Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Overestimated performance of auditory attention decoding caused by experimental design in EEG recordings

Yujie Yan[1,3], Xiran Xu[2,3], Haolin Zhu[2,3], Songyi Li[2,3], Bo Wang[2,3], Xihong Wu[2,3], Jing Chen[1,2,3]

[1]Center for BioMed-X Research, Academy for Advanced Interdisciplinary Studies, Peking University, China, [2]Speech and Hearing Research Center, School of Intelligence Science and Technology, Peking University, China, [3]National Key Laboratory of General Artificial Intelligence, China

2301111611@stu.pku.edu.cn

Abstract

Auditory Attention Decoding (AAD) identifies a listener's focus in complex auditory scenes based on cortical neural responses. High decoding performance using DNN-based methods has been achieved with public EEG datasets. However, performance may be overestimated as models might learn temporal-autocorrelation features rather than auditory attention-related features. While data splitting risks have been discussed, experimental design risks have not. In this work, we collected a non-block design (NBD) scalp-EEG and ear-EEG joint dataset and compared it to previous block design (BD) datasets using DNN-based models. Results show a significant accuracy drop from BD to NBD dataset, while a linear stimulus reconstruction model remains robust. Inter-trial phase coherence analysis confirms stronger neural phase-locking to attended speech in BD dataset. These findings suggest BD enhances coherence of neural response but risks overestimating AAD accuracy. Code and data are released.

Index Terms: auditory attention decoding, EEG, experimental design

1. Introduction

In complex auditory scenes with multiple simultaneous speakers, listeners can focus on a target speaker while ignoring others, a phenomenon known as the "cocktail party problem" [1]. Neuroscience research has shown that auditory selective attention enhances cortical neural responses to the attended stream, enabling the development of auditory attention decoding (AAD) techniques that use EEG or MEG to identify the attended object [2]. It has also been demonstrated that auditory spatial attention is neurally encoded as brain lateralization, and the locus of attended location can be identified via EEG, when target and competing streams are spatially separated [3, 4, 5]. Based on these findings, auditory spatial attention decoding (ASAD) methods have been developed[6, 7, 8], with potential applications in cognitively controlled hearing devices [9, 10].

In EEG recordings for AAD studies, the trial serves as the fundamental experimental unit. During each trial, usually over 60-second, participants are required to maintain continuous auditory attention to a pre-defined target [11, 12, 13]. However, emerging evidence raises critical concerns in AAD regarding the effects of temporal autocorrelations (TAs) in EEG signals [14]. As adjacent temporal segments inherently exhibit higher signal similarity than distant ones, conventional "within-trial" data splitting strategies - when continuous EEG signals of a given class are segmented into samples and further split into training and test sets - risk inducing model overfitting to TAs features rather than auditory attention features [15, 16, 17, 18]. Although splitting strategies like "leave-trials-out" have been proposed to mitigate the overestimated performance [15], recent studies suggest that the persistence of TAs (spanning hundreds of seconds) might influence larger experimental units, such as blocks in block design experiment [17].

The block design paradigm, in which participants maintain identical task conditions across consecutive trials, is a common experimental design in cognitive research [19]. It indicates that this design enhances signal-to-noise ratio (SNR) through condition repetition and stabilizes behavioral responses, thereby facilitating neuroimaging data classification [20]. Block design is also employed in many AAD studies (e.g., [21, 22, 23, 24]). However, its impact in AAD research remains unclear:

1) In other BCI tasks, such as image decoding or imagined speech decoding, block design has been proposed to enable models to learn block-level TAs rather than stimulus-driven neural activity[17, 25]. Whether similar effects exist in AAD remains unexplored, potentially compromising the validity of AAD models.

2) When conditions are presented in blocks, humans tend to react more consistently and respond faster [26, 27]. However, whether block design offers similar benefits for AAD tasks—particularly for ASAD—remains ambiguous.

