Towards Telepathic Communication: A Multi-Band EEG Model for Imaginary Speech Decoding

Yifan Zhang, Yuting Ding and Fei Chen
Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, China
12332132@mail.sustech.edu.cn, 12332166@mail.sustech.edu.cn, fchen@sustech.edu.cn

***Abstract*—Imaginary speech decoding from EEG remains a challenging task, particularly under subject-independent settings, due to high inter-subject variability and limited labeled data. To address this, we propose TeleProNet, a meta-learning framework based on prototypical networks that enables few-shot classification of imagined speech across subjects. The model extracts spatial–temporal features from Mu, Beta, Gamma, and Delta frequency bands, and integrates complementary cross-band information via a learnable band-fusion module. A hybrid distance metric combining Euclidean and cosine similarities is employed to enhance prototype alignment, facilitating effective cross-subject generalization. Experiments on the Kara One dataset show that TeleProNet achieves 93.84% accuracy on an 11-class task under a leave-one-subject-out setting, validating its robustness and generalization ability. These findings suggest a promising step toward EEG-based telepathic communication systems.**

I. INTRODUCTION

Language is the primary mode of human communication. However, for each person, most of the communication in daily life occurs not through spoken words, but silently, as inner, or imaginary speech. Many people suffering from conditions such as locked-in syndrome, amyotrophic lateral sclerosis (ALS), or other diseases that make them unable to speak, in such cases, brain–computer interface (BCI) technologies that decode imagined speech from neural signals offer a potential lifeline for them—opening a window to communication. This vision can be traced back to Hans Berger, the inventor of electroencephalography (EEG), who was fascinated by the brain's silent signals and imagined that they might one day support telepathic communication [1].

Early imagined speech decoding approaches relied on handcrafted features and conventional machine learning techniques such as Support Vector Machines (SVMs), Random Forests [2], utilizing statistical descriptors and spatial filters like Common Spatial Patterns (CSP). [3] To enhance performance in EEG-based decoding tasks, many studies have adopted deep learning techniques to extract more robust and abstract representations of brain signals. Commonly used models include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid architectures that combine multiple neural components. For instance, hierarchical CNN–autoencoder models have been proposed to capture features linked to phonological articulation [4], while EEGNet variants combined with bilinear transformations have achieved improved performance [5]. Other approaches combine handcrafted features (e.g., Mel-Frequency Cepstral Coefficients (MFCCs)) with Artificial Neural Networks (ANNs) to leverage both domain knowledge and data-driven representations, thereby boosting overall accuracy [6].

Recent advances in imagined speech decoding have explored spectral dynamics based on Hilbert transform [7], functional connectivity such as phase locking value (PLV) method [8], and various classification strategies. Bakhshali et al. [9] analyzed EEG-based functional connectivity during imagined speech using Correntropy spectral density, they achieved 81.1% accuracy on the Kara One dataset. Research has also shown that Broca and Wernicke and the auditory cortex play a key role in imaginary speech and visual imagery [10]. Notably, Alizadeh and Omranpour achieved 97.34% accuracy on the Kara One dataset using a multiclass CSP-based approach with binary decomposition (One-vs-All, One-vs-One) and ensemble learning [11].

However, decoding imagined speech from EEG remains challenging due to limited data, high inter-subject variability, and the inherently low signal-to-noise ratio. Conventional deep learning models often overfit and generalize poorly in such scenarios. Meta-learning has emerged as a promising paradigm by enabling models to learn generalizable representations from few-shot learning episodes, making it particularly suitable for EEG-based imagined speech decoding. Among various meta-learning methods, Prototypical Networks have demonstrated strong effectiveness for EEG classification [12]. Prototypical Networks represent each class by a prototype, computed as the mean of embedded support examples, and classify new samples based on their distance to these prototypes in the embedding space. The distance-based structure and meta-learning nature make them well-aligned with the challenges of EEG-based BCI tasks. For instance, Hernandez-Galvan et al. employed Prototypical Networks on the Kara One dataset to achieve 91.51% accuracy under subject-independent settings [13].

