception are insufficient to effectively distinguish between real and AIGC. This challenge becomes particularly evident when confronting increasingly realistic DeepFake audio, as modern deep learning and machine learning algorithms struggle to differentiate between fake and real-world generated audio (Salehi et al., 2024). This issue has serious societal implications, as in the digital age, online communication is susceptible to being infiltrated by counterfeit content, especially when face-to-face interaction is not involved. This challenge extends beyond audio and affects images as well (Duffy et al., 2024). For instance, the widespread concerns surrounding AI-based face-swapping and facial synthesis technologies have not only caused distress to individuals but have also undermined public trust in the authenticity of information (Nirkin et al., 2021; Rana et al., 2022; Huang et al., 2023). Therefore, in addition to enhancing the discriminative capabilities of machine learning or neural network models, exploring alternative approaches for DeepFake detection becomes increasingly crucial.

Electroencephalography (EEG) enables real-time monitoring of electrical activity with millisecond-level temporal resolution. By placing electrodes on the scalp, EEG captures the electrical activity of cortical, providing an important tool for investigating human cognitive processes (Cohen, 2017). EEG has garnered attention for its potential application in DeepFake detection (Eiserbeck et al., 2023).

In 2022, Moshel combined DeepFake images with real images and designed a novel single-subject EEG paradigm. In their study, they presented faces, bedrooms, and bicycle images, and successfully decoded the brain activity related to AIGC faces (Moshel et al., 2022). In 2023, Tarchi took a different approach by designing DeepFake images with emotional content, not only exploring faces and non-faces but also investigating the EEG differences between DeepFake and real-world materials from an emotional perspective (Tarchi et al., 2023). Furthermore, another study highlighted that, despite the limitations of current deep learning and machine learning techniques in distinguishing real and fake content, EEG can be effectively used to differentiate between genuine and fake audio (Salehi et al., 2024).

It is worth noting that most existing studies are based on single-subject experiments, with limited diversity in the materials used, which restricts the generalizability of their findings. The present research employs the EEG hyperscanning technique to investigate whether there are differences in brain activity between two individuals when exposed to real and AIGC. Hyperscanning has been widely applied in multi-agent interaction scenarios, such as joint decision-making, cooperation-competition tasks, and verbal interaction (Chen et al., 2024; Balconi & Vanutelli, 2017; Balconi et al., 2018; S. Hu, Zhang, et al., 2023; Li et al., 2024), but there is limited research utilizing naturalistic stimuli in hyperscanning paradigms (Babiloni & Astolfi, 2014). Naturalistic stimuli, compared to highly controlled artificial stimuli, have greater ecological validity and are better suited to evoke real-world perception and cognition processes. For instance, visual stim-

uli often contain rich color and texture information, while auditory stimuli may involve complex tonal, rhythmic, and semantic content (Sonkusare et al., 2019). This research aims to explore the interpersonal neural response differences to real versus fake content, aligning more closely with real-world scenarios.

Given the scarcity of relevant datasets and the lack of prior research on EEG hyperscanning studies in this domain, we first designed an experimental paradigm incorporating different conditions and recorded EEG to construct a dataset. This research proposed two approaches for analysis. First, an inter brain synchrony based approach was employed for DeepFake detection, utilizing appropriate synchrony metrics such as the inter-subject correlation analysis introduced by Hasson et al. (2004). These synchrony measures were analyzed to uncover the cognitive neural mechanisms underlying the viewing of DeepFake and real materials. Second, a deep separable neural network based on the squeeze-and-excitation attention module (SEDSC) was constructed to predict the authenticity of content using hyperscanning data. In Wang's study (Wang et al., 2023), the four input channels were stacked without considering the distinct neurocognitive contributions of different frequency bands. To address this, our study introduced channel weighting during feature extraction, allowing the network to learn the relative importance of each frequency band. The weighted channels were then spatially stacked and fed into the network for classification, enhancing the model's ability to distinguish real from AIGC based on EEG synchrony. This study constructed a hyperscanning EEG dataset and provided robust neurocognitive evidence that dyads exhibited obvious divergences in inter-brain synchrony when processing real versus AIGC stimuli, with real stimuli consistently inducing higher levels of neural synchrony. Furthermore, leveraging the proposed SEDSC model, we successfully achieved binary classification between real and AIGC conditions, attaining a high classification accuracy, thereby demonstrating the model's efficacy in distinguishing between authentic and AIGC based on inter-brain synchrony dynamics.

