Title: A Comprehensive Review of Imagined Speech Decoding in Brain-Computer Interfaces: Utilizing EEG and fNIRS Technologies

Authors: Monireh Motaqi[1], Majid Moallemi[2], Abolfazl Mirani[3], Yeganeh Aboutorabi[4], Boshra Hatef[5,*]

1. *Physiotherapy Research Center, School of Rehabilitation, Shahid Beheshti University of Medical Sciences, Tehran, Iran.*
2. *Department of Biomedical Engineering, Faculty of Engineering, Islamic Azad University, Mashhad Branch, Mashhad, Razavi Khorasan, Iran.*
3. *Biomedical Engineering Research Center, New Health Technologies, Baqiyatallah University of Medical Sciences, Tehran, Iran.*
4. *Department of Microbiology, Faculty of Basic Sciences, Islamic Azad University, Mashhad Branch, Mashhad, Razavi Khorasan, Iran.*
5. *Neuroscience Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.*

***Corresponding Author**: Boshra Hatef, Neuroscience Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran. Email: boshrahatef@bmsu.ac.ir

To appear in: **Basic and Clinical Neuroscience**

Received date: 2025/08/07

Revised date: 2025/09/25

Accepted date: 2025/12/24

Please cite this article as:

Motaqi, M., Moallemi, M., Mirani, A., Aboutorabi, Y., Hatef, B. (In Press). A Comprehensive Review of Imagined Speech Decoding in Brain-Computer Interfaces: Utilizing EEG and fNIRS Technologies. Basic and Clinical Neuroscience. Just Accepted publication Jul. 10, 2026. Doi: http://dx.doi.org/10.32598/bcn.2026.8254.1

DOI: http://dx.doi.org/10.32598/bcn.2026.8254.1

Abstract

The use of brain–computer interfaces (BCIs) to decode imagined speech has significant clinical and assistive potential. Twenty-six studies investigated covert speech decoding between 2009 and 2025 using EEG, fNIRS, or hybrid EEG–fNIRS systems. Early research (2009–2012) primarily focused on analyzing phonemes and syllables with EEG, achieving accuracy rates around 75%. From 2013 to 2017, CNN-based phoneme decoding produced highly variable results (40%–83%), with more complex multiclass tasks occasionally performing poorly (as low as 26.7%). Since 2018, binary paradigms such as yes/no responses have reached 64%–100% accuracy. CNN variants (about 83.4%), AlexNet (90.3%), and LSTM-RNNs (92.5%) demonstrated notable improvements, whereas architectures like EEGNet and SPDNet often underperformed (24.79%–66.93%). In hybrid EEG–fNIRS methods, convolutional neural networks (CNNs) achieved roughly 53% accuracy, while traditional classifiers like SVM and LDA performed better, reaching 78–79%. These results indicate that although deep learning and multimodal systems have potential for enhancing imagined speech decoding, there are still major challenges related to generalization, variability, and robustness.

Keywords: Brain-computer interface, imagined speech decoding, EEG, fNIRS, machine learning, multimodal fusion.

Introduction

The brain-computer interface (BCI) enables direct interaction between the human cerebral cortex and external devices [1, 2]. Individuals control devices via cognitive intent, translating brain impulses or blood flow into signals through BCIs. This is crucial in neurorehabilitation, assistive tech for motor disabilities, and restoring communication in severe impairments. A key research area is decoding imagined speech [3, 4]. The methodology involves individuals silently thinking words or phrases while the system tries to identify their mental speech. Those unable to speak or move can use imagined speech as a natural, voluntary communication method. Reading these covert cognitive processes is challenging due to their weak neural activity often masked by noise, complicating investigation [5, 6]. Two non-invasive techniques are commonly employed by researchers to solve this challenge, Electroencephalography (EEG) and Functional Near-Infrared Spectroscopy (fNIRS) [7, 8]. Integrating EEG and fNIRS into a multimodal framework offers an excellent opportunity. Scientists can leverage each technique's strengths: EEG captures rapid brain activity, while fNIRS precisely targets specific regions [9]. An investigation utilized a configuration to classify imagery related to hand movements, involving three EEG electrodes (C3, C4, and Cz) and ten fNIRS channels [10-12]. The system achieved 81.2% accuracy, outperforming EEG at 74.7% and fNIRS at 56.8% [12]. The success of combining different BCI approaches shows these technologies can improve the field and inspire hope. However, merging these data types is very challenging [13, 14]. The differing temporal scales of EEG and fNIRS—instantaneous signals versus delayed blood flow changes—require researchers to use advanced methods like phase-space reconstruction, Common Spatial Patterns (CSP), and hemodynamic features (e.g., the Hurst exponent) to synchronize and combine the data [12]. Multimodal brain-computer interfaces can restore communication and autonomy for critical individuals by combining EEG and fNIRS advantages [15, 16].

In recent years, there has been a growing scholarly interest in decoding internally generated speech with EEG signals, which goes beyond basic binary classification tasks like yes/no responses, where neural networks applied to 60-channel EEG data achieved accuracy ranging from 70% to nearly 100% [17-19]. Researchers quickly adopted CSP analysis and Support Vector Machines (SVMs) as methods that provided comparable performance [20]. This review paper explains why specific machine learning and signal processing methods are suitable for speech decoding by concluding valid studies. While previous reviews mainly focus on EEG for decoding imagined speech, this study emphasizes evidence from fNIRS and hybrid EEG–fNIRS systems alongside EEG. Although reports on fNIRS-only [21] and hybrid [12, 22-25] approaches are limited, these findings offer valuable insights. This review investigates whether combining or replacing these modalities can enhance decoding imagined speech by using fNIRS' ability to detect hemodynamic changes and complement EEG's limitations.

Methodology

We conducted a narrative review of empirical studies on decoding imagined speech with EEG, fNIRS, or hybrid EEG–fNIRS systems from 2009 to 2025, across PubMed, IEEE Xplore, Scopus, and Google Scholar. Keywords included combinations like 'EEG' and 'imagined speech', etc. The search yielded approximately 15,900 records. After eliminating duplicates, we examined titles and abstracts. The studies had to meet three criteria to be included: (1) the articles had to be original or reviews in English, (2) they had to investigate imagined speech decoding using EEG, fNIRS, or both, and (3) the methods used had to be clearly explained. If studies were not in English, did not involve imagined speech, or did not have clear methodological details, they were excluded. We found 26 studies that met the criteria after reviewing the full texts.

To compare, the research studies were divided into three distinct categories:

1. Laboratory-based investigations, with a focus on data acquisition and designing experimental tasks.
2. Algorithmic investigations emphasize the importance of feature extraction and the effectiveness of classification methods.
1. Utilization of integrated EEG-fNIRS methodologies for investigating the synthesis of multimodal signal convergence.

Preprocessing methods were consistently used across studies, mainly focusing on bandpass filtering to retain signals within EEG (0.5–40 Hz) or fNIRS frequency ranges, artifact removal via ICA, and phase-space reconstruction in EEG to reveal nonlinear attributes. Feature extraction included PSD, DWT, CSP, Riemannian features, and fNIRS indicators like hemoglobin concentration changes. Connectivity metrics like PDC and PLV studied motor, Broca's, and prefrontal areas during imagined speech tasks. Algorithms such as SVM, CNN, RNN, and hybrid models like CSP EEGNet were compared for their effectiveness in handling high-dimensional, small-sample EEG and fNIRS data. SVM offered robust decision boundaries, while CNN and RNN excelled in feature learning and temporal modeling. Performance was evaluated based on classification accuracy and error rates for comparison.

