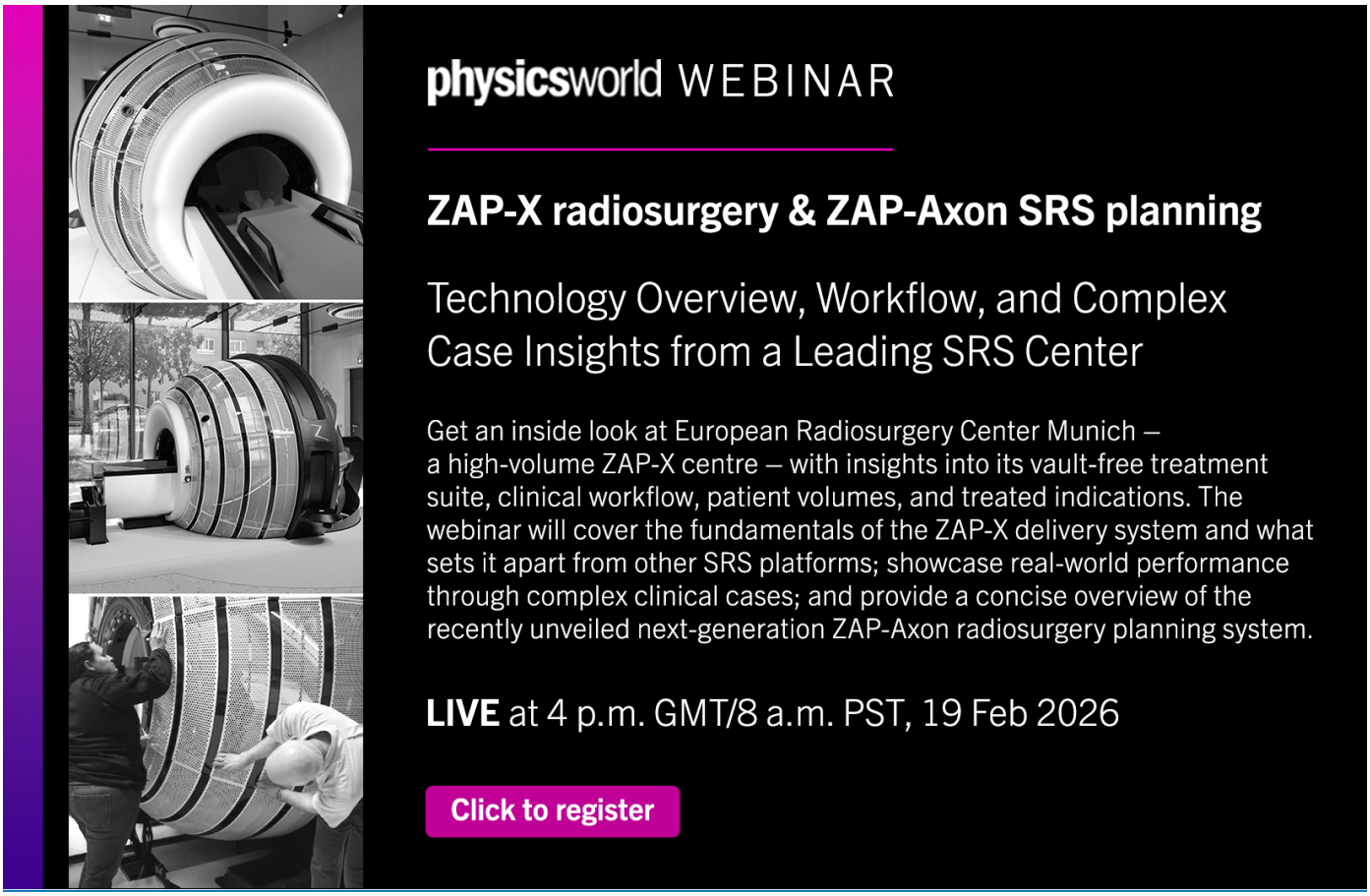
Journal of Neural Engineering

PAPER • **OPEN ACCESS**

Dynamic graph representation of EEG signals for speech imagery recognition

To cite this article: Cengiz Selcuk and Nikolaos V Boulgouris 2025 *J. Neural Eng.* **22** 066043



Journal of Neural Engineering

PAPER

OPEN ACCESS

RECEIVED
9 May 2025

REVISED
14 November 2025

ACCEPTED FOR PUBLICATION
15 December 2025

PUBLISHED
30 December 2025

Dynamic graph representation of EEG signals for speech imagery recognition

Cengiz Selcuk and Nikolaos V Boulgouris*

Department of Electronic and Electrical Engineering, Brunel University of London, Uxbridge, London UB8 3PH, United Kingdom
* Author to whom any correspondence should be addressed.

E-mail: nikolaos.boulgouris@brunel.ac.uk

Keywords: brain–computer interfaces, electroencephalography, graph signal processing, speech imagery

Abstract

Objective. Speech imagery recognition from electroencephalography (EEG) signals is an emerging challenge in brain-computer interfaces, and has important applications, such as in the interaction with locked-in patients. In this work, we use graph signal processing for developing a more effective representation of EEG signals in speech imagery recognition. *Approach*. We propose a dynamic graph representation that uses multiple graphs constructed based on the time-varying correlations between EEG channels. Our methodology is particularly suitable for signals that exhibit fluctuating correlations, which cannot be adequately modeled through a static (single graph) model. The resultant representation provides graph frequency features that compactly capture the spatial patterns of the underlying multidimensional EEG signal as well as the evolution of spatial relationships over time. These dynamic graph features are fed into an attention-based long short-term memory network for speech imagery recognition. A novel EEG data augmentation method is also proposed for improving training robustness. *Main results*. Experimental evaluation using a range of experiments shows that the proposed dynamic graph features are more effective than conventional time-frequency features for speech imagery recognition. The overall system outperforms current state-of-the-art approaches, yielding accuracy gains of up to 10%. *Significance*. The dynamic graph representation captures time-varying spatial relationships in EEG signals, overcoming limitations of static graph models and conventional feature extraction. Combined with data augmentation and attention-based classification, it demonstrates substantial improvements over existing methods in speech imagery recognition.

1. Introduction

The analysis of brain signals and activity is a compelling research area which can provide insight into cognitive activities and affective states. Among the methods that can help monitor brain activity, electroencephalography (EEG) is the most popular way to capture brain signals. EEG has advantages over other methods as it is non-invasive, does not require expensive medical devices, and can be used for applications where the subject is mobile [1]. Processing and analyzing EEG data have various applications [2, 3], such as diagnosis of neurological disorders, sleep analysis, silent speech applications [4, 5], brain–computer interfaces (BCI), or general human–machine interfaces [6].

Graph signal processing (GSP) [7–10] is an emerging field that provides insight into the spatial features and relationships within a graph. The application of GSP on brain signals aims to better analyze and model brain waves and interconnectivities between different regions of the brain and give a more accurate model for the representation of signals that have been acquired from a network of sensors. Such models are particularly suited for brain signals, which tend to have specific spatial relationships between different regions of the brain. Therefore, instead of analyzing brain signals individually, GSP provides a framework

to analyze the network through graph signals and graph frequencies.

A particular EEG application, which has received relatively less attention, is speech imagery recognition, i.e. the interpretation of the specific words *imagined* by a subject [11–13]. This application can have important applications, such as communication for coma patients or other BCIs. Although some methods focusing on this application have been proposed in the past [14–19], these methods typically rely on signal analysis approaches that extract features from each EEG channel separately. Such analysis neglects the structure of the network of sensors as well as the spatial relationships between them. One major challenge in processing and analyzing EEG signals is dealing with its multi-channel nature. The EEG data reside on an irregular domain and analyzing channels separately neglects the spatial features of the signals. To jointly assess EEG signals, one would need a different representation, based on graphs.

A particular challenge in the case of speech imagery recognition is that speech is a sequence of multiple words which are produced and expressed (covertly or overtly) within a short period of time. This fact suggests that the reflection of specific words or notions on the shape of a brain signal is momentary and hard to detect. GSP can represent jointly the ensemble of brain signals acquired from a subject and can potentially offer features with increased discriminatory capacity, which will be useful for the speech imagery recognition. Despite the above, the dynamic nature of speech raises questions as to whether graph representation, which relies on a fixed graph construction and focuses on spatial frequencies, can deal with the challenges of speech imagery recognition.

Graph modeling for brain signals acquired in EEG has been used in several past works [20–25]. All such works construct a graph by determining the graph weights on the basis of the spatial proximity between the respective vertices, where vertices that are farther away from each other are linked through edges that have smaller weights. However, since the topology of the graph in the modeling of EEG signals matches the location (on the head of a subject) of the EEG electrodes, which remains fixed, the resultant graphs are static and cannot model different behaviors in situations where signals from different vertices exhibit varying correlation over time. To overcome this limitation, in our proposed EEG modeling using graphs, we use weights that are calculated based on the temporally varying correlations between the respective vertex pair signals. As those correlations vary over time, our approach naturally leads to a more flexible modeling, based on multiple graphs that are jointly used for the representation of brain signals.