Methods

Design

The experiment employed a within-subjects design to explore how dyadic EEG signals could differentiate between real and DeepFake materials. The experimental conditions were manipulated across three key factors: (1) Stimuli modality, (2) Valence, and (3) Animacy. The selection of modality, valence and animacy as research factors aims to explore the differences in neural activity between real and AIGC from multiple dimensions. The primary task for participants was to assess the authenticity of various materials presented to them. These materials included images, texts, and audio clips, which were either real or AIGC. Participants were asked to identify whether each stimulus was real or fake based on their subjective evaluation. The Modality factor included three levels:images, texts, and audio. The Valence factor encompassed three levels: positive, neutral, and negative. The Animacy factor distinguished between objects with humans and objects without humans (OWH and OWOH) based on the nature of stimuli. OWH included human faces, speech sounds, and texts involving conversations. OWOH included photos of everyday objects, natural sounds, and texts without any descriptions of people or conversations. Thus, the experiment featured a 3 (Modality: Image, Text, Audio) × 3 (Valence: Positive, Neutral, Negative) × 2 (Animacy: OWH, OWOH) design.

Participants

30 pairs of participants (N = 60) were initially recruited for this study, with two pairs excluded due to data quality issues, resulting in a final sample of 28 pairs. Participants were all right-handed, and aged 24.3 ± 3.2 years. The samples were balanced in terms of gender, with 15 males and 15 females. All participants had normal or corrected-to-normal vision, no history of neurological disorders, and were not under the influence of alcohol or drugs prior to the experiment. Informed consent was obtained from each participant, in accordance with ethical guidelines approved by the institutional review board. Each participant underwent the experiment in pairs, and the sessions were conducted in a controlled quiet environment. All participants were master degree candidates, with only minimal differences in their ability to distinguish AIGC.

Stimuli

To enhance the ecological validity of experimental paradigm and better approximate a naturalistic stimulus framework, the selected materials were grounded in real-world contexts, ensuring alignment with daily cognitive and perceptual experiences. The AIGC models were evaluated by a panel of twenty individuals to ensure the plausibility of their outputs and comparability to real materials.

Image The image stimuli were categorized by animacy (OWH vs. OWOH) and Valence (positive, neutral, negative). For real OWH images, we used the Radboud Faces Database (Langner et al., 2010), which provides high-quality, emotionally expressive human faces, balanced for gender and varied in lighting and pose. For real OWOH images, we selected photos from the Göttingen Affective Picture Database (Dan-Glauser & Scherer, 2011), which contains high-resolution images of objects and natural scenes, categorized by emotional valence. For AIGC OWH images, we used the Generated Photos platform to generate realistic, customizable human faces matching the required emotional valence (positive, neutral, negative). AIGC OWOH images were generated using Stable Diffusion (Rombach et al., 2022), a latent text-to-image diffusion model, to create photorealistic images of OWOH subjects that aligned with the desired emotional tone. All images, both real and generated, were carefully balanced for visual attributes such as gender, background color, size, lighting, and resolution to ensure consistency across conditions.