To investigate whether experimental design affects AAD-related experimental results, we collected a non-block design scalp-EEG and ear-EEG joint dataset under a multi-speaker condition. The experimental stimuli were identical to those used in previous work [24, 28], with the only difference being that, in our design, the direction of the target speaker was randomly varied for each trial, rather than remaining constant across the 10 trials within a block. This dataset was jointly analyzed with previous public block design datasets [24, 28]. We used representative ASAD models, including CNN [6] and STANet [7] to compare decoding performance between datasets. Given that these DNN-based models may be prone to overfitting on specific features, we also applied a linear stimulus reconstruction algorithm for further comparison. Inter-Trial Phase Coherence (ITPC, [29, 30]) was used to assess neural response phase coherence. The main contributions of this paper are summarized as follows:

1) We collect and release a non-block design scalp-EEG and ear-EEG joint dataset, which provides a resource for investigating AAD in more complex and realistic experimental settings.

2) We investigate the impact of experimental design on the performance of ASAD models, specifically exploring how block design and non-block design conditions influence decoding accuracy and neural feature extraction.

The implementation code and dataset are available on Github: `https://github.com/shirleyyan0120/BDandNBDcode-IS25` and Zenodo: `https://zenodo.org/records/14887886`.

10.21437/Interspeech.2025-85

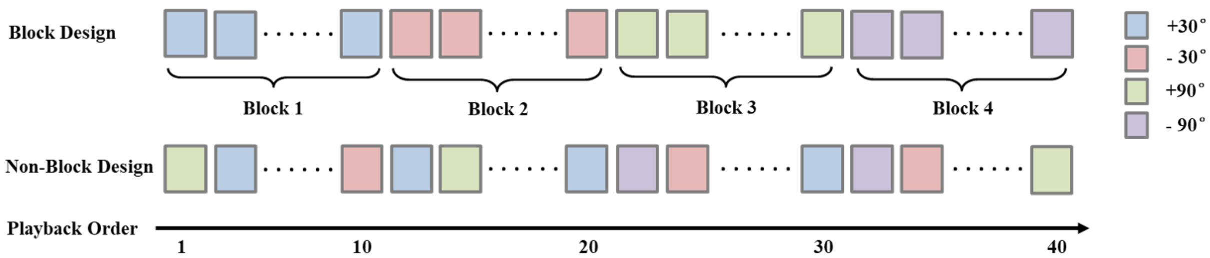

Figure 1: *Schemetic of Block Design and Non-Block Design in the experiment. Each square represents a trial, with different colors indicating different auditory spatial attention conditions.*

2. Materials and Methods

2.1. Participants

Sixteen university students with normal hearing, aged between 18 and 26 years, participated in the experiment. All participants were native Mandarin speakers and right-handed, with twelve of them being male. Prior to the experiment, they were briefed on the procedure and objectives. The study design followed ethical guidelines and was approved by the Peking University Institutional Review Board.

2.2. Stimuli and experimental procedure

The audio stimuli used in this experiment were drawn from the Chinese speech materials (*Twenty Thousand Leagues under the Sea* by Jules Verne) as described in previous work [21]. The preprocessing of these stimuli followed the same procedure as in [24, 28]. To ensure comparability, the experimental environment was kept the same as in previous works[24, 28]. Participants were instructed to focus on the audio from the target spatial location while ignoring the others. During stimulus presentation, subjects maintained visual fixation on a white crosshair displayed on the computer screen in front of them, keeping their head fixed.

Unlike previous studies [24, 28], this experiment employed a non-block design, where the sequence of the 40 trials was randomly shuffled before each experiment, meaning adjacent trials could involve different spatial attention conditions, as shown in Fig. 1. To ensure an even distribution of experimental conditions, each spatial attention condition was attended to in exactly 10 trials for a total of 40 trials.

2.3. EEG data acquisition and preprocessing

During the experiment, subjects were seated comfortably in an anechoic room. Scalp-EEG and ear-EEG were recorded simultaneously using two different devices while the subjects performed the auditory task. Scalp-EEG was collected with a 64-channel EEG amplifier (NeuSen Wireless EEG/ERP, Neuracle, China), and EEG signals were recorded at a sampling rate of 1000 Hz using the NeuSen Recorder software (Neuracle, China). The electrodes were placed using an EEG cap (Neuracle, China), following the international 10-20 system. Continuous ear-EEG data were recorded with a SAGA 32+/64+ amplifier (TMSi, Oldenzaal, the Netherlands) and two cEEGrids [31], acquired via Polybench 1.34 software (TMSi, Oldenzaal, the Netherlands). The cEEGrids were positioned around both ears, with an additional wrist electrode used as the ground. Ear-EEG data were re-referenced to a common average reference, sampled at 500 Hz, and stored for offline analysis.

