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

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

1. CNER 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
shrividya.shashidhara@gmail.com, saneesh@bmsit.in, picone@temple.edu

Imagined speech is defined as the internal simulation of speaking without producing audible sound [1]. Brain–computer interfaces (BCIs) that decode imagined speech into text promise a communication channel for individuals with severe speech impairments. While most efforts have targeted word or phoneme level classification using electroencephalography (EEG), magnetoencephalography (MEG) and functional near infrared spectroscopy (fNIRS) as modalities, the capacity to decode continuous, coherent, and contextually relevant imagined speech remains as an active area of research. This review examines literatures that focuses on neuro-cognitive basis of imagined speech, non-invasive neural acquisition modalities, surveys signal processing and decoding methodologies, and scrutinize fluency-specific challenges and metrics, outlining benchmarks, current limitations. While prior reviews have addressed word-level and phoneme-level classification [3, 4], in this review we focus on fluency-specific challenges.

This review synthesizes research on non-invasive imagined speech decoding with emphasis on fluency. The search terms such as "imagined speech," "imagined speech BCI," "EEG/MEG/fNIRS decoding," and "continuous speech decoding" were used to gather the literature. Papers emphasizing recent advances in deep learning architectures, transfer learning, and language model integration were commonly used in this process [3, 4]. Current findings are drawn from 109 papers, organized into five thematic sections. The number of papers reviewed, and their corresponding focus areas are presented in Figure 1.

Analysis of the neurocognitive foundations literature reveals that imagined speech is a motor-grounded, hierarchically organized cognitive process. The phenomenon represents a truncated form of overt speech, sharing similar neural pathways while lacking the final articulatory execution phase. The interdependence of motor/articulatory and auditory-perceptual components demonstrate that imagined speech is not merely "silent speech" but a distinct phenomenon whose multidimensional variability, temporal processing constraints, and individual differences in strategy have critical implications for imagined speech BCI development. Functional neuroimaging and neurolinguistics research highlight a core network for imagined 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 imagined speech requires rapid lexical retrieval, seamless syntax assembly, and dynamic working memory updates to manage serial order and semantic coherence. Break-downs in any component can manifest as hesitations or incoherent output [2].

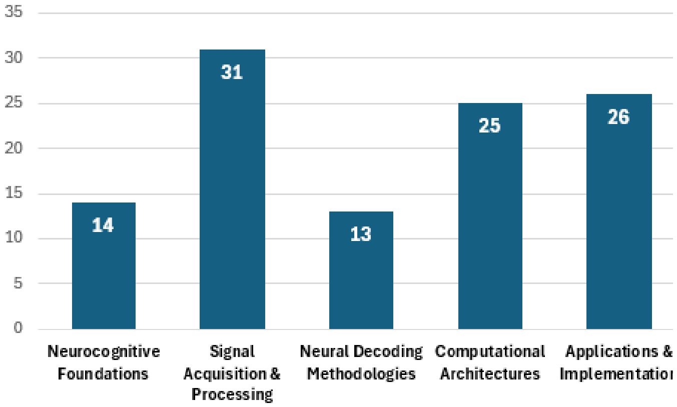


Figure 1. Overview of the literature based on focus area

We reviewed 13 research articles on non-invasive neural decoding. These studies predominantly EEG or MEG often in combination with fNIRS, to classify imagined phonemic or word level speech commands. The superior temporal resolution of an EEG makes it the predominant modality for imagined speech BCIs. Reviews report classification

979-8-3315-7370-6/25/$31.00 ©2025 IEEE

accuracies for small vocabularies (3–5 words) between 60–95% using feature extraction and deep learning [3]. However, EEG suffers from low spatial resolution and high susceptibility to noise and artifacts, impairing continuous decoding of fluent speech. MEG captures brain activity with millisecond temporal precision comparable to EEG, while providing better spatial resolution for localizing neural sources. However, MEG systems require magnetically shielded rooms and expensive sensor arrays, limiting accessibility compared to EEG, and generate large data volumes that complicate real-time processing [4]. Despite these constraints, subject-independent MEG decoding (where models generalize to new users without retraining) has achieved accuracies approaching those of subject-dependent systems through domain adaptation and curriculum learning [5], demonstrating MEG's potential for practical, fluent BCIs for speech decoding. fNIRS offers greater spatial resolution in comparison with EEG but limited by latency that constrains real-time fluency [6].