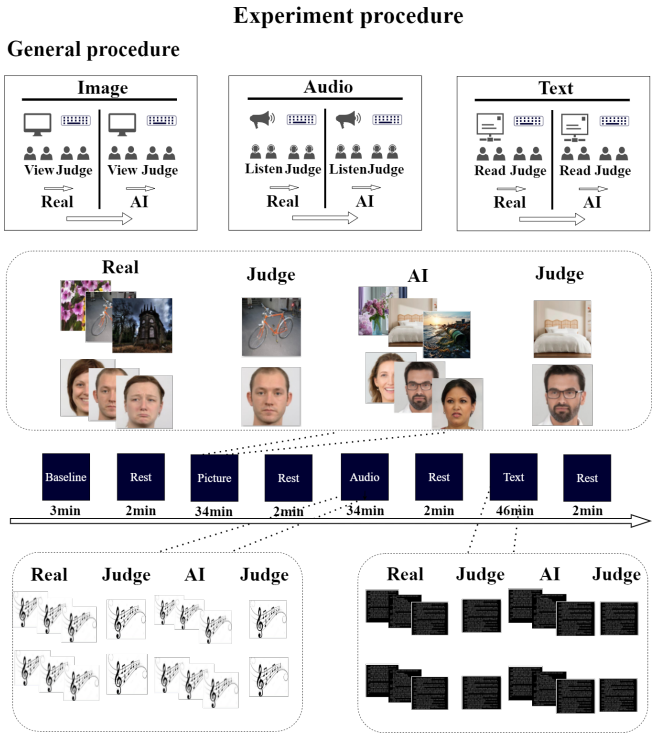

Figure 1: Experimental Paradigm. The experiment first conducts an image task, followed by an audio and text task. Each task consists of twelve groups, with the first six groups being real materials and the last six groups being AIGC materials. Each stimuli task followed by a real vs. fake judgement task.

Audio For real OWH audio, we selected three monologues reflecting different emotional valences (positive, neutral, and negative), each designed to replicate natural, real-life scenarios. This approach ensured that the content and emotional expression were aligned with everyday speech habits. In contrast to conversational dialogues, we focused exclusively on monologues to maintain consistency across conditions. The emotional content and tone of these monologues were evaluated through ratings from a panel of 20 individuals to ensure accurate emotional expression. For real OWOH audio, we selected natural or environmental sounds, such as traffic noise, drilling, and flowing water, and similarly evaluated them for emotional valence. For AIGC OWH audio, we used Reecho AI, a sophisticated model capable of generating human-like music. Emotional speech matching the valences of the real stimuli was produced to ensure consistency. AIGC OWOH audio was created using Stability AI, a platform that generates realistic audio from real sound samples. Both AIGC audio types were carefully matched with the real stimuli in terms of tone, emotional content, and speaker characteristics. All audio stimuli were balanced for key factors such as speaker number, tone, and content, ensuring consistency across conditions.

Text For real OWH text, we selected excerpts from three short novels by Mo Yan, a Nobel laureate known for his distinctive blend of magical realism and Chinese folklore. The excerpts were categorized into different emotional valences (positive, neutral, negative) based on ratings from a panel of 20 individuals. For real OWOH text, we used works from three literary masters: Gu Qingsheng, Yu Gongjin, and Lao She. These texts were similarly rated for emotional valence. AIGC OWH and OWOH texts were written by ChatGPT-4.0, instructed to imitate the selected authors' writing styles while ensuring the content were original and different from the real texts. We also maintained consistency in factors such as word count and stylistic elements, matching the original works' tone and structure.

Procedure

The experimental paradigm is illustrated in Fig 1. The experiment was conducted in a quiet, comfortable environment, where two participants sat side by side. And after signing the informed consent, both participants underwent a 3 mins resting EEG recording. The first task involved the presentation of image stimuli. Participants viewed a series of 12 sets of images, each containing 20 images. The images were categorized into real OWH and OWOH, with further subdivisions based on emotional valence. Each image was displayed for 3s, with each set shown for 1 min. After each set, participants were instructed to evaluate the authenticity of the presented images by pressing a designated key on the keyboard—'1' for real and '2' for fake—ensuring a standardized response format. Following each judgment, a 2 mins rest period was provided. The audio and text tasks followed the same general procedure. For the audio task, participants listened to a 1 min audio clip, and for the text task, they read a 2 mins passage. After each task, they made a judgment on whether the content was real or fake, followed by another 2 mins rest period.

Data Preprocessing

Due to issues with Fz electrode, its data was interpolated by averaging signals from adjacent F3, FCz, F4 electrodes. The EEG data were re-referenced to a common average to eliminate reference bias (S. Hu, Lai, et al., 2018; S. Hu et al., 2019; S. Hu, Yao, & Valdes-Sosa, 2018; S. Hu, Karahan, & Valdes-Sosa, 2018). A 48–52Hz notch filter was applied to remove power-line noise, followed by a 1–30Hz band-pass filter to exclude low-frequency drift and high-frequency noise. The data were downsampled to 250Hz for efficiency, and ICA was applied to remove artifacts related to eye movements, muscle activity, and other noise (S. Hu et al., 2025; S. Hu, Ruan, Langer, et al., 2023; S. Hu, Ruan, Zhang, et al., 2023).