Results

This section outlines 26 studies decoding imagined speech with EEG, fNIRS, or both, highlighting key patterns, comparing findings, and exploring reasons for variances.

Data distribution methodology

EEG-exclusive investigations (20 out of 26): EEG has been established as the premier modality due to its remarkable temporal resolution and non-invasive characteristics. The reported classification accuracy in EEG-exclusive investigations exhibit considerable variability, ranging from approximately 34.2% to 99%, depending on the task complexity, the number of electrodes used, and the preprocessing methodologies employed [6, 20].

fNIRS-exclusive investigations (1 out of 26): Despite delays in the hemodynamic response, fNIRS is effective in binary mental communication tasks, with past accuracy rate around 71–75% [26]. This study introduced a three-class imagined speech BCI, allowing participants to communicate by thinking "yes", "no", or resting. The average online accuracy over three blocks was 64.1% ± 20.6, and nine of twelve participants were above chance. Results varied due to signal-to-noise ratio, mental task performance, and channel setup, especially over the left temporal and temporoparietal areas, which provided the most discriminative information[21].

Hybrid EEG–fNIRS investigations (5 out of 26): Through the integration of EEG characteristics (e.g., CSP, Root Mean Square RMS)) with fNIRS indicators (e.g., average concentrations of oxyhemoglobin, Hurst exponent), hybrid systems have demonstrated superior performance relative to their single modality counterparts by margins of up to 20% in numerous studies. Nevertheless, challenges related to synchronization (the alignment of rapid EEG intervals with the more gradual fNIRS segments) and the increased complexity of the equipment occasionally introduce extraneous noise, which partially mitigates these advantages [12, 22-25, 27, 28].

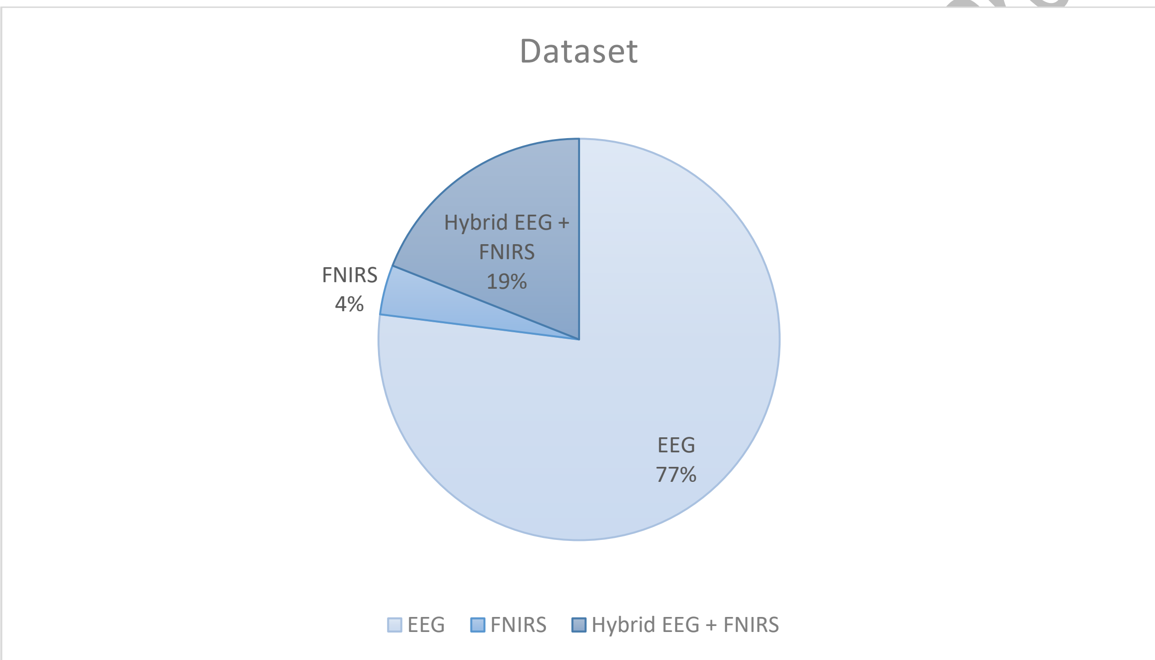

Figure 1: The distribution of data modalities utilized in the reviewed studies.

Comparative analysis and classification of speech stimuli

Various scholarly studies have used a variety of imagined speech tasks that vary in complexity and number of categories.

1. **Binary Tasks (Yes/No):**
   - **EEG-only:** Accuracy reached approximately 99%, particularly with methods such as Common Spatial Pattern and SVM. Simple two-category experiments remain consistently robust and stable [29].
   - **fNIRS-only:** In three-class classification, the average accuracy over the last three online blocks was about 64.1% ±20.6%. In the final block, nine of twelve participants reached 83.8% ±9.4%, with each mental task lasting 15 seconds to ensure an adequate hemodynamic response [21].

2. **Phonemic Stimuli (e.g., /a/ or /u/) in Closed-Set Environments:**
    - **EEG-based:** Accuracy metrics using methods like the Hilbert Transform with CSP, predictive models like matched filters or CNN, range from 57% to 85%. One study achieved 75% accuracy with a real-time matched filter and PSD features. CNN architectures like CNNeeg1-1 reached up to 85% accuracy on balanced datasets[18, 30-34].

3. **Word-Based Tasks (e.g., "up," "down," "left," "right," "help," "stop"):**
    - **EEG-only:** Accuracy metrics for tasks involving six imaginary Persian words range from 85% to 97%, especially when using FFTs with SVM or CNN/LSTM. An AlexNet study on ten imaginary words achieved about 90.03% [23, 35-40].

4. **Multi-Class Phrases (up to 13 categories, sometimes multilingual):**
    - **EEG-only:** For multilingual setups with 6–12 lexemes in languages such as English, Arabic, Persian, and Spanish, the accuracy of CNN, rLDA, and RF models varies from 23.7% (12-class) / 34.2% (13-class) to 62.37%, depending on the features and validation methods used [17, 35-37, 41-50].
    - **Hybrid EEG–fNIRS:** EEG and fNIRS data are combined using features like Discrete Wavelet Transform coefficients from Symlet-10 across six levels, the EEG signal's RMS value, and the average [HbO] concentration from fNIRS. One study reported a classification accuracy of 89.4% for imagined speech[24]. These findings demonstrate that the hybrid approach surpasses single-modality methods, with EEG alone achieving accuracy between 34.2% and 70.33%, and fNIRS alone generally showing weaker results compared to the combined technique [22-24, 41, 46, 51, 52].

Feature Extraction Techniques

To extract meaningful features from neural signals, the studies discussed in this review employed a variety of signal processing techniques. The methodology used was primarily determined by the type of data available, whether it was EEG, fNIRS, or a hybrid system. The following are the main techniques discussed:

EEG-Based Features

1. **Root mean square (RMS)**: The temporal-domain parameter measures EEG signal amplitude via RMS, reflecting signal energy sensitive to cerebral activity, especially during motor imagery or movement [32, 53, 54].
2. **Power Spectral Density (PSD):** PSD assesses power across frequency bands—delta, theta, alpha, beta, gamma—linked to specific cognitive functions [47].