The data preprocessing was identical to the procedures used in our previous work [24, 28]. For ASAD, the EEG data was down-sampled to 128 Hz, bandpass filtered between 14 and 31 Hz, and normalized. For subsequent analysis, both scalp-EEG and ear-EEG data were band-pass filtered between 2 and 8 Hz, as low-frequency neural activity (< 8 Hz) in the auditory cortex has been reported to be phase-locked to speech envelopes [21, 22]. The data were then baseline-corrected and downsampled to 64 Hz. These steps were performed using the EEGLAB toolbox in MATLAB.

2.4. ASAD models

The CNN [6] model and STANet [7] were utilized for all datasets to evaluate the performance of ASAD.

We trained a model for all subjects in one specific dataset, employing a 5-fold cross-validation. For each fold in each dataset, 60% of trials per subject were used for training, 20% for testing, and 20% for validation. These trials were further segmented into corresponding decision windows: 0.5-second, 1-second, and 2-second. The final results were calculated as the mean and standard deviation across these 5-folds.

2.5. Speech stimulus reconstruction

The SR method was the same as previous work [21, 24, 28, 32]. When extracting the speech temporal envelope, each clean speech waveform was processed through a filter bank of 8 bandpass filters to simulate the basilar membrane's bandpass characteristics. The center frequencies of these filters ranged from 150 Hz to 8000 Hz, evenly spaced on the equivalent rectangular bandwidth (ERB) scale. A Hilbert transform was then applied to the output of each filter to obtain the corresponding analytical envelope. To account for cochlear compression nonlinearity, the signals were raised to a power of 0.3, followed by low-pass filtering at a cutoff frequency of 8 Hz. Finally, the envelopes from all frequency bands were averaged.

A set of spatio-temporal filters $g(\tau, n)$ (also called decoders) represents the linear backward mapping from the neural response EEG $r(t, n)$ (sampled at discrete time $t = 1, , T$ at EEG channel $n = 1, , N$) to the temporal envelope of speech stimulus $s(t)$, as shown in equation 1:

$$\hat{s}(t) = \sum_{n} \sum_{\tau} r(t + \tau, n) g(\tau, n) \tag{1}$$

where $\hat{s}(t)$ represents the reconstructed speech envelope. The filters integrate the neural responses across a certain range of time lag τ, then the integrated signals are summed across channels to obtain the reconstructed speech envelope. For each trial, the reconstructed envelope was calculated through the decoder, and the correlation (Pearson's r) between the reconstructed and attended speech envelopes was computed. A speech stream was considered correctly identified if the correlation with the attended envelope was higher than with the other three unattended envelopes. The decoding accuracies of scalp-EEG and ear-EEG were evaluated by the percentage of correctly identified trials for each subject.

3. Results

3.1. ASAD models

As shown in Table 1, the results revealed a significant drop in decoding accuracy for both scalp-EEG and ear-EEG under non-block design conditions compared to block design conditions across all decision window lengths.

Table 1: *Comparison of ASAD accuracy (%) between datasets with different experimental designs achieved by CNN [6] and STANet [7] for 3 decision windows (0.5s, 1s, 2s). S-BD: Scalp-EEG with Block Design [24]; S-NBD: Scalp-EEG with Non-Block Design (Ours); E-EEG BD: Ear-EEG with Block Design [28]; E-NBD: Ear-EEG with Non-Block Design (Ours).*

Model	**Dataset**	**Decision Window**		
		0.5s	**1s**	**2s**
CNN	S-BD	86.8±2.4	82.5±2.0	76.6±3.1
	S-NBD	29.3±3.2	29.4±3.4	29.4±3.6
	E-BD	85.3±1.7	83.5±2.2	82.8±2.3
	E-NBD	32.0±2.1	33.2±2.4	34.0±2.6
STANet	S-BD	91.7±1.9	92.3±1.5	92.2±1.6
	S-NBD	28.8±2.9	29.1±2.8	27.5±1.6
	E-BD	91.1±0.8	91.4±0.8	90.4±0.8
	E-NBD	26.4±1.2	26.4±0.8	25.4±0.4

To further testify that the model is more likely to overfit to "block" features, we replaced the original trial labels of non-block design datasets (representing four spatial-directions, labeled 0–3) with pseudo-labels based on the trial's playback order within its 'block' (e.g., trials with playback order 1–10 were assigned label 0, trials with playback order 11–20 were assigned label 1, and so on), while keeping all other procedures identical to those in Table 1. For block design datasets, assigning pseudo-labels based on the block effectively mirrors the original true labels, since in block design experiments, all trials within the same block share the same label. As a result, the ASAD pseudo-label accuracy achieved by the CNN and STANet for block design data was identical to the results in Table 1 for block design. These results are presented in Table 2. The decoding accuracy for these pseudo-labels remained high in the non-block design condition, compared to the results in Table 1 for non-block design data.