Moreover, many existing models still overlook the complementary information across multiple bands, limiting their ability to extract generalizable representations. EEG signals span multiple frequency bands, each associated with different cognitive processes. Delta waves (0.5–4 Hz) are the slowest brainwaves, associated with deep rest and unconscious processing, recent studies have shown that Delta-band cortical tracking is strongly associated with speech comprehension [14]. Theta waves (4–8 Hz) reflect a dreamy, intuitive state linked to creativity and emotional depth, it has been found that theta-band cortical tracking is primarily associated with speech clarity [14]. Alpha waves (8–13 Hz) represent relaxed alertness and act as a bridge between conscious and subconscious states. Alpha oscillations in the occipital cortex are modulated by visual input and attention [15]. Although Alpha and Mu

979-8-3315-7206-8/25/$31.00 ©2025 IEEE

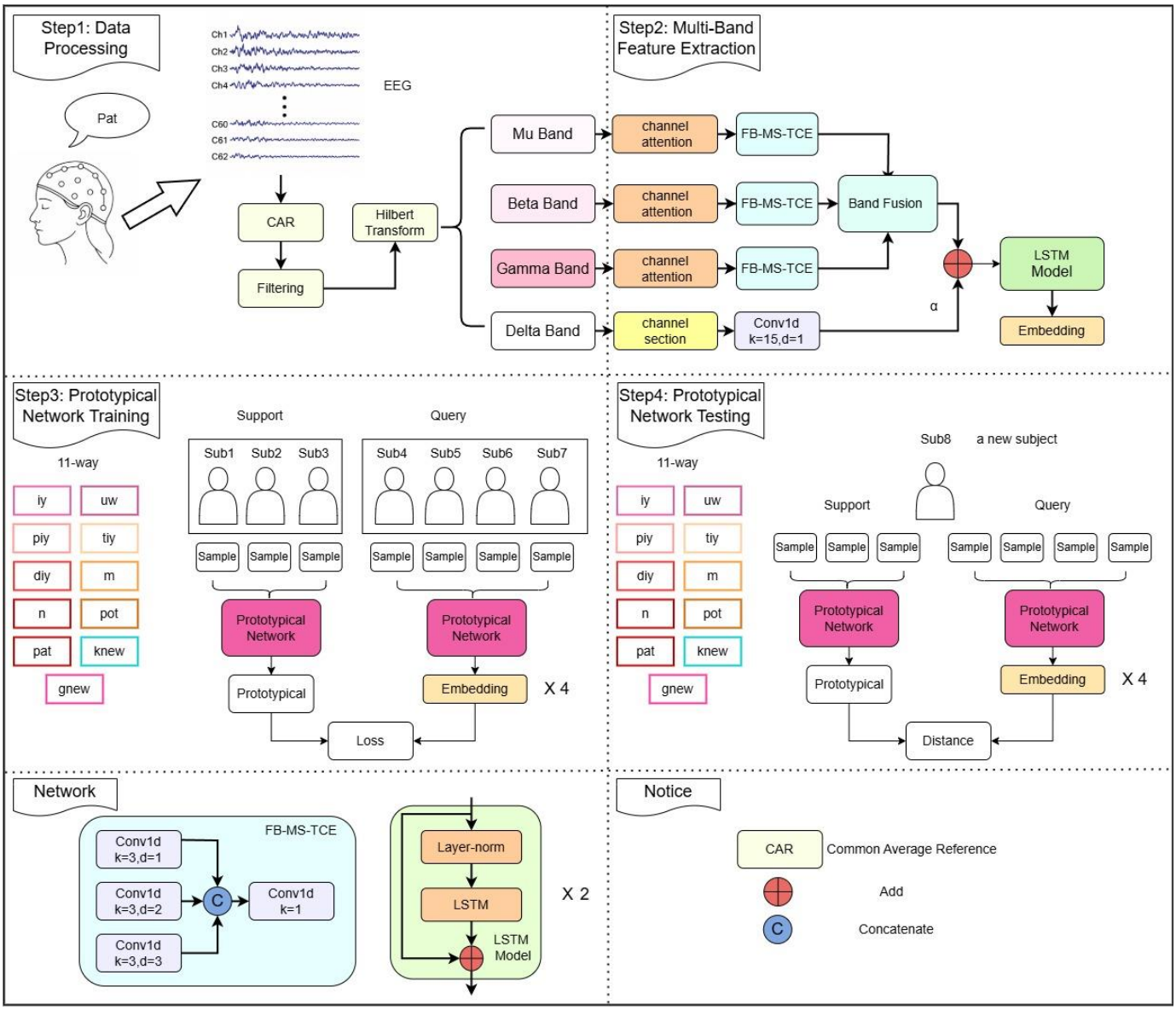

Fig. 1 The Proposed EEG-Based Multi-Band Prototypical Network for Imagined Speech Decoding.

rhythms (8–12 Hz) share similar frequency ranges, they differ slightly in their functional significance, Mu rhythms are more related to production and perception of speech [15]. Beta (13–30 Hz) reflects motor planning and top-down language control [16]. Gamma (>30 Hz) is tied to high-level cognition, learning, and perceptual integration. For imagined speech event detection, gamma-band activity in the motor cortex significantly contributes to classification performance [17]. These studies collectively highlight the value of multi-band modeling, with each frequency band providing distinct yet complementary information for imaginary speech decoding.

Motivated by prior studies, we propose TeleProNet, a multi-band prototypical learning framework for imagined speech classification. Our key contributions are:

(1) Multi-Band Feature Modeling: We design a unified architecture that extracts temporal representations from the Mu, Beta, and Gamma bands using frequency-specific convolutional modules. A gated band-fusion mechanism is introduced to adaptively integrate complementary features across bands. To further improve generalization across subjects, Delta-band information is incorporated post-fusion to provide low-frequency contextual modulation.

(2) Hybrid Prototypical Loss Function: To improve class separability in the meta-learning setting, we propose a hybrid loss that combines Euclidean and cosine similarities to enhance prototype alignment, capturing both magnitude and directional relationships in the embedding space.

II. Materials and Methods

A. Dataset

The Kara One dataset, released by the University of Toronto in 2015 [18], is a multimodal imagined speech dataset containing EEG, facial tracking, and audio data. It includes 7 phonemic/syllabic prompts (iy, uw, piy, tiy, diy, m, n) and 4 phonetically similar words (pat, pot, know, gnaw). Each trial consists of rest, stimulus, imagined speech, and actual speech phases. Due to issues such as occasional participant inattention

and poor electrode grounding, only EEG data from 8 out of 14 subjects were retained for analysis.

B. EEG Preprocessing and Feature Extraction

EEG signals were first re-referenced using Common Average Reference (CAR) and filtered with a 60 Hz notch filter to remove power-line noise. We then applied a 0.5–164 Hz bandpass filter to retain relevant EEG activity. Electrooculography (EOG) artifacts were removed via ICA with automated EOG detection (MNE toolbox), EOG refers to the electrical activity recorded around the eyes, while electromyography (EMG) artifacts were identified by high-frequency energy ratios. EEG signals during the imagined speech phase were baseline-corrected using a 2 s pre-task interval.

Signals were decomposed into four bands: Mu, Beta, Gamma, and Delta—using FIR filters. The analytic envelope of each band was obtained via Hilbert Transform and down sampled to 256 Hz. For data augmentation and temporal modeling, we applied a sliding window (600 samples, 200-sample overlap) separately within the training and testing sets to avoid data leakage.