3. **Discrete Wavelet Transform (DWT):** The DWT enables multi-resolution EEG decomposition, making it ideal for identifying time-frequency features crucial for interpreting imagined speech [6, 41].
4. **Common Spatial Patterns (CSP):** CSP is used in binary classification to improve class discrimination by fine-tuning spatial filters across EEG channels [29, 46, 49, 52, 55].
5. **Phase-Based Connectivity Measures:** The use of techniques such as PLV and coherence to examine synchronization among cerebral regions can provide insight into the neural dynamics involved in the generation of internal speech [44, 56].

fNIRS-Based Features Key features derived from fNIRS data include:

1. Mean changes in the concentrations of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) [21].
2. Variations in slope of hemodynamic response curves [57].
3. Hurst exponent values are used to capture the complexity and long-term memory characteristics of fNIRS signals [51].

Multimodal Features (EEG–fNIRS Hybrid Systems):

EEG and fNIRS signals were used in studies where features were either combined or integrated before classification. This method leverages EEG's high temporal resolution along with the spatial and hemodynamic information from fNIRS to create more complete and insightful data representations. As a result, this approach helps the audience understand the potential of multimodal features for improving their knowledge of brain function [58].

Classification Algorithms

In this research, there have been various classification algorithms utilized, each chosen based on the data's specific nature, task difficulty, and dataset size.

Support Vector Machine (SVM)

The reviewed research shows that this classification algorithm, commonly used, has yielded promising results, especially in complex tasks and large datasets. In EEG-specific setups, SVM is the most used classifier. For example, EEG features like band power and CSP extracted from specific filters achieved subject-specific accuracies of 70% to 99.9% when distinguishing "Yes/No" responses in occupational questions. Under the mixed-time scenario, the SVM achieved up to 97% accuracy in binary classification for six imagined Persian words [35]. Other research has also found accuracy between 80.2% and approximately 86% for tasks using simple yes/no words [29, 59, 60]. SVM demonstrated efficiency with integrated EEG and fNIRS features, achieving 81.2% accuracy in distinguishing left and right-hand motor imagery. This highlights SVM's ability to handle high-dimensional multimodal data [51].

Linear Discriminant Analysis (LDA) and Regularized LDA (rLDA)

In fNIRS tasks, rLDA achieved a mean accuracy of 64.1% ± 20.6% for a dichotomous 'Yes/No' task; however, the considerable variance indicated a susceptibility to minor datasets and noise [21]. Regularized rLDA achieved about 61% accuracy in decoding six Spanish words from an EEG. Using EEG and fNIRS, shrinkage-based LDA can categorize MI, MA, and rest states with an accuracy of around 78–79%. The accuracy was 77.5 ±12.1% in a small setup with 2 EEG channels and 2 fNIRS pairs. LDA's effectiveness decreases if Gaussian assumptions are violated, even though it is both simple and interpretable [21, 22, 61, 62].

Random Forest (RF)

The RF algorithm's accuracy in multiclass EEG word classification ranged from 68.18 to 70.33% for **five Spanish words** (‘arriba, abajo, izquierda, derecha, seleccionar’) together with their **English equivalents**, depending on features and subjects [41]. A separate study with twelve words or phrases, plus a rest class (creating a 13-class setup), showed accuracy dropping to 26.7% for visual imagery and 34.2% for imagined speech [41, 46, 63]. RF achieved moderate accuracy with features like wavelet transforms or statistical analyses but generally performed worse than deep learning [64].

Naive Bayes (NB)

Some EEG studies show that this method is used mainly with low-dimensional or discrete attributes. A Naive Bayes classifier using a Bag-of-Features approach achieved an average accuracy of 65.65% ±13.39 when classifying five Spanish words ('arriba,' 'abajo,' 'izquierda,' 'derecha,''seleccionar'). In transfer experiments without calibration, the accuracy decreased to 58.74% ± 13.39 for the word *Up* and 61.38% ± 12.47 for the word *Down*. Compared to SVM or deep learning, neural networks (NB) tend to perform worse when features are correlated, despite their speed and simplicity [65, 66].

LDA Variants

Regularized adaptations like ridge-regularized LDA or shrinkage-based LDA have been used alongside conventional LDA. These techniques improved weight stability, especially with limited data or contaminated EEG channels [65, 66].

Adaptive and Kalman Filters

In phoneme recognition, implemented adaptive filtering to differentiate /ku/ and /ba/ using EEG data, achieving 75% accuracy [47]. In EEG/ECoG vowel decoding, Kalman filtering estimated the first two formants (F1, F2) for vowels (AA, IY, UW), with about 71% accuracy. These methods leverage signal continuity but need careful parameter tuning [32, 67, 68].

Convolutional Neural Networks (CNN)

A two-stage system was built that included Convolutional Neural Networks, Spatial CNN, Temporal CNN (TCNN), Denoising Autoencoder (DAE), and XGBoost. Initially, this approach classified six binary phonological categories with an accuracy of 83.42%. Then, it used these features for token recognition, reaching an accuracy of 53.36% across 11 tokens. A baseline method that relied solely on raw covariance features achieved 28.08%. AlexNet classified ten English words with a accuracy of 90.3% [36, 43, 69].

• **CNN with attentional mechanisms**

The EEGNet model incorporating attention mechanisms was able to achieve a 57% accuracy in identifying four syllables (/Ba/, /Ku/, /He/, /Li/). Even if attention is below 60%, it can still effectively extract salient spatial-spectral features from noisy EEG signals [40, 70].

• **In the domain of EEG and fNIRS fusion**

A CNN used to integrate EEG and fNIRS features achieved accuracy ranges between 53% and 87.18%, depending on the subjects and tasks (text, image, audio). This variability shows that CNNs can leverage multimodal synergy but are subject to individual differences [23, 51, 52].

EEGNet Variants and CTC

Three categories (/a/, /u/, and rest) were classified using a model combining EEGNet-inspired CNN, RNN (LSTM), and CTC loss. The study used character-level edit distance instead of accuracy. Despite no attention mechanism, results showed that compact depthwise convolutions effectively captured spatial–temporal features in EEG signals [34, 71].

Recurrent Neural Networks (RNNs)

Four-directional EEG classification (Up, Down, Left, Right) was investigated by using LSTM-RNNs with a accuracy of 92.5%. These models are adept at capturing intricate, temporally varying characteristics, especially in tasks that involve motor imagery [71].

Graph Neural Networks (GNNs)

The GraphIS method, combining classical signal processing, graph processing, and graph learning features with a two-stage SVM (RBF kernel), achieved a 50.1% accuracy in decoding 'Rock, Paper, Scissors, Rest' imagined speech EEG, surpassing chance (25%) and the CSP baseline (47.1%), highlighting the benefit of feature fusion [49].

EEGNet-SPDNet

The researchers evaluated the EEGNet–SPDNet architecture by combining EEGNet's temporal features with SPDNet's covariance representations for two imaginable speech EEG tasks. BCI 2020 gave it 66.93% and Kara One gave it 24.79%. Results suggest that Riemannian geometry features may outperform Euclidean methods in EEG classification [50].

