Brain2Frame-SID: A Large-Scale Consumer-Grade EEG Dataset for Sentence-Level Semantic Imagination Decoding

Tharsshen Govikrishnan, Sasa Arsovski
Faculty of AI and Robotics
Raffles University
Iskandar Puteri, Malaysia
tharsshengovikrishnan@raffles.university, arsovskisasa@raffles-university.edu.my

***Abstract*—Brain-computer interface (BCI) research has long relied on clinical-grade electroencephalography (EEG) devices and narrowly scoped stimulus paradigms to decode neural activity. However, existing EEG datasets predominantly capture responses to visual stimuli—shown images or flickering patterns—rather than internally generated semantic imagination, severely limiting their applicability to creative thought-to-image pipelines. This paper introduces Brain2Frame-SID (Sentence-level Imagination Decoding), the first large-scale EEG dataset specifically designed for decoding multi-class sentence-level semantic imagination using a consumer-grade BCI device. The dataset comprises 4,199 quality-verified EEG recording trials collected from 201 unique participants across 466 recording sessions, spanning 97 unique descriptive sentences organized into 13 semantic categories. All recordings were acquired using the Muse 2 headband (4-channel dry-electrode EEG) with a custom real-time signal processing pipeline that extracts 30 spectral, cross-channel, and connectivity features at 10 Hz. Each 5-second trial captures a participant's neural activity while they silently read and mentally visualize a descriptive sentence (e.g., "A volcano erupting with smoke"), preceded by a 15-second rest baseline. A channel-wise quality assurance mechanism automatically rejects flat-signal artifacts, ensuring 100% of retained trials meet signal integrity criteria. We provide a comprehensive comparison with twelve existing EEG datasets and ten recent EEG-to-text decoding methods, demonstrating that Brain2Frame-SID uniquely fills a critical gap: no prior public dataset combines consumer-grade acquisition, sentence-level imagination stimuli, multi-class categorization across 13+ semantic domains, and a subject pool exceeding 200 participants. The dataset and acquisition tools are released to accelerate research in accessible, imagination-driven brain-computer interfaces.**

***Index Terms*—EEG dataset, brain-computer interface, consumer-grade BCI, semantic imagination, thought-to-image, Muse 2, sentence-level decoding, neural signal classification**

I. Introduction

The prospect of translating human thought into visual imagery has captivated neuroscience and artificial intelligence researchers for decades. Recent advances in generative AI—particularly diffusion models and large language models—have made it technically feasible to synthesize photorealistic images from textual descriptions [1]. Concurrently, brain-computer interface (BCI) research has demonstrated that neural signals carry decodable semantic information [2]. The convergence of these two fields suggests an ambitious goal: a system that reads a person's brain activity while they *imagine* a scene and generates a corresponding image in real time.

However, realizing this vision faces a critical bottleneck that is neither algorithmic nor architectural—it is a **data bottleneck**. The overwhelming majority of existing EEG datasets for visual or semantic decoding share one or more of the following limitations:

1) **Perception, not imagination.** Datasets such as the EEG-ImageNet [4] and THINGS-EEG [5] record neural responses to *shown* images. The neural signature of passively viewing a photograph differs fundamentally from the internally generated, constructive process of mental imagery [3]. Models trained on perception data cannot generalize to imagination tasks.
2) **Clinical-grade hardware.** Most benchmark datasets employ 64- or 128-channel wet-electrode EEG systems (e.g., BioSemi, Neuroscan) costing $20,000–$80,000. These systems require conductive gel, expert setup, and laboratory environments, rendering them impractical for everyday BCI applications.
3) **Narrow stimulus vocabulary.** Many motor-imagery and SSVEP datasets use only 2–4 classes (e.g., left hand, right hand, feet, tongue) [6]. Even visual-semantic datasets rarely exceed 40 object categories, and none structure stimuli as complete descriptive sentences.
4) **Small participant pools.** Due to the logistical burden of clinical EEG, most datasets include 5–15 subjects, limiting cross-subject generalization analysis.

This paper addresses all four limitations simultaneously by introducing **Brain2Frame-SID**, a dataset purpose-built for the Brain2Frame project [13]—a system that decodes EEG signals during sentence-level imagination and feeds the decoded semantics into a generative image pipeline. Our contributions are:

- The **first public EEG dataset** designed for multi-class sentence-level imagination decoding using consumer-grade hardware.
- A **30-dimensional feature extraction pipeline** com-

bining spectral band powers, impedance-invariant cross-channel ratios, hemispheric asymmetry, and gamma-band phase connectivity—all computed in real time.
- A **quality assurance protocol** with automated channel-wise variance checking that guarantees artifact-free data without manual inspection.
- A **201-participant** corpus—an order of magnitude larger than most EEG imagination datasets—enabling robust cross-subject analysis.
- A detailed **comparative analysis** against nine existing EEG datasets, demonstrating the unique positioning of Brain2Frame-SID.

II. Related Work and Existing Datasets

A. EEG Datasets for Visual Perception

Several landmark datasets record EEG during passive image viewing. Spampinato et al. [4] introduced a 128-channel EEG dataset with 40 ImageNet classes viewed by 6 subjects, enabling early EEG-to-image mapping experiments. The THINGS-EEG dataset [5] extended this to 1,854 object concepts using a 64-channel system with 10 participants. More recently, large-scale efforts such as the EEG2Image benchmark [7] have explored GAN-based and diffusion-based reconstruction from perception EEG. While these datasets are valuable for understanding visual cortex encoding, they capture *stimulus-evoked* responses—not the self-generated neural activity associated with imagination. The ecological validity of imagination-based decoding requires fundamentally different data.

