Review Paper

A Comprehensive Review of Imagined Speech Decoding in Brain-computer Interfaces: Utilizing Electroencephalography and Functional Near-infrared Spectroscopy

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, Ma.C., Islamic Azad University, Mashhad, Iran.
3. Biomedical Engineering Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.
4. Department of Microbiology, Ma.C., Islamic Azad University, Mashhad, Iran.
5. Neuroscience Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.

Use your device to scan and read the article online

Citation Motaqi, M., Moallemi, M., Mirani, A., Aboutorabi, Y., & Hatef, B. (2026). A Comprehensive Review of Imagined Speech Decoding in Brain-computer Interfaces: Utilizing Electroencephalography and Functional Near-infrared Spectroscopy. *Basic and Clinical Neuroscience, 17*(2), 175-190. http://dx.doi.org/10.32598/bcn.2026.8254.1

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

Article info:
Received: 07 Aug 2025
First Revision: 05 Sep 2025
Accepted: 10 Jan 2026
Available Online: 01 Mar 2026

Keywords:
Brain-computer interface (BCI), Imagined speech decoding, Electroencephalography (EEG), Functional near-infrared spectroscopy (fNIRS), Machine learning, Multimodal fusion

ABSTRACT

Introduction: The use of brain–computer interfaces (BCIs) to decode imagined speech has significant clinical and assistive potential.

Methods: Twenty-four studies investigated covert speech decoding between 2009 and 2025 using electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), or hybrid EEG–fNIRS systems.

Results: Early research (2009–2012) primarily focused on analyzing phonemes and syllables with EEG, achieving accuracy rates around 75%. From 2013 to 2017, convolutional neural network (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 studies, reported accuracy ranged from approximately 53% with CNN-based decoding to 70.45±19.19% with an RLDA-based decision-level fusion approach, although direct comparisons are limited by differences in tasks, fusion strategies, and validation settings.

Conclusion: Although deep learning and multimodal systems have potential for enhancing imagined speech decoding, there are still major challenges related to generalization, variability, and robustness.

*** Corresponding Author:**
Boshra Hatef, Associate Professor.
Address: *Neuroscience Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran.*
Tel: *+98 (21) 88040060*
E-mail: *boshrahatef@bmsu.ac.ir*

Highlights

- EEG was the dominant modality for imagined-speech decoding.
- Binary EEG paradigms achieved accuracies approaching 99%.
- Multiclass accuracy declined as vocabulary size and task complexity increased.
- Deep-learning models achieved 90.3% with AlexNet and 92.5% with LSTM-RNNs.
- Hybrid EEG–fNIRS performance varied across tasks, fusion strategies, and validation settings.

Plain Language Summary

People who cannot speak because of severe neurological or motor conditions may still be able to imagine words and sentences. Brain–computer interfaces aim to detect this silent speech and convert it into commands or messages. This review examined studies published between 2009 and 2025 that used two non-invasive methods to record brain activity. Electroencephalography, or EEG, measures rapid electrical activity through sensors placed on the scalp. Functional near-infrared spectroscopy, or fNIRS, measures slower changes in blood oxygenation in the outer regions of the brain. A small number of studies combined both methods. EEG was the most frequently used approach. Studies asking participants to choose between two options, such as "yes" and "no," generally reported better results than studies involving multiple words or phrases. Performance became less reliable as vocabulary size and task complexity increased. Newer machine-learning methods, including deep neural networks, achieved high accuracy in some experiments, but results varied considerably across datasets and participants. Systems often performed well when trained and tested on the same individuals, but their performance generally declined when applied to new individuals. Combining EEG and fNIRS may provide complementary information. However, only a small number of imagined-speech studies have evaluated this approach, and their findings remain inconsistent. These results suggest that imagined-speech technology is promising but is not yet ready for routine clinical use. Larger shared datasets, standardized evaluation methods, and systems that can adapt to individual users are needed. Progress in these areas could eventually support more natural communication for people who have lost the ability to speak or move.

Introduction

The brain-computer interface (BCI) enables direct interaction between the human cerebral cortex and external devices (Mak & Wolpaw, 2009; McFarland & Wolpaw, 2011). 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 (Panachakel & Ramakrishnan, 2021; Pichiorri & Mattia, 2020). 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 (Diego Lopez-Bernal et al., 2022; Proix et al., 2022). Two non-invasive techniques are commonly employed by researchers to solve this challenge, electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) (Ferrari & Quaresima, 2012). 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 (Li et al., 2022). An investigation utilized a configuration to classify imagery related to hand movements, involving three EEG electrodes (C3, C4, and Cz) and ten fNIRS channels (Ge et al., 2017; Rupawala et al., 2018; Wang et al., 2022). The system achieved 81.2% accuracy, outperforming EEG at 74.7% and fNIRS at 56.8% (Ge et al., 2017). The success of combining different BCI approaches shows that these technologies can improve the field and inspire hope. However, merging these data types is very challenging (Ahn & Jun, 2017; Deligani et al., 2021). The differing temporal scales of EEG and fNIRS—instantaneous signals versus delayed blood flow changes—require researchers to use advanced methods such as phase-space reconstruction, common spatial patterns (CSP), and hemodynamic features (e.g. the Hurst exponent) to synchronize and combine the data (Ge et al., 2017). Multimodal brain-computer interfaces can restore communication and autonomy for critical individuals by combining the advantages of EEG and fNIRS (Chen et al. 2023; Guo et al., 2024).

In recent years, there has been 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% (Abdulghani et al., 2023; Bakhshali et al., 2022; Choi & Kim, 2019b). Researchers quickly adopted CSP analysis and support vector machines (SVMs) as methods that provide comparable performance (Alzahrani et al., 2024). This review paper explained 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 emphasized evidence from fNIRS and hybrid EEG–fNIRS systems alongside EEG. Although reports on fNIRS-only (Sereshkeh et al., 2018) and hybrid (Arif et al., 2024; Cooney et al., 2021; Ge et al., 2017; Kwon et al., 2020; Rezazadeh Sereshkeh et al., 2019) approaches are limited, these findings offer valuable insights. This review investigated whether combining or replacing these modalities can enhance decoding imagined speech by leveraging fNIRS' ability to detect hemodynamic changes and complement EEG's limitations.