C. Frequency Band Selection

Based on prior neurophysiological studies, we select the Mu (8–12 Hz), Beta (13–30 Hz), and Gamma (30–70 Hz) bands as the primary frequency bands, as they are closely associated with speech-related cognitive processes. Among them, Mu rhythms are notably active during covert reading, implicating their role in internal speech generation [19]. To enhance cross-subject generalization, we additionally incorporate the Delta-band (0.5–4 Hz), which has demonstrated higher discriminative power and stability in subject-independent EEG classification [8].

For feature extraction, we select 48 electrodes from the original 62 channels, these electrodes primarily cover the frontal, temporal, central, and parietal regions, which are widely recognized as key areas involved in speech planning, production, and auditory-phonological processing. A channel attention mechanism based on the Squeeze-and-Excitation (SE) module [20] is applied to the Mu, Beta, and Gamma bands to enhance spatial feature representation.

In contrast, Delta-band feature were extracted from a separate subset of 10 electrodes (P1, P2, PZ, POZ, CPZ, CP3, C1, C2, C3, C4), mainly distributed across the parietal and central regions, which are involved in integrative cognitive processing.

D. Structure of the Prototype Network

As illustrated in Figure 1, our network consists of three main components: (1) a multi-band feature extraction module (2) a gated Band-Fusion module, and (3) a long-term temporal modeling module. Each part is described as follows:

(1) Multi-band Feature Extraction Module: As shown in Figure 1, following channel attention, we extract features from the Mu, Beta, and Gamma bands using the Frequency-Band Multi-Scale Temporal Convolution Encoder (FB-MS-TCE), which applies parallel 1D convolutions with dilation rates of 1, 2, and 3 to capture multi-scale temporal patterns, each with 64 output channels. For the Delta band, we apply a separate 1D CNN with a larger kernel size of 15 to model long-range low-frequency dynamics.

(2) Band-Fusion: This module is inspired by the gating mechanism in Long Short-term Memory Networks (LSTM). It selectively fuses complementary features and integrate them with the previously combined features.

$$gate = \sigma(W_1([Mu, Beta, Gamma])), \quad (1)$$

$$candidate = tanh\,(W_2([Mu, Beta, Gamma]), \quad (2)$$

where $Mu, Beta, Gamma$ are the features of each band after the encoder, W_i is 1D convolution layer, W_1 , W_2 are convolution layers for band-wise compression, The operator [·] denotes channel-wise concatenation. These two outputs are combined via element-wise multiplication to generate the gated update feature:

$$update = gate \odot candidate, \quad (3)$$

where, $update$ stands for complementary features, $\odot$ indicates element-wise multiplication. Meanwhile, another branch is computed by summing the three frequency bands, which passing them through a forget gate to control contribution of the combined features:

$$forget = \sigma\big(W_3(Mu + Beta + Gamma)\big), \quad (4)$$

where, $forget$ stands for combined features, W_3 is a normal convolution for feature enhancement. The complementary and combined features are summed:

$$Fusion = update + forget, \quad (5)$$

this mechanism allows the model to selectively integrate complementary frequency features while preserving useful original information. The subject-independent task also requires the Delta-band features to be added to the fusion result, with the result:

$$Final = Fusion + \alpha \times Delta, \quad (6)$$

where $Delta$ is the feature of Delta-band after 1D convolution, α is the learnable parameter.

(3) Long-Term Temporal Modeling: The final fused features are max pooled with a kernel size of 2, then passed through two residual bidirectional LSTM layers (64 units per direction), each preceded by Layer Normalization. A Global Average Pooling (GAP) layer is applied over the temporal dimension to obtain a fixed length embedding.

