



GTAnet: Geometry-Guided Temporal Attention for EEG-Based Sound Source Tracking in Cocktail Party Scenarios

Saurav Pahuja¹, Gabriel Ivucic², Siqi Cai^{*3}, Dashanka De Silva¹, Haizhou Li^{1,4}, Tanja Schultz²

¹Machine Listening Lab (MLL), University of Bremen, Germany; ²Cognitive Systems Lab (CSL), University of Bremen, Germany; ³Department of ECE, National University of Singapore, Singapore; ⁴School of Data Science, The Chinese University of Hong Kong, Shenzhen, China

saurav@uni-bremen.de, caisiqi@ieee.org

Abstract

EEG-based sound source tracking in cocktail party scenarios is a challenging task due to the complex acoustic environment. To address this, we propose a novel approach, Geometry-guided Temporal Attention (GTAnet), which integrates the spatial configuration of multi-channel EEG signals with the temporal dynamics of neural activity. GTAnet constructs a geometry-based graph to capture the spatial relationships between electrodes while employing a temporal attention mechanism to highlight key intervals in auditory processing. The results show that GTAnet outperforms baseline models and offers interpretable insights into the neural mechanisms underlying auditory scene analysis.

Index Terms: Auditory attention, Trajectory reconstruction, EEG, Cocktail party problem

1. Introduction

Humans have the remarkable ability to focus their attention on a specific speaker or sound source in environments filled with competing speakers or background noise. This cognitive phenomenon, known as the cocktail party effect [1], has been widely studied in neuroscience, revealing strong connections between cortical activity and key features of attended speech, such as spectrotemporal representations [2]. Additionally, humans can perceive and track sound source positions and detect motion shifts [3]. This ability is largely facilitated by the Inferior Colliculus (IC), which processes Interaural Time Differences (ITDs) to estimate sound direction with high precision [4, 5].

Prior research has demonstrated that EEG can be used to reconstruct the azimuth of sound sources, even in complex auditory environments where multiple speakers are present [6, 7]. However, these studies mainly rely on linear trajectory mappings, which fail to capture the inherently non-linear characteristics of EEG signals [8]. This limitation is reflected in the relatively low Pearson correlation between actual and reconstructed auditory trajectories [6, 7].

With the rise of deep learning, selective auditory attention detection has seen significant advancements, particularly in static environments [8, 9]. Convolutional neural networks (CNNs) have been commonly employed to extract meaningful EEG features [9], but they do not fully account for the inter-channel relationships in EEG signals. To address this, Graph neural networks (GNNs) have been introduced, incorporating predefined adjacency matrices to model EEG channel connectivity [10]. However, existing GNN-based approaches suffer from two key limitations:

- EEG electrodes are typically arranged on the scalp in a non-planar (3D) configuration, as opposed to being on a flat 2D

surface. GNNs fail to capture the complex spatial relationships.

- Temporal dependencies of EEG signals, crucial for reconstructing the sound source trajectories during selective listening, are often not explicitly addressed.

Recent advances in the use of 3D geometric information have significantly improved performance of GNN-based regression tasks across various domains. For instance, in molecular chemistry, the Geometry-enhanced molecular representation learning method improves property prediction accuracy by capturing the intricate spatial relationships between atoms [11]. In neuroscience, GNNs have been used to model the relationships between EEG electrodes based on their 3D spatial configuration, enhancing brain activity analysis [12]. Furthermore, temporal attention mechanisms in EEG-based models—showcased by TAnet [13]—have been shown to decode salient temporal features from neural signals within short decision windows, suggesting a promising avenue for enhancing trajectory prediction by leveraging fine-grained temporal dynamics. In a related study, recent work on geometry-guided facial expression recognition leverages spatial-temporal graphs plus transformers to capture both local and non-local dependencies in facial dynamics, demonstrating the potential of 3D geometric data and attention-based GNNs [14]. These approaches highlight the potential of combining 3D geometric data with attention-enhanced GNNs to improve predictive performance in complex, dynamic systems.

Inspired by these findings, we propose GTAnet, a novel Geometry-Guided Temporal Attention Network designed to enhance EEG-based sound source tracking for both attended and unattended sound sources. Our model introduces two major novelties:

- **3D Geometry-Based Graph Construction:** We use the three-dimensional coordinates of EEG electrodes [15, 16] to generate a biologically plausible graph topology, reflecting volume conduction effects and ensuring closer electrodes have stronger connections.
- **Temporal Attention for Sound Source Tracking:** We employ a temporal attention mechanism across EEG sequences, allowing the model to prioritize the most relevant moments for distinguishing between attended and unattended trajectories.