The distribution of the included studies by data modality is shown in Figure 1. This re-view investigated whether combining or re-placing these modalities could enhance imag-ined-speech decoding by leveraging fNIRS' ability to detect hemodynamic changes and compensate for EEG's limitations.

Materials and Methods

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', among others. 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 research or reviews published 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. Studies were excluded if they were not in English, did not involve imagined speech, or lacked clear methodological details. After reviewing the full texts, we identified 24 studies that met the criteria.

For comparison, the research studies were divided into three distinct categories:

1. Laboratory-based investigations, focusing on data acquisition and designing experimental tasks.

2. Algorithmic investigations, emphasizing the importance of feature extraction and the effectiveness of classification methods.

3. 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 24 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 24): 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 (Alzahrani et al., 2024; Diego Lopez-Bernal et al., 2022).

fNIRS-exclusive investigations (1 out of 24): Despite delays in the hemodynamic response, fNIRS is effective in binary mental communication tasks, with past accuracy rate around 71–75% (Abdalmalak et al., 2020). 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 performed 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 (Sereshkeh et al., 2018).

Hybrid EEG–fNIRS investigations (3 out of 24): 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 the remaining hybrid studies reported variable performance depending on the experimental task, fusion strategy, and validation protocol. 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 (Ali et al., 2023; Arif et al., 2024; Bourguignon et al., 2022; Cooney et al., 2021; Ge et al., 2017; Kwon et al., 2020; Rezazadeh Sereshkeh et al., 2019).

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.

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 (Choi & Kim, 2019a).

fNIRS-only: In three-class classification, the average accuracy over the last three online blocks was about 64.1±20.6%. By the final online block, 9 of the 12 participants performed above chance, with a mean three-class accuracy of 83.8±9.4% among these nine participants (Sereshkeh et al., 2018).

Phonemic stimuli (e.g. /a/ or /u/) in closed-set environments

EEG-based: Accuracy metrics using methods, like the Hilbert transform with CSP, and 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 (Bakhshali et al., 2022; Biswas & Sinha, 2022; Guenther & Brumberg, 2011; Sarmiento et al., 2021; Wang et al., 2017).

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% (Asghari Bejestani et al., 2022; Chengaiyan & Anandhan, 2015; Chengaiyan et al., 2020; Chinta & Moorthi, 2022; Cooney et al., 2021; García-Salinas et al., 2019; Lee et al., 2022).

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 (Abdulghani et al., 2023; Asghari Bejestani et al., 2022; Chengaiyan & Anandhan, 2015; Chinta & Moorthi, 2022; Cooney et al., 2019; D'Zmura et al., 2009; Einizade et al., 2022; Lee et al., 2019; Lee et al., 2021; Rousis et al., 2024; Saha et al., 2019a; Saha et al., 2019b; Torres-García et al., 2016).

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 70.45±19.19% for imagined speech (Rezazadeh Sereshkeh et al., 2019). These findings demonstrate that the hybrid approach surpasses single-modality methods, with EEG alone achieving accuracy between 34.2% and

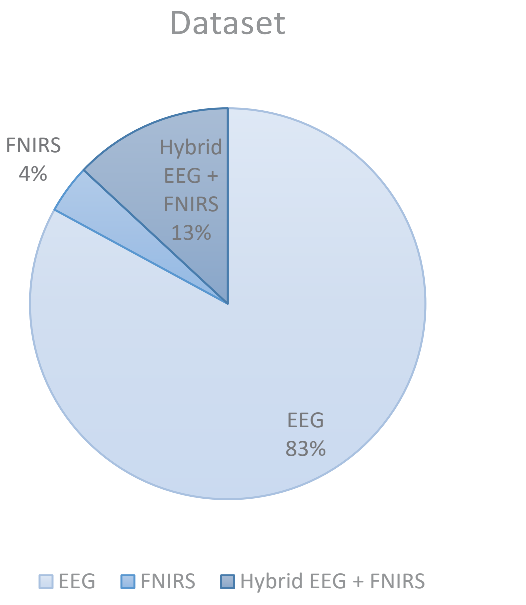

Figure 1. The distribution of data modalities utilized in the reviewed studies

70.33%, and fNIRS alone generally showing weaker results compared to the combined technique (Cooney et al., 2021; Khan et al., 2021; Kwon et al., 2020; Lee et al., 2019; Rezazadeh Sereshkeh et al., 2019; Shin et al., 2018; Torres-García et al., 2016).

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 EEG, fNIRS, or a hybrid system. The following are the main techniques discussed:

EEG-based features

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 (Guenther & Brumberg, 2011; Ismail & Karwowski, 2020; Tavakolan et al., 2016).

Power spectral density (PSD): PSD assesses power across frequency bands—delta, theta, alpha, beta, gamma—linked to specific cognitive functions (D'Zmura et al., 2009).

Discrete wavelet transform (DWT): DWT enables multi-resolution EEG decomposition, making it ideal for identifying time-frequency features crucial for interpreting imagined speech (Diego Lopez-Bernal et al., 2022; Torres-García et al., 2016).

Common spatial patterns (CSP): CSP is used in binary classification to improve class discrimination by fine-tuning spatial filters across EEG channels (Choi & Kim, 2019a; Einizade, Mozafari, Jalilpour, Bagheri, & Hajipour Sardouie, 2022; Einizade, Mozafari, Jalilpour, Bagheri, & Sardouie, 2022; Lee et al., 2019; Shin et al., 2018).

Phase-based connectivity measures: Techniques such as PLV and coherence are used to examine synchronization among cerebral regions, providing insight into the neural dynamics involved in the generation of internal speech (Lee et al., 2021; Mohammadi et al., 2023).

fNIRS-based features

Key features derived from fNIRS data

Mean changes in the concentrations of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) (Sereshkeh et al., 2018).

Variations in the slope of hemodynamic response curves (Albinet et al., 2014).

Hurst exponent values, used to capture the complexity and long-term memory characteristics of fNIRS signals (Khan et al., 2021).

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 knowledge of brain function (Fazli et al., 2012).