3.2. Stimulus reconstruction

After observing a drop in decoding accuracy under the non-block design using ASAD models, we hypothesized that direct classification of labels in ASAD models could be one of the influencing factors. To further investigate this, we employed a simple linear reconstruction model to reconstruct the speech envelope and assess its impact. The speech temporal envelope

Table 2: *ASAD pseudo-label accuracy (%) achieved by CNN [6] and STANet [7] in our datasets for 3 decision windows (0.5s, 1s, 2s). S-pNBD: Scalp-EEG with Non-Block Design when giving pseudo-labels; E-pNBD: Ear-EEG with Non-Block Design when giving pseudo-labels.*

Model	**Dataset**	**Decision Window**		
		0.5s	**1s**	**2s**
CNN	S-pNBD	62.7±3.4	64.1±3.1	64.4±2.9
	E-pNBD	61.7±2.8	65.0±2.5	64.5±2.5
STANet	S-pNBD	63.2±3.2	59.0±4.7	44.2±5.2
	E-pNBD	57.8±3.7	53.4±4.0	40.4±4.6

was reconstructed from both scalp-EEG and ear-EEG data for each trial. The correlation coefficient (Pearson's r) between the reconstructed envelope and the envelope of each speech stream was computed to assess cortical envelope tracking for attended and unattended speech. Decoding accuracy, averaged across subjects for each decision window, is presented in Fig. 2.

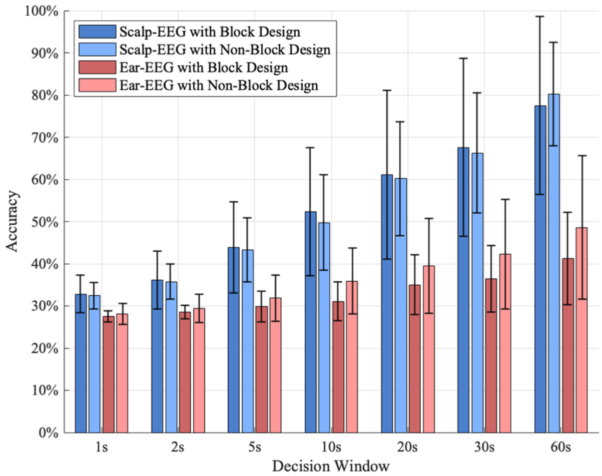

Figure 2: *Decoding accuracy of Stimulus Reconstruction for different decision window, signals and block/non-block design (chance level = 25%)). Bars in different colors represent results from different datasets.*

Our results were compared with those obtained under block design conditions for both scalp-EEG and ear-EEG, aligning with previous findings: the reconstructed envelope exhibited a stronger correlation with the attended speech compared to the unattended speech. Furthermore, the decoding accuracies were significantly higher than the chance level (25%), demonstrating robust neural tracking of the attended speech stream, validating the effectiveness of our dataset. Surprisingly, these results indicate that the decoding performance of the SR method showed no significant differences across different experimental designs. We hypothesize that this robustness stems from the fundamental nature of the SR method. Unlike ASAD methods, which classify discrete labels, SR focuses on reconstructing the time-varying speech envelope.