The remainder of this paper is organized as follows. Section 2 introduces the proposed GTAnet framework, detailing its 3D graph construction and temporal attention mechanism. Section 3 describes the dataset and experimental settings used for evaluation. Section 4 presents and discusses the results of GTAnet on the sound source tracking task, highlighting its advantages over existing methods. Finally, Section 5 concludes the paper.

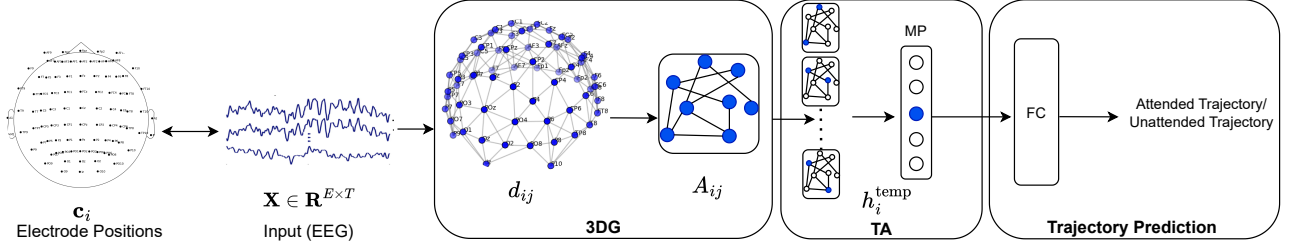


Figure 1: The overall architecture of GTAnet, which consists of three key components: 3D-Geometry Graph (3DG), temporal attention (TA), and trajectory prediction module. GTAnet is trained to track auditory attention by estimating the spatial trajectories of moving sound sources from EEG inputs. Specifically, 3DG extracts spatial features from EEG, while TA captures temporal dynamics. These features are then fed into the trajectory prediction module to generate attended/unattended trajectories.

2. Method

Following previous studies [6, 7], sound source tracking is realized by trajectory estimation. Specifically, the azimuth of a moving sound source (ranging between -90° and $+90^\circ$) is reconstructed from EEG signals for each short-time window in our study. A trial is defined as a continuous period during which a participant selectively attends to one of two moving sound sources. To enable dynamic tracking, each trial is segmented into short-time windows, each of which comprises a number of EEG samples.

Given one such window of EEG input, denoted $\mathbf{X} \in \mathbb{R}^{E \times T}$ —with E representing the number of EEG channels (nodes) and T the number of temporal samples (time steps)—GTAnet predicts the azimuth angle of the sound source for the window based on previous and current EEG window. In this work, the model tracks the sound source movement throughout the trial.

Our method capitalizes on the intrinsic spatial information provided by the electrode layout and the rich temporal dynamics inherent in EEG recordings. As illustrated in Figure 1, the approach is built upon three major components: (i) 3DG: 3D Geometry-Based Graph Construction, (ii) TA: Temporal Attention for temporal dynamics Extraction, and (iii) Trajectory Prediction via Neural Networks.

2.1. 3DG: 3D Geometry-Based Graph Construction

The spatial arrangement of EEG electrodes is valuable prior information for understanding neural signal propagation. We utilize the standardized BioSemi 128-channel montage, which specifies the 3D coordinates of each electrode. Let the coordinate of electrode i be denoted by

$$\mathbf{c}_i = [x_i, y_i, z_i] \quad (1)$$

with $i = 1, \dots, 128$. Our goal is to construct a graph where each electrode represents a node, and edges encode the spatial relationships between them.

To achieve this, we first compute the Euclidean distance between every pair of electrodes:

$$d_{ij} = \|\mathbf{c}_i - \mathbf{c}_j\|_2 \quad (2)$$

This distance quantifies the physical proximity between electrodes i and j in a 3-dimensional space. Closer electrodes are expected to capture more correlated neural signals due to volume conduction effects.

To convert these distances into a measure of connectivity, we apply a Gaussian kernel:

$$A_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right) \quad (3)$$

where σ is a hyperparameter controlling the rate of decay. This step produces a dense similarity matrix $\mathbf{A}$, which encodes the likelihood that two electrodes share meaningful information. However, many of these connections may be weak or noisy. To focus on the most significant spatial relationships and reduce computational complexity, we apply a threshold τ to sparsify the matrix. Only connections with $A_{ij} > \tau$ are retained, resulting in a sparsified adjacency matrix $\hat{\mathbf{A}}$. This matrix serves as a blueprint for subsequent processing, ensuring that our network leverages only the most reliable spatial connections.