Using a *static* graph is a common method to model and represent EEG signals. A popular way to assign the weights for the graph adjacency matrix is using the thresholded Gaussian kernel weighting function [7]. For example, [20] uses this method to create a static graph for all the subjects and ignores any possible variation. Similarly, [24] creates a static graph based on Cartesian 3D coordinates of the EEG sensors and assigns the weights based on the distances between sensor pairs. As a further example, authors in [23] use a graph-based spatio-temporal attention neural network for emotion recognition using EEG data. The connections for the graph are based on the inverse of the Euclidean distances for sensor pairs. This approach leads to the same static graph for each subject and cannot model dynamic processes where different graphs would be suitable at different times.

Speech imagery recognition [26], which is the focus of the present work, is a research area receiving increasing attention. A widely cited study [14] used hand-crafted statistical features and classified imagined words and phonemes using a deep-belief network and support vector machines. Classification was based on EEG data from subjects who imagined talking those words without making any sound or movements. The work in [14] presented five different binary classification experiments, and reported 18.08% accuracy for the *vowel-only vs consonant (C/V)* classification experiment, and 79.16% accuracy for the presence of a vowel. This approach has limited applicability, as in a real-life scenario it is not feasible to manually choose features for each subject/application.

The specific use of EEG-based GSP for speech imagery has not been widely explored. The work in [27] uses a graph approach based on a static graph. The database in [15] was used for training and testing. The works in [24] and [28] used GSP for speech imagery classification. However, both of them used static graphs with fixed parameters, calculated based on spatial proximities between sensor pairs. Other works [16–19] used the popular '*Kara One*' database [14], but did not use graph representations.

In this work, we address the issues arising from the use of static graphs by proposing a dynamic graph signal representation methodology for speech imagery recognition from EEG signals. Our approach utilizes multiple graphs to represent a set of EEG signals by combining different graph models which are jointly used for the formation of a novel EEG signal representation. This approach is particularly efficient for representing signals that exhibit varying behavior which cannot be accurately modeled through a static graph model. Further, we classify the EEG signals represented using the proposed graph representation using a neural network architecture that is customized to the underlying representation. To boost classification performance, we also introduce a novel EEG data augmentation method to expand the training

dataset without modifying the underlying connectivity or channel-wise relationships of EEG sensors.

Specifically, the contributions of this work are:

- The construction of multiple graphs, with distinct properties, that are jointly used for the modeling of multidimensional EEG signals. The proposed dynamic graph representation takes into account the time-varying interdependencies between channels and achieves EEG signal representations that are more accurately reflecting the underlying brain functions. The number of graphs and eigenvectors are fine-tuned to the underlying signals.
- An application of the above dynamic graph representation on speech imagery EEG data followed by a frequency transformation of the resultant graph signals, which yields a powerful EEG representation with high discriminatory capacity.
- An long short-term memory (LSTM)-based machine learning architecture for the classification of imagined words, combined with a novel graph-based data augmentation method to create augmented EEG data and increase the number of samples for the training without changing the connectivity structure and channel-wise relationships between different channels.
- A thorough comparison of the resulting system with current state-of-the-art systems, which shows the superiority of the proposed system in speech imagery recognition.

The structure of the paper is as follows. In section 2, we briefly present the fundamentals of GSP and the representation of EEG signals using graphs. The dynamic graph representation of EEG signals is presented in section 3. In section 4.1, we present the spatiotemporal representation of graph EEG signals. In section 4.2, we present a neural network architecture that is suitable for our method. In section 5, we present a novel method for EEG data augmentation. In section 6, we present an experimental assessment of the dynamic graph representation of EEG signals and its effectiveness in speech imagery recognition. Discussion is presented in section 7 and conclusions are drawn in section 8.

2. GSP theory and graph representation of EEG signals

2.1. GSP theory and notation

A graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ can be described by its edges $\mathcal{E}$ and vertices $\mathcal{V} = \{v_0, v_1, \ldots, v_{N-1}\}$, where N is the cardinality of the vertex set. A *graph signal* $\boldsymbol{s} = \begin{bmatrix} s_0 & s_1 & \ldots & s_{N-1} \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^N$ is an N-dimensional signal whose values at each time instant are indexed using a graph $\mathcal{G}$ as shown in figure 1.

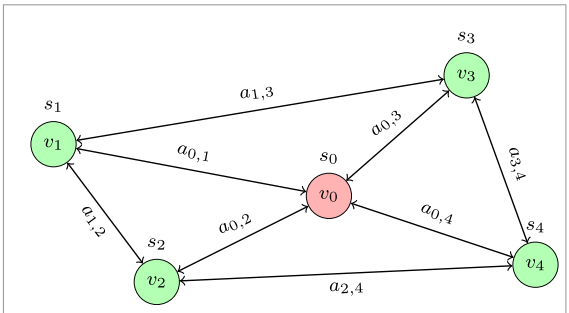


Figure 1. Signal $\boldsymbol{s} = [s_0, s_1, s_2, s_3, s_4]^{\mathrm{T}}$ is shown on graph $\mathcal{G}$ with vertex set $\mathcal{V} = \{v_0, v_1, v_2, v_3, v_4\}$. Weight $a_{i,j}$ represents the weight for the edge from vertex j to vertex i.

The adjacency matrix $\boldsymbol{A}$ is an $N \times N$ matrix that encodes the connections between the vertices of the graph. The element a_{ij}, $i,j = 0, 1, \ldots, N-1$, of the adjacency matrix indicates whether there is a connection from vertex j to i; a_{ij} will be 1 if there is a connection and 0 when there is no incoming edge. If the graph is weighted, then $a_{ij} = w_{ij}$, where w_{ij} is the weight for that edge. If the graph is undirected, then $\boldsymbol{A}$ will be symmetric, with $a_{ij} = a_{ji}$ and $w_{ij} = w_{ji}$.

Assuming $\boldsymbol{A}$ has a complete set of eigenvectors, spectral decomposition of $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$. The matrix $\boldsymbol{V}$ holds the eigenvectors of $\boldsymbol{A}$ on its columns as $\boldsymbol{V} = [\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N-1}]$ and the matrix $\boldsymbol{\Lambda}$ holds the eigenvalues of $\boldsymbol{A}$ on its diagonal as $\boldsymbol{\Lambda} = diag(\lambda_0, \lambda_1, \ldots, \lambda_{N-1})$. Degree matrix $\boldsymbol{D}$ is a diagonal matrix with each element d_{ii} on its diagonal showing the total number of incoming edges (or weights) to vertex i from other vertices, i.e. $d_{i,i} = \sum_j a_{i,j}$. As a_{ij} is an edge from j to i, each element on the diagonal of $\boldsymbol{D}$ shows the number of incoming connections and it has all 0 s on the off-diagonal elements.

As an operator, degree matrix $\boldsymbol{D}$ amplifies the signal at a vertex by the total weight around the neighborhood, and adjacency matrix $\boldsymbol{A}$ provides the weighted sum of the neighbor channels. Laplacian matrix $\boldsymbol{L}$, which is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, gives a sense of variance for a vertex around its neighborhood. This factors in the number of vertices in the vertex neighborhood as it is not normalized. Therefore, when the Laplacian operator is applied to a graph signal, a vertex with higher degree will have a greater value when the signal at that vertex is different than the neighbor vertices, which gives a sense of variance.

We can get the eigendecomposition of the Laplacian matrix as $\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian operation and $\boldsymbol{U}$ is the matrix containing the eigenvectors of $\boldsymbol{L}$ in its columns. The Graph Fourier matrix is given by $\boldsymbol{F} = \boldsymbol{U}^{\mathrm{H}}$ and $\hat{\boldsymbol{s}} = \boldsymbol{F}\boldsymbol{s}$ is the graph Fourier transform (GFT) of a graph signal $\boldsymbol{s}$.

2.2. Graph representation of EEG signals

Let $\boldsymbol{S}$ be a matrix representing EEG data with N channels, where each row of $\boldsymbol{S}$ includes the signal captured by an EEG electrode.