Across 31 studies in Signal Acquisition and Processing, it is noted that the focus is clearly on improving the quality of the data in non-invasive methods. The typical steps followed are signal collection, preprocessing, and artifact reduction. Independent component analysis and related techniques are commonly used for artifact removal. This is supported by regression and adaptive filters when reference signals are available. Building on these preprocessing techniques, recent work emphasizes the design of effective feature representations for fluent decoding. Traditional approaches segment signals into fixed windows, extracting spectral features such as power spectral density, band power. Sliding windows combined with deep learning methods such as Convolutional Neural Networks and Bidirectional Long Short-Term Memory networks are used in the recent approaches to capture temporal dependencies [7]. Transfer learning strategies, further refine feature extraction by leveraging simpler binary tasks (imagined speech vs. rest) to improve multi-class decoding performance [8].

Most of the papers that discuss computational architectures are focused on deep learning for EEG-based BCIs. CNNs and RNNs outperform traditional methods but face challenges in data scale and interpretability. In imagined-speech decoding, CNN variants and transformer-based models such as EEGformer [9] achieve strong multiclass performance by modeling temporal and frequency dependencies. Hybrid architectures like CTNet enhance generalization and efficiency, while transfer learning and cross-subject adaptation address data scarcity.

Integrating pre-trained NLP models and RNN-based models, enables contextual smoothing in EEG-to-text decoding. Techniques like beam search and probabilistic decoding can re-rank output hypotheses, while GPT-style autocorrection adjusts word sequences for coherence. Post processing pipelines leveraging these models typically include constrained re-ranking, confidence threshold calibration, and strategies to ground outputs and mitigate hallucinations. Such hybrid approaches can transform word-level predictions into coherent sentence strings. A summary of these post-processing techniques is shown in Figure 2.

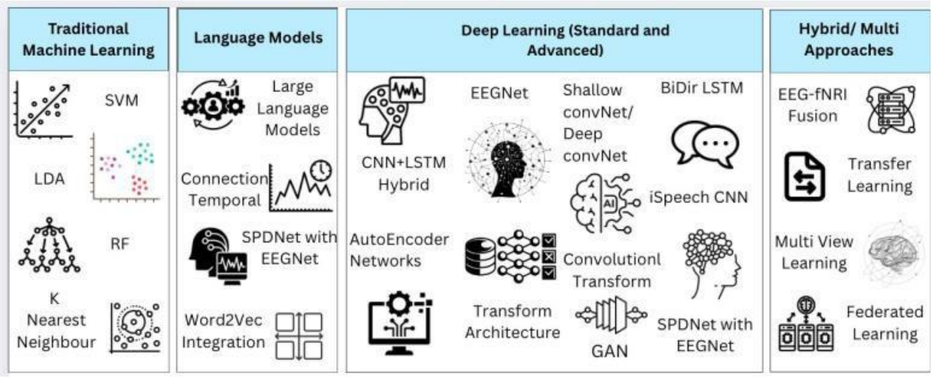


Figure 2. Machine learning and Deep Learning Approaches.

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 of high-density EEG recordings of imagined speech from healthy adults facilitate large-scale model training [10].

Decoding of imagined speech using non-invasive BCIs remains a challenge. Progress from word-level classification to continuous text generation requires high quality neural recording, advanced decoding architectures, and powerful language models, supported by standardized datasets and fluency-focused evaluation frameworks. Authors are creating a standard dataset while reviewing the state of the art and updates to this review, will be available at https://github.com/cnerlab/brain-to-text/. We acknowledge the Vision Group of Science and Technology, Karnataka, India, for supporting this review by providing facilities under GRD1116.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Wan Z, Li M, Liu S, Huang J, Tan H and Duan W (2023) EEGformer: A transformer–based brain activity classification method using EEG signal. *Front. Neurosci.* 17:1148855. doi: 10.3389/fnins.2023.1148855

[10] 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.