Hybrid and Ensemble Approaches

Many studies examine classifier combinations or layered frameworks, like using CSP or wavelet features with SVM or RF, or LDA followed by ensemble voting with CNNs or SVMs over deep features. These methods improve robustness but add complexity [19, 42, 46]. Evaluation of all classification models used subject-dependent or subject-independent protocols, with methods like k-fold cross-validation and leave-one-subject-out (LOSO) to assess generalizability [72]. Additional details, including stimulus types, class labels, and accuracy ranges, are presented in Table 1.

Table 1: Summary of classification algorithms and reported accuracies

Ref.	Modality	Number of Subjects	Task type	Signal Acquisition Setup	Stimuli / Classes	Fusion strategy	Method	Accuracy
[19]	EEG	23 subjects	binary yes/no	60-channel	Yes/No (Occupation-based Qs)	early	CSP + SVM	70–99.9%
[21]	fNIRS	12 subjects	Three Classes(Yes,No,Rest)	44-channel	Yes/No/Rest	-	mean value of [HbO] + RLDA	64.1% ± 20.6%
[47]	EEG	4 subjects	multiclass	128-channel	/ba/, /ku/	early	power spectral densities + Adaptive Filter	75%
[32]	EEG/ECoG	2 subjects	phoneme	Two-channel/48-channel	AA, IY, UW	early	root-mean-squared (RMS) + Kalman Filter	~71%
[37]	EEG	6 subjects	word-level	19-channel	8 English Words	early	_	Qualitative
[41]	EEG	27 subjects	multiclass	14-channel	five Spanish words (“arriba,” “abajo,” “izquierda,” “derecha,” “seleccionar”)	early	DWT + Random Forest	68.18–70.33%
[34]	EEG	Not mentioned	phoneme	8-channel	/a/, /u/, rest	-	EEGNet-inspired CNN combined with RNN (LSTM) and CTC	-
[38]	EEG	2 subjects	word-level	14-channel	CVC Words	early	PSD and Relative Power, PLV	Qualitative
[42]	EEG	15 subjects	Binary classification of imagined *word-pairs*	six-channel	6 Spanish Words	-	Deep CNN and Shallow CNN (proposed); baseline: rLDA with FBCSP. No decision-level fusion	CNN-Deep: 62.37% CNN-Shallow: 60.88% rLDA (FBCSP): 57.80%
[48]	EEG	14 subjects	Two Classes	64-channel	7 syllables, four words (Kara One)	early	CNN+TCNN+DAE+XGBoost	83.42% (phonological features, binary tasks); 53.36% (11 tokens); 28.08% (raw covariance)
[39]	EEG	27 subjects	word-level	14-channel	five Spanish words (“arriba,” “abajo,” “izquierda,” “derecha,” “seleccionar”)	early	Bag of Features (BoF) + Naive Bayes	65.5%
[40]	EEG	10 subjects	word-level	58-channel	/Ba/, /Ku/, /He/, /Li/	early	EEGNet + Attention	57%
[33]	EEG	15 subjects	phoneme	18-channel/14-channel	/a/, /e/, /i/, /o/, /u/	decision-level	CNNeeg1-1	BD1 = 65.62% BD2 = 85.66%
[46]	EEG	7 subjects	multiclass	64-channel	12 English Words + Rest	decision-level	CSP + RF	34.2%
[44]	EEG	16 subjects	multiclass	64-channel	words/phrases (ambulance, clock, hello, help me, light, pain, stop, thank you, toilet, TV, water, and yes)	-	PLV	Qualitative
[35]	EEG	5 subjects	multiclass	19-channel	six Persian words for this experiment: { خیر، بله، راست، چپ، پایین، بالا }. They are pronounced as {(bʌlʌ), (pʌyɪn), (chæp), (rʌst), (bæle), (kheɪr)}	decision-level	FFTs + SVM	**97%** (binary tasks) **88.2%** (7-class: six words + rest)
[49]	EEG	15 subjects	multiclass	64-channels	Rock, Paper, Scissors, Rest	decision-level	GraphIS (Fusion of classical + GSP/GL features; 2-stage SVM)	50.1% (significant improvement over CSP = 47.1%)
[18]	EEG	14 and 8 subjects	phoneme	64-channel	KARA ONE/FUM	early	Hilbert transform + SVM	81.1%
[12]	EEG + fNIRS	12 subjects	Two classes	**EEG:** 64-channel **fNIRS:** 52-channel	Imagined left- versus right-hand movement	early	Eeg: Phase-Space Reconstruction (PSR),CSP Fnirs: Moving Average,Hurst Exponent + SVM	81.2%
[22]	EEG + fNIRS	18 subjects	Three classes	**EEG:** 11-channel **fNIRS:** 16-channel	MI, MA, IS	decision-level	CSP + sLDA	≈78–79%
[23]	EEG + fNIRS	19 subjects	word-level	**EEG:** 64-channel **fNIRS:** 16-channels	**Action-Text (AT):** *squeeze, jump, kiss, smile* **Action-Image (AI):** *squeeze, jump, kiss, smile* **Action-Audio (AA):** *squeeze, jump, kiss, smile* **Combinations-Text (CT):** *red ball, green hat, red green, ball hat*	late	CNN	53%

					Combinations-Image (CI): *red ball, green hat, red green, ball hat* **Combinations-Audio (CA):** *red ball, green hat, red green, ball hat*			
[24]	EEG + fNIRS	11 subjects	Three Classes (Yes,No,Rest)	**EEG:** 32-channel **fNIRS:** 44-channels	Yes, No, Rest	decision-level	DWT (Symlet-10, 6 levels); RMS (EEG); Mean [HbO] (fNIRS) + RLDA	%70.45 ± 19.19
[25]	EEG + fNIRS	**26 subjects**	Two Classes	**EEG:** 30-channel **fNIRS:**72-channel	**Word Generation – WG/Baseline – BL(Rest)**	dual-stream	EF-Net	Subject-dependent: fNIRS only: 99.69% fNIRS + EEG: 99.36% Subject semi-dependent: fNIRS only: 98.50% fNIRS + EEG: 98.31% Subject independent: fNIRS only: 63.80% fNIRS + EEG: 65.05%
[36]	EEG	10 subjects	word-level	16-channel with 2-reference	10 Words(Up, Down, Left, Right,…)	early	Morlet Continuous wavelet transform + AlexNet	90.3%
[17]	EEG	4 subjects	multiclass	8-channel	Up, Down, Left, Right	early	Wavelet scattering transform + LSTM-RNN	92.5%
[50]	EEG	11 and 15 subjects	multiclass	64-channel	Kara One and BCI 2020	dual-stream	EEGNet - SPDNet	2020 BCI Comp = 66.93% / Kara One = 24.79%

Classification Performance

The accuracy of classification, as documented in various studies, was influenced by variables such as the complexity of the imagined speech tasks, the number of stimulus categories, and the data modalities used. The following is a synthesis of performance trends categorized accordingly:

- **Binary EEG Tasks:** The tasks above typically yielded elevated accuracy rates, ranging from 75% to exceeding 99%, especially when employing feature extraction methodologies such as CSP in conjunction with SVM classifiers. The effectiveness of this performance was further enhanced by the implementation of simplified task designs and a reduction in the number of classes [29].
- **Multi-Class EEG Tasks:** Accuracy seemed to decrease as the complexity of the tasks increased. Depending on the number of imagined words or phrases and the specific combination of features and classifiers used, the documented performance could range from 34.2% to 85% [32, 33, 36-43, 45, 46, 73].
- **fNIRS-Only Investigations:** The accuracy of classification was typically moderate in studies that relied solely on fNIRS data. For instance, an accuracy of 64.1% ±20.6% was achieved in one study using the mean value of [HbO] as a feature and the rLDA algorithm in a binary (Yes/No) classification task [21].
- **Hybrid EEG–fNIRS Systems:** Even with the same experimental settings, multimodal methods consistently surpass single-modality systems. One study showed that integrating EEG and fNIRS features for imagined speech achieved a peak classification accuracy of 53%, notably higher than EEG alone (about 30–37%) or fNIRS alone (around 28–31%) [23]. While hybrid methods have the ability to improve accuracy from 65–72% (single modality) to about 87%, these numbers are obtained from different tasks with different class numbers and difficulty levels. They are not directly comparable across studies and should be viewed as a general indicator of the potential of multimodal systems, not as definitive proof of superiority.