An EEG signal can be represented as a sequence of time frames (vectors) that include the signal samples from all channels at time t. A time frame representing the signal at time t is defined as

$$\boldsymbol{s}^{(t)} = \begin{bmatrix} s_0^{(t)} & s_1^{(t)} & \dots & s_{N-1}^{(t)} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^N \quad (1)$$

where $t = 0, \dots \zeta - 1$, is the temporal index of the respective time frame and ζ is the total number of time frames. The sequence of signal vectors for all time indices can be represented as

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}^{(0)} & \dots & \boldsymbol{s}^{(t)} & \dots & \boldsymbol{s}^{(\zeta-1)} \end{bmatrix}. \quad (2)$$

For each time frame t, the respective GFT frequency vector is obtained as $\hat{\boldsymbol{s}}^{(t)} = \boldsymbol{F}\boldsymbol{s}^{(t)}$. Therefore, the final GFT representation of the entire EEG data is a temporal sequence of consecutive graph frequency vectors, which can be compactly expressed as

$$\hat{\boldsymbol{S}} = \boldsymbol{F}\boldsymbol{S} = \begin{bmatrix} \hat{\boldsymbol{s}}^{(0)} & \dots & \hat{\boldsymbol{s}}^{(t)} & \dots & \hat{\boldsymbol{s}}^{(\zeta-1)} \end{bmatrix}. \quad (3)$$

The GFT does not capture temporal variations at a time instant. In the ensuing section, we describe our proposed method for dynamic graph representation and modeling of the brain signals. Such modeling and representation is suitable for speech imagery recognition.

3. Dynamic graph representation of EEG signals using multiple adjacencies

The definition of a graph $\mathcal{G}$ requires that the connections and weights between graph vertices be specified. There are several methods to construct a graph and define its connections, such as using a thresholded Gaussian kernel weighting function [7]. This approach can model a system where the neighboring vertices exhibit high correlations, such as temperature variations in adjacent geographical locations. However, such modeling would not be suitable in the case of EEG signals where the connections and correlations between different vertices do not depend on proximity between vertices. For this reason, modeling the network structure of the brain requires a different approach.

Apart from determining connections and weights of the graph $\mathcal{G}$, another modeling challenge is the possibility that the underlying signals may exhibit variations that cannot be adequately represented by a single graph. A static graph can model a system where the connections between the vertices remain unchanged over time. However, when a graph is used to model brain signals, one has to consider that brain activity depends on the brain task undertaken and, therefore, different graphs may be needed for different underlying brain activities. For this reason, we do not rely on one static graph to model brain signals, instead we use a dynamic graph model that is based on multiple graphs for the representation of the brain signals.

The modeling of an EEG signal using K graphs requires the identification of K distinct correlation patterns within the graph signal. To this end, we use a sliding window to slice the multi-channel EEG signal into M temporal segments $\tau_m, m = 0, \dots, M-1$. For the EEG matrix $\boldsymbol{S} \in \mathbb{R}^{N \times L}$, where N is the number of EEG channels and L is the number of time samples, the mth temporal segment is defined as $\tau_m = [t_1^{(m)}, t_2^{(m)}]$. The mth signal segment from channel i is $\boldsymbol{z}_i^{(m)} \in \mathbb{R}^n$, with $\boldsymbol{z}_i^{(m)} = (\boldsymbol{S}_{i,\tau_m})^{\mathsf{T}}$, where $n = t_2^{(m)} - t_1^{(m)} + 1$.

Subsequently, for each temporal segment, we calculate correlations between channel pairs. In our graph modeling, the correlation values between channels are directly used as weights in adjacency matrices. If $\boldsymbol{z}_i, \boldsymbol{z}_j$ represent signals from two channels, the (i,j) element of the normalized correlation matrix $\boldsymbol{P}^{(m)}$ for the mth segment is calculated as

$$\rho_{ij}^{(m)} \triangleq \operatorname{ncorr}\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right) = \frac{1}{2}\left(\frac{\boldsymbol{z}_i \boldsymbol{z}_j}{|\boldsymbol{z}_i||\boldsymbol{z}_j|} + 1\right), 0 \leqslant \rho_{ij}^{(m)} \leqslant 1 \quad (4)$$

and represents the normalized (non-negative) correlation between channel i and channel j for temporal segment τ_m. As each N-channel signal is divided into M segments, M normalized correlation[1] matrices with size $N \times N$ are obtained from each multidimensional signal, representing cross correlations between different channels for each of the temporal segments.

The varying correlations between the channels of a multidimensional signal (e.g. an EEG signal) indicate that multiple adjacencies may need to be taken into account for the graph modeling of EEG signals. To capture the varying correlation between channels over time, we cluster the correlation matrices from different temporal segments using the k-means algorithm [29], and calculate K correlation centers $\boldsymbol{C}^{(k)}, k = 0, \dots, K-1$. For each experiment, we use all the available training data (all subjects and trials) to create K global graphs. In particular, in intra-subject experiments, we use all the training data from the subject, and in cross-subject experiments, we use all the training data from all subjects to create K global graphs based on correlation patterns.

To create multiple graphs, we use the correlation centers to define K adjacency matrices $\boldsymbol{A}^{(k)}, k = 0, \dots, K-1$, as

$$\boldsymbol{A}^{(k)} = \boldsymbol{C}^{(k)} - \boldsymbol{I} \quad \text{for } k = 0, \dots, K-1 \quad (5)$$

[1] Henceforth, normalized correlation will be referred to as 'correlation' for simplicity.

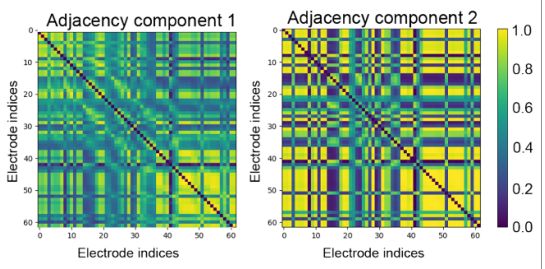


Figure 2. Adjacency matrices in a dynamic graph model with two ($K = 2$) correlation clusters. The differences between the two adjacencies show that the correlations between channels in the multidimensional EEG signal are varying over time. Therefore, the use of multiple adjacencies offers advantages in the graph modeling of the signal as the multiple adjacencies can capture more accurately the varying correlations between EEG channels over time.

where diagonal elements have been set to zero to prevent self-loops in the graphs. To analyze the graph spectrum of different graphs, we use the Laplacian matrices

$$\boldsymbol{L}^{(k)} = \boldsymbol{D}^{(k)} - \boldsymbol{A}^{(k)} \quad \text{for } k = 0, \ldots, K-1 \tag{6}$$

where the diagonal matrix $\boldsymbol{D}^{(k)}$ is the degree matrix for correlation cluster k.

We present a graph model example with two adjacency components ($K = 2$) in figure 2 and an example with four adjacency components ($K = 4$) in figure 3. Through empirical evaluation, we determined that using shorter temporal segments results in decreased similarity among adjacency components, which helps capture instantaneous variations in correlation between channels. A segment length of 16 ms was selected because it provides sufficient variability. We also observed that overlaps did not affect the results but increased computational cost, and therefore were not used.

For each graph, we get the eigendecomposition

$$\boldsymbol{L}^{(k)} = \boldsymbol{U}^{(k)} \boldsymbol{\Lambda}^{(k)} \left(\boldsymbol{U}^{(k)}\right)^{\mathrm{H}} \quad \text{for } k = 0, \ldots, K-1 \tag{7}$$

with eigenvectors

$$\boldsymbol{U}^{(k)} = \left[\mathbf{u}_0^{(k)} \mathbf{u}_1^{(k)} \ldots \mathbf{u}_{N-1}^{(k)}\right] \quad \text{for } k = 0, \ldots, K-1. \tag{8}$$

An example of graph eigenvectors (8) is shown in figure 4 (generated using the PyGSP package [30]). In general, our dynamic graph modeling involves $K(N-1)+1$ eigenvectors. This is because each of the K different graphs $\mathcal{G}^{(k)}$ produces N eigenvectors. However, we discard the constant vectors from all graphs except the first one resulting in $K(N-1)+1$ eigenvectors overall. The larger set of eigenvectors can capture a wider set of graph frequencies (spatial patterns) because the dynamic representation takes into account the varying channel-wise correlations over time.

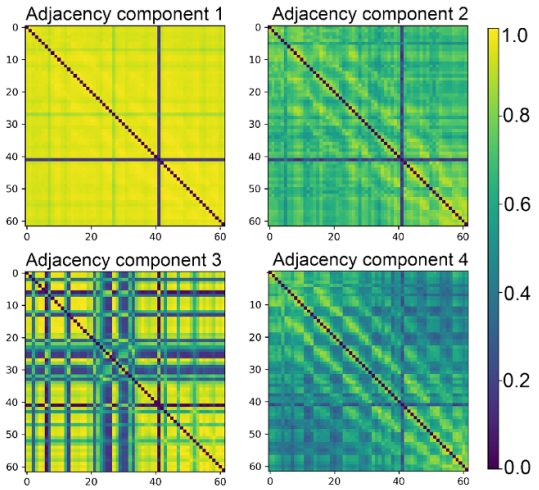


Figure 3. Adjacency matrices in a dynamic graph model with four ($K = 4$) correlation clusters. Horizontal and vertical labels represent electrode indices. When $K = 4$, the correlations between EEG channels are assumed to exhibit four distinct patterns over time.