Classification algorithms

Various classification algorithms were utilized in this research, each chosen based on the specific nature of the data, task difficulty, and dataset size.

SVM

The reviewed studies show that commonly used classification algorithms have yielded promising results, especially in complex tasks and large datasets. In EEG-

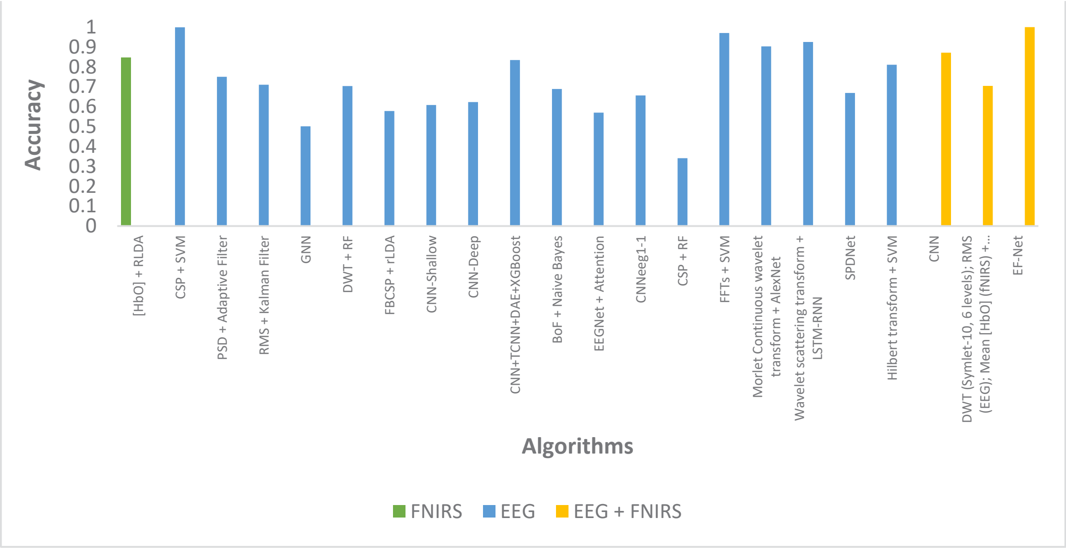

Figure 2. Classification accuracy by modality and classifier

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. In a mixed-time scenario, SVM achieved up to 97% accuracy in binary classification for six imagined Persian words (Asghari Bejestani et al., 2022). Other research has also reported accuracies between 80.2% and approximately 86% for tasks using simple yes/no words (Batres-Mendoza et al., 2017; Choi & Kim, 2019a; Li & Liu, 2013). This highlights SVM's ability to handle high-dimensional multimodal data (Khan et al., 2021).

Linear discriminant analysis (LDA) and regularized LDA (rLDA)

In an fNIRS task, rLDA achieved a mean three-class accuracy of 64.1±20.6% for the classification of imagined "yes," imagined "no," and rest; however, the considerable variance indicated a susceptibility to minor datasets and noise (Sereshkeh et al., 2018). An rLDA classifier combined with FBCSP achieved an accuracy of 57.80% in imagined spoken word-pair classification (Cooney et al., 2019). Although LDA is simple and interpretable, its effectiveness decreases if Gaussian assumptions are violated (Bourguignon et al., 2022b; Bourguignon et al., 2022a; Kwon et al., 2020; Sereshkeh et al., 2018).

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') along with their English equivalents, depending on features and subjects (Torres-García et al., 2016). 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 (Lee et al., 2019; Park et al., 2024; Torres-García et al., 2016). RF achieved moderate accuracy with features, like wavelet transforms or statistical analyses but generally performed worse than deep learning (Roy et al., 2019).

Naive bayes (NB)

Some EEG studies show that this method is used mainly with low-dimensional or discrete attributes. A NB 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, NB tends to perform worse when features are correlated, despite its speed and simplicity (García-Salinas et al., 2019).

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 (Lotte et al., 2018; Xin et al., 2017).

Table 1. Summary of classification algorithms and reported accuracies

Ref.	Modality	Sample Size	Task Type	Signal Acquisition Setup	Stimuli / Classes	Fusion strategy	Method	Accuracy
Choi & Kim (2019b)	EEG	23	Binary yes/no	60-channel	Yes/no (occupation-based Qs)	early	CSP + SVM	70–99.9%
Sereshkeh et al. (2018)	fNIRS	12	Three classes (yes, no, rest)	44-channel	Yes/no/rest	-	Mean value of [HbO] + RLDA	64.1±20.6%
D'Zmura et al. (2009)	EEG	4	Multiclass	128-channel	/ba/, /ku/	Early	Power spectral densities + adaptive filter	75%
Guenther & Brumberg (2011)	EEG/ECoG	2	Phoneme	Two-channel/48-channel	AA, IY, UW	Early	Root-mean-squared (RMS) + Kalman filter	~71%
Chengaiyan & Anandhan (2015)	EEG	6	Word-level	19-channel	8 English words	Early	_	Qualitative
Torres-García et al. (2016)	EEG	27	Multiclass	14-channel	Five Spanish words ("arriba," "abajo," "izquierda," "derecha," "seleccionar")	Early	DWT + random forest	68.18–70.33%
Wang et al. (2017)	EEG	Not mentioned	Phoneme	8-channel	/a/, /u/, rest	-	EEGNet-inspired CNN combined with RNN (LSTM) and CTC	-
Chengaiyan et al. (2020)	EEG	2	Word-level	14-channel	CVC Words	Early	PSD and Relative Power, PLV	Qualitative
Cooney et al. (2019)	EEG	15	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%
Saha et al. (2019b)	EEG	14	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)
García-Salinas et al. (2019)	EEG	27	Word-level	14-channel	Five Spanish words ("arriba," "abajo," "izquierda," "derecha," "seleccionar")	Early	Bag of features (BoF) + naive bayes	65.65±13.39%
Lee et al. (2022)	EEG	10	Word-level	58-channel	/Ba/, /Ku/, /He/, /Li/	Early	EEGNet + attention	57%
Sarmiento et al. (2021)	EEG	15 (BD1) and 50 (BD2)	Phoneme	18-channel/14-channel	/a/, /e/, /i/, /o/, /u/	Decision-level	CNNeeg1-1	BD1 = 65.62% BD2 = 85.66%

