Fluency in Imagined Speech Decoding Using Non-Invasive Techniques: A Review

S. Shrividya[1,2], *S. Thundiyil*[1] *and J. Picone*[3]

1. Computational Neuroscience and Engineering Research Lab, BMS Institute of Technology & Management, Bengaluru, India
2. College of Engineering, Northeastern University, Boston, USA
3. Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA

Imagined speech is defined as the internal simulation of speaking without producing audible sound [1]. Brain–computer interfaces (BCIs) that convert imagined speech promise a communication channel for individuals with severe speech impairments. While most efforts have targeted word or phoneme level classification using electroencephalography (EEG) and magnetoencephalography (MEG) as modalities, the capacity to decode continuous, coherent, and contextually relevant inner speech remains as an active area of research. This review examines the neuro-cognitive basis of imagined speech. We compare non-invasive neural acquisition modalities, surveys signal processing and decoding methodologies, and scrutinize fluency-specific challenges and metrics, outlining benchmarks, current limitations. We discuss emerging solutions such as large language model (LLM) integration that offer significant promise in revolutionizing this field. We discuss personalized pipelines for real-world, fluent BCI applications in assistive technology and silent communication.

The cognitive process of imagined speech involves mentally hearing one's own voice while thinking in the form of sound, without intentional movement of articulators such as lips, tongue, or hands. The phenomenon represents a truncated form of overt speech, sharing similar neural pathways while lacking the final articulatory execution phase. Functional neuroimaging and neurolinguistics research highlight a core network for inner speech involving the inferior frontal gyrus (Broca's area), supplementary motor area, and superior temporal gyrus, interacting via the phonological loop to support lexical access and syntax generation. Fluent inner speech requires rapid lexical retrieval, seamless syntax assembly, and dynamic working memory updates to manage serial order and semantic coherence. Breakdowns in any component can manifest as hesitations or incoherent output [2].

The superior temporal resolution of an EEG makes it the predominant modality for imagined speech BCIs. Reviews report classification accuracies for small vocabularies (3–5 words) between 60–95% using feature extraction and deep learning [3, 4]. However, an EEG suffers from low spatial resolution and high susceptibility to noise and artifacts, impairing continuous decoding of fluent speech. MEG offers comparable spatio-temporal resolution with improved spatial specificity. Subject-independent imagined speech decoding with MEG has achieved near subject-dependent accuracy via domain adaptation and curriculum learning, underscoring MEG's potential for generalized fluent decoding [5, 6]. Functional near infrared spectroscopy (fNIRS) captures hemodynamic responses linked to imagined speech, offering greater spatial specificity than EEG but limited by latency that constrains real-time fluency [7]. Figure 1 shows various modalities preprocessing techniques covered in this review

Fluent decoding demands continuous feature representations. Traditional approaches segment signals into fixed windows, extracting spectral features (power spectral density, band power) and time–frequency maps (e.g., SPWVD) to feed into classifiers [8]. More recent pipelines employ sliding windows combined with deep representation learning—such as CNNs and Bi-LSTMs—to capture temporal dependencies and handle pauses or hesitations [9]. Transfer learning strategies (e.g., FSFTL) leverage simpler binary tasks (imagined speech vs. rest) to refine feature extractors for multi-class decoding [10].

Machine learning paradigms range from support vector machines and ensemble classifiers to end-to-end deep neural networks. Attention-based architecture has demonstrated improved local feature learning for multi-word classification albeit at modest accuracies (56%) [11].

Integrating pre-trained NLP models (transformers, RNN-based language models) enables contextual smoothing: beam search and probabilistic decoding can re-rank output hypotheses, while GPT-style autocorrection adapts word sequences for coherence. Such hybrid pipelines hold promise for elevating word-level predictions into fluid sentence strings. A summary of various post-processing techniques used is shown in Figure 2.

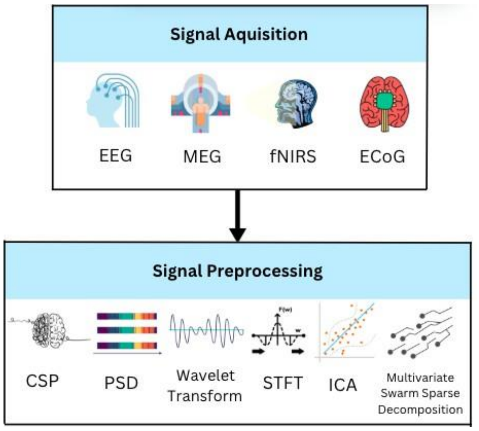

Figure 1. Modalities and Preprocessing Techniques.