ಬಿ.ಎಂ.ಎಸ್. ತಾಂತ್ರಿಕ ಮತ್ತು ವ್ಯವಸ್ಥಾಪನಾ ಮಹಾವಿದ್ಯಾಲಯ
BMS INSTITUTE OF TECHNOLOGY & MANAGEMENT
(Autonomous Institution Under VTU)
Yelahanka, Bengaluru -560119

https://bmsit.ac.in

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
2.College of Engineering, Northeastern University
3.Neural Engineering Data Consortium, Temple University

NEURAL ENGINEERING DATA CONSORTIUM
www.nedcdata.org

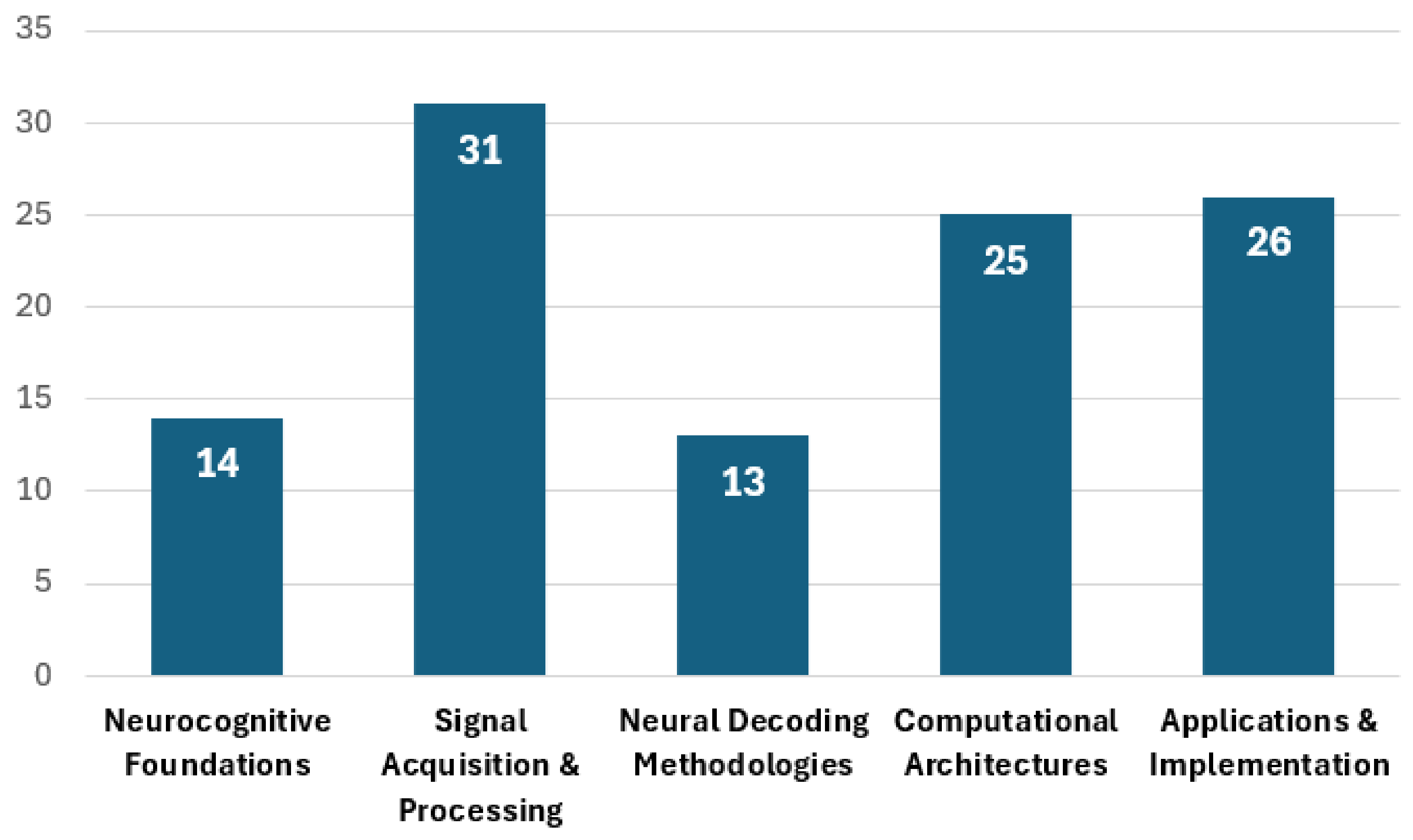
COMPUTATIONAL NEUROSCIENCE AND ENGINEERING RESEARCH LABORATORY
https://cnerlab.github.io/

Abstract

- Imagined speech is the internal simulation of speaking without producing audible sound, offering a promising communication channel for individuals with severe speech impairments through brain-computer interfaces (BCIs).
- Current approaches predominantly target word or phoneme level classification using EEG, MEG, and fNIRS modalities, achieving 60-95% accuracy for small vocabularies (3-5 words) .
- Research gap: Continuous, coherent, and contextually relevant imagined speech decoding remains challenging, particularly achieving fluency.
- Review scope: This paper synthesizes 109 research articles across neurocognitive foundations, non-invasive neural modalities, signal processing methodologies, computational architectures, and fluency-specific evaluation metrics.
- Key focus: This work emphasizes fluency-specific challenges including lexical continuity, syntactic coherence, and real-time decoding constraints.