The scientific community increasingly agrees that hybrid systems decode better by combining EEG's temporal resolution with fNIRS's hemodynamic sensitivity (Fig. 2).

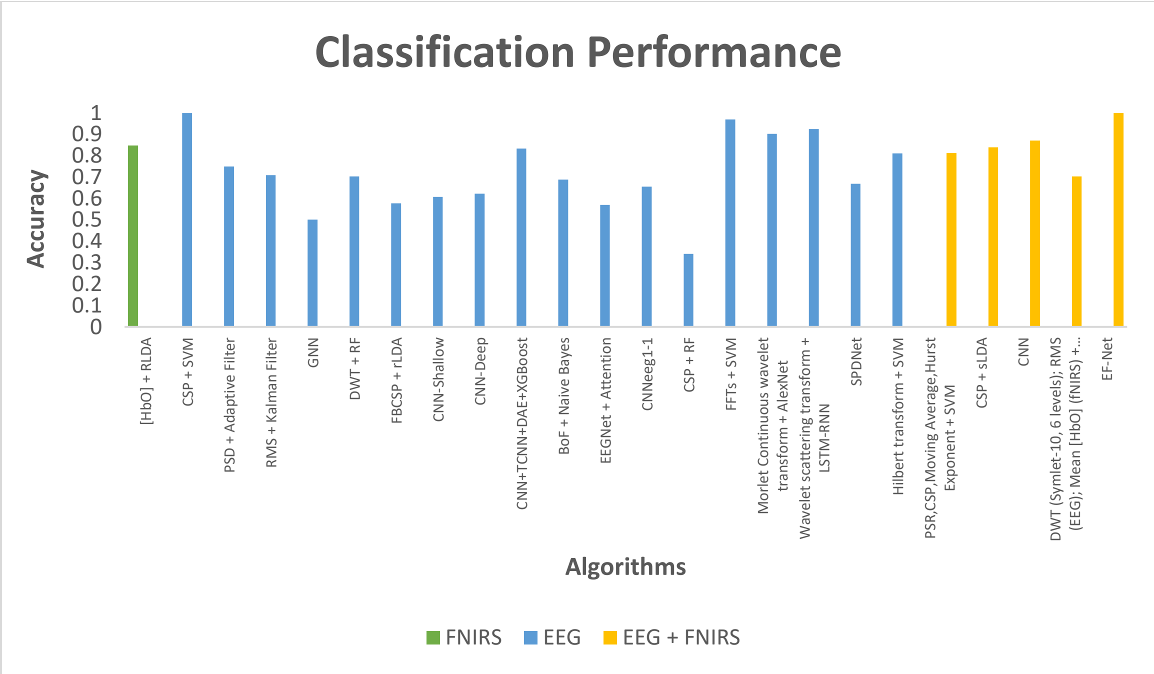

Figure 2. Classification accuracy by modality and classifier.

EEG–fNIRS Hybrid Data

Two studies examined hybrid EEG and fNIRS for decoding imagined speech. EEG electrodes targeted motor and speech-related areas, while fNIRS optodes focused on the prefrontal cortex connected to language functions. Using feature fusion of CSP EEG features and fNIRS HbO means, they classified with SVM and Structured LDA. Both studies showed accuracy increases of up to 20%, demonstrating the potential of multimodal systems for imagined speech decoding [12, 23, 52].

Experimental Configurations

Although research focus and design varied, many studies shared similar experimental setups, including the main aspects listed below:

- **Participant Demographics**

Most studies involved healthy, right-handed participants, often aged 20-30, with gender data provided for demographics.

- **Signal Acquisition**
    - EEG systems vary from consumer headsets to research-grade devices with 14-128 electrodes, offering different spatial resolutions [74, 75].
    - Most fNIRS setups use 10-16 optodes, with short-separation channels to improve spatial specificity and reduce interference [74].
- **Session Structure**

Experimental sessions involved multiple blocks with randomized imagined speech trials. To avoid fatigue, trials lasted 2 to 10 seconds, followed by inter-trial intervals.

- **Data Preprocessing**

Most studies use standard preprocessing like band-pass filtering (e.g., 0.5–45 Hz for EEG), ICA for artifact removal in EEG, and baseline correction for fNIRS signals.

- **Validation Methods**

Studies used cross-validation methods like 10-fold, leave-one-out, or trial-wise validation to assess model generalizability, based on dataset size and design. This trend toward standardizing protocols is key for comparability and reproducibility in imagined speech decoding.

Discussion

The growing research on decoding imagined speech with EEG and fNIRS, and their combined use, indicates major progress in BCI technology [23, 76]. This review summarizes advances in methods to acquire cerebral signals, extract patterns, and classify them to understand silent speech intentions, all within carefully controlled laboratory settings [6, 77].

Model performance and key insights

EEG remains the main tool for decoding imagined speech due to its high temporal resolution and ease of use [78]. EEG classification accuracy varies from 60% to over 90%, influenced by task complexity, features, and algorithms [6, 20, 76]. Although fNIRS has been less used, it shows good performance (65-85%) in simpler binary tasks [21, 26, 79]. The integration of EEG and fNIRS into a hybrid system often improves performance, sometimes by as much as 25% over using either method alone [23, 24]. Each modality offers advantages: EEG's rapid response and fNIRS's spatial resolution work together for a better understanding of imagined speech activity [80, 81].

Strengths and Weaknesses of EEG and fNIRS

This review's limitations include reliance on English publications and specific search terms, which may potentially omit relevant research. Larger and standardized datasets are needed for better comparisons, as highlighted by the few hybrid EEG–fNIRS studies. In order to measure performance accurately, future reviews should include systematic methods and meta-analyses. EEG has excellent temporal resolution, recording changes in cerebral activity in milliseconds, making it ideal for capturing rapid neural responses in inner speech. [82]. Model generalizability was assessed through cross-validation methods such as 10-fold, leave-one-out, or trial-wise validation, based on dataset size and design. The trend towards standardizing protocols is crucial for achieving comparability and reproducibility in imagined speech decoding. [83] and fNIRS is slower, unable to measure blood flow changes that take time to become evident due to real-time limitations [83, 84]. Furthermore, the flow of dermis can cause noise during measurements. Although the reported advancements are significant, there are significant limitations in many of the studies reviewed. For instance, [24]only achieved around 70% ternary accuracy in classifying imagined speech using a hybrid EEG–fNIRS approach, and performance varied significantly from person to person. similarity, [23]reported accuracies of about 34% for imagined speech using deep learning, showing that decoding speech without invasive methods remains highly uncertain. Recent CNN-based models, such as EF-Net, were unable to surpass 65% F1 in subject-independent settings, emphasizing that many models only perform well in controlled lab conditions [25]. The gap between promising results and practical clinical applications is highlighted by these differences.