Ref.	Modality	Sample Size	Task Type	Signal Acquisition Setup	Stimuli / Classes	Fusion strategy	Method	Accuracy
Lee et al. (2019)	EEG	7	Multiclass	64-channel	12 English words + Rest	Decision-level	CSP + RF	34.2%
Lee et al. (2021)	EEG	16	Multiclass	64-channel	Words/phrases (ambulance, clock, hello, help me, light, pain, stop, thank you, toilet, TV, water, and yes)	-	PLV	Qualitative
Asghari Bejestani et al. (2022)	EEG	5	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)
Einizade et al. (2022)	EEG	15	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%)
Bakhshali et al. (2022)	EEG	14 and 8	Phoneme	64-channel	KARA ONE/FUM	Early	Hilbert transform + SVM	81.1%
Cooney et al. (2021)	EEG + fNIRS	19	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 Combinations-Image (CI): red ball, green hat, red green, ball hat Combinations-audio (CA): red ball, green hat, red green, ball hat	Late	CNN	53%
Rezazadeh Sereshkeh et al. (2019)	EEG + fNIRS	11	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

Ref.	Modality	Sample Size	Task Type	Signal Acquisition Setup	Stimuli / Classes	Fusion strategy	Method	Accuracy
Arif et al. (2024)	EEG + fNIRS	26	Binary mental-state recognition: word generation versus baseline	EEG: 30- channel fNIRS:72-channel	Word generation (WG) versus baseline/rest (BL)	Dual-stream	EF-Net	SF1 score: Subject-dependent: fNIRS only: 99.69% fNIRS + EEG: 99.36% Subject-semidependent: fNIRS only: 98.50% fNIRS + EEG: 98.31% Subject-independent: fNIRS only: 63.80% fNIRS + EEG: 65.05%
Chinta & Moorthi, (2022)	EEG	10	Word-level	16-channel with 2-reference	10 words (up, down, left, right, ...)	Early	Morlet continuous wavelet transform + AlexNet	90.3%
Abdulghani et al. (2023)	EEG	4	Multiclass	8-channel	Up, down, left, right	Early	Wavelet scattering transform + LSTM-RNN	92.5%
Rousis et al. (2024)	EEG	11 and 15	Multiclass	64-channel	Kara One and BCI 2020	Dual-stream	EEGNet - SPDNet	2020 BCI Comp = 66.93% / Kara One = 24.79%

Adaptive and Kalman filters

In phoneme recognition, adaptive filtering was implemented to differentiate /ku/ and /ba/ using EEG data, achieving 75% accuracy (D'Zmura et al., 2009). 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 (Guenther & Brumberg, 2011; Guenther et al., 2009; Sinai et al., 2005).

Convolutional neural networks (CNN)

A two-stage system was built that included CNNs, 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 an accuracy of 90.3% (Alotaibi, 2023; Chinta & Moorthi, 2022; Saha et al., 2019).

CNN with attentional mechanisms

The EEGNet model incorporating attention mechanisms was able to achieve 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 (Cisotto et al., 2020; Lee et al., 2022).

• 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 (Cooney et al., 2021; Khan et al., 2021; Shin et al., 2018).

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 (Lee & Lee, 2022; Wang et al., 2017).

Recurrent neural networks (RNNs)

EEG classification (Up, Down, Left, Right) was investigated using LSTM-RNNs, achieving an accuracy of 92.5%. These models are adept at capturing intricate, temporally varying characteristics, especially in tasks involving motor imagery (Lee & Lee, 2022).

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 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 (Einizade et al., 2022).

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 reported an accuracy of 66.93%, and Kara One reported 24.79%. The results suggest that Riemannian geometry features may outperform Euclidean methods in EEG classification (Rousis et al., 2024).

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 (Choi & Kim, 2019b; Cooney et al., 2019; Lee et al., 2019). The evaluation of all classification models employed subject-dependent or subject-independent protocols, using methods like k-fold cross-validation and leave-one-subject-out (LOSO) to assess generalizability (Dos Santos et al., 2023). Additional details, including stimulus types, class labels, and accuracy ranges, are presented in Table 1.

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 high accuracy rates, ranging from 75% to over 99%, especially when employing feature extraction methodologies, such as CSP in conjunction with SVM classifiers. The effectiveness of this performance was further enhanced by simplified task designs and a reduction in the number of classes (Choi & Kim, 2019a).

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 ranged from 34.2% to 85% (Asghari Bejestani et al., 2022; Chengaiyan & Anandhan, 2015; Chengaiyan et al., 2020; Chinta & Moorthi, 2022; Cooney et al., 2019; García-Salinas et al., 2019; Guenther & Brumberg, 2011; Lee et al., 2022; Lee et al., 2019; Lee, 2021; Saha et al., 2019; Sarmiento et al., 2021; Torres-García et al., 2016).

fNIRS-only investigations: The accuracy of classification was typically moderate in studies relying solely on fNIRS data. For instance, one study achieved a mean three-class accuracy of 64.1±20.6% using the mean value of [HbO] as a feature and the rLDA algorithm to distinguish imagined "yes," imagined "no," and rest (Sereshkeh et al., 2018).

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%) (Cooney et al., 2021). While hybrid methods can improve accuracy from 65–72% (single modality) to about 87%, these numbers come from different tasks with varying class numbers and difficulty levels. Therefore, they are not directly comparable across studies and should be viewed as general indicators of the potential of multimodal systems, rather than definitive proof of superiority.

Hybrid systems may provide complementary information from EEG and fNIRS, although their performance advantage is not consistent across tasks, datasets, and validation protocols (Figure 2).

EEG–fNIRS hybrid data

The remaining hybrid EEG–fNIRS studies specifically addressing imagined-speech-related tasks used different feature- and decision-level fusion strategies. Their findings suggest that EEG and fNIRS may provide complementary information; however, performance varied across participants, experimental tasks, classifiers, and validation protocols. Therefore, the available findings do not yet establish consistent superiority over single-modality systems (Cooney et al., 2021; Ge et al., 2017; Shin et al., 2018).