Background and Methods

- Imagined speech is a motor-grounded, hierarchically organized cognitive process.
- Core brain regions involved in silent speech decoding include the inferior frontal gyrus (Broca's area), supplementary motor area, and superior temporal gyrus which interact via the phonological loop to support imagined speech production.
- Fluency requirements include rapid lexical retrieval, seamless syntax assembly, and dynamic working memory updates are essential for continuous, coherent imagined speech generation.
- Papers emphasizing recent advances in deep learning architectures, transfer learning, and language model integration were commonly used in this process.
- Current findings are drawn from 109 papers, organized into five thematic sections. The number of papers reviewed, and their corresponding focus areas are presented in the figure below.



Methodology

Signal Acquisition

- 13 studies reviewed on non-invasive neural decoding using EEG, MEG, and fNIRS modalities.
- EEG predominates due to superior temporal resolution, portability, and cost-effectiveness compared to MEG (expensive, requires shielded rooms) and fNIRS (latency constraints).

Signal Processing (31 studies reviewed):

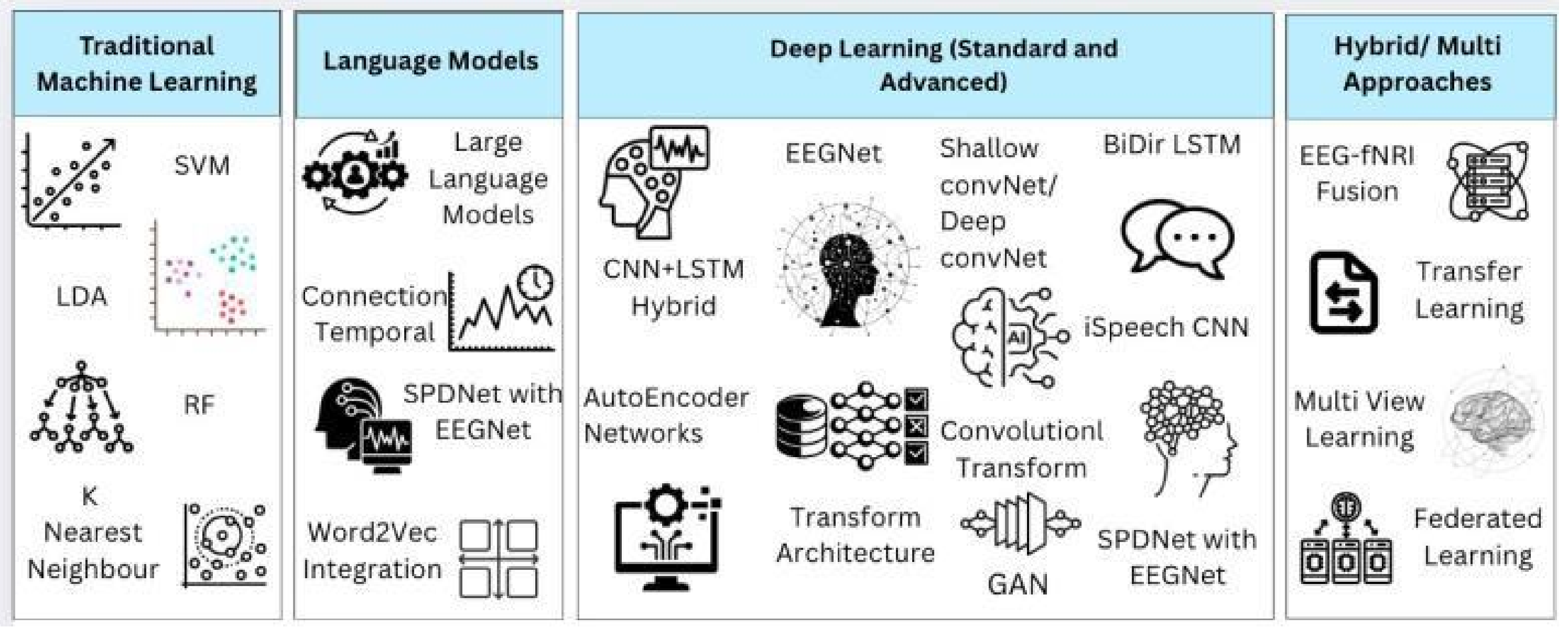
- Pre-processing pipeline: Signal collection > Artifact reduction using ICA > Regression and adaptive filters.
- Feature extraction: Traditional approaches use fixed windows with spectral features (power spectral density, band power).
- Modern approaches: Sliding windows with deep learning to capture temporal dependencies.