E. Prototypical Classification

In each meta-learning episode, which includes a support and a query set, all samples are passed through the entire network to obtain embeddings, the prototype p_k for class k is computed as the mean of its support embeddings:

$$p_k = \frac{1}{|S_k|} \sum_{s_i \in S_k} f_\theta(s_i)\,, \quad (7)$$

where S_k is all the samples of the k^{th} class support set, in this

paper we adopt a 3-shot configuration, so here it is 3, $f_\theta(\cdot)$ denotes the network introduced above, i.e., the prototype network, with θ learnable parameters.

In the classification stage of the prototypical network, we first compute the distance between each query sample q and the prototype p_k of each class k. To capture both magnitude and directional differences between the embeddings, we construct a weighted distance function:

$$D(q, p_k) = \lambda \cdot |f_\theta(q) - p_k|_2^2 + (1-\lambda) \cdot \left(1 - \frac{f_\theta(q) \cdot p_k}{|f_\theta(q)| \cdot |p_k|}\right), \quad (8)$$

where $D(q, p_k)$ denotes the distance between the embedding of the query sample q and the prototype p_k of class k , λ=0.5 is the weighting factor of the Euclidean distance to the cosine distance. After obtaining the distances, the distances are converted into classification probabilities through the SoftMax function:

$$P_\theta\left(y^{(q)} = k^* \mid q\right) = \frac{exp\left(-D(q, p_{k^*})\right)}{\sum_{j=1}^{K} exp\left(-D(q, p_j)\right)}, \quad (9)$$

where $P_\theta\left(y^{(q)} = k^* \mid q\right)$ denotes the predicted probability that the query sample q belongs to class k^* , $y^{(q)}$ represents the category of q samples, which belonging to k .

The loss function in the training phase is calculated from the classification probability as follows:

$$\mathcal{L}_{proto_\theta} = -log\, P_\theta\left(y^{(q)} = k \mid q\right). \quad (10)$$

where $\mathcal{L}_{proto_\theta}$ represents the prototype classification loss for the query sample q, encouraging the model to minimize the distance to the correct class prototype p_k while maximizing the distance to others.

III. Experiments and Results

A. Subject-Independent Evaluation Setup

We evaluate our method under a subject-independent setting using the Leave-One-Subject-Out (LOSO) cross-validation strategy. In the subject-dependent setting, models are trained and tested on data from the same subject, whereas in the subject-independent setting, models are trained on data from a subset of subjects and tested on unseen subjects. In each fold, one subject is held out as the test set, while the remaining seven are used for training. Final performance is reported as the average accuracy over all 8 subjects.

B. Training Strategy

As shown in Fig. 1, we adopt a novel few-shot training strategy based on episodes. Each training episode is designed as an 11-way, 3-shot, 4-query task. For each class, 3 support samples are selected from 3 different subjects among the training data. This ensures that the support set captures diverse cross-subject representations. The query set consists of 4 samples per class, drawn from the remaining 4 training subjects. A total of 14,900 training episodes are generated to enable robust learning of subject-invariant representations.

Table I

Subject-Wise Performance Metrics Under Subject-Independent Evaluation (LOSO Evaluation)

Subject for test	Precision	Recall	F1-score	Accuracy	Kappa
MM05	0.9311	0.9270	0.9268	0.9270	0.9197
MM10	0.9978	0.9977	0.9977	0.9977	0.9975
MM11	0.9940	0.9939	0.9939	0.9939	0.9932
MM16	0.8325	0.8295	0.8293	0.8295	0.8125
MM18	0.9788	0.9784	0.9784	0.9784	0.9762
MM19	0.9143	0.9107	0.9104	0.9107	0.9018
MM21	0.9823	0.9820	0.9820	0.9820	0.9802
P02	0.8932	0.8875	0.8871	0.8875	0.8762
average	0.9405 ± 0.0586	0.9384 ± 0.0603	0.9382 ± 0.0604	0.9384 ± 0.0603	0.9322 ± 0.0630