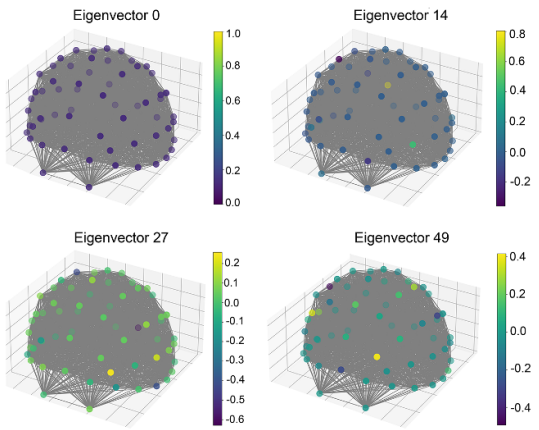


Figure 4. Different eigenvectors of the first adjacency component of the dynamic graph for the first subject of the database [14]. The first eigenvector shows the DC component where all the elements have the same value. Higher graph frequencies (represented by the other eigenvectors) exhibit greater variation in the elements of respective eigenvectors.

To simplify the above representation, from each correlation cluster we select Δ eigenvectors, indexed $e_0^{(k)}, e_1^{(k)}, \ldots, \ldots, e_{\Delta-1}^{(k)}$. Then we use the selected eigenvectors from each cluster to form the dynamic graph eigenvector matrix $\boldsymbol{\Omega}$,

$$\boldsymbol{\Omega} = \Big[\underbrace{\boldsymbol{u}_{e_0}^{(0)} \boldsymbol{u}_{e_1}^{(0)} \ldots \boldsymbol{u}_{e_{\Delta-1}}^{(0)}}_{\text{from first cluster}} \ldots \underbrace{\boldsymbol{u}_{e_1}^{(K-1)} \ldots \boldsymbol{u}_{e_{\Delta-1}}^{(K-1)}}_{\text{from last cluster}}\Big]. \tag{9}$$

The eigenvectors in $\boldsymbol{\Omega}$ represent different graph frequencies from all graphs within a dynamic graph representation. These graphs capture the dynamic relationships between vertices. By projecting each original N-dimensional signal $\boldsymbol{s}$ onto the eigenvectors of the dynamic graph eigenvector matrix $\boldsymbol{\Omega}$, we obtain

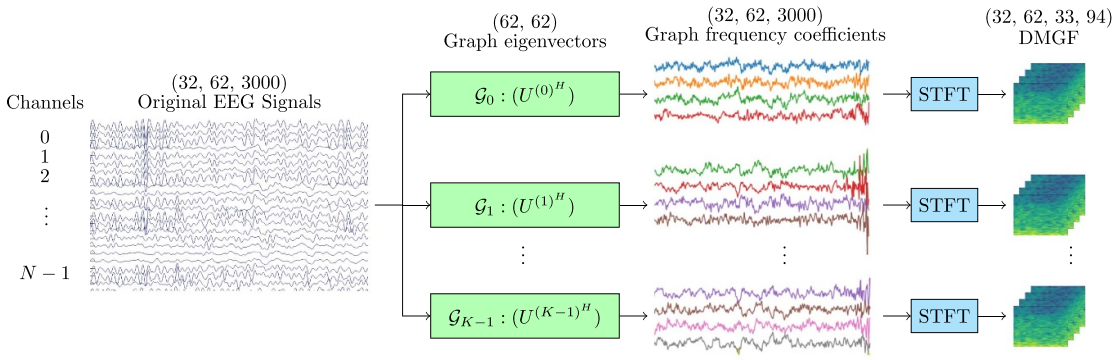


Figure 5. Spatiotemporal representation of EEG signals using a dynamic graph with K constituent graphs. EEG signals are projected onto the eigenvectors of different graphs and spatial frequency components are obtained from each graph. We calculate the short-time Fourier transform (STFT) of each component separately to get the dynamic multiple graph Fourier (DMGF) representation that captures frequencies both in graph space and time. Tensor shapes are shown for the Kara One [14] database with a batch size of 32, 62 channels, and a signal length of 3000.

the new representation $\hat{\boldsymbol{s}} = \boldsymbol{\Omega}^{\mathrm{H}}\boldsymbol{s}$. For the entire original EEG signal (N-dimensional time series), the new representation is

$$\hat{\boldsymbol{S}} = \boldsymbol{\Omega}^{\mathrm{H}}\boldsymbol{S} \tag{10}$$

where $\hat{\boldsymbol{S}}$ is a $K\Delta$-dimensional series of graph frequency coefficients. In the following sections, we present how we use these features in neural network-based speech imagery recognition.

4. Speech imagery recognition system

4.1. Spatiotemporal representation of the graph EEG signals

The above graph-based modeling produces time series of coefficients representing spatial frequencies (spatial patterns) over time. To take into account temporal variations of those spatial frequencies, we calculate separately the short-time Fourier transform (STFT) magnitude of the time series of each graph frequency coefficient, as shown in figure 5. The resultant *dynamic multiple graph Fourier* (DMGF) representation expresses the original multidimensional signal with respect to the way in which the signal's spatial patterns are changing over time.

In order to prepare the DMGF for neural network classification, we offset STFT magnitude values by 1 to make the smallest magnitude equal to 1. We then apply log-normalization to keep the STFT magnitude values within a range that can be handled by the neural network. In this way, the minimum log-magnitude is 0 while small magnitudes, such as 10^{-20}, are represented as approximately 0. The resultant representation is graphically shown in figure 6.

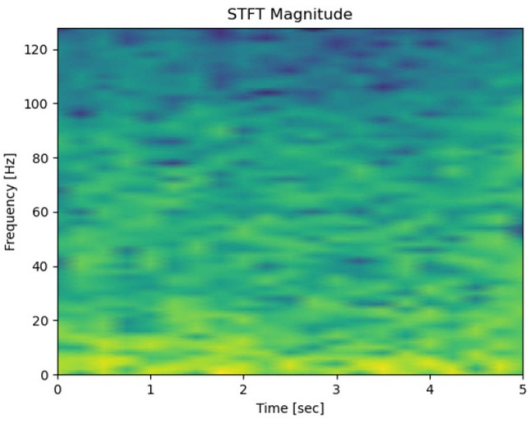


Figure 6. A sample DMGF (dynamic multiple graph Fourier) representation that is input to the neural network. Multiple such representations are input simultaneously (one per graph frequency coefficient sequence).

4.2. Neural network architecture for speech imagery classification

For the classification task we used a three-layer LSTM network with attention mechanism as shown in figure 7. The input signal is projected onto the eigenvectors of the dynamic graph representation, yielding a set of graph frequencies for each of the different graphs. The subsequent STFT of each graph frequency time series captures the variations of dynamic graph frequencies through time. The resultant DMGF coefficients are used as inputs to the LSTM network. To simplify the representation and reduce the number of features input to the LSTM network, we calculate the average of every four temporal frequency components of each graph frequency. The number of features input to the LSTM network is the number of graph frequencies multiplied by the number of temporal frequencies.

As seen in figure 7, we use three bidirectional LSTM (BiLSTM) layers. We chose the LSTM for such modeling as LSTMs are robust at learning long-term and short-term dependencies in long temporal sequences, like the ones used in the present work. The small number of LSTM layers limits the number of parameters in the neural network, reduces

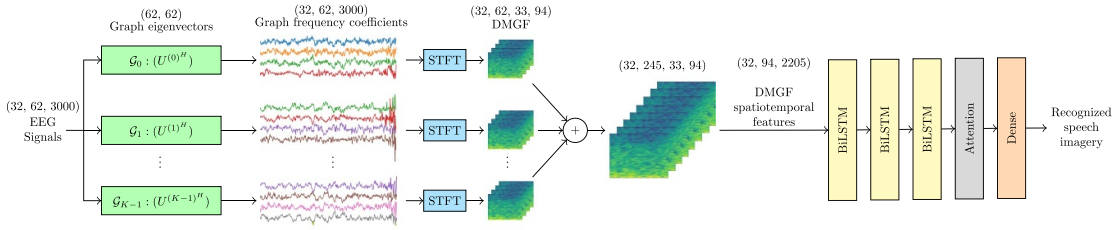


Figure 7. Main architecture for classification of speech imagery using a dynamic graph with K constituent graphs. Multi-channel EEG signals are projected onto the eigenvectors of each graph, yielding time series of graph (spatial) frequency coefficients. These are then subjected to short-time Fourier transform (STFT), producing the proposed DMGF representation. The resulting DMGF coefficients are reshaped into a 2D tensor of shape (*timesteps, features*) and fed into an attention-based BiLSTM network for speech imagery recognition. Tensor shapes are shown for the Kara One [14] database with a batch size of 32, 62 channels, and a signal length of 3000.

the computational cost, and prevents overfitting. We used 256 hidden units per direction in each layer and did not use any dropout to regulate overfitting, as the model itself converges well and dropout induces oscillations in the loss function on small validation sets. Each spatiotemporal graph frequency in the DMGF representation is used as a different channel/feature which is input to the first LSTM layer, while the cascaded LSTM layer learns how those features evolve through time. The attention mechanism at the end of the last LSTM layer is used to ensure that the neural network focuses on the most relevant features at the end of the cascaded LSTM layers.