4. Discussions

4.1. Inter-Trial Phase Coherence analysis

To investigate whether block design enhances neural responses in AAD tasks, we employed Inter-Trial Phase Coherence (ITPC). ITPC measures the phase consistency of neural responses to speech at each frequency. We hypothesize that ITPC values under a block design paradigm will be higher than those under a non-block design. For a given spatial attention condition s, suppose there are K_s trials (10 trials per subject per condition). For each EEG channel and each trial k, the instantaneous phase $\varphi_k(t, f)$ was estimated using Morlet wavelet decomposition in logarithmic steps between 1 and 20 Hz, with wavelet cycles increasing from 4 to 10. The Inter-Trial Phase Coherence (ITPC) at each time-frequency bin was calculated by averaging phase angles across the K_s trials, as shown in equation 2:

$$\text{ITPC}_{t,f,s} = \frac{1}{K_s} \left| \sum_{k=1}^{K_s} e^{i \cdot \varphi_k(t,f)} \right| \tag{2}$$

$\varphi_k(t, f)$ represents the instantaneous phase for trial k at time t and frequency f, and K_s is the number of trials corresponding to the given spatial attention condition s. This procedure was repeated for all 4 spatial attention conditions, and for each attention condition, the within-condition ITPC was obtained by averaging the ITPC values across trials and time. We calculated the ITPC for each experimental design by averaging the within-condition ITPC value. The difference between block design and non-block design ITPC was then analyzed and reported.

To investigate the impact of block design and non-block design on the phase consistency of neural responses to speech, we computed the Inter-Trial Phase Coherence (ITPC) for controlled spatial attention conditions across four databases and averaged the results to enable cross-dataset comparisons. 3 shows representative ITPC values for scalp-EEG electrode (Cz, C3).

ITPC values for electrodes on both sides of the brain were higher under block design than those under the non-block design within 8-14 Hz which is associated with auditory spatial attention [3, 4, 5]. This suggests that block design enhances the detectability of neural responses related to auditory spatial attention, likely due to more consistent subject behavior under block design conditions, as previously conclusion [19]. Our ITPC results further support the notion that block design facilitates the observation of neural responses associated with auditory spatial attention.

4.2. Overestimated decoding performance in ASAD

Some researchers have noted overestimated decoding performance in ASAD, and the “leave-trials-out” splitting strategy has been proposed as a solution [15]. However, Xu et al. [17] suggested that overestimated decoding performance could occur whenever continuous EEG signals of a given class are segmented into samples and further split into training and test sets. When using a block design experimental setup, even with the “leave-trials-out” splitting strategy, models may still achieve overestimated decoding performance [24, 33]. Our study further supports this conclusion. Future ASAD work should exercise greater caution to avoid overestimating decoding performance due to experimental design and data splitting.

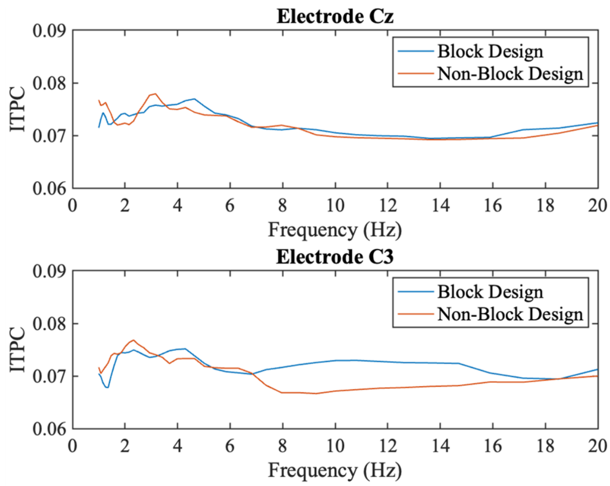

Figure 3: *ITPC for representative electrodes (Cz, C3) between Block Design and Non-Block Design. Lines in different colors represent results from different experimental design datasets.*

4.3. Limitations

While this study highlights the affect of experimental design in AAD, several limitations warrant consideration. Although the linear SR model demonstrated robustness across paradigms, it remains unclear whether other AAD reconstruction models exhibit similar performance. Moreover, our non-block design dataset, while ecologically valuable, did not fully replicate real-world auditory dynamics.

5. Conclusions

This study demonstrates that experimental design critically impacts auditory attention decoding (AAD). Results reveals a significant drop of decoding accuracy using typical DNN-based models from block design datasets to non-block design datasets, e.g. STANet: 92.3% vs. 29.1% for scalp-EEG in the 1s decision window. A further experiment classifying pseudo-labels (splitting into ‘block’ based on the trial’s playback order) achieved 64.1% accuracy in 1s. This indicates that block design risks overestimating ASAD accuracy as models might learn temporal-autocorrelation features rather than auditory attention-related features. In contrast, stimulus reconstruction (SR) methods showed robust performance across experimental designs. Inter-Trial Phase Coherence (ITPC) analysis further revealed stronger neural phase-locking to attended speech in block design, supporting its role in improving the coherence of neural responses to the target speech.