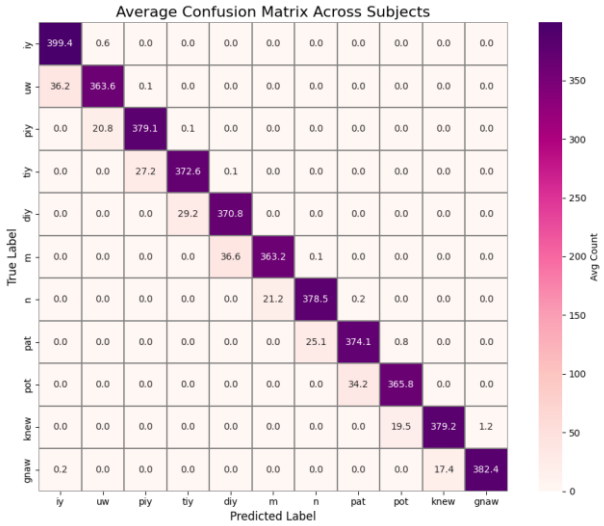

Fig. 2 Average Confusion Matrix of 11-Class Imagined Speech Classification Across 8 Subjects (LOSO Evaluation)

C. Testing Strategy

During testing, both the support and query samples are drawn exclusively from the held-out subject. For each class, 3 samples are randomly selected to form the support set, and 4 samples are used as the query set. Each fold includes 100 test episodes, covering all available test data. We aggregate predictions across all test episodes and the final accuracy is computed by averaging results across all subjects.

D. Subject-Independent Evaluation Result

As shown in Table I, under the subject-independent setting, our method achieves an average accuracy of 93.84%, with precision and F1-score all above 93%, indicating strong and consistent performance. The reported macro recall and accuracy values are identical because each testing episode samples an equal number of query trials per class, resulting in a perfectly balanced test set. The average Kappa score of 0.9322 suggests high agreement beyond chance. Most subjects achieved consistently high accuracy, though Subject MM16

TABLE II

BENCHMARKING TELEPRONET AGAINST EXISTING METHODS FOR 11-CLASS IMAGINED SPEECH DECODING (KARA ONE DATASET)

Work	Method	Subject dependent/independent	Problem	Accuracy
Mini et al. (2021) [6]	SMRT/MFCC/LPCC + PCA/ANN	dependent	11-class	77%
Bakhshali et al. (2022) [9]	pCBCSD + SVM	dependent	11-class	81.6%
Alizadeh et.al (2023) [11]	EM-CSP + KNN	dependent	11-class (one vs. all)	97.34%
Hernandez-Galvan et al. (2023) [13]	A prototypical network (CNN+GRU)	independent	11-class	91.5%
TeleProNet (Ours)	A prototypical network (multi-band and Delta modulation)	independent	11-class	**93.84%**

exhibited lower accuracy (82.95%). Figure. 2 presents the average confusion matrix across 8 subjects under the LOSO evaluation. The matrix shows strong diagonal dominance, indicating that most samples are correctly classified. Off-diagonal values are minimal, suggesting low confusion between classes.

E. Comparison with Existing Methods

Table II summarizes the performance of recent methods on the 11-class imagined speech classification task using the Kara One dataset. Prior studies have mainly focused on subject-dependent settings, relying on handcrafted features (e.g., MFCC) or spatial filtering methods (e.g., CSP). Among these, Alizadeh et al. [11] achieved a high accuracy of 97.34% through a one-vs-all strategy based on EM-CSP and KNN. However, their method, along with Mini et al. [6], used data from 13 subjects, whereas later studies [9,13] and ours adopted data from 8 subjects. This adjustment follows prior findings that reported data quality issues in 6 out of 14 subjects of the Kara One dataset.

Hernandez-Galvan et al. introduced a prototypical network for subject-independent classification, reaching 91.5% accuracy. However, their model did not incorporate multi-band fusion. Inspired by recent advances in domain-guided contrastive adversarial learning for EEG workload decoding [21], we aim to improve imagined speech recognition under subject-independent settings.