5. Data augmentation

EEG datasets generally have limited samples compared to other types of datasets. Even with EEG epoching, whereby a trial is divided into multiple trials, the number of available trials is still small. For intra-subject experiments, which use data from only one subject, the risk of overfitting increases. For cross-subject experiments, data from different subjects are used, i.e. a higher number of trials per class is available. However, cross-subject experiments are challenging as signals from different subjects may vary too much.

There are different methods to augment EEG data, i.e. adding noise, combining different EEG sequences, delaying an EEG sequence and combining it with different EEG sequences. However, those techniques have limitations [31]. In the present work, instead of randomly combining different channels with randomly selected patches of sequences, we use graph operators to combine the channel signals with signals from neighboring vertices. It should be noted that to create an artificial instance of a channel signal, all other channels also need to be taken into account to preserve channel-wise correlations in the artificial instance. For that reason, we first use the original multi-channel EEG signal to create an artificial multi-channel EEG signal and then extract the specific channel from the artificial signal.

If $\boldsymbol{S}$ is a multi-channel signal in the form of a matrix with rows representing channels, we create an artificial sample as

$$\boldsymbol{S}^{'} = \alpha \boldsymbol{S} + (1-\alpha)\,\boldsymbol{Q}\boldsymbol{S} + \theta \boldsymbol{\mathcal{L}}\boldsymbol{N} \tag{11}$$

where $\boldsymbol{Q}$ is the *random walk matrix* defined as $\boldsymbol{Q} = \boldsymbol{D}^{-1}\boldsymbol{A}$, $\boldsymbol{\mathcal{L}}$ is the *symmetric normalized Laplacian* defined as $\boldsymbol{\mathcal{L}} = \boldsymbol{D}^{-1}\boldsymbol{L}\boldsymbol{D}^{-1}$, and $\boldsymbol{N}$ is a white noise matrix. The product $\boldsymbol{Q}\boldsymbol{S}$ provides the weighted average of the neighboring vertices using the weights for each pair of vertices [32]. Constants α and θ are parameters to fine-tune the augmentation. Specifically, α is the weight of the channel signal itself, while $1-\alpha$ is the weight of the weighted average of the neighboring channels, and θ is the weight of the noise. As seen, the white noise is multiplied with the *symmetric normalized Laplacian* operator $\boldsymbol{\mathcal{L}}$ to amplify the noise by the average variance between the vertex and its neighbors. This approach adds more noise when there is higher normalized variance between neighboring vertices. This ensures the preservation of the relationships between neighboring channels. Figure 8 presents the creation of an artificial sample based on the pairwise relationships of vertices of a graph.

6. Experimental assessment

6.1. Databases

For the experimental evaluation of our method, we used the databases in [14] and [15], which include EEG data for speech imagery. The first database that we used is the Kara One database [14]. Although the database also includes video and audio data, we only used the EEG signals collected while subjects were performing speech imagery. There are 14 subjects performing seven phonemic/syllabic prompts, which are /iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/, and four different words *'pat'*, *'pot'*, *'knew'* and *'gnaw'*,

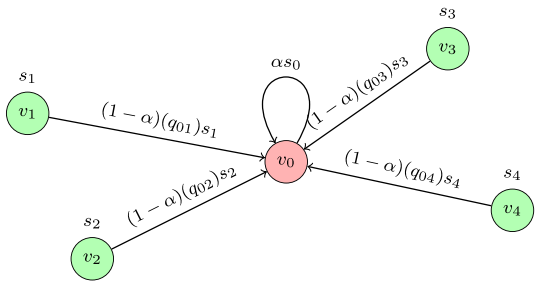


Figure 8. Creation of an artificial EEG signal by using graph operators as formulated in (11). Signal $\boldsymbol{s} = [s_0, s_1, s_2, s_3, s_4]^\mathsf{T}$ is shown on a graph with vertex set $\mathcal{V} = \{v_0, v_1, v_2, v_3, v_4\}$. For each vertex i, neighbor j contributes with a value of $(1-\alpha)(q_{i,j})s_j$ alongside αs_i, which is a portion of the original signal value s_i, for vertex i.

with around 12 samples per class. To enable a direct comparison, we performed the same five binary classification experiments as in [14, 16–18].

The second database was introduced in [15] and includes four different experiments, namely, *short word* classification, *long word* classification, *vowel* classification and *short word vs long word* classification. Out of the four experiments, the *long word* classification and *short vs long word* classification experiments are binary, while the other two experiments involve three classes. In *short word vs long word* experiment, the subject sees the *short word 'in'* or the *long word 'cooperate'* as visual cues on a computer monitor and the subject imagines speaking those words without moving their lips or any other muscles. Similar protocols were applied to other classification tasks. As only a subset of the subjects contributed trials for the *short word vs long word* experiment, we represent them as subjects A, B, C, D, E, F instead of using their IDs for clarity.

6.2. Experiments

We applied our proposed dynamic graph representation to the EEG signals in the above databases. To model brain signal activity during speech imagery tasks, we created multiple graphs by clustering correlation matrices, calculating multiple adjacency matrices, and creating a separate graph for each adjacency matrix. To calculate the correlations, as described in (4), we used segments of 16 ms from each trial to compute channel-wise correlations. The correlation matrices were then clustered and the cluster centers yielded four ($K = 4$) distinct adjacency matrices. We empirically found that using 4 clusters gives the best performance, while higher numbers of clusters do not yield consistent gains. After computing the eigendecomposition of each adjacency matrix to obtain the graph Fourier matrices of each graph, as in (7), we used all eigenvectors from all graphs to construct the dynamic graph eigenvector matrix, as described in (9). Subsequently, for each EEG trial, we calculated the respective graph frequencies and we applied STFT to the respective temporal sequences of graph frequencies, producing a DMGF representation. An example of our DMGF representation is shown in figure 6.

For classification, we concatenated the spatial-frequency and temporal-frequency dimensions of the DMGF data structure and used the resultant structure as input to the LSTM layers. The architecture of our main scheme is shown in figure 7. We applied the STFT with a 64-point Hann window and 50% overlap. The 4D output of the STFT is in the form of (trials × graph frequencies × time frequencies × time) and after concatenation it acquires a 3D shape (trials × graph/time frequency combinations × time). We used this representation as an input to a three-layer LSTM network [33, 34] with attention to classify brain signal activity during speech imagery. The tensor shapes for the experiments conducted on Kara One database [14] are presented in figures 5 and 7, where 32 is the batch size, 62 is the number of channels, and 3000 is the signal length per channel. Tensor shapes for the dataset [15] are the same except for the number of channels, which is 64 instead of 62. As shown in figure 7, the shape of the tensor fed into the neural network is $(32, 94, 2205)$ with 32 being the batch size, 94 the number of time steps, and 2205 the number of frequency features.

For the experiments conducted on Kara One database [14], we used a leave-one-subject-out cross-validation scheme as it was conducted in [14, 16–18] to have the identical setups for a fair comparison. For intra-subject (per-subject) experiments conducted on database [15], we used a nested 10-fold cross-validation [35] scheme to validate our results.

6.3. Exploration of different numbers of adjacencies

For fine-tuning our dynamic graph representation, we explored the impact of K, the number of correlation clusters, on classification accuracy. To this end, we used the database in [15] and ran different experiments using one, two, and four clusters. We repeated the experiments on Subject A and Subject D, as Subject A has a lower classification accuracy than other subjects and Subject D has higher classification accuracy. We present the results in table 1. As seen, using a higher number of correlation clusters generally improves performance, as multiple graphs capture more accurately the varying correlations between EEG channels over time. As stated in the preceding subsection, we empirically found that a practical choice for the number of constituent graphs is $K = 4$.

6.4. Exploration of different eigenvector selection strategies

For the purpose of assessing the discrimination capacity of graph frequencies in speech imagery recognition, we explored the use of subsets of graph frequencies. We present the results in table 2, where

Table 1. Classification accuracies using different numbers of correlation clusters using the database [15]. As seen, using a higher number of clusters helps with classification accuracy. Results are averaged for all subjects. Our results are shown in bold.