Temporal decoding

As shown in Fig 2, we used Intersubject Correlation (ISC) to measure the synchronization of brain responses across participants while facing the same stimuli, which is particularly suited for naturalistic stimuli paradigms, as it allows the measurement of neural response consistency across individuals

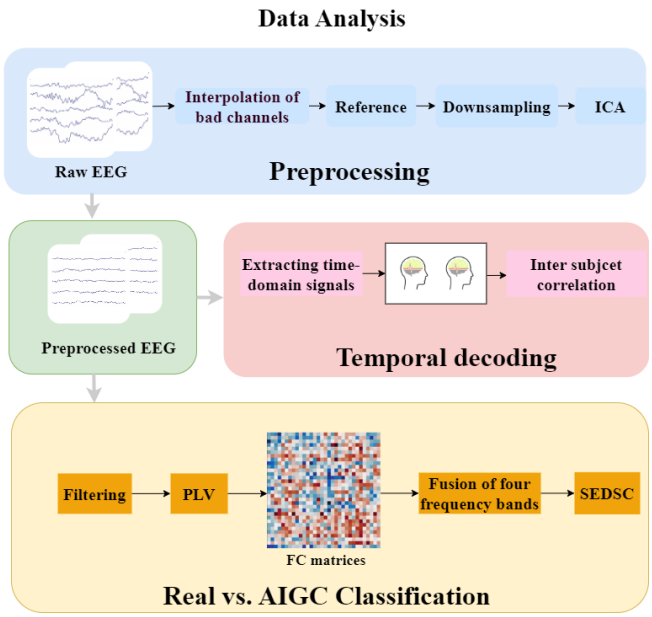

Figure 2: The data processing workflow is as follows:(1) Initially, the EEG data are preprocessed. (2) Subsequently, an ISC analysis is caculated on the EEG time series to quantify the degree of synchronization between subjects. (3) Finally, the EEG signals are filtered into four frequency bands, and the Phase-Locking Value (PLV) metric is computed to model the functional connectivity for each frequency band. The resulting four connectivity matrices are then used as input to a neural network for classification.

without the need for a pre-defined response model (Hasson et al., 2004). First, for each pair of participants k and l, we compute the cross-electrode covariance matrix C_{kl}, which is given by:

$$C_{kl} = \frac{1}{T}\sum_{t=1}^{T}(X_k(t)-\mu_k)(X_l(t)-\mu_l)^T \tag{1}$$

Here, T is the total number of time points, $X_k(t)$ and $X_l(t)$ represent the EEG signals of participants k and l at time point t, and μ_k and μ_l are their mean signals over time. This covariance matrix reflects the dyadic brain synchronization.

Next, we computed the ISC value for each pair of participants using the following formula:

$$ISC_{kl} = \frac{\sum_{t=1}^{T}(X_k(t)\cdot v_i)(X_l(t)\cdot v_i)}{\sqrt{\sum_{t=1}^{T}(X_k(t)\cdot v_i)^2\sum_{t=1}^{T}(X_l(t)\cdot v_i)^2}} \tag{2}$$

Here, i denotes the index of the principal component, v_i denotes the principal component vector extracted through eigenvalue decomposition of the matrix C_{kl}, which captures the dominant spatial pattern of neural activity shared between participants. The formula computes the normalized covariance between the dual EEG signals projected onto v_i, effectively quantifying the trial-to-trial response consistency (synchronization) along this neurophysiologically meaningful component.

Real vs. AIGC classification

As shown in Fig 2, we first computed the functional connectivity (FC) matrix using PLV, then constructed a dyadic interbrain network from this matrix, which was subsequently input into our proposed SEDSC.

PLV PLV measures the phase synchrony between two signals within a specific frequency band (S. Hu et al., 2024). The PLV between two subjects is used to assess the phase coupling of their EEG signals when processing the same stimulus. The formula is as follows:

$$\mathrm{PLV}(t) = \left|\frac{1}{N}\sum_{n=1}^{N}\exp\left(j\left[\theta_1(t,n)-\theta_2(t,n)\right]\right)\right| \tag{3}$$

where $\theta_1(t,n)$ and $\theta_2(t,n)$ are the instantaneous phases for the n-th channel of each subject, and N is the number of channels. Finally, the overall PLV for the experiment is averaged over all time points, which ranges from 0 to 1, with higher values indicating stronger synchrony.