Classification and Decoding:

- Traditional methods: 60-95% accuracy for small vocabularies (3-5 words) using spectral features.
- Deep learning architectures: CNNs, Bidirectional LSTMs, and transformers (EEGformer) for multiclass performance.
- Hybrid approaches: CTNet combines CNN-transformer strengths. Transfer learning leverages binary tasks to improve multi-class decoding.

Post-Processing:

- NLP model integration: Pre-trained language models and RNNs enable contextual smoothing.
- Techniques: Beam search, probabilistic decoding, GPT-style autocorrection for coherent sentence generation.



Computational Theme	Key Approaches & Examples	Application Insight
Traditional Machine Learning	SVM, LDA, Random Forests, KNN etc	Used for simpler, low-vocabulary classification tasks.
Deep Learning (Standard and Advanced)	EEGNet, CNNs, BiDiR LSTM,, Transformer Architecture, GAN	Strong multiclass performance by modeling temporal and frequency dependencies.
Language Models / Post-Processing	Large Language Models, Connection Temporal, Word2Vec Integration, GPT-style autocorrection.	Address fluency challenges, Helping transform word-level predictions into coherent sentence strings.
Hybrid/Multi Approaches	EEG-fNIRS Fusion, Multi View Learning	Enhance generalization and efficiency

Performance and Evaluation

Current Classification Metrics:
- Small vocabulary tasks (3-5 words): 60-95% accuracy using feature extraction and deep learning methods.
- Subject-independent MEG decoding approaches subject-dependent accuracy levels through domain adaptation and curriculum learning.

Fluency Metrics:
- Words per Minute (WPM) - measures decoding speed.
- Sentence Coherence Score - evaluates syntactic and semantic continuity.
- Latency - real-time response delay.
- Perplexity - language model confidence in predictions.

Evaluation Challenges:
- Subjective assessments (user satisfaction, perceived naturalness) complement objective metrics but lack standardization across studies.
- Developing benchmark tasks and combined fluency indices is critical for rigorous evaluation.

Benchmark Datasets:
- Chinese Imagined Speech Corpus (Chisco): 20,000+ sentences of high-density EEG recordings from healthy adults.
- Facilitates large-scale model training and standardized evaluation.

Current Limitations:
- Most systems confined to word-level or phoneme-level classification.
- Continuous, coherent, and contextually relevant speech decoding remains challenging.
- Gap between isolated word accuracy and fluent sentence generation.

Summary

- Current achievement: Non-invasive BCIs using EEG achieve 60-95% accuracy for small vocabulary (3-5 words) imagined speech classification through deep learning and hybrid architectures.
- Critical gap: Transitioning from isolated word-level classification to continuous, fluent, and contextually coherent imagined speech decoding remains the primary challenge.
- Requirements for fluency: Progress demands integration of high-quality neural recording, advanced decoding architectures (CNNs, transformers), powerful language models for post-processing, and standardized evaluation frameworks.
- Evaluation needs: Standardized benchmark datasets (like Chisco with 20,000+ sentences) and combined fluency metrics (WPM, coherence, latency, perplexity) are critical for rigorous assessment.
- Future direction: Decoding imagined speech using non-invasive BCIs requires continued advancement in signal processing, cross-subject generalization, real-time processing optimization, and NLP integration to enable practical communication systems.
- Dataset and future updates will be available at: github.com/cnerlab/brain-to-text

Acknowledgements

- We acknowledge CNER Lab at BMS Institute of Technology & Management, Bengaluru, for providing facilities that are funded by the Vision Group of Science and Technology, Karnataka, India (Grant No. GRD1116). Thanks to the Neural Engineering Data Consortium at Temple University for collaboration and support.