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 (Flanagan & Saikia, 2023; Saha et al., 2015).

Most fNIRS setups used 10-16 optodes, with short-separation channels to improve spatial specificity and reduce interference (Flanagan & Saikia, 2023).

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 used 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, as well as their combined use, indicates significant progress in BCI technology (Abdulghani et al., 2023; Cooney et al., 2021). This review summarized advances in methods to acquire cerebral signals, extract patterns, and classify them to understand silent speech intentions, all within carefully controlled laboratory settings (Diego Lopez-Bernal et al., 2022; Rahman et al., 2024).

Model performance and key insights

EEG remains the primary tool for decoding imagined speech due to its high temporal resolution and ease of use (Lopez-Bernal et al., 2022). EEG classification accuracy varies from 60% to over 90%, influenced by task complexity, features, and algorithms (Abdulghani et al., 2023; Alzahrani et al., 2024; Diego Lopez-Bernal et al., 2022). Although fNIRS has been used less frequently, it shows good performance (65–85%) in simpler binary tasks (Abdalmalak et al., 2020; Herff et al., 2012; Sereshkeh et al., 2018). The integration of EEG and fNIRS into a hybrid system often improves performance, sometimes by as much as 25% over using either method alone (Cooney et al., 2021; Rezazadeh Sereshkeh et al., 2019). Each modality offers advantages: EEG's rapid response and fNIRS's spatial resolution work together to provide a better understanding of imagined speech activity (Ghosh et al., 2024; Li et al., 2017).

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 (Parvizi & Kastner, 2018). Model generalizability was assessed through cross-validation methods, such as 10-fold, leave-one-out, or trial-wise validation, depending on dataset size and design. The trend toward standardizing protocols is crucial for achieving comparability and reproducibility in imagined speech decoding (Cutini & Brigadoi, 2014). In contrast, fNIRS is slower and unable to measure blood flow changes immediately, as these take time to become evident due to real-time limitations (Arredondo, 2023; Cutini & Brigadoi, 2014). Furthermore, blood flow in the dermis can cause noise during measurement.

Although the reported advancements are significant, many of the reviewed studies have notable limitations. For instance, a study (Rezazadeh Sereshkeh et al., 2019) 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, another study (Cooney et al., 2021) 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, achieved a subject-independent F1 score of 65.05%, emphasizing that many models only perform well in controlled lab conditions (Arif et al., 2024). These differences highlight the gap between promising results and practical clinical applications.

Classical vs deep learning methods

SVMs, LDA, and RF are still widely used in machine learning algorithms (Asghari Bejestani et al., 2022; Batres-Mendoza et al., 2017; Choi & Kim, 2019a; Khan et al., 2021; Li & Liu, 2013; Lee et al., 2019; Lotte et al., 2018; Park et al., 2024; Roy et al., 2019; Torres-García et al., 2016; Xin et al., 2017). These methods are simple, transparent, and effective for small datasets, especially with features, like PSD, CSP, or wavelet transforms (Choi & Kim, 2019a; D'Zmura et al., 2009; Einizade et al., 2022; Lee et al., 2019; Diego Lopez-Bernal et al., 2022; Shin et al., 2018). Recent advancements favor deep learning, especially CNNs and RNNs, which can independently identify data patterns, removing the need for manual feature engineering (Alotaibi, 2023; Chinta & Moorthi, 2022; Lee & Lee, 2022; Saha et al., 2019). Despite their potential to improve accuracy, these advanced models face challenges. They need large training data, which are 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 do not 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.

Regarding 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 performance outside the lab due to lack of personalization; and small datasets that 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 to enhance trust and clinical applicability. Recent studies that combined EEG and fNIRS (Arif et al., 2024; Cooney et al., 2021; Ge et al., 2017; Kwon et al., 2020; Rezazadeh Sereshkeh et al., 2019) further demonstrate the potential of integrating these technologies. These studies suggest that multimodal fusion may improve performance in some experimental settings; however, its benefit is task- and validation-dependent and is not consistently observed across all studies. 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 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. For EF-Net, the multimodal F1 score reached 99.36% in the subject-dependent setting but decreased to 65.05% in the subject-independent setting. This 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 focus 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. Overcoming these challenges is crucial to transform imagined speech BCIs from controlled laboratory experiments into practical clinical and assistive devices that provide reliable, scalable communication for users.

Ethical Considerations

Compliance with ethical guidelines

This article is a narrative review with no human or animal sample.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

Conceptualization: Monireh Motaqi and Boshra Hatef; Data collection: Abolfazl Mirani, Yeganeh Aboutorabi, and Monireh Motaqi; Writing: All authors.

Conflict of interest

The authors declared no conflict of interest.

References

Abdalmalak, A., Milej, D., Yip, L. C., Khan, A. R., Diop, M., & Owen, A. M., et al. (2020). Assessing time-resolved fNIRS for brain-computer interface applications of mental communication. *Frontiers in Neuroscience, 14*, 105. [DOI:10.3389/fnins.2020.00105] [PMID]

Abdulghani, M. M., Walters, W. L., & Abed, K. H. (2023). Imagined speech classification using EEG and deep learning. *Bioengineering (Basel, Switzerland), 10*(6), 649. [DOI:10.3390/bioengineering10060649] [PMID]

Ahn, S., & Jun, S. C. (2017). Multi-modal integration of EEG-fNIRS for brain-computer interfaces-current limitations and future directions. *Frontiers in Human Neuroscience, 11*, 503. [DOI:10.3389/fnhum.2017.00503] [PMID]

Albinet, C. T., Mandrick, K., Bernard, P. L., Perrey, S., & Blain, H. (2014). Improved cerebral oxygenation response and executive performance as a function of cardiorespiratory fitness in older women: A fNIRS study. Frontiers in Aging Neuroscience, 6, 272. [DOI:10.3389/fnagi.2014.00272] [PMID]

Ali, M. U., Kim, K. S., Kallu, K. D., Zafar, A., & Lee, S. W. (2023). OptEF-BCI: An optimization-based hybrid EEG and fNIRS-brain computer interface. *Bioengineering, 10*(5), 608. [DOI:10.3390/bioengineering10050608] [PMID]