SEDSC In the cognitive neuroscience-based DeepFake detection using ISC, evident differences across four frequency bands under different conditions have been observed. This indicated that each frequency band played distinct roles in processing real versus DeepFake materials. Accordingly, we extracted PLV from these four frequency bands and, did not simply stack the channels, which is different from the RFDSC model (Wang et al., 2023). We developed an advanced neural architecture, the Squeeze-and-Excitation-Enhanced Depthwise Separable Convolutional Network (SEDSC), as depicted in Fig 3, which integrated the Squeeze-and-Excitation (J. Hu et al., 2018) attention mechanism with a Depthwise Separable Convolutional (Howard et al., 2017) framework to process these PLV features. The Squeeze-and-Excitation (SE) module first took the feature vector X_1 with dimensions $32\times32\times4$ which is the input vector of the model and performs global average pooling (Squeeze operation), followed by the excitation operation. This process learned and reweighted the importance of each channel, resulting in a refined feature vector X_2 with the same dimensions, $32\times32\times4$. The reweighted feature vector X_2 is then passed through depthwise separable convolutional layers. Specifically, each of the four frequency bands in X_2 is convolved with four 1×32 kernels, producing a feature vector X_3 with dimensions $32\times1\times4$. Next, X_3 undergoes pointwise convolution with four 1×1 kernels to yield X_4, a feature vector with dimensions $32\times1\times4$. Finally, X_4 was flattened and inputted into a softmax for binary classification. However, the previous RFDSC model directly passed the raw input through depthwise convolution and pointwise convolution, followed by classification. In contrast, our model incorporated the SE module, which enhanced feature representation before classification. Following SE-driven reweighting, the refined feature maps are fed into

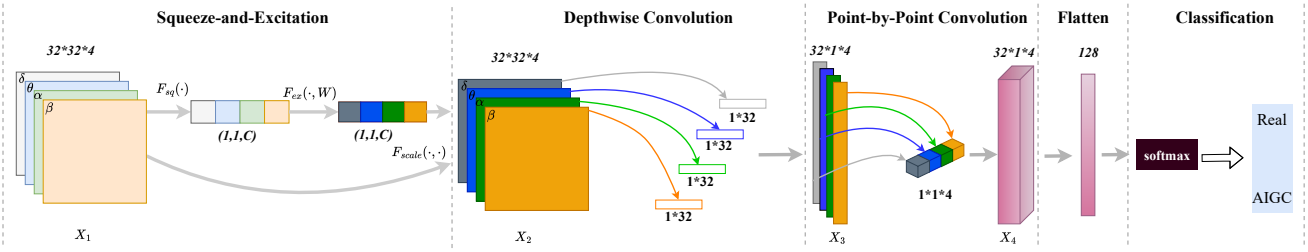

Figure 3: Model structure of SEDSC. These four frequency bands are spatially stacked and processed by an attention module to assign weights, highlighting the most important features for classification. The weighted features are passed through a depthwise separable convolution network to capture spatial and spectral information, then flattened and fed into fully connected layers for classification.

the depthwise separable convolutional layers. This efficient convolutional architecture disentangled spatial and channel-wise filtering, enhancing the network's ability to capture subtle inter-frequency dependencies while maintain- ing computational efficiency.

Result

Temporal decoding

The data were segmented to extract the necessary segments, and ISC was then calculated. Fig 4 reflects the ISC values of 28 pairs of participants under different conditions in the hyperscanning experiment. Paired-sample t-tests were applied to perform statistical analysis, with all tests corrected using False Discovery Rate (FDR) method.

Modality Panels a–d of Fig 4 reveals obvious differences in ISC between the real task and the AIGC task, with distinct patterns across modalities. In the text and audio modalities, ISC was consistently higher during the real task compared to the AIGC task. Specifically, in the β bands, the ISC values for the real task were significantly greater ($p < 0.0001$). In the audio task, ISC differences were observed across both low and high-frequency bands. The δ band showed a significant difference ($p < 0.01$), the β band also exhibited significant differences ($p < 0.05$). However, in the image modality, the pattern reversed: ISC was significantly higher during the AIGC task compared to the real, specifically in the β band ($p < 0.01$).