2.2. TA: Temporal Attention for Feature Extraction

EEG signals exhibit dynamic temporal patterns, which are critical for accurate trajectory prediction. Therefore, it is crucial for the model to identify and emphasize these informative temporal segments to improve prediction accuracy.

For each time step t , the EEG data is represented as a vector $\mathbf{x}_t \in \mathbb{R}^{128}$, corresponding to the readings from all electrodes. For a batch of recordings, we represent the data as a tensor

$$\mathbf{X} \in \mathbb{R}^{B \times T \times 128} \quad (4)$$

where B is the batch size and T the number of time steps. At every time step, the electrodes (nodes) are connected by the previously constructed sparsified graph $\hat{\mathbf{A}}$.

To capture the spatial dependencies at each time step, we apply a graph attention mechanism. In essence, this mechanism allows the network to learn which neighboring electrodes are most informative for each node. For a given electrode i and its neighbor j , we compute an attention score:

$$e_{ij} = \text{LeakyReLU}\left(a^\top [W\mathbf{x}_i \parallel W\mathbf{x}_j]\right) \quad (5)$$

where W is a learnable weight matrix, a is a learnable attention vector, and $\parallel$ denotes feature concatenation. The scores are normalized over the neighborhood of each electrode using a softmax function, leading to adaptive weights α_{ij} . With these weights, we aggregate the features from neighboring electrodes:

$$h_{i,t} = \text{ELU}\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} W\mathbf{x}_j\right) \quad (6)$$

where $\text{ELU}(\cdot)$ is a nonlinear activation function. This process enhances the spatial representation at every time step by emphasizing more informative regions of the scalp.

Temporal dynamics are captured by processing the sequence of spatial features $\{h_{i,t}\}_{t=1}^T$ through a temporal attention mechanism. This module learns to assign different weights to different time steps, effectively highlighting those moments that are most predictive of the target trajectory. By integrating the spatially enhanced features over time, we obtain a comprehensive spatiotemporal representation for each electrode:

$$h_i^{\text{temp}} = \text{Aggregation} \left(\{h_{i,t}\}_{t=1}^T \right) \quad (7)$$

Here, ‘‘aggregation’’ refers to the learned temporal pooling process, which ensures that both short-term fluctuations and long-term trends are captured.

To enhance robustness and capture critical activations, we apply max pooling after the temporal attention mechanism. This operation preserves peak responses across time steps, complementing the weighted aggregation from temporal attention to emphasize both temporal importance and the strongest neural signals.

2.3. Trajectory Prediction

Once the spatiotemporal features have been extracted from the EEG data, these features must be transformed into a prediction of the target trajectory. The enhanced node representations h_i^{temp} (for all electrodes) are stacked into a tensor

$$\mathbf{H} \in \mathbb{R}^{B \times 128 \times d} \quad (8)$$

where d is the feature dimension after attention processing. This tensor is then flattened into a single feature vector for each instance:

$$\mathbf{h} = \text{Flatten}(\mathbf{H}) \quad (9)$$

The flattened features are passed through a series of fully connected (dense) layers. These layers serve two main purposes: they condense the high-dimensional features into a more compact representation and they introduce further non-linearity to capture complex dependencies. The final dense layer produces the predicted trajectory $\hat{y}$ via a simple linear transformation:

$$\hat{y} = W_{\text{dense}} \mathbf{h} + b_{\text{dense}} \quad (10)$$

where W_{dense} and b_{dense} are trainable parameters.

2.4. Overall Framework

Our proposed GTAnet integrates spatial priors (3D) with dynamic temporal modeling to predict trajectories from EEG data. The method consists of the following key steps:

1. **3D Geometry-Based Graph Construction:** Constructing a sparsified graph from the 3D coordinates of EEG electrodes using a Gaussian kernel and thresholding.
2. **Temporal Attention:** Enhancing the EEG signals by applying spatial attention at each time step, followed by a temporal attention mechanism to aggregate dynamic features.
3. **Trajectory Prediction:** Transforming the resulting spatiotemporal features through fully connected layers to produce the final trajectory prediction.

The model is trained end-to-end by minimizing the mean squared error (MSE) between the predicted trajectory $\hat{y}$ and the ground truth y . We employ the Adam optimizer, which adapts

the learning rate during training for improved convergence. To further enhance robustness and generalization, we incorporate batch normalization and dropout within our network. These techniques help mitigate overfitting and ensure that the network learns a representation that generalizes well to unseen data.