This work underscores the complexity of AAD tasks, showing that evaluation on limited data may be insufficient. Future research should consider broader scenarios and relevant factors, integrating experimental design awareness into AAD frameworks to improve real-world generalizability.

6. Acknowlegement

This work was supported by the STI 2030—Major Projects (No. 2021ZD0201500) and the High-performance Computing Platform of Peking University.

7. References

[1] E. C. Cherry, “Some Experiments on the Recognition of Speech, with One and with Two Ears,” *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

[2] J. A. O’Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, “Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG,” *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–1706, Jul. 2015.

[3] A. Bednar and E. C. Lalor, “Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG,” *NeuroImage*, vol. 205, p. 116283, Jan. 2020.

[4] Y. Deng, I. Choi, and B. Shinn-Cunningham, “Topographic specificity of alpha power during auditory spatial attention,” *NeuroImage*, vol. 207, p. 116360, Feb. 2020.

[5] M. Wöstmann, J. Vosskuhl, J. Obleser, and C. S. Herrmann, “Opposite effects of lateralised transcranial alpha versus gamma stimulation on auditory spatial attention,” *Brain Stimulation*, vol. 11, no. 4, pp. 752–758, Jul. 2018.

[6] S. Vandecappelle, L. Deckers, N. Das, A. H. Ansari, A. Bertrand, and T. Francart, “EEG-based detection of the locus of auditory attention with convolutional neural networks,” *eLife*, vol. 10, p. e56481, Apr. 2021.

[7] E. Su, S. Cai, L. Xie, H. Li, and T. Schultz, “STAnet: A Spatiotemporal Attention Network for Decoding Auditory Spatial Attention From EEG,” *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 7, pp. 2233–2242, Jul. 2022.

[8] X. Xu, B. Wang, Y. Yan, X. Wu, and J. Chen, “A densenet-based method for decoding auditory spatial attention with eeg,” in *International Conference on Acoustics, Speech and Signal Processing*, Seoul, Korea, Apr. 2024, p. 1946–1950.

[9] E. Ceolini, J. Hjortkjær, D. D. E. Wong, J. O’Sullivan, V. S. Raghavan, J. Herrero, A. D. Mehta, S.-C. Liu, and N. Mesgarani, “Brain-informed speech separation (BISS) for enhancement of target speaker in multitalker speech perception,” *NeuroImage*, vol. 223, p. 117282, Dec. 2020.

[10] S. Van Eyndhoven, T. Francart, and A. Bertrand, “EEG-Informed Attended Speaker Extraction From Recorded Speech Mixtures With Application in Neuro-Steered Hearing Prostheses,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1045–1056, May 2017.

[11] N. Das, J. Vanthornhout, T. Francart, and A. Bertrand, “Stimulus-aware spatial filtering for single-trial neural response and temporal response function estimation in high-density EEG with applications in auditory research,” *NeuroImage*, vol. 204, p. 116211, Jan. 2020.

[12] S. A. Fuglsang, T. Dau, and J. Hjortkjær, “Noise-robust cortical tracking of attended speech in real-world acoustic scenes,” *NeuroImage*, vol. 156, pp. 435–444, Aug. 2017.

[13] Y. Zhang, H. Ruan, Z. Yuan, H. Du, X. Gao, and J. Lu, “A Learnable Spatial Mapping for Decoding the Directional Focus of Auditory Attention Using EEG,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1–5.

[14] V. V. Nikulin and T. Brismar, “Long-range temporal correlations in alpha and beta oscillations: Effect of arousal level and test–retest reliability,” *Clinical Neurophysiology*, vol. 115, no. 8, pp. 1896–1908, Aug. 2004.

[15] C. Puffay, B. Accou, L. Bollens, M. J. Monesi, J. Vanthornhout, H. Van Hamme, and T. Francart, “Relating EEG to continuous speech using deep neural networks: A review,” *Journal of Neural Engineering*, vol. 20, no. 4, p. 041003, Aug. 2023.

[16] I. Rotaru, S. Geirnaert, N. Heintz, I. Van De Ryck, A. Bertrand, and T. Francart, “What are we really decoding? Unveiling biases in EEG-based decoding of the spatial focus of auditory attention,” *Journal of Neural Engineering*, vol. 21, no. 1, p. 016017, Feb. 2024.