Valence Panels e–h of Fig 4 reveals the ISC results across different stimuli valence. In the negative valence, ISC induced by real task showed obvious differences across all frequency bands, with real materials producing consistently higher ISC compared to AIGC materials. Specifically, the ISC difference in the β band was highly significant ($p < 0.01$), and significant ISC differences were also observed in the other three bands (α, θ, and δ), all with $p < 0.05$. In the positive valence, ISC induced by real materials showed strong significances in both the θ and α bands ($p < 0.01$), while the ISC induced by AIGC materials were weaker. In the neutral valence, no significant ISC differences were observed between real and AIGC materials ($p > 0.05$). However, across all tasks, ISC induced by real materials was consistently higher than that induced by AIGC materials.

Animacy Panels i–j of Fig 4 shows the ISC values based on the animacy as OWH vs. OWOH. In the OWH group, ISC induced by real materials showed significant differences in both the θ and α bands, with differences observed at $p < 0.05$ in both bands. As there were no significant results in the δ and β bands, these results were omitted in the figure.

Real vs. AIGC classification

Our proposed SEDSC was compared with RFDSC and the traditional CNN with kernal 3×3, utilizing five-fold cross-validation. As shown in Table1, our model achieved much higher classification accuracy when inputting individual frequency band or all frequency bands, consistently outperforming both CNN and RFDSC, with the highest accuracy reaching 92.42%. Compared to RFDSC, our SEDSC model achieved improvements of 2.27%, 11.11%, 5.99%, 7.10%, and 9.69% under the δ, θ, α, β, and all-frequency conditions, respectively. Relative to CNN, SEDSC exhibited gains of 4.21%, 13.79%, 6.77%, 11.67%, and 15.52% across the five input conditions. This demonstrated that the inclusion of the attention mechanism enhanced the representation of the frequency bands, thereby improving the performance of SEDSC in classification tasks.

Discussion and Conclusion

In this study, we designed an innovative experimental paradigm, and for the first time integrated DeepFake detection with EEG hyperscanning. We conducted an experimental investigation and constructed a dual-subject EEG dataset, proposing two novel approaches for detecting AIGC. These approaches leverage both neuroscientific cognitive methodologies and advanced neural network for detection, both yielding considerable results. In the first method, leveraging neuroscientific principles, we calculated the interbrain neural synchrony by applying ISC analysis. Under our experimental conditions, real materials consistently elicited higher ISC across frequency bands, with the exception that AIGC images evoked greater ISC. Takagi proposed that AIGC image is ca-

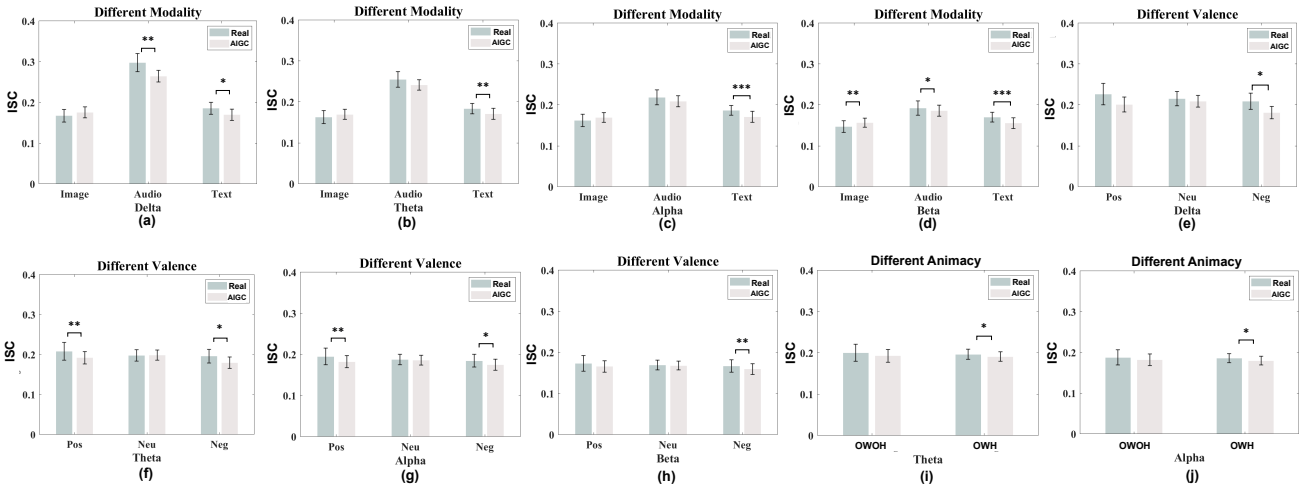

Figure 4: ISC differences as to different factors. (1) a-d is the ISC differences as to Stimuli Modality. (2) e-h is the ISC differences as to Stimuli Valence. (3) i-j is the ISC differences as to Stimuli Animacy. (*: P<0.05, **: P<0.01, ***: P<0.001)