Classification accuracies using different number of correlation clusters.				
Method	Short words	Vowels	Long words	Short vs long words
[15]	50.07 ± 7.60%	48.96 ± 6.21%	66.18 ± 5.78%	80.05 ± 5.80%
[26]	N/A	N/A	62.99 ± 4.78%	N/A
[27]	44.2%	46%	71%	70%
DMGF with 1 cluster ($K = 1$)	**55.06** ± 3.32%	**52.54** ± 3.44%	**71.92** ± 5.27%	**85.77** ± 5.87%
DMGF with 2 clusters ($K = 2$)	**56.28** ± 2.83%	**53.50** ± 3.71%	**73.08** ± 5.43%	**86.83** ± 4.51%
DMGF with 4 clusters ($K = 4$)	**57.33** ± 2.52%	**55.08** ± 2.59%	**74.75** ± 5.36%	**88.58** ± 3.91%
DMGF with 8 clusters ($K = 8$)	**56.78** ± 2.65%	**54.71** ± 2.55%	**74.33** ± 5.03%	**87.79** ± 5.14%

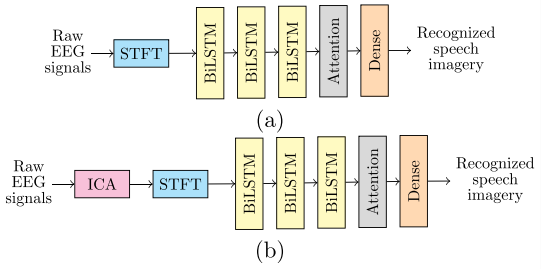


Figure 9. Baseline (conventional) methods without graph representation, for comparison purposes. (a) Baseline method using only short-time Fourier transform (STFT). (b) Baseline method with independent component analysis (ICA) and STFT representation.

the representation of EEG signals using the lowest n eigenvectors is denoted as GSP_n, i.e. GSP_{20} is a representation using the first/lowest 20 eigenvectors (out of a total of 64). As seen in table 2, using a subset of eigenvectors yields comparable results to using all eigenvectors. This means that the coefficients representing low spatial frequencies carry most discriminatory capacity and can be used reliably for classification of speech imagery. However, the best performance is achieved when high graph frequencies are also taken into account, which suggests that high graph frequencies are also useful in speech imagery.

To confirm that the dynamic graph representation of EEG signals leads to improved speech imagery classification performance, we deployed an experimental setup without any graph representations, which served as a baseline system (shown in figure 9(a)). The baseline system directly uses raw EEG signals as input to the STFT, i.e. no graph representation and no graph frequencies are used. In addition, we tested a further baseline system where the raw EEG signals are first subjected to independent component analysis (ICA) using the algorithm defined in [36]. Subsequently, STFT is applied to the time series of independent components, and the resulting frequency features are fed into the neural network (figure 9(b)).

Results using the two baseline methods are presented in table 2. As seen, even using only five eigenvectors with our 4-cluster dynamic graph representation, it is possible to outperform the two baseline methods that do not use graph signal representations. This confirms the suitability and effectiveness of our graph representation for speech imagery classification.

6.5. Exploration of different neural network variations

We repeated the experiments using three alternative neural network architectures for classification. These are explained below.

6.5.1. Multi-channel 3-BiLSTM with attention (figure 10(a))

We tested our methodology using two separate graphs and separate graph eigenvector sets. We calculated the STFT of the resulting graph frequencies and provided them as inputs to different LSTM layers. This approach is similar to the main method, shown in figure 7, but processes the features of different graphs separately. The attention layers at the end of the LSTM layers facilitate the selection of the most relevant features for the classification task. The concatenation of the resulting outputs was fed into a dense layer with a softmax function that calculates the probability of each class.

6.5.2. Multi-channel 5-Conv2D with attention (figure 10(b))

This architecture uses five 2D convolution layers instead of LSTM layers. The low number of LSTM layers and convolution layers helps keep computational cost low and avoid overfitting.

6.5.3. 1-channel 5-Conv2D with attention (figure 10(c))

This architecture uses 2D convolutional layers on a single channel, taking as input the same graph representation used by the main architecture.

Table 2. Classification accuracies using DMGF constructed with different subsets of eigenvectors for the *short word vs long word* experiment using database [15]. Results for traditional baselines and the method in [15] are also reported for reference. The best results are shown in bold.

Classification accuracies using a subset of the eigenvectors.		
Method	Subject A	Subject D
Benchmark [15]	$70.3 \pm 5.5\%$	$88.0 \pm 6.4\%$
Baseline (Raw signal + STFT)	$61.50 \pm 3.20\%$	$70.63 \pm 4.00\%$
Baseline (Raw signal + ICA + STFT)	$63.00 \pm 4.00\%$	$72.50 \pm 5.00\%$
Dynamic GSP_5 + STFT	$68.00 \pm 3.32\%$	$76.88 \pm 4.88\%$
Dynamic GSP_{10} + STFT	$71.50 \pm 5.02\%$	$80.00 \pm 3.75\%$
Dynamic GSP_{20} + STFT	$75.00 \pm 3.87\%$	$86.25 \pm 3.75\%$
Dynamic GSP_{40} + STFT	$78.50 \pm 3.20\%$	$91.88 \pm 4.00\%$
Dynamic GSP_{64} + STFT	$\mathbf{81.50} \pm 3.91\%$	$\mathbf{93.75} \pm 4.84\%$

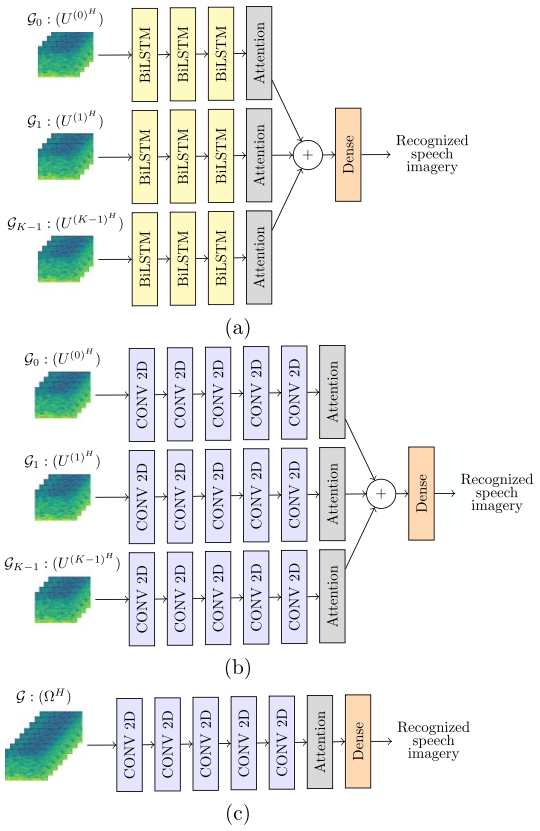


Figure 10. Neural network architecture variants for the classification of our DMGF representation. (a) Multi-channel 3-BiLSTM with attention, (b) Multi-channel 5-Conv2D with attention and (c) 1-channel 5-Conv2D with attention.

Results obtained using the above architectures are presented in table 3. As seen in table 3, the main scheme (figure 7) outperforms the alternative architectures, providing evidence that the dynamic graph EEG representation combined with a single-channel LSTM-based neural network is the architecture that yields the best recognition results.

6.6. Hyperparameter tuning

Our neural network architecture was based on the model proposed in [37], with adjustments made to prevent overfitting and improve training efficiency. We used three layers of BiLSTM, each with 256 hidden units per direction, resulting in an effective output dimensionality of 512 units per layer. A single

Table 3. Classification performance using our DMGF representation combined with different neural network architectures for *short word vs long word* classification using database [15]. The best results are shown in bold.

Classification accuracies using alternative neural networks.		
Neural network variation	Subject A	Subject D
Benchmark [15]	70.3 ± 5.5%	88.0 ± 6.4%
2-channel 3-BiLSTM + attention	79.00 ± 4.36%	91.25 ± 4.15%
2-channel 5-Conv2D + attention	80.00 ± 3.87%	92.50 ± 5.45%
1-channel 5-Conv2D + attention	78.00 ± 5.10%	89.38 ± 5.63%
1-channel 3-BiLSTM + attention	**81.50** ± 3.91%	**93.75** ± 4.84%

dense layer was used at the output, with the number of neurons equal to the number of target classes. This was followed by a softmax activation function to produce class probabilities.

All hyperparameters were selected through empirical tuning on the validation set within a nested 10-fold cross-validation scheme [35]. The model was trained using the Adam optimizer for 3000 epochs. We employed a learning rate scheduler that reduced the learning rate when the validation loss plateaued, with a patience parameter set to 400 epochs. Learning rates of 10^{-3} and 10^{-4} were initially explored, but resulted in stagnation of the training loss. A learning rate of 10^{-5} yielded more stable convergence and was therefore used in all experiments.

We did not apply dropout regularization, as the model converged well without it, and dropout was observed to introduce oscillations in the loss function on small validation sets.

6.7. Data augmentation

Due to the scarcity of EEG data relevant to the task of speech imagery analysis and recognition, we introduced artificial trials using the novel methodology presented in section 5. To generate artificial trials, we set $\alpha = 0.9$ to combine 90% of the original signal with 10% of the weighted average of signals from the proximity of each vertex. We also used $\theta = 0.05$ to introduce white noise based on the signals within the proximity of each vertex. We used the augmented data exclusively for training the neural network. To assess performance, we utilized the original test samples. To evaluate the impact of data augmentation, we applied the same experimental setup for speech imagery classification and compared the classification accuracies with and without using augmented data. Results are presented in the ensuing sections.