Lexical continuity and syntactic coherence define fluency in BCI outputs. Key matrices that are used to measure fluency are Words per Minute (WPM), Sentence Coherence Score, Latency and Perplexity. While these matrices provide quantitative measures of fluency, subjective assessments such as user satisfaction, perceived naturalness etc., complement objective metrics but lack standardization. Developing benchmark tasks and combined fluency indices is critical for rigorous evaluation.

Public EEG datasets such as the Chinese Imagined Speech Corpus (Chisco) with more than 20,000 sentences per subject facilitate large-scale model training [11]. Paradigms include sentence recall, free thought, and question-answering tasks that simulate real communication. However, continuous imagined speech datasets remain scarce, impeding fluent decoding research.

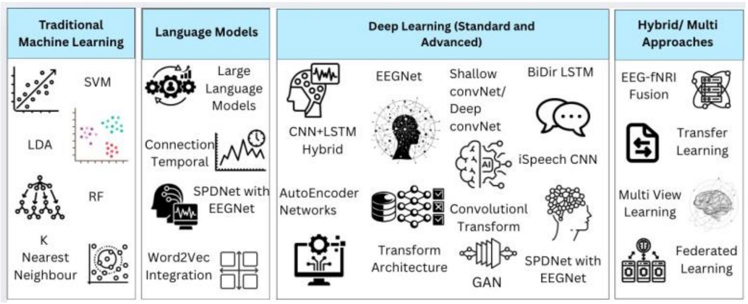

Figure 2. Machine learning and Deep Learning Approaches.

Some of the current challenges and limitations include low SNR in EEG and latency in fNIRS, which hinder real-time continuous decoding; limited sentence-level datasets that restrict model generalization; and high individual differences that necessitate speaker adaptive approaches. This may be overcome by integrating LLMs to post-process decoded word streams, which offers dynamic context prediction and error correction, and smoothing disjoint outputs into coherent text. Leveraging unlabeled data and cross-subject transfer techniques can mitigate data scarcity, enabling models to adapt to new users with minimal calibration. Feedback loops in which decoded text informs subsequent decoding via reinforcement learning can progressively refine fluency through closed-loop training. Customized models that learn individual neural signatures of inner speech can enhance decoding accuracy and fluency, particularly for clinical populations.

Decoding imagined speech fluently via non-invasive BCIs remains a formidable challenge at the intersection of neuroscience, signal processing, and natural language processing. Progress from word-level classification to continuous, coherent text generation requires harmonizing high-fidelity neural recording, advanced decoding architectures, and powerful language models, underpinned by standardized datasets and fluency focused evaluation frameworks. Achieving this milestone will unlock transformative applications in assistive technology, silent communication, and neurorehabilitation.

REFERENCES

[1] Kraemer, David JM, C. Neil Macrae, Adam E. Green, and William M. Kelley. "Sound of silence activates auditory cortex." Nature 434, no. 7030 (2005): 158-158.

[2] Cooney, Ciaran, Raffaella Folli, and Damien Coyle. "Neurolinguistics research advancing development of a direct-speech brain-computer interface." IScience 8 (2018): 103-125.

[3] Lopez-Bernal, Diego, David Balderas, Pedro Ponce, and Arturo Molina. "A state-of-the-art review of EEG-based imagined speech decoding." Frontiers in human neuroscience 16 (2022): 867281.

[4] L. Zhang, Y. Zhou, P. Gong, and D. Zhang, “Speech imagery decoding using eeg signals and deep learning: A survey,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, no. 1, pp. 22–39, 2025.

[5] D. Dash, P. Ferrari, A. Babajani-Feremi, D. Harwath, A. Borna, and J. Wang, “Subject generalization in classifying imagined and spoken speech with meg,” in *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2023, pp. 1–4.

[6] D. Dash, P. Ferrari, A. Babajani-Feremi, A. Borna, P. D. D. Schwindt, and J. Wang, “Magnetometers vs gradiometers for neural speech decoding,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 6543–6546.

[7] 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.

[8] A. Kamble, P. H. Ghare, and V. Kumar, “Deep-learning-based bci for automatic imagined speech recognition using spwvd,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–10, 2023.

[9] M. Bisla and R. S. Anand, “Optimized cnn-bi-lstm–based bci system for imagined speech recognition using foa-dwt,” *Advances in Human-Computer Interaction*, vol. 2024, no. 1, p. 8742261, 2024.

[10] A. Li, Z. Wang, X. Zhao, T. Xu, T. Zhou, and H. Hu, “Enhancing word-level imagined speech bci through heterogeneous transfer learning,” in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–4.

[11] 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)*. IEEE, 2022, pp. 1–4.