Table 1: Classification accuracy based on different classifiers

Classifier	Rhythm Wave				
	δ (1-3Hz)	θ (4-7Hz)	α (8-12Hz)	β (13-30Hz)	All Frequency
CNN	78.23% ±3.15%	69.78% ±3.28%	74.56% ±3.42%	71.94% ±3.13%	74.89% ±3.27%
RFDSC	80.17% ±2.88%	72.46% ±3.05%	75.34% ±3.12%	76.51% ±3.19%	80.72% ±2.95%
SEDSC	**82.44% ±2.41%**	**83.57% ±2.63%**	**81.33% ±3.62%**	**83.61% ±2.69%**	**90.41% ±2.01%**

pable of reconstructing visual representations closely aligned with human brain activity (Takagi & Nishimoto, 2023), and may consequently elicit stronger neural responses. Although we did not delve deeply into the underlying neural mechanisms, the results demonstrated differences across all four frequency bands during the experiment, indicating distinct neural processing characteristics between brains. Based on these findings, we performed weighted fusion of the four frequency bands to extract the learning weights of neural network. We then designed a novel depth-separable convolutional neural network based on the Squeeze-and-Excitation attention mechanism (SEDSC), which weighted feature extraction from the four frequency bands for classification. Compared with individual frequency bands and spatial stacking of all frequency bands, our network achieved the highest classification accuracy of 92.42%, which improved the accuracy by 9.69% compared to the RFDSC, further validating the existence of significantly different neural responses. This study successfully employed two methods for DeepFake detection. Both at the level of neuroscience and in terms of neural networks, distinct patterns were successfully demonstrated. It provides new ideas and methods for the detection of AIGC-based DeepFakes.

However, while this study successfully proposed two methods for AIGC detection based on EEG synchronization and lightweight neural networks, its scope remains somewhat limited, as it did not delve deeply into potential influencing factors. Variables such as different modalities, valence, and animacy could play important roles in the underlying neural mechanisms. Moreover, our study has focused solely on images, audio, and text, and future research could expand to include video as a new modality as it carries more complex and richer information. And the participant group in this study was relatively small and homogenous. Future research should consider exploring more diverse populations. Furthermore, in addition to ISC analysis, future research could explore alternative metrics such as Inter-Subject Functional Connectivity, weighted Phase Lag Index, and mutual information, which are capable of capturing more complex nonlinear relationships and information flow structures. And they can partially suppress volume conduction effects, thus enhancing the comprehensive decoding of inter-brain neural synchrony. Morever there is potential for advancing deep learning models, such as by incorporating latest network architectures or optimization algorithms, to improve classification performance and adaptability to diverse data.

Acknowledgements

This work is supported by the National Natural ScienceFoundation of China (62101003), and the Chinese Ministry of Education (MOE) Project of Humanities and Social Sciences (22YJC190007)

References

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neuroscience & Biobehavioral Reviews*, *44*, 76–93.

Balconi, M., & Vanutelli, M. E. (2017). Cooperation and competition with hyperscanning methods: review and future application to emotion domain. *Frontiers in computational neuroscience*, *11*, 86.

Balconi, M., Vanutelli, M. E., et al. (2018). Eeg hyperscanning and behavioral synchronization during a joint actions. *Neuropsychological Trends*, *2018*(24), 23–47.

Busacca, A., & Monaca, M. A. (2023). Deepfake: Creation, purpose, risks. In *Innovations and economic and social changes due to artificial intelligence: The state of the art* (pp. 55–68). Springer.

Chen, Y. Y., Rosenbaum, G. M., Fan, H., Li, T., Flournoy, J., Mair, P., & Somerville, L. (2024). How beliefs around peers' risk preferences get incorporated into adolescents' decision making. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Cohen, M. X. (2017). Where does eeg come from and what does it mean? *Trends in neurosciences*, *40*(4), 208–218.

Dan-Glauser, E. S., & Scherer, K. R. (2011). The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, *43*, 468–477.