[17] X. Xu, B. Wang, B. Xiao, Y. Niu, Y. Wang, X. Wu, and J. Chen, “Beware of Overestimated Decoding Performance Arising from Temporal Autocorrelations in Electroencephalogram Signals,” *arXiv preprint arXiv:2405.17024*, 2024.

[18] Z. Qiu, J. Gu, D. Yao, J. Li, and Y. Yan, “Enhancing spatial auditory attention decoding with wavelet-based prototype training,” *Biomedical Signal Processing and Control*, vol. 100, p. 107130, Feb. 2025.

[19] P. Simone, S. Concetto, S. Joseph, K. Isaak, G. Daniela, and S. Mubarak, “Correct block-design experiments mitigate temporal correlation bias in EEG classification,” *arXiv preprint arXiv:2012.03849*, 2020.

[20] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, “Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence,” *Scientific Reports*, vol. 6, no. 1, p. 27755, Jun. 2016.

[21] Z. Fu, X. Wu, and J. Chen, “Congruent audiovisual speech enhances auditory attention decoding with EEG,” *Journal of Neural Engineering*, vol. 16, no. 6, p. 066033, Nov. 2019.

[22] B. Wang, X. Xu, Y. Niu, C. Wu, X. Wu, and J. Chen, “Eeg-based auditory attention decoding with audiovisual speech for hearing-impaired listeners,” *Cerebral Cortex*, vol. 33, no. 22, pp. 10 972–10 983, 2023.

[23] E. Alickovic, E. H. N. Ng, L. Fiedler, S. Santurette, H. Innes-Brown, and C. Graversen, “Effects of hearing aid noise reduction on early and late cortical representations of competing talkers in noise,” *Frontiers in neuroscience*, vol. 15, p. 636060, 2021.

[24] Y. Yan, X. Xu, H. Zhu, P. Tian, Z. Ge, X. Wu, and J. Chen, “Auditory Attention Decoding in Four-Talker Environment with EEG,” in *Interspeech 2024*, Kos, Greece, Sep. 2024, pp. 432–436.

[25] R. Li, J. S. Johansen, H. Ahmed, T. V. Ilyevsky, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, “The Perils and Pitfalls of Block Design for EEG Classification Experiments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

[26] N. A. Roque, T. J. Wright, and W. R. Boot, “Do different attention capture paradigms measure different types of capture?” *Attention, Perception, & Psychophysics*, vol. 78, no. 7, pp. 2014–2030, Oct. 2016.

[27] L. E. Ethridge, S. Brahmbhatt, Y. Gao, J. E. Mcdowell, and B. A. Clementz, “Consider the context: Blocked versus interleaved presentation of antisaccade trials,” *Psychophysiology*, vol. 46, no. 5, pp. 1100–1107, Sep. 2009.

[28] H. Zhu, Y. Yan, X. Xu, Z. Ge, P. Tian, X. Wu, and J. Chen, “Using Ear-EEG to Decode Auditory Attention in Multiple-speaker Environment,” in *International Conference on Acoustics, Speech and Signal Processing*, Hyderabad, India, Apr. 2025, pp. 1–5.

[29] H. Luo and D. Poeppel, “Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex,” *Neuron*, vol. 54, no. 6, pp. 1001–1010, Jun. 2007.

[30] R. M. Van Diepen and A. Mazaheri, “The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience,” *Scientific Reports*, vol. 8, no. 1, p. 2990, Feb. 2018.

[31] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, “Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear,” *Scientific Reports*, vol. 5, no. 1, p. 16743, Nov. 2015.

[32] P. J. Schäfer, F. I. Corona-Strauss, R. Hannemann, S. A. Hillyard, and D. J. Strauss, “Testing the Limits of the Stimulus Reconstruction Approach: Auditory Attention Decoding in a Four-Speaker Free Field Environment,” *Trends in Hearing*, vol. 22, p. 233121651881660, Jan. 2018.

[33] M. A. Tanveer, M. A. Skoglund, B. Bernhardsson, and E. Alickovic, “Deep learning-based auditory attention decoding in listeners with hearing impairment*,” *Journal of Neural Engineering*, vol. 21, no. 3, p. 036022, Jun. 2024.