Classical vs. Deep Learning methods.

SVMs, LDA, and RF are still widely used in machine learning algorithms. [29, 35, 51, 59, 60] [41, 46, 63-66]. These methods are simple, transparent, and effective for small datasets, especially with features like PSD, CSP, or wavelet transforms [6, 29, 46, 47, 49, 52, 55]. Recent advancements favor deep learning, especially CNNs and RNNs, which can independently identify data patterns, removing the need for manual feature engineering [36, 43, 69, 71]. Despite their potential to improve accuracy, these advanced models face challenges. They need large training data, which is often limited, and act as 'black boxes', making their outputs hard to interpret—an issue in clinical settings. Progress in decoding imagined speech has been made, but challenges remain. Most studies use small vocabularies and offline tests that don't mimic real interactions. Small sample sizes, variability, and limited data hinder generalization, with little use of transfer learning. Models also lack neurophysiological interpretability, reducing trust and usefulness. Additionally, their robustness to artifacts like motion, visual, and environmental noise in real-time remains problematic. In regard to challenges for real-world BCI deployment, transitioning from controlled experiments to real-world imagined speech BCIs faces major obstacles: weak, variable signals; decoding delays with fNIRS; poor outside performance due to lack of personalization; small datasets hinder robust model training. Solutions include developing better algorithms, standard methods, and richer, multimodal datasets. Future brain activity research should focus on multimodal systems (like EEG and fNIRS) for better understanding, creating larger and shared datasets, using hybrid deep learning models (CNN-RNNs, Transformers), developing personalized, adaptive systems, and employing interpretable AI for trust and clinical use. Recent studies that combined EEG and fNIRS [12, 22-25] further demonstrate the potential of integrating these technologies. These studies show that deep multimodal architectures consistently outperform unimodal EEG or fNIRS systems, achieving higher accuracy and robustness across subjects. The inclusion of their information emphasizes the importance of hybrid frameworks for future BCI development, particularly in clinical and assistive communication settings.

Conclusion

Research increasingly confirms that EEG and fNIRS can decode imagined speech, confirming this field's potential to enhance brain–computer interface (BCI) technologies. While traditional machine learning offers strong benchmarks, advancements in deep learning and multimodal integration have opened new opportunities, leading to a significant boost in decoding accuracy. However, there is still a significant obstacle to overcome. Studies show that classification accuracy can reach nearly 99% in subject-dependent setups, but drops to around 65% in subject-independent contexts. The gap highlights the challenge of creating models that perform well for new users and real-world scenarios, where robustness and adaptability are crucial. Future efforts should be concentrated on the creation of hybrid systems, the expansion of shared datasets, the development of adaptive and interpretable deep learning models, and the establishment of standardized evaluation methods to address this issue. To transform imagined speech BCIs from controlled laboratory experiments into practical clinical and assistive devices that provide reliable, scalable communication for users overcoming these challenges is crucial.

Disclosure of interests

Conflict of Interest Statement

The authors declared no conflict of interest

Fundings

The authors declared no financial support

Author Contribusion

All authors were involved in the drafting, which was conceived by M.M. and B.H., and data was collected by A.M., Y.A., and M.M.

References

[1] D. J. McFarland and J. R. Wolpaw, "Brain-computer interfaces for communication and control," *Communications of the ACM,* vol. 54, no. 5, pp. 60–66, 2011.

[2] J. N. Mak and J. R. Wolpaw, "Clinical applications of brain-computer interfaces: current state and future prospects," *IEEE reviews in biomedical engineering,* vol. 2, pp. 187–199, 2009.

[3] F. Pichiorri and D. Mattia, "Brain-computer interfaces in neurologic rehabilitation practice," *Handbook of clinical neurology,* vol. 168, pp. 101–116, 2020.

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

[5] T. Proix *et al.*, "Imagined speech can be decoded from low-and cross-frequency intracranial EEG features," *Nature communications,* vol. 13, no. 1, p. 48, 2022.

[6] D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, "A state-of-the-art review of EEG-based imagined speech decoding," *Frontiers in human neuroscience,* vol. 16, p. 867281, 2022.

[7] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *Neuroimage,* vol. 63, no. 2, pp. 921–935, 2012.

[8] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, *Experience sampling method: Measuring the quality of everyday life*. Sage, 2007.

[9] R. Li, D. Yang, F. Fang, K.-S. Hong, A. L. Reiss, and Y. Zhang, "Concurrent fNIRS and EEG for brain function investigation: a systematic, methodology-focused review," *Sensors,* vol. 22, no. 15, p. 5865, 2022.

[10] M. Rupawala, H. Dehghani, S. J. Lucas, P. Tino, and D. Cruse, "Shining a light on awareness: a review of functional near-infrared spectroscopy for prolonged disorders of consciousness," *Frontiers in neurology,* vol. 9, p. 350, 2018.

[11] J. Wang, Y.-H. Chen, J. Yang, and M. Sawan, "Intelligent classification technique of hand motor imagery using EEG Beta rebound follow-up pattern," *Biosensors,* vol. 12, no. 6, p. 384, 2022.

[12] S. Ge *et al.*, "A brain-computer interface based on a few-channel EEG-fNIRS bimodal system," *IEEE Access,* vol. 5, pp. 208–218, 2017.

[13] S. Ahn and S. C. Jun, "Multi-modal integration of EEG-fNIRS for brain-computer interfaces–current limitations and future directions," *Frontiers in human neuroscience,* vol. 11, p. 503, 2017.

[14] R. J. Deligani, S. B. Borgheai, J. McLinden, and Y. Shahriari, "Multimodal fusion of EEG-fNIRS: a mutual information-based hybrid classification framework," *Biomedical optics express,* vol. 12, no. 3, pp. 1635–1650, 2021.

[15] M. Guo, L. Feng, X. Chen, M. Li, and G. Xu, "A novel strategy for differentiating motor imagination brain-computer interface tasks by fusing EEG and functional near-infrared spectroscopy signals," *Biomedical Signal Processing and Control,* vol. 95, p. 106448, 2024.

[16] J. Chen, Y. Xia, X. Zhou, R. Vidal, A. Thomas, and R. Loureiro, "fNIRS-EEG BCIs for motor rehabilitation: a review. Bioengineering 2023; 10: 1393," ed.

[17] M. M. Abdulghani, W. L. Walters, and K. H. Abed, "Imagined speech classification using EEG and deep learning," *Bioengineering,* vol. 10, no. 6, p. 649, 2023.

[18] 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.

[19] J. W. Choi and K. H. Kim, "Covert Intention to Answer “Yes” or “No” Can Be Decoded from Single-Trial Electroencephalograms (EEGs)," *Computational Intelligence and Neuroscience,* vol. 2019, no. 1, p. 4259369, 2019.

[20] S. Alzahrani, H. Banjar, and R. Mirza, "Systematic Review of EEG-Based Imagined Speech Classification Methods," *Sensors,* vol. 24, no. 24, p. 8168, 2024.

[21] A. R. Sereshkeh, R. Yousefi, A. T. Wong, and T. Chau, "Online classification of imagined speech using functional near-infrared spectroscopy signals," *Journal of neural engineering,* vol. 16, no. 1, p. 016005, 2018.