Alotaibi, F. M., & Fawad (2023). An AI-inspired spatio-temporal neural network for EEG-based emotional status. *Sensors, 23*(1), 498. [DOI:10.3390/s23010498] [PMID]

Alzahrani, S., Banjar, H., & Mirza, R. (2024). Systematic review of EEG-based imagined speech classification methods. *Sensors, 24*(24), 8168. [DOI:10.3390/s24248168] [PMID]

Arif, A., Wang, Y., Yin, R., Zhang, X., & Helmy, A. (2024). EF-Net: mental state recognition by analyzing multimodal EEG-fNIRS via CNN. *Sensors (Basel, Switzerland), 24*(6), 1889. [DOI:10.3390/s24061889] [PMID]

Arredondo, M. M. (2023). 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, 29*(1), 106-117. [DOI:10.1037/cdp0000469] [PMID]

Asghari Bejestani, M. R., Mohammad Khani, G. R., Nafisi, V. R., & Darakeh, F. (2022). EEG-based multiword imagined speech classification for persian words. *BioMed Research International, 2022*, 8333084. [DOI:10.1155/2022/8333084] [PMID]

Bakhshali, M. A., Khademi, M., & Ebrahimi-Moghadam, A. (2022). Investigating the neural correlates of imagined speech: An EEG-based connectivity analysis. *Digital Signal Processing, 123*, 103435. [DOI:10.1016/j.dsp.2022.103435]

Batres-Mendoza, P., Ibarra-Manzano, M. A., Guerra-Hernandez, E. I., Almanza-Ojeda, D. L., Montoro-Sanjose, C. R., & Romero-Troncoso, R. J., et al. (2017). Improving EEG-based motor imagery classification for real-time applications using the QSA method. *Computational Intelligence and Neuroscience, 2017*, 9817305. [DOI:10.1155/2017/9817305] [PMID]

Biswas, S., & Sinha, R. (2022). Wavelet filterbank-based EEG rhythm-specific spatial features for covert speech classification. *IET Signal Processing, 16*(2), 92-105. [DOI:10.1049/sil2.12059]

Bourguignon, N. J., Bue, S. L., Guerrero-Mosquera, C., & Borragán, G. (2022). Bimodal EEG-fNIRS in neuroergonomics. Current evidence and prospects for future research. *Frontiers in Neuroergonomics, 3*, 934234. [DOI:10.3389/fnrgo.2022.934234] [PMID]

Chen, J., Xia, Y., Zhou, X., Vidal Rosas, E., Thomas, A., & Loureiro, R., et al. (2023). fNIRS-EEG BCIs for motor rehabilitation: A review. *Bioengineering (Basel, Switzerland), 10*(12), 1393. [DOI:10.3390/bioengineering10121393] [PMID]

Chengaiyan, S., & Anandhan, K. (2015). Analysis of speech imagery using functional and effective EEG based brain connectivity parameters. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 9*(4), 33-48. [DOI:10.4018/IJCINI.2015100103]

Chengaiyan, S., Balathayil, D., Anandan, K., & Thomas, C. B. (2020). Effect of power and phase synchronization in multi-trial speech imagery. In *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications* (pp. 1654-1673). Pennsylvania: IGI Global Scientific Publishing. [DOI:10.4018/978-1-7998-1204-3.ch082]

Chinta, B., & Moorthi, M. (2022). Brain computer interface-EEG based imagined word prediction using convolutional neural network visual stimuli for speech disability. [Preprint] [DOI:10.21203/rs.3.rs-1143834/v2]

Choi, J. W., & Kim, K. H. (2019a). Covert Intention to Answer "Yes" or "No" Can Be Decoded from Single-Trial Electroencephalograms (EEGs). *Computational Intelligence and Neuroscience, 2019*, 4259369. [DOI:10.1155/2019/4259369] [PMID]

Choi, J. W., & Kim, K. H. (2019b). Covert intention to answer "yes" or "no" can be decoded from single-trial electroencephalograms (EEGs). *Computational Intelligence and Neuroscience, 2019*, 4259369. [DOI:10.1155/2019/4259369] [PMID]

Cisotto, G., Zanga, A., Chlebus, J., Zoppis, I., Manzoni, S., & Markowska-Kaczmar, U. (2020). Comparison of attention-based deep learning models for EEG classification. arXiv preprint arXiv:2012.01074. [DOI:10.21203/rs.3.rs-279263/v1]

Cooney, C., Folli, R., & Coyle, D. (2022). A bimodal deep learning architecture for EEG-fNIRS decoding of overt and imagined speech. *IEEE Transactions on Biomedical Engineering, 69*(6), 1983-1994. [DOI:10.1109/TBME.2021.3132861] [PMID]

Cooney, C., Korik, A., Raffaella, F., & Coyle, D. (2019). Classification of imagined spoken word-pairs using convolutional neural networks. *Proceedings of the 8th Graz Brain-Computer Interface Conference, 2019,* 338-343. [DOI:10.3217/978-3-85125-682-6-62]

Cutini, S., & Brigadoi, S. (2014). Unleashing the future potential of functional near-infrared spectroscopy in brain sciences. *Journal of Neuroscience Methods, 232,* 152–156. [DOI:10.1016/j.jneumeth.2014.05.024] [PMID]

D'Zmura, M., Deng, S., Lappas, T., Thorpe, S., Srinivasan, R. (2009). Toward EEG sensing of imagined speech. In: Jacko, J.A. (eds), *Human-computer interaction. New trends.* HCI 2009. Lecture notes in computer science, vol 5610. Berlin: Springer. [Link]

Deligani, R. J., Borgheai, S. B., McLinden, J., & Shahriari, Y. (2021). Multimodal fusion of EEG-fNIRS: A mutual information-based hybrid classification framework. *Biomedical Optics Express, 12*(3), 1635-1650. [DOI:10.1364/BOE.413666] [PMID]

Dos Santos, E. M., San-Martin, R., & Fraga, F. J. (2023). Comparison of subject-independent and subject-specific EEG-based BCI using LDA and SVM classifiers. *Medical & Biological Engineering & Computing, 61*(3), 835-845. [DOI:10.1007/s11517-023-02769-3] [PMID]