3. Experiments

3.1. Dataset

We conducted our experiments using the SEED corpus, a dataset provided by Trinity College Dublin [7], referred to as the Trinity dataset hereafter. This dataset consists of EEG recordings from 10 participants, each wearing headphones while listening to two simultaneously moving speech streams. These speech streams were read by two different male speakers, with participants instructed to focus on only one of them in a trial.

The auditory stimuli were spatially rendered using the Oculus Audio SDK, simulating dynamic and unpredictable movements within the frontal horizontal plane. Both speech streams moved continuously and randomly within this plane, covering angles between -90° (left) and $+90^\circ$ (right) relative to the participant. The motion followed an angular velocity of 50° per second. Each participant completed a speech comprehension task consisting of 40 trials, with attention alternating between the two speakers after every 20 trials. Each trial lasted 60 seconds, and for analysis, we segmented the data into 60 one-second windows.

EEG signals were recorded using a 128-channel Biosemi ActiveTwo system [17], bandpass filtered between 0.02 and 30 Hz, and downsampled to 64 Hz. To remove artifacts caused by eye movements, we applied Independent Component Analysis (ICA) along with the ADJUST algorithm [18].

3.2. Experimental Settings

We implemented the 3D Geometry-Based Graph and Temporal Attention components using TensorFlow. The training process aimed to minimize the MSE loss function with a learning rate of 10^{-3} , optimized using Adam. Each training batch contained 60 windows of EEG signals and each window consist of 64 samples per second as hop size, and the model was trained for 100 iterations. The implementation was built using TensorFlow 2.14 in Python 3.8, running on a GeForce RTX 3090 GPU and an Intel i9-12900K CPU.

For model evaluation, we conducted leave-one-trial-out cross-validation [19] in a subject-dependent setting. The EEG data for each subject was divided into training, validation, and test sets. In each cross-validation fold, the model was trained on 38 trials, fine-tuned on one trial, and tested on the remaining trial. This process was repeated 40 times across all trials for every subject, and the results were averaged to assess the overall model performance.

Following prior studies [7, 20], we evaluated the model using two key metrics:

- **Mean Squared Error (MSE):** Measuring the prediction error between the reference and predicted trajectory.
- **Pearson Correlation Coefficient (PCC):** Assessing the linear relationship between the actual and predicted trajectory.

These metrics provide insight into how effectively the model tracks attended and unattended speech trajectories.

4. Results and Discussion

4.1. Comparative Study

Due to the lack of prior research on EEG-based sound source tracking with deep learning, we compare GTAnet with established benchmark models used in EEG-based kinematics and motor trajectory reconstruction [21, 22]. Specifically, we evaluate four models: (1) Linear Regression (LR) [7], (2) CNN-LSTM [23] for learning temporal features, (3) GCN-LSTM [24] for static spatial-temporal learning, and (4) GTAnet, which incorporates 3D electrode geometry and temporal attention to model EEG signals propagation dynamically.

As summarized in Table 1, our proposed GTAnet consistently achieves the highest PCC across all trials and subjects, indicating that GTAnet effectively reconstructing both attended and unattended sound source trajectories. Additionally, GTAnet yields the lowest Mean Squared Error (MSE), particularly in reconstructing attended trajectories, which are critical for selective auditory attention. Notably, GTAnet has only 5,528 parameters—fewer than GCN-LSTM (8,686) and CNN-LSTM (9,256). Despite its lower parameter count, GTAnet’s 3D graph-based processing and temporal attention mechanisms enhance both efficiency and tracking performance. As illustrated in Figure 2, the predicted trajectory (orange curve) aligns more closely with the ground truth (blue curve) in the attended condition compared to the unattended one. Notably, GTAnet effectively tracks left and right movements in the frontal space, demonstrating its robustness in spatial sound source tracking.

Compared to LR [7], GTAnet’s ability to incorporate temporal features is critical for handling dynamic environments with unpredictable trajectories. In contrast to CNN-LSTM, GTAnet excels in capturing spatiotemporal dependencies, as CNN-LSTM primarily focuses on temporal learning and lacks structural connectivity. While both GTAnet and GCN-LSTM model spatial and temporal dependencies, GTAnet outperforms GCN-LSTM by utilizing 3D graph geometry instead of a static adjacency matrix, which reduces complexity and improves interpretability. Figure 3 highlights this advantage, showing high attention weights in temporal and parietal channels, which are known to play a critical role in auditory attention during sound source tracking tasks [25, 26].