[22] J. Kwon, J. Shin, and C.-H. Im, "Toward a compact hybrid brain-computer interface (BCI): Performance evaluation of multi-class hybrid EEG-fNIRS BCIs with limited number of channels," *PloS one,* vol. 15, no. 3, p. e0230491, 2020.

[23] C. Cooney, R. Folli, and D. Coyle, "A bimodal deep learning architecture for EEG-fNIRS decoding of overt and imagined speech," *IEEE Transactions on Biomedical Engineering,* vol. 69, no. 6, pp. 1983–1994, 2021.

[24] A. Rezazadeh Sereshkeh, R. Yousefi, A. T. Wong, F. Rudzicz, and T. Chau, "Development of a ternary hybrid fNIRS-EEG brain–computer interface based on imagined speech," *Brain-Computer Interfaces,* vol. 6, no. 4, pp. 128–140, 2019.

[25] A. Arif, Y. Wang, R. Yin, X. Zhang, and A. Helmy, "EF-Net: Mental State Recognition by Analyzing Multimodal EEG-fNIRS via CNN," (in eng), *Sensors (Basel),* vol. 24, no. 6, Mar 15 2024, doi: 10.3390/s24061889.

[26] A. Abdalmalak *et al.*, "Assessing time-resolved fNIRS for brain-computer interface applications of mental communication," *Frontiers in Neuroscience,* vol. 14, p. 105, 2020.

[27] N. J. Bourguignon, S. L. Bue, C. Guerrero-Mosquera, and G. Borragán, "Bimodal EEG-fNIRS in neuroergonomics. Current evidence and prospects for future research," *Frontiers in Neuroergonomics,* vol. 3, p. 934234, 2022.

[28] M. U. Ali, K. S. Kim, K. D. Kallu, A. Zafar, and S. W. Lee, "OptEF-BCI: an optimization-based hybrid EEG and fNIRS–brain computer interface," *Bioengineering,* vol. 10, no. 5, p. 608, 2023.

[29] J. W. Choi and K. H. Kim, "Covert Intention to Answer “Yes” or “No” Can Be Decoded from Single-Trial Electroencephalograms (EEGs)," *Computational Intelligence and Neuroscience,* vol. 2019, pp. 1–11, 07/10 2019, doi: 10.1155/2019/4259369.

[30] S. Biswas and R. Sinha, "Wavelet filterbank-based EEG rhythm-specific spatial features for covert speech classification," *IET Signal Processing,* vol. 16, no. 1, pp. 92–105, 2022.

[31] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch II, "Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG," *Neurocomputing,* vol. 30, no. 1-4, pp. 201–218, 2000.

[32] F. H. Guenther and J. S. Brumberg, "Brain-machine interfaces for real-time speech synthesis," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011: IEEE, pp. 5360–5363.

[33] L. C. Sarmiento, S. Villamizar, O. López, A. C. Collazos, J. Sarmiento, and J. B. Rodríguez, "Recognition of EEG Signals from Imagined Vowels Using Deep Learning Methods," *Sensors,* vol. 21, no. 19, p. 6503, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/19/6503.

[34] K. Wang, X. Wang, and G. Li, "Simulation experiment of bci based on imagined speech eeg decoding," *arXiv preprint arXiv:1705.07771,* 2017.

[35] M. Asghari Bejestani, G. R. Mohammad Khani, V. Nafisi, and F. Darakeh, "EEG-Based Multiword Imagined Speech Classification for Persian Words," *BioMed Research International,* vol. 2022, no. 1, p. 8333084, 2022.

[36] B. Chinta and M. Moorthi, "Brain Computer Interface-EEG based Imagined Word Prediction Using Convolutional Neural Network Visual Stimuli for Speech Disability," 2022.

[37] S. Chengaiyan and K. Anandhan, "Analysis of speech imagery using functional and effective EEG based brain connectivity parameters," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI),* vol. 9, no. 4, pp. 33–48, 2015.

[38] S. Chengaiyan, D. Balathayil, K. Anandan, and C. B. Thomas, "Effect of power and phase synchronization in multi-trial speech imagery," in *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications*: IGI Global Scientific Publishing, 2020, pp. 1654–1673.

[39] J. S. García-Salinas, L. Villaseñor-Pineda, C. A. Reyes-García, and A. A. Torres-García, "Transfer learning in imagined speech EEG-based BCIs," *Biomedical Signal Processing and Control,* vol. 50, pp. 151–157, 2019.

[40] D.-H. Lee, S.-J. Kim, and K.-W. Lee, "Decoding high–level imagined speech using attention–based deep neural networks," in *2022 10th International Winter Conference on Brain-Computer Interface (BCI)*, 2022: IEEE, pp. 1–4.

[41] A. A. Torres-García, C. A. Reyes-García, L. Villaseñor-Pineda, and G. García-Aguilar, "Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification," *Expert Systems with Applications,* vol. 59, pp. 1–12, 2016.

[42] C. Cooney, A. Korik, F. Raffaella, and D. Coyle, "Classification of imagined spoken word-pairs using convolutional neural networks," in *The 8th Graz BCI Conference, 2019*, 2019: Verlag der Technischen Universitat Graz, pp. 338–343.

[43] P. Saha, S. Fels, and M. Abdul-Mageed, "Deep learning the EEG manifold for phonological categorization from active thoughts," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019: IEEE, pp. 2762–2766.

[44] S.-H. Lee, M. Lee, and S.-W. Lee, "Functional connectivity of imagined speech and visual imagery based on spectral dynamics," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 2021: IEEE, pp. 1–6.

[45] M. R. Asghari Bejestani, G. R. Mohammad Khani, V. R. Nafisi, and F. Darakeh, "EEG-Based Multiword Imagined Speech Classification for Persian Words," (in eng), *Biomed Res Int,* vol. 2022, p. 8333084, 2022, doi: 10.1155/2022/8333084.

[46] S.-H. Lee, M. Lee, J.-H. Jeong, and S.-W. Lee, "Towards an EEG-based intuitive BCI communication system using imagined speech and visual imagery," in *2019 IEEE international conference on systems, man and cybernetics (SMC)*, 2019: IEEE, pp. 4409–4414.

[47] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward EEG sensing of imagined speech," in *Human-Computer Interaction. New Trends: 13th International Conference, HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings, Part I 13*, 2009: Springer, pp. 40–48.

[48] 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.

[49] A. Einizade, M. Mozafari, S. Jalilpour, S. Bagheri, and S. H. Sardouie, "Neural decoding of imagined speech from EEG signals using the fusion of graph signal processing and graph learning techniques," *Neuroscience Informatics,* vol. 2, no. 3, p. 100091, 2022.

[50] 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 *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024: IEEE, pp. 1531–1535.

[51] H. Khan, N. Naseer, A. Yazidi, P. K. Eide, H. W. Hassan, and P. Mirtaheri, "Analysis of human gait using hybrid EEG-fNIRS-based BCI system: a review," *Frontiers in Human Neuroscience,* vol. 14, p. 613254, 2021.

[52] J. Shin, J. Kwon, and C.-H. Im, "A ternary hybrid EEG-NIRS brain-computer interface for the classification of brain activation patterns during mental arithmetic, motor imagery, and idle state," *Frontiers in neuroinformatics,* vol. 12, p. 5, 2018.