Duffy, S., August, A., & Wisniewski, K. (2024). Discriminating real from ai-generated faces: Effects of emotion, gender, and age. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Eiserbeck, A., Maier, M., Baum, J., & Abdel Rahman, R. (2023). Deepfake smiles matter less—the psychological and neural impact of presumed ai-generated faces. *Scientific Reports*, *13*(1), 16111.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *science*, *303*(5664), 1634–1640.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... Adam, H. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications (2017). *arXiv preprint arXiv:1704.04861*, *126*.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7132–7141).

Hu, S., Karahan, E., & Valdes-Sosa, P. A. (2018). Restate the reference for eeg microstate analysis. *arXiv preprint arXiv:1802.02701*.

Hu, S., Lai, Y., Valdes-Sosa, P. A., Bringas-Vega, M. L., & Yao, D. (2018). How do reference montage and electrodes setup affect the measured scalp eeg potentials? *Journal of neural engineering*, *15*(2), 026013.

Hu, S., Ruan, J., Langer, N., Bosch-Bayard, J., Lv, Z., Yao, D., & Valdes-Sosa, P. A. (2023). Spectral homogeneity cross frequencies can be a quality metric for the large-scale resting eeg preprocessing. *arXiv preprint arXiv:2310.11994*.

Hu, S., Ruan, J., Valdes-Sosa, P. A., & Lv, Z. (2025). How do the resting eeg preprocessing states affect the outcomes of postprocessing? *NeuroImage*, *310*, 121122.

Hu, S., Ruan, J., Zhang, Z., Langer, N., Bosch-Bayard, J., Yao, D., & Valdes-Sosa, P. (2023). The impacts of preprocessed eeg quality on the fidelity of spectra and functional connectivity. *International Journal of Psychophysiology*, *188*, 85–86.

Hu, S., Yao, D., Bringas-Vega, M. L., Qin, Y., & Valdes-Sosa, P. A. (2019). The statistics of eeg unipolar references: derivations and properties. *Brain topography*, *32*, 696–703.

Hu, S., Yao, D., & Valdes-Sosa, P. A. (2018). Unified bayesian estimator of eeg reference at infinity: rrest (regularized reference electrode standardization technique). *Frontiers in Neuroscience*, *12*, 297.

Hu, S., Zhang, P., Gong, X., Liu, C., Lv, Z., & Hou, J. (2023). Withdrawn: The impact of different forms of verbal interaction on neural synchronization in romantic couples.

Hu, S., Zhang, Z., Zhang, X., Wu, X., & Valdes-Sosa, P. A. (2024). ξ-π: A nonparametric model for neural power spectra decomposition. *IEEE Journal of Biomedical and Health Informatics*, *28*(5), 2624-2635. doi: 10.1109/JBHI.2024.3364499

Huang, B., Wang, Z., Yang, J., Ai, J., Zou, Q., Wang, Q., & Ye, D. (2023). Implicit identity driven deepfake face swapping detection. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 4490–4499).

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and emotion*, *24*(8), 1377–1388.

Li, Q., Kompatsiari, K., Dabranau, A., & Konvalinka, I. (2024). The role of friendship in dynamic group coordination. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Moshel, M. L., Robinson, A. K., Carlson, T. A., & Grootswagers, T. (2022). Are you for real? decoding realistic ai-generated faces from neural activity. *Vision Research*, *199*, 108079.

Nirkin, Y., Wolf, L., Keller, Y., & Hassner, T. (2021). Deepfake detection based on discrepancies between faces and their context. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 6111–6121.

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. *IEEE access*, *10*, 25494–25513.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 10684–10695).

Salehi, M., Stefanov, K., & Shareghi, E. (2024). Hu-

man brain exhibits distinct patterns when listening to fake versus real audio: Preliminary evidence. *arXiv preprint arXiv:2402.14982*.

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: critically acclaimed. *Trends in cognitive sciences*, *23*(8), 699–714.

Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 14453–14463).

Tarchi, P., Lanini, M. C., Frassineti, L., & Lanatà, A. (2023). Real and deepfake face recognition: An eeg study on cognitive and emotive implications. *Brain Sciences*.

Tidler, Z. R., & Catrambone, R. (2021). Individual differences in deepfake detection: Mindblindness and political orientation. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Wang, X., Shi, R., Wu, X., & Zhang, J. (2023). Decoding human interaction type from inter-brain synchronization by using eeg brain network. *IEEE Journal of Biomedical and Health Informatics*.