In summary, GTAnet’s superior performance in capturing spatiotemporal dependencies highlights its potential as a powerful tool for continuous EEG-based sound source tracking. Its ability to efficiently model dynamic environments, combined with its compact architecture and interpretable design, positions GTAnet as a promising approach in this field.

Table 1: Average Mean Squared Error (MSE) and Pearson Correlation Coefficient (PCC) scores on the Trinity dataset with 1-second windows. Models marked with * show a significant difference compared to the proposed GTAnet (paired t -test ($p < 0.05$)).

Model	Attended		Unattended	
	MSE	PCC	MSE	PCC
LR (baseline) [7]	0.41*	0.06*	0.44*	0.02*
CNN-LSTM [23]	0.32*	0.09*	0.36*	0.06*
GCN-LSTM [24]	0.29*	0.12*	0.31*	0.09*
GTAnet (ours)	0.21	0.23	0.25	0.18

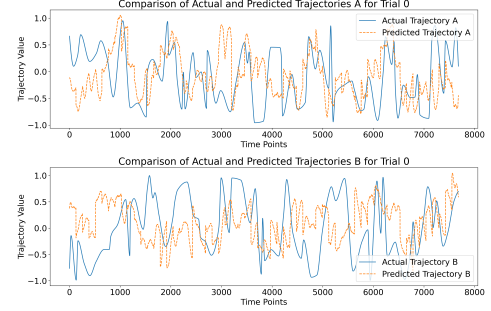


Figure 2: Comparison of actual (blue) and predicted (orange) sound source trajectories for both attended (top) and unattended (bottom) conditions. Here the trajectory value represents the angle of the azimuth.

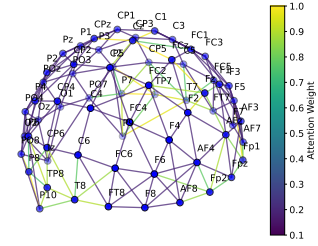


Figure 3: Visualization of GTAnet, generated by averaging data from 40 trials across 10 subjects. This spatiotemporal representation combines spatial and temporal attention weights. Node size indicates electrode importance, while edge color intensity and thickness represent spatial attention weight strength.

4.2. Ablation Analysis

We performed ablation experiments on GTAnet using the Trinity dataset with a 1-second window, including: (i) replacing the TA module with dense layers, which resulted in the Gnet model, and (ii) removing the 3DG module, which resulted in the TAnet model.

As shown in Table I, CNN-LSTM achieved a PCC of 0.09 (attended) and 0.06 (unattended), and for GCN-LSTM with PCC of 0.12 (attended) and 0.09 (unattended) while Gnet outperformed it with 0.18 and 0.15, respectively, showing a 0.09 PCC improvement. TAnet also surpassed CNN-LSTM and GCN-LSTM, achieving 0.16 (attended) and 0.13 (unattended). GTAnet achieved the highest performance, with a PCC of 0.23 (attended) and 0.18 (unattended), demonstrating the effectiveness of combining 3D graph processing and temporal attention.

5. Conclusion

In this study, we introduced GTAnet, a novel model that integrates 3D electrode geometry, temporal attention, and end-to-end regression to predict sound source trajectories from EEG data. By leveraging spatial priors based on the 3D proximity of electrodes and temporal attention, our model enhances the feature representation of EEG signals. Results confirm its effectiveness in capturing spatiotemporal EEG dynamics, with GTAnet outperforming other baseline models. Furthermore, the interpretability of GTAnet offers valuable insights into the underlying neural mechanisms involved in tracking moving sound sources.

6. Acknowledgements

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (University Allowance, EXC 2077, University of Bremen, Germany).