Our method achieved 93.84% accuracy under subject independent LOSO evaluation. This highlights our model's superior ability to extract subject-invariant representations for imagined speech decoding.

F. Ablation Study

To assess the contribution of each module in TeleProNet, we conducted ablation experiments under the subject-independent setting:

(1) w/o Band-Fusion (Direct Addition): The features extracted from the Mu, Beta, and Gamma bands are directly added together without using the fusion gate module.

(2) w/o Delta: The Delta-band branch is removed, and only the Mu, Beta, and Gamma bands are used in the main encoding path.

TABLE III

MODULE ABLATION EXPERIMENT RESULTS

Number	Model	Accuracy (%) ↑	F1-score (%) ↑
1	**w/o Band-Fusion (add)**	9.20% ± 8.96%	8.88% ± 8.70%
2	**w/o Delta**	74.82% ± 21.44%	74.55% ± 21.98%
3	**w/o channel attention**	88.65% ± 3.91%	88.62% ± 3.91%
4	**w/o cosine distance**	89.60% ± 10.45%	89.54% ± 10.56%
5	**Only Beta**	21.52% ± 21.91%	21.10% ± 21.76%
6	**Only Mu**	11.47% ± 1.92%	10.66% ± 2.80%
7	**Only Gamma**	27.10% ± 21.43%	26.52% ± 21.35%
8	**Full Model**	93.84% ± 6.03%	93.82% ± 6.04%

(3) w/o Channel Attention: The channel attention modules are removed, and all channels are treated with equal importance.

(4) w/o Cosine Distance: The cosine component is removed from the hybrid prototype distance function, using only Euclidean distance for similarity measurement.

(5) Single Band (Mu / Beta / Gamma): To evaluate the stand-alone contribution of each main frequency band, the model is trained separately using only one of the Mu, Beta, or Gamma bands, excluding other bands.

Table III demonstrates the contribution of each module in the proposed model. Removing the band fusion module and using simple addition led to a sharp performance drop, with accuracy reduced to 9.2%, showing that direct summation cannot learn complementary information between frequency bands.

When using individual frequency bands, the Gamma-only model outperforms Mu and Beta, indicating that high-frequency activity carries more discriminative information for imagined speech. However, the limited performance of single-band models suggests that complementary spectral features are required for robust decoding.

Excluding the Delta band reduced accuracy by 19.02%, indicating its role in enhancing cross-subject generalization. Removing channel attention caused a smaller drop (5.19%),

suggesting that the mechanism enables each frequency band to emphasize distinct neural regions associated with its cognitive role. Replacing the hybrid distance metric with Euclidean-only also slightly lowered performance (−4.42%), suggesting that both magnitude and directional similarity in the embedding space are critical for accurate prototype alignment.

IV. Conclusion

In this study, we propose a multi-band prototypical network framework for imaginary speech decoding. A gated fusion module is employed to integrate complementary information across frequency bands. Experiments on the Kara One dataset achieved 93.84% accuracy under the 11-class classification task under subject-independent setting, demonstrating strong cross-subject generalization.

Notably, our results suggest that Delta-band activity—traditionally associated with sleep and unconscious processing—also plays a significant role during conscious imagined speech. Overall, our work demonstrates the feasibility of cross-subject imagined speech decoding and provides a foundation for future research toward telepathic communication—a greater goal of imaginary speech brain–computer interface system. Code is available at https://github.com/XDUANN/TeleProNet.git

References

[1] J. T. Panachakel and A. G. Ramakrishnan, "Decoding covert speech from EEG-a comprehensive review," Frontiers in neuroscience, vol. 15, p. 642251, 2021.

[2] Y. V. Varshney and A. Khan, "Imagined speech classification using six phonetically distributed words," Frontiers in Signal Processing, vol. 2, p. 760643, 2022.