6.8. Results on Kara One database [14]

We performed five distinct binary classification tasks on the Kara One database [14], following the same experimental setups used in [14, 16–18]. These classification tasks are: *vowel-only vs consonant (C/V)*, *presence of nasal (±Nasal)*, *presence of bilabial (±Bilabial)*, *presence of high-front vowel (±/iy/)*, and *presence of high-back vowel (±/uw/)*. We focused on these binary tasks to make sure that our experimental setups are identical to those in [14, 16–18] and ensure that our comparisons to the respective methods are conclusive. We tested our DMGF system in subject-independent classification, where one subject is kept aside for testing, while the remaining subjects are used for training. The results are averaged over all subjects and presented in table 4 in comparison to results from relevant benchmarks [14, 16–18]. As seen, in most experiments our results exceed the benchmark accuracies by margins that range from 4% to over 10%. This shows that our novel methodology can achieve substantial improvements in speech imagery recognition.

6.9. Results on database [15]

We used the database in [15] to test our DMGF method for speech imagery classification involving the *short word 'in'* and *long word 'cooperate'*. Results are presented in table 5. As seen, our method outperforms the benchmark method by up to 10%, and exhibits lower standard variation across subjects. In table 5 we also report results with our training data augmentation method. As seen, augmentation leads to a small additional increase in recognition performance accompanied by a decrease in variability. These experiments suggest that data augmentation is a viable strategy for addressing the problem of lack of training data. The overall consistency and robustness of DMGF are further evidenced in table 6 where relevant performance metrics are reported.

In addition to the above assessment, we evaluated our system in cross-subject (subject-independent) classification where data from all subjects are used for system training. In general, EEG classification is more difficult when signals from different subjects are used for training and testing [37], as opposed to using signals from one subject. Our method is inherently

Table 4. Phonological imagery classification accuracies using our proposed DMGF method on KARA One database [14]. The experiments conducted are *vowel-only vs consonant (C/V)*, *presence of nasal (±Nasal)*, *presence of bilabial (±Bilabial)*, *presence of high-front vowel (±/iy/)*, and *presence of high-back vowel (±/uw/)*. The best results are shown in bold.

Classification accuracies on KARA One database [14].					
Method	C/V	±Nasal	±Bilabial	±/iy/	±/uw/
[14]	18.08%	63.50%	56.64%	59.60%	79.16%
[16]	25.00%	47.00%	53.00%	53.00%	74.00%
[17]	85.23%	73.45%	**75.55%**	73.30%	81.99%
[18]	86.52%	72.10%	69.08%	75.27%	83.98%
DMGF	**90.06** ± 1.98%	**76.73** ± 1.48%	71.71 ± 1.60%	**89.56** ± 1.47%	**91.08** ± 1.81%

Table 5. *Short word vs long word* classification accuracy using our proposed DMGF method, applied with and without data augmentation. Performance is reported for our proposed 4-cluster method ($K = 4$) with 3-layer LSTM network and attention mechanism. The best results are shown in bold.

Classification accuracies for *short word vs long word* on database [15].			
Subject	Benchmark [15]	DMGF	DMGF + augm
Subject A	70.3 ± 5.5	**81.50** ± 3.91%	**82.50** ± 3.35%
Subject B	71.5 ± 5.0	**89.50** ± 3.50%	**90.00** ± 3.16%
Subject C	81.9 ± 6.5	**91.25** ± 5.00%	**91.88** ± 4.00%
Subject D	88.0 ± 6.4	**93.75** ± 4.84%	**94.38** ± 4.38%
Subject E	79.3 ± 7.9	**81.50** ± 3.20%	**82.00** ± 2.45%
Subject F	89.3 ± 3.5	**94.00** ± 3.00%	**94.00** ± 2.00%
Average	80.1 ± 5.8	**88.58** ± 3.91%	**89.13** ± 3.22%

Table 6. Performance metrics for each subject for *short vs long word* experiment. The proposed system exhibits consistent precision, recall, *F*1 score and accuracy across all subjects, demonstrating its robust performance.

Performance metrics for *short vs long word* experiment.				
Subject	Precision	Recall	*F*1 Score	Accuracy
Subject A	0.82	0.80	0.81	0.82
Subject B	0.88	0.92	0.90	0.90
Subject C	0.89	0.94	0.91	0.91
Subject D	0.96	0.91	0.94	0.94
Subject E	0.83	0.79	0.81	0.82
Subject F	0.96	0.92	0.94	0.94

Table 7. Classification accuracies for the *short word vs long word* experiment using our proposed DMGF method, applied with augmented EEG data. The best results are shown in bold.

Classification accuracies for *short word vs long word* classification.		
Task	Benchmark [15]	DMGF
Intra-subject classification	80.1 ± 5.8	**89.13** ± 3.22%
Cross-subject classification	N/A	**76.52** ± 1.65%

robust in subject-independent classification tasks because the clustering of correlation matrices from different subjects models task-dependent relationships between different graph vertices instead of subject-specific traits. Therefore, to emphasize features relevant to speech imagery, we create dynamic graphs by combining training data from all subjects and imagined words. Results are presented in table 7 and show that our method sustains a 75% accuracy in the cross-subject scenario, which confirms the effectiveness of our approach. We also conducted cross-subject experiments using augmented data.

7. Discussion

7.1. Methodological innovation and architectural design

The conventional approach in EEG signal analysis and classification is based on applying feature extraction on each EEG channel separately, usually by means of a frequency transformation, such as STFT. The extracted features from all EEG channels are then classified through a neural network. Our approach, based on GSP, calculates multiple spatial representations, which take into account all EEG channels simultaneously. The deployment of multiple graphs

for the dynamic representation of EEG signals enables our system to capture the dynamic nature of EEG signals, which exhibit fluctuating correlations between EEG channels and would render a fixed graph (single-graph) representation less effective. The experimental results and comparisons we report in tables 1–3 show that the representation of EEG signals using multiple graphs leads to clear performance gains. While an increasing number of constituent graphs generally improves performance, as seen in table 1, we have found that using a dynamic graph representation with four ($K = 4$) graphs is a good practical choice, as higher numbers of graphs did not yield any consistent performance gains. Each constituent graph contributes to the EEG representation, with best performance achieved if all eigenvectors are used with each of these graphs. As seen in table 2, the best-performing version of our system, which forms EEG representations by using all eigenvectors from four graphs, outperforms reference methods and traditional approaches (e.g. based on ICA or STFT) by a significant margin. Our method is architecturally completed by using a suitable attention-based BiLSTM network for classification, as shown in table 3.

7.2. Summative evaluation of results

A full experimental evaluation of our DMGF methodology is presented in tables 4–7, where comparisons to other methods are also presented. We tested our DMGF method in a range of scenarios involving phoneme and imagined word recognition. As seen in tables 4 and 5, using DMGF achieves up to 10% gains in recognition accuracy, including a smaller gain from our data augmentation approach which helped with increasing the robustness of the training. These results provide evidence of the merits of our methodological approach and the ability of the proposed dynamic graph approach in the representation of EEG signals for speech imagery recognition. The results on cross-subject experiments (reported in table 7) are also of particular interest as they show that it is possible to achieve speech imagery recognition even if multiple subjects are used for the training of the system. We believe that further overall gains are possible both from further developing the dynamic graph representation and also from formulating an optimized augmentation approach.

7.3. Statistical significance of results

To assess whether improvements over benchmark methods were statistically significant and lead to safe conclusions, we performed two-proportion z-tests [38] on classification accuracies for the results presented in tables 1, 4, 5 and 7 where our method is compared against the benchmark methods [14–18, 26, 27]. The z-tests confirmed that the gains over prior work were statistically significant ($p < 0.05$). Overall, these analyses indicate that the performance gains achieved by our dynamic graph system are numerically robust and statistically significant.

7.4. Relation to hypergraphs and multigraphs

It should be noted that our dynamic graph modeling, which relies on multiple graphs, is different from *hypergraphs* [39, 40], where a single edge can connect more than two vertices, and therefore edges can have any cardinality. Although hypergraphs have received increasing attention [41, 42], they have not yet been widely adopted. Further, our framework is unrelated to *'multigraphs'*, i.e. graphs that can have multiple edges between the same pairs of vertices to represent pairwise relationships [40]. By contrast, our approach focuses on addressing the limitations of static graphs. Our proposed DMGF representation achieved this by expressing the original multidimensional EEG signal through a broader set of graph frequency components, which were computed by factoring in variations across subjects, tasks, and time.