Einizade, A., Mozafari, M., Jalilpour, S., Bagheri, S., & Sardouie, S. H. (2022). Neural decoding of imagined speech from EEG signals using the fusion of graph signal processing and graph learning techniques. *Neuroscience Informatics, 2*(3), 100091. [DOI:10.1016/j.neuri.2022.100091]

Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., & Müller, K. R., et al. (2012). Enhanced performance by a hybrid NIRS-EEG brain computer interface. *Neuroimage, 59*(1), 519-529. [DOI:10.1016/j.neuroimage.2011.07.084] [PMID]

Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage, 63*(2), 921-935. [DOI:10.1016/j.neuroimage.2012.03.049] [PMID]

Flanagan, K., & Saikia, M. J. (2023). Consumer-grade electroencephalogram and functional near-infrared spectroscopy neurofeedback technologies for mental health and wellbeing. *Sensors, 23*(20), 8482. [DOI:10.3390/s23208482] [PMID]

García-Salinas, J. S., Villaseñor-Pineda, L., Reyes-García, C. A., & Torres-García, A. A. (2019). Transfer learning in imagined speech EEG-based BCIs. *Biomedical Signal Processing and Control, 50,* 151-157. [DOI:10.1016/j.bspc.2019.01.006]

Ge, S., Yang, Q., Wang, R., Lin, P., Gao, J., & Leng, Y., et al. (2017). A brain-computer interface based on a few-channel EEG-fNIRS bimodal system. *IEEE Access, 5,* 208-218. [DOI:10.1109/ACCESS.2016.2637409]

Ghosh, S., Máthé, D., Harishita, P. B., Sankarapillai, P., Mohan, A., & Bhuvanakantham, R., et al. (2024). Review of Multimodal Data Acquisition Approaches for Brain-Computer Interfaces. *BioMed, 4*(4), 548-587. [DOI:10.3390/biomed4040041]

Guenther, F. H., & Brumberg, J. S. (2011). Brain-machine interfaces for real-time speech synthesis. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2011,* 5360–5363. [DOI:10.1109/IEMBS.2011.6091326] [PMID]

Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., & Panko, M., et al. (2009). A wireless brain-machine interface for real-time speech synthesis. *Plos One, 4*(12), e8218. [DOI:10.1371/journal.pone.0008218] [PMID]

Guo, M., Feng, L., Chen, X., Li, M., & Xu, G. (2024). 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, 95,* 106448. [DOI:10.1016/j.bspc.2024.106448]

Herff, C., Putze, F., Heger, D., Guan, C., & Schultz, T. (2012). Speaking mode recognition from functional near infrared spectroscopy. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2012,* 1715–1718. [DOI:10.1109/EMBC.2012.6346279] [PMID]

Ismail, L. E., & Karwowski, W. (2020). Applications of EEG indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis. *Plos One, 15*(12), e0242857. [DOI:10.1371/journal.pone.0242857] [PMID]

Khan, H., Naseer, N., Yazidi, A., Eide, P. K., Hassan, H. W., & Mirtaheri, P. (2021). Analysis of human gait using hybrid EEG-fNIRS-based BCI system: A review. *Frontiers in Human Neuroscience, 14,* 613254. [DOI:10.3389/fnhum.2020.613254] [PMID]

Kwon, J., Shin, J., & Im, C. H. (2020). Toward a compact hybrid brain-computer interface (BCI): Performance evaluation of multi-class hybrid EEG-fNIRS BCIs with limited number of channels. Plos One, 15(3), e0230491. [DOI:10.1371/journal.pone.0230491] [PMID]

Lee, D. H., Kim, S. J., & Lee, K. W. (2022). *Decoding high-level imagined speech using attention-based deep neural networks.* Paper presented at 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Korea, 21-23 February 2022. [DOI:10.1109/BCI53720.2022.9734310]

Lee, S.-H., Lee, M., Jeong, J. H., & Lee, S. W. (2019). *Towards an EEG-based intuitive BCI communication system using imagined speech and visual imagery.* Paper presented at IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 06-09 October 2019. [DOI:10.1109/SMC.2019.8914645]

Lee, S. H., Lee, M., & Lee, S. W. (2021). *Functional connectivity of imagined speech and visual imagery based on spectral dynamics.* Paper presented at 9th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea, 22-24 February 2021. [DOI:10.1109/BCI51272.2021.9385302]

Lee, Y. E., & Lee, S. H. (2022). *EEG-transformer: Self-attention from transformer architecture for decoding EEG of imagined speech.* Paper presented at 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Korea, 21-23 February 2022. [DOI:10.1109/BCI53720.2022.9735124]

Li, R., Potter, T., Huang, W., & Zhang, Y. (2017). Enhancing performance of a hybrid EEG-fNIRS system using channel selection and early temporal features. *Frontiers in Human Neuroscience, 11,* 462. [DOI:10.3389/fnhum.2017.00462] [PMID]

Li, R., Yang, D., Fang, F., Hong, K. S., Reiss, A. L., & Zhang, Y. (2022). Concurrent fNIRS and EEG for brain function investigation: A systematic, methodology-focused review. *Sensors, 22*(15), 5865. [DOI:10.3390/s22155865] [PMID]

Li, Z. G., & Liu, G. Z. (2013). Research of" yes" and" no" responses by auditory stimuli in human EEG. *Applied Mechanics and Materials, 310*, 660-664. [Link]

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2022). A state-of-the-art review of EEG-based imagined speech decoding. *Frontiers in Human Neuroscience, 16*, 867281. [DOI:10.3389/fnhum.2022.867281] [PMID]

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., & Rakotomamonjy, A., et al. (2018). A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. *Journal of Neural Engineering, 15*(3), 031005. [DOI:10.1088/1741-2552/aab2f2] [PMID]

Mak, J. N., & Wolpaw, J. R. (2009). Clinical applications of brain-computer interfaces: Current state and future prospects. *IEEE Reviews in Biomedical Engineering, 2*, 187-199. [DOI:10.1109/RBME.2009.2035356] [PMID]

McFarland, D. J., & Wolpaw, J. R. (2011). Brain-computer interfaces for communication and control. *Communications of the ACM, 54*(5), 60-66. [DOI:10.1145/1941487.1941506] [PMID]