[53] M. Tavakolan, X. Yong, X. Zhang, and C. Menon, "Classification scheme for arm motor imagery," *Journal of medical and biological engineering,* vol. 36, pp. 12–21, 2016.

[54] L. E. Ismail and W. Karwowski, "Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis," *Plos one,* vol. 15, no. 12, p. e0242857, 2020.

[55] A. Einizade, M. Mozafari, S. Jalilpour, S. Bagheri, and S. Hajipour Sardouie, "Neural decoding of imagined speech from EEG signals using the fusion of graph signal processing and graph learning techniques," *Neuroscience Informatics,* vol. 2, no. 3, p. 100091, 2022/09/01/ 2022, doi: https://doi.org/10.1016/j.neuri.2022.100091.

[56] Y. Mohammadi, C. Graversen, J. Østergaard, O. K. Andersen, and T. Reichenbach, "Phase-locking of neural activity to the envelope of speech in the delta frequency band reflects differences between word lists and sentences," *Journal of Cognitive Neuroscience,* vol. 35, no. 8, pp. 1301–1311, 2023.

[57] C. T. Albinet, K. Mandrick, P. L. Bernard, S. Perrey, and H. Blain, "Improved cerebral oxygenation response and executive performance as a function of cardiorespiratory fitness in older women: a fNIRS study," *Frontiers in aging neuroscience,* vol. 6, p. 272, 2014.

[58] S. Fazli *et al.*, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage,* vol. 59, no. 1, pp. 519–529, 2012.

[59] Z. G. Li and G. Z. Liu, "Research of" Yes" and" No" Responses by Auditory Stimuli in Human EEG," *Applied Mechanics and Materials,* vol. 310, p. 660, 2013.

[60] P. Batres-Mendoza *et al.*, "Improving EEG-Based Motor Imagery Classification for Real-Time Applications Using the QSA Method," *Computational intelligence and neuroscience,* vol. 2017, no. 1, p. 9817305, 2017.

[61] N. J. Bourguignon, et al. , ""Bimodal EEG-fNIRS in neuroergonomics. Current evidence and prospects for future research."," *Frontiers in Neuroergonomics 3 (2022): 934234.*

[62] N. J. Bourguignon, et al. , ""Bimodal EEG-fNIRS in neuroergonomics. Current evidence and prospects for future research." " *Frontiers in Neuroergonomics 3 (2022): 934234.*

[63] H. Park, Y. Cho, T. Lee, and H.-H. Kim, "Enhancing Word Recognition in Imagined Speech Using Non-Invasive EEG for Improved BCI Applications," presented at the 2024 International Conference on Cyberworlds (CW), 2024/10/29, 2024. [Online]. Available: http://dx.doi.org/10.1109/cw64301.2024.00074.

[64] Y. Roy, et al., ""Deep learning-based electroencephalography analysis: a systematic review." " *Journal of neural engineering 16.5 (2019): 051001.*

[65] F. Lotte, et al. , ""A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update." " *Journal of neural engineering 15.3 (2018): 031005.*

[66] Y. Xin, et al., " "Semi-supervised regularized discriminant analysis for EEG-based BCI system." " *Intelligent Data Engineering and Automated Learning–IDEAL 2017: 18th International Conference, Guilin, China, October 30–November 1, 2017, Proceedings 18. Springer International Publishing, 2017.*

[67] F. H. Guenther *et al.*, "A wireless brain-machine interface for real-time speech synthesis," *PloS one,* vol. 4, no. 12, p. e8218, 2009.

[68] A. Sinai *et al.*, "Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming," *Brain,* vol. 128, no. 7, pp. 1556–1570, 2005.

[69] F. M. Alotaibi, "An AI-inspired spatio-temporal neural network for EEG-based emotional status," *Sensors,* vol. 23, no. 1, p. 498, 2023.

[70] G. Cisotto, A. Zanga, J. Chlebus, I. Zoppis, S. Manzoni, and U. Markowska-Kaczmar, "Comparison of attention-based deep learning models for eeg classification," *arXiv preprint arXiv:2012.01074,* 2020.

[71] Y.-E. Lee and S.-H. Lee, "EEG-transformer: Self-attention from transformer architecture for decoding EEG of imagined speech," in *2022 10th International winter conference on brain-computer interface (BCI)*, 2022: IEEE, pp. 1–4.

[72] E. M. Dos Santos, R. San-Martin, and F. J. Fraga, "Comparison of subject-independent and subject-specific EEG-based BCI using LDA and SVM classifiers," *Medical & biological engineering & computing,* vol. 61, no. 3, pp. 835–845, 2023.

[73] S.-H. Lee, Minji Lee, and Seong-Whan Lee. , ""Functional connectivity of imagined speech and visual imagery based on spectral dynamics."," *9th International Winter Conference on Brain-Computer Interface (BCI). IEEE, 2021.,* 2021

[74] K. Flanagan and M. J. Saikia, "Consumer-grade electroencephalogram and functional near-infrared spectroscopy neurofeedback technologies for mental health and wellbeing," *Sensors,* vol. 23, no. 20, p. 8482, 2023.

[75] S. Saha, Y. I. Nesterets, M. Tahtali, and T. E. Gureyev, "Evaluation of spatial resolution and noise sensitivity of sLORETA method for EEG source localization using low-density headsets," *Biomedical physics & engineering express,* vol. 1, no. 4, p. 045206, 2015.

[76] M. Abdulghani, W. Walters, and K. Abed, "Imagined speech classification using EEG and deep learning. Bioengineering 10 (6), 649 (2023)," ed.

[77] N. Rahman, D. M. Khan, K. Masroor, M. Arshad, A. Rafiq, and S. M. Fahim, "Advances in brain-computer interface for decoding speech imagery from EEG signals: a systematic review," *Cognitive Neurodynamics,* pp. 1–19, 2024.

[78] D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, "A State-of-the-Art Review of EEG-Based Imagined Speech Decoding," (in eng), *Front Hum Neurosci,* vol. 16, p. 867281, 2022, doi: 10.3389/fnhum.2022.867281.

[79] C. Herff, F. Putze, D. Heger, C. Guan, and T. Schultz, "Speaking mode recognition from functional near infrared spectroscopy," in *2012 annual international conference of the IEEE engineering in medicine and biology society*, 2012: IEEE, pp. 1715–1718.

[80] S. Ghosh *et al.*, "Review of Multimodal Data Acquisition Approaches for Brain–Computer Interfaces," *BioMed,* vol. 4, no. 4, pp. 548–587, 2024.

[81] R. Li, T. Potter, W. Huang, and Y. Zhang, "Enhancing performance of a hybrid EEG-fNIRS system using channel selection and early temporal features," *Frontiers in human neuroscience,* vol. 11, p. 462, 2017.

[82] J. Parvizi and S. Kastner, "Human intracranial EEG: promises and limitations," *Nature neuroscience,* vol. 21, no. 4, p. 474, 2018.

[83] S. Cutini and S. Brigadoi, "Unleashing the future potential of functional near-infrared spectroscopy in brain sciences," *Journal of Neuroscience Methods,* vol. 232, pp. 152–156, 2014.

[84] M. M. Arredondo, "Shining a light on cultural neuroscience: Recommendations on the use of fNIRS to study how sociocultural contexts shape the brain," *Cultural diversity & ethnic minority psychology,* vol. 29, no. 1, p. 106, 2023.