8. Conclusion

We proposed a dynamic graph representation for speech imagery recognition from EEG signals. The proposed dynamic graph represents brain signals using different graph frequencies obtained through multiple graphs that can capture the varying correlations among EEG signals. This approach is particularly suitable for representing signals which cannot be adequately modeled through a simple static graph model. We used the resulting dynamic graph representation of EEG signals as input to an attention-based LSTM network, trained using a novel EEG data augmentation process. The resulting system was shown to be superior to current state-of-the-art methods for speech imagery recognition.

Data availability statement

No new data were created or analysed in this study.

Author contributions

Cengiz Selcuk 0009-0007-9309-3590
Conceptualization (equal), Data curation (lead), Formal analysis (equal), Investigation (lead), Methodology (equal), Software (lead), Validation (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (supporting)

Nikolaos V Boulgouris 0000-0002-5382-6856
Conceptualization (equal), Formal analysis (equal), Investigation (supporting), Methodology (equal), Supervision (lead), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (lead)

References

[1] Michel C M, Murray M M, Lantz G, Gonzalez S, Spinelli L and De Peralta R G 2004 EEG source imaging *Clin. Neurophysiol.* **115** 2195–222

[2] Abiri R, Borhani S, Sellers E W, Jiang Y and Zhao X 2019 A comprehensive review of EEG-based brain–computer interface paradigms *J. Neural Eng.* **16** 011001

[3] Khosla A, Khandnor P and Chand T 2020 A comparative analysis of signal processing and classification methods for different applications based on EEG signals *Biocybern. Biomed. Eng.* **40** 649–90

[4] Tabar Y R and Halici U 2017 Brain computer interfaces for silent speech *Eur. Rev.* **25** 208–30

[5] Lee W, Seong J J, Ozlu B, Shim B S, Marakhimov A and Lee S 2021 Biosignal sensors and deep learning-based speech recognition: a review *Sensors* **21** 1399

[6] Gu X, Cao Z, Jolfaei A, Xu P, Wu D, Jung T-P and Lin C-T 2021 EEG-based brain-computer interfaces (BCIs): a survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18** 1645–66

[7] Shuman D I, Narang S K, Frossard P, Ortega A and Vandergheynst P 2013 The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains *IEEE Signal Process. Mag.* **30** 83–98

[8] Sandryhaila A and Moura J M F 2013 Discrete signal processing on graphs *IEEE Trans. Signal Process.* **61** 1644–56

[9] Ortega A, Frossard P, Kovacevic J, Moura J M F and Vandergheynst P 2018 graph signal processing: overview, challenges and applications *Proc. IEEE 106* 808–28

[10] Leus G, Marques A G, Moura J M, Ortega A and Shuman D I 2023 Graph signal processing: history, development, impact and outlook *IEEE Signal Process. Mag.* **40** 49–60

[11] Alzahrani S, Banjar H and Mirza R 2024 Systematic review of EEG-based imagined speech classification methods *Sensors* **24** 8168

[12] Rahman N, Khan D M, Masroor K, Arshad M, Rafiq A and Fahim S M 2024 Advances in brain-computer interface for decoding speech imagery from EEG signals: a systematic review *Cogn. Neurodyn.* **18** 3565–83

[13] Haresh M V and Begum B S 2025 Towards imagined speech: identification of brain states from EEG signals for BCI-based communication systems *Behav. Brain Res.* **477** 115295

[14] Zhao S and Rudzicz F 2015 Classifying phonological categories in imagined and articulated speech *2015 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 992–6 (available at: https://ieeexplore.ieee.org/document/7178118)

[15] Nguyen C H, Karavas G K and Artemiadis P 2017 Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features *J. Neural Eng.* **15** 016002

[16] Sun P and Qin J 2017 Neural networks based EEG-speech models (arXiv:1612.05369)

[17] Saha P, Fels S and Abdul-Mageed M 2019 Deep learning the EEG manifold for phonological categorization from active thoughts *ICASSP 2019—2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 2762–6 (available at: https://ieeexplore.ieee.org/document/8682330)

[18] Bakhshali M A, Khademi M, Ebrahimi-Moghadam A and Moghimi S 2020 EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density *Biomed. Signal Process. Control* **59** 101899

[19] Datta S and Boulgouris N V 2021 Recognition of grammatical class of imagined words from EEG signals using convolutional neural network *Neurocomputing* **465** 301–9

[20] Mortaheb S, Annen J, Chatelle C, Cassol H, Martens G, Thibaut A, Gosseries O and Laureys S 2019 A graph signal processing approach to study high density EEG signals in patients with disorders of consciousness *2019 41st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 4549–53 (available at: https://ieeexplore.ieee.org/document/8856436/)

[21] Georgiadis K, Laskaris N, Nikolopoulos S and Kompatsiaris I 2019 Connectivity steered graph Fourier transform for motor imagery BCI decoding *J. Neural Eng.* **16** 056021

[22] Einizade A, Mozafari M, Rezaei-Dastjerdehei M, Aghdaei E, Mijani A M and Sardouie S H 2020 Detecting ADHD children based on EEG signals using graph signal processing techniques *2020 27th National and 5th Int. Iranian Conf. on Biomedical Engineering (ICBME)* pp 264–70 (available at: https://ieeexplore.ieee.org/document/9319456)

[23] Sartipi S, Torkamani-Azar M and Cetin M 2021 EEG Emotion Recognition via graph-based spatio-temporal attention neural networks *2021 43rd Annual Int. Conf. IEEE Engineering in Medicine & Biology Society (EMBC) (Mexico)* (IEEE) pp 571–4 (available at: https://ieeexplore.ieee.org/document/9629628/)

[24] Einizade A, Mozafari M, Jalilpour S, Bagheri S and Sardouie S H 2022 Neural decoding of imagined speech from EEG signals using the fusion of graph signal processing and graph learning techniques *Neurosci. Inform.* **2** 100091

[25] Sartipi S, Torkamani-Azar M, and Cetin M 2023 A hybrid end-to-end spatio-temporal attention neural network with graph-smooth signals for EEG emotion recognition *IEEE Trans. Cogn. Dev. Syst.* **16** 1–7

[26] Jiménez-Guarneros M and Gómez-Gil P 2021 Standardization-refinement domain adaptation method for cross-subject EEG-based classification in imagined speech recognition *Pattern Recognit. Lett.* **141** 54–60

[27] Mohan A and Anand R S 2025 Classification of imagined speech signals using functional connectivity graphs and machine learning models *Brain Topogr.* **38** 25

[28] Alharbi Y F and Alotaibi Y A 2024 Imagined speech recognition and the role of brain areas based on topographical maps of EEG signal *2024 47th Int. Conf. on Telecommunications and Signal Processing (TSP) (Prague, Czech Republic)* (IEEE) pp 274–9 (available at: https://ieeexplore.ieee.org/document/10605970/)

[29] Lloyd S 1982 Least squares quantization in PCM *IEEE Trans. Inf. Theory* **28** 129–37

[30] Defferrard M, Martin L, Pena R, and Perraudin N 2017 PyGSP: graph signal processing in python *Zenodo* (available at: https://zenodo.org/records/1003158)

[31] Lashgari E, Liang D and Maoz U 2020 Data augmentation for deep-learning-based electroencephalography *J. Neurosci. Methods* **346** 108885

[32] Ortega A 2022 *Introduction to Graph Signal Processing* 1st ed. (Cambridge University Press) (available at: www.cambridge.org/core/product/identifier/9781108552349/type/book)

[33] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80

[34] Greff K, Srivastava R K, Koutnik J, Steunebrink B R and Schmidhuber J 2017 LSTM: a search space odyssey *IEEE Trans. Neural Netw. Learn. Syst.* **28** 2222–32

[35] Vabalas A, Gowen E, Poliakoff E and Casson A J 2019 Machine learning algorithm validation with a limited sample size *PLoS One* **14** e0224365

[36] Hyvärinen A and Oja E 2000 Independent component analysis: algorithms and applications *Neural Netw.* **13** 411–30
[37] Zhang G, Davoodnia V, Sepas-Moghaddam A, Zhang Y and Etemad A 2020 Classification of hand movements from EEG using a deep attention-based LSTM network *IEEE Sens. J.* **20** 3113–22
[38] Lock R H, Lock P F, Morgan K L, Lock E F and Lock D F 2020 *Statistics: Unlocking the Power of Data* (Wiley)
[39] Berge C 1976 *Graphs and Hypergraphs* [2d rev. edn] (transl. by Edward Minieka ed) (North-Holland)
[40] Bickle A E 2020 *Fundamentals of Graph Theory* (Pure and Applied Undergraduate Texts) (American Mathematical Society)
[41] Gao Y, Zhang Z, Lin H, Zhao X, Du S and Zou C 2021 Hypergraph learning: methods and practices *IEEE Trans. Pattern Anal. Mach. Intell.* (available at: https://ieeexplore.ieee.org/docum ent/9264674/) p 1
[42] Antelmi A, Cordasco G, Polato M, Scarano V, Spagnuolo C and Yang D 2023 A survey on hypergraph representation learning *ACM Comput. Surv.* **56** 24:1–38