Mohammadi, Y., Graversen, C., Østergaard, J., Andersen, O. K., & Reichenbach, T. (2023). 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, 35*(8), 1301-1311. [DOI:10.1162/jocn_a_02016] [PMID]

Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding covert speech from EEG-A comprehensive review. *Frontiers in Neuroscience, 15*, 642251. [DOI:10.3389/fnins.2021.642251] [PMID]

Park, H., Cho, Y., Lee, T., & Kim, H. H. (2024). *Enhancing word recognition in imagined speech using non-invasive EEG for improved BCI applications*. Paper presented at International Conference on Cyberworlds (CW), Kofu, Japan, 29-31 October 2024. [DOI:10.1109/CW64301.2024.00074]

Parvizi, J., & Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nature Neuroscience, 21*(4), 474–483. [PMID]

Pichiorri, F., & Mattia, D. (2020). Brain-computer interfaces in neurologic rehabilitation practice. *Handbook of clinical neurology, 168*, 101-116. [DOI:10.1016/B978-0-444-63934-9.00009-3] [PMID]

Proix, T., Delgado Saa, J., Christen, A., Martin, S., Pasley, B. N., & Knight, R. T., et al. (2022). Imagined speech can be decoded from low-and cross-frequency intracranial EEG features. *Nature Communications, 13*(1), 48. [DOI:10.1038/s41467-021-27725-3] [PMID]

Rahman, N., Khan, D. M., Masroor, K., Arshad, M., Rafiq, A., & Fahim, S. M. (2024). Advances in brain-computer interface for decoding speech imagery from EEG signals: A systematic review. *Cognitive Neurodynamics, 18*(6), 3565–3583. [DOI:10.1007/s11571-024-10167-0] [PMID]

Rezazadeh Sereshkeh, A., Yousefi, R., Wong, A. T., Rudzicz, F., & Chau, T. (2019). Development of a ternary hybrid fNIRS-EEG brain-computer interface based on imagined speech. *Brain-Computer Interfaces, 6*(4), 128-140. [DOI:10.1080/2326263X.2019.1698928]

Rousis, G., Kalaganis, F. P., Nikolopoulos, S., Kompatsiaris, I., & Petrantonakis, P. C. (2024). *Combining EEGNet with SPDNet towards an end-to-end architecture for imagined speech decoding*. Paper presented at 32nd European Signal Processing Conference (EUSIPCO), Lyon, France, 26-30 August 2024. [DOI:10.23919/EUSIPCO63174.2024.10715364]

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering, 16*(5), 051001. [DOI:10.1088/1741-2552/ab260c] [PMID]

Rupawala, M., Dehghani, H., Lucas, S. J. E., Tino, P., & Cruse, D. (2018). Shining a light on awareness: A review of functional near-infrared spectroscopy for prolonged disorders of consciousness. *Frontiers in Neurology, 9*, 350. [DOI:10.3389/fneur.2018.00350] [PMID]

Saha, P., Abdul-Mageed, M., & Fels, S. (2019). Speak your mind! towards imagined speech recognition with hierarchical deep learning [Preprint]. [DOI:10.48550/arXiv.1904.05746]

Saha, P., Fels, S., & Abdul-Mageed, M. (2019). *Deep learning the EEG manifold for phonological categorization from active thoughts*. Paper presented at ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12-17 May 2019. [DOI:10.1109/ICASSP.2019.8682330]

Saha, S., Nesterets, Y. I., Tahtali, M., & Gureyev, T. E. (2015). Evaluation of spatial resolution and noise sensitivity of sLORETA method for EEG source localization using low-density headsets. *Biomedical physics & engineering express, 1*(4), 045206. [DOI:10.1088/2057-1976/1/4/045206]

Sarmiento, L. C., Villamizar, S., López, O., Collazos, A. C., Sarmiento, J., & Rodríguez, J. B. (2021). Recognition of EEG signals from imagined vowels using deep learning methods. *Sensors, 21*(19), 6503. [DOI:10.3390/s21196503] [PMID]

Sereshkeh, A. R., Yousefi, R., Wong, A. T., & Chau, T. (2019). Online classification of imagined speech using functional near-infrared spectroscopy signals. *Journal of Neural Engineering, 16*(1), 016005. [DOI:10.1088/1741-2552/aae4b9] [PMID]

Shin, J., Kwon, J., & Im, C. H. (2018). 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, 12*, 5. [DOI:10.3389/fninf.2018.00005] [PMID]

Sinai, A., Bowers, C. W., Crainiceanu, C. M., Boatman, D., Gordon, B., & Lesser, R. P., et al. (2005). Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. *Brain, 128*(7), 1556-1570. [DOI:10.1093/brain/awh491] [PMID]

Tavakolan, M., Yong, X., Zhang, X., & Menon, C. (2016). Classification scheme for arm motor imagery. *Journal of Medical and Biological Engineering, 36*, 12-21. [DOI:10.1007/s40846-016-0102-7] [PMID]

Torres-García, A. A., Reyes-García, C. A., Villaseñor-Pineda, L., & García-Aguilar, G. (2016). Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification. *Expert Systems with Applications, 59*(C), 1-12. [DOI:10.1016/j.eswa.2016.04.011]

Wang, J., Chen, Y. H., Yang, J., & Sawan, M. (2022). Intelligent classification technique of hand motor imagery using EEG Beta rebound follow-up pattern. *Biosensors, 12*(6), 384. [DOI:10.3390/bios12060384] [PMID]

Wang, K., Wang, X., & Li, G. (2017). Simulation experiment of bci based on imagined speech eeg decoding. *arXiv preprint arXiv:1705.07771*. [DOI:10.48550/arXiv.1705.07771]

Xin, Y., Wu, Q., Zhao, Q., & Wu, Q. (2017). "Semi-supervised regularized discriminant analysis for EEG-based BCI system. In H. Yin, Y. Gao, S. Chen, Y. Wen, G. Cai &T. Gu (Eds.), *Intelligent data engineering and automated learning-IDEAL 2017. Lecture Notes in Computer Science, vol 10585*. Cham: Springer. [Link]