7. References

- [1] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *Journal of the Acoustical Society of America*, vol. 25, pp. 975–979, 1953. [Online]. Available: <https://doi.org/10.1121/1.1907229>
- [2] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, pp. 233–236, 2012. [Online]. Available: <https://doi.org/10.1038/nature11020>
- [3] Y. Deng, I. Choi, and B. Shinn-Cunningham, "Topographic specificity of alpha power during auditory spatial attention," *NeuroImage*, vol. 207, p. 116360, 2020. [Online]. Available: <https://doi.org/10.1016/j.neuroimage.2019.116360>
- [4] P. X. Joris, M. Van der Heijden, and P. H. Smith, "Temporal aspects of coincidence detection in the medial superior olive," *Journal of Neuroscience*, vol. 18, no. 22, pp. 10 131–10 141, 1998.
- [5] M. D. Grothe B, Pecka M, "Mechanisms of sound localization in mammals," *Physiol Rev.*, vol. 90(3), pp. 983–1012, 2010.
- [6] A. Bednar and E. C. Lalor, "Neural tracking of auditory motion is reflected by delta phase and alpha power of eeg," *NeuroImage*, vol. 181, pp. 683–691, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:51726606>
- [7] 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, 2020.
- [8] S. Pahuja, S. Cai, T. Schultz, and H. Li, "XAnet: Cross-Attention Between EEG of Left and Right Brain for Auditory Attention Decoding," in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2023, pp. 1–4.
- [9] 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, 2022.
- [10] S. Cai, T. Schultz, and H. Li, "Brain Topology Modeling With EEG-Graphs for Auditory Spatial Attention Detection," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 1, pp. 171–182, 2024.
- [11] X. Fang, L. Liu, J. Lei, D. He, S. Zhang, J. Zhou, F. Wang, H. Wu, and H. Wang, "Geometry-enhanced molecular representation learning for property prediction," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 127–134, 2022.
- [12] Z. Li, K. Hwang, K. Li, J. Wu, and T. Ji, "Graph-generative neural network for eeg-based epileptic seizure detection via discovery of dynamic brain functional connectivity," *Scientific Reports*, vol. 12, no. 1, p. 18998, 2022.
- [13] Y. Ding and F. Chen, "Tanet: A new temporal attention network for eeg-based auditory spatial attention decoding with a short decision window," 2024. [Online]. Available: <https://arxiv.org/abs/2401.05819>
- [14] R. Zhao, T. Liu, Z. Huang, D. P. Lun, and K.-M. Lam, "Spatial-temporal graphs plus transformers for geometry-guided facial expression recognition," *IEEE Transactions on Affective Computing*, vol. 14, no. 4, pp. 2751–2767, 2023.
- [15] V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," *NeuroImage*, vol. 34, no. 4, pp. 1600–1611, 2007. [Online]. Available: <https://doi.org/10.1016/j.neuroimage.2006.09.024>
- [16] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.
- [17] BioSemi, *ActiveTwo System*, 2023, available online: <http://www.biosemi.com/products.htm>. [Online]. Available: <http://www.biosemi.com/products.htm>
- [18] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011. [Online]. Available: <https://doi.org/10.1111/j.1469-8986.2010.01061.x>
- [19] G. Ivucic, S. Pahuja, F. Putze, S. Cai, H. Li, and T. Schultz, "The impact of cross-validation schemes for eeg-based auditory attention detection with deep neural networks," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.
- [20] B. Su and E. Gutierrez-Farewik, "Gait Trajectory and Gait Phase Prediction Based on an LSTM Network," *Sensors*, vol. 20, no. 24, p. 7127, 2020.
- [21] P. Wang, X. Cao, Y. Zhou, P. Gong, M. Yousefnezhad, W. Shao, and D. Zhang, "A comprehensive review on motion trajectory reconstruction for EEG-based brain-computer interface," *Frontiers in neuroscience*, vol. 17, p. 1086472, 2023. [Online]. Available: <https://doi.org/10.3389/fnins.2023.1086472>
- [22] A. Jain and L. Kumar, "Subject-independent trajectory prediction using pre-movement EEG during grasp and lift task," *Biomedical Signal Processing and Control*, vol. 86, p. 105160, 2023. [Online]. Available: <https://doi.org/10.1016/j.bspc.2023.105160>
- [23] D. Borra, V. Mondini, E. Magosso, and G. R. Müller-Putz, "Decoding movement kinematics from EEG using an interpretable convolutional neural network," *Computers in Biology and Medicine*, vol. 165, p. 107323, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482523007886>
- [24] Z. Wu, M. Huang, A. Zhao, and Z. Lan, "Traffic prediction based on GCN-LSTM model," in *Journal of Physics: Conference Series*, vol. 1972. IOP Publishing, 2021, p. 012107, international Conference on Internet of Things and Smart City (IoTSC 2021), 4–6 June 2021, Kunming, China.
- [25] 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. 1, pp. 1697–1706, 2015.
- [26] J. R. Kerlin, A. J. Shahin, and L. M. Miller, "Attentional gain control of ongoing cortical speech representations in a "cocktail party"," *Journal of Neuroscience*, vol. 30, no. 2, pp. 620–628, 2010.