[3] P. Agarwal, R. K. Kale, M. Kumar, and S. Kumar, "Silent speech classification based upon various feature extraction methods," in Proc. 7th Int. Conf. Signal Process. Integr. Netw. (SPIN), Feb. 2020, pp. 16–20.

[4] P. Saha, M. Abdul-Mageed, and S. Fels, "Speak your mind! Towards imagined speech recognition with hierarchical deep learning," arXiv preprint arXiv:1904.05746, 2019.

[5] G. Rousis, F. P. Kalaganis, S. Nikolopoulos, I. Kompatsiaris, and P. C. Petrantonakis, "Combining EEGNet with SPDNet towards an end-to-end architecture for imagined speech decoding," in Proc. 32nd Eur. Signal Process. Conf. (EUSIPCO), Sep. 2024, pp. 1531–1535.

[6] P. Mini, T. Thomas, and R. Gopikakumari, "EEG based direct speech BCI system using a fusion of SMRT and MFCC/LPCC features with ANN classifier," Biomedical Signal Processing and Control, vol. 68, p. 102625, 2021.

[7] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward EEG sensing of imagined speech," in Proc. 13th Int. Conf. Human-Computer Interaction (HCI), Jul. 2009, pp. 40–48.

[8] M. Haresh and B. S. Begum, "Towards imagined speech: Identification of brain states from EEG signals for BCI-based communication systems," Behavioural Brain Research, vol. 477, p. 115295, 2025.

[9] M. A. Bakhshali, M. Khademi, and A. Ebrahimi-Moghadam, "Investigating the neural correlates of imagined speech: An EEG-based connectivity analysis," Digital Signal Processing, vol. 123, p. 103435, 2022.

[10] S.-H. Lee, J.-H. Park, and D.-S. Kim, "Imagined Speech and Visual Imagery as Intuitive Paradigms for Brain-Computer Interfaces," International Conference on Brain-Computer Interface (BCI), IEEE, 2025, pp. 1–4.

[11] D. Alizadeh and H. Omranpour, "EM-CSP: an efficient multiclass common spatial pattern feature method for speech imagery EEG signals recognition," Biomedical Signal Processing and Control, vol. 84, p. 104933, 2023.

[12] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," Advances in neural information processing systems, vol. 30, 2017.

[13] A. Hernandez-Galvan, G. Ramirez-Alonso, and J. Ramirez-Quintana, "A prototypical network for few-shot recognition of speech imagery data," Biomedical Signal Processing and Control, vol. 86, p. 105154, 2023.

[14] O. Etard and T. Reichenbach, "Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise," J. Neurosci., vol. 39, no. 29, p. 5750.

[15] H. M. Hobson and D. V. Bishop, "The interpretation of mu suppression as an index of mirror neuron activity: past, present and future," Royal Society Open Science, vol. 4, no. 3, p. 160662, 2017.

[16] M. Pefkou, L. H. Arnal, L. Fontolan, and A.-L. Giraud, "θ-Band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech," Journal of Neuroscience, vol. 37, no. 33, pp. 7930–7938, 2017.

[17] A. de Borman et al., "Imagined speech event detection from electrocorticography and its transfer between speech modes and subjects," Communications Biology, vol. 7, no. 1, p. 818, 2024.

[18] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," presented at the 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 2015, pp. 992–996.

[19] T. Tamura et al., "Audio-vocal monitoring system revealed by mu-rhythm activity," Frontiers in psychology, vol. 3, p. 225, 2012.

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7132–7141.

[21] R. Zhan, D. Li, S. Wang, and Q. Liu, "D2CAN: Domain-guided contrastive adversarial network for EEG-based cross-subject cognitive workload decoding," in Human Brain and Artificial Intelligence, Q. Liu, Y. Qu, H. Wu, Y. Qi, A. Zeng, and D. Pan, Eds., Springer, 2025, pp. 250–264.