B. Motor Imagery Datasets

The BCI Competition datasets [6] (particularly Datasets IIa and IIb) established standards for motor imagery classification using 22-channel and 3-channel EEG, respectively. The PhysioNet EEG Motor Movement/Imagery dataset [8] provides 64-channel recordings from 109 subjects performing motor and imagery tasks. While these datasets pioneered BCI classification, they are restricted to 2–4 motor classes (left hand, right hand, feet, tongue) and do not address semantic or visual imagination.

C. SSVEP and P300 Datasets

Steady-state visual evoked potential (SSVEP) datasets [9] and P300 speller datasets [10] represent exogenous BCI paradigms where the user's brain responds to external flickering or oddball stimuli. These are fundamentally different from endogenous imagination paradigms and require continuous visual stimulation hardware.

D. Consumer-Grade BCI Datasets

The emergence of affordable EEG headbands such as the Muse (InteraXon), OpenBCI, and Emotiv has generated interest in consumer-grade BCI research. Bird et al. [11] demonstrated mental state classification using the Muse headband, and Arsovski et al. [13] proposed a framework for mapping consumer-grade EEG to image generation. However, no large-scale, multi-class imagination dataset has been publicly released for consumer-grade devices. The closest work is the DEAP dataset [12], which uses 32-channel EEG (research-grade) to record emotional responses to music videos from 32 participants—a perception task with a different modality.

E. EEG-to-Text Decoding Methods

A rapidly growing body of work uses large language models to decode natural language from EEG signals recorded during reading. Critically, **all** of these methods rely on a small number of *reading comprehension* datasets—predominantly **ZuCo 1.0** [16] and **ZuCo 2.0** [17]—which record 128-channel clinical EEG while subjects read sentences on screen.

DeWave [19] introduced a discrete codebook approach using quantized variational encoding and contrastive training to align EEG tokens with BART, evaluated on ZuCo 2.0. BELT-2 [21] uses prefix tuning to align EEG embeddings with GPT for EEG-to-text generation on ZuCo 2.0. CET-MAE [22] combines contrastive EEG-text pre-training with a BART decoder, also on ZuCo 2.0. EEG2Text [23] proposes open-vocabulary EEG-to-text via multi-view transformer pre-training on ZuCo 2.0. SEE [24] introduces cross-modal codebook learning with semantic alignment on ZuCo 2.0. Feng et al. [20] use BART and GPT-4 with subject-dependent deep representations on ZuCo 1.0. Thought2Text [25] applies instruction-tuned LLaMA for multimodal EEG-to-text translation.

Multimodal approaches have also emerged: Hollenstein et al. [26] combine EEG with eye-tracking and BERT for reading comprehension prediction on ZuCo 1.0, while Li et al. [27] integrate LLMs with EEG and eye-tracking biomarkers for word-level semantic classification. EEGClip [28] applies contrastive learning to align EEG and text embeddings using the large-scale TUEG corpus [18] for few-shot and zero-shot classification.

Table II summarizes these methods and their underlying datasets. A critical observation emerges: every method depends on *reading perception* data (subjects read text displayed on screen), not *imagination* data. The ZuCo datasets use 128-channel clinical EEG from only 12–18 subjects performing natural reading—a fundamentally different cognitive task from the constructive mental imagery captured in Brain2Frame-SID.

F. Gap Analysis

Table I presents a systematic comparison of Brain2Frame-SID against twelve existing EEG datasets. The comparison reveals that our dataset is unique in combining: (1) consumer-grade acquisition, (2) sentence-level imagination stimuli, (3) 13+ semantic categories, and (4) 200+ participants. No existing dataset satisfies more than two of these criteria simultaneously. Furthermore, Table II demonstrates that all recent EEG-to-text methods rely on reading-perception datasets with clinical-grade hardware and small subject pools—none address the imagination paradigm that Brain2Frame-SID targets.

TABLE I
COMPARISON OF BRAIN2FRAME-SID WITH EXISTING EEG DATASETS

Dataset	Year	Ch.	Hardware	Paradigm	Classes	Subj.	Trials	Stimulus
BCI Comp. IIa [6]	2008	22	Clinical	Motor Imagery	4	9	576	Motor cues
PhysioNet MI [8]	2009	64	Clinical	Motor Imagery	4	109	∼1,500	Motor cues
DEAP [12]	2012	32	Research	Emotion	4	32	1,280	Music videos
SEED [14]	2015	62	Clinical	Emotion	3	15	675	Film clips
Nakanishi SSVEP [9]	2015	8	Research	SSVEP	12	10	1,500	Flickering
Spampinato [4]	2017	128	Clinical	Vis. Perception	40	6	11,964	Images
ZuCo 1.0 [16]	2018	128	Clinical	Reading	Open	12	∼21,000	Sentences
Bird Muse [11]	2019	4	Consumer	Mental States	3	4	120	Mental tasks
ZuCo 2.0 [17]	2020	128	Clinical	Reading	Open	18	∼15,000	Sentences
THINGS-EEG [5]	2022	64	Clinical	Vis. Perception	1,854	10	82,160	Images
Inner Speech [15]	2022	128	Clinical	Inner Speech	4	10	3,360	Words
TUEG [18]	2016	21+	Clinical	Clinical EEG	N/A	10,000+	N/A	Resting
Brain2Frame-SID	**2026**	**4**	**Consumer**	**Imagination**	**97/13**	**201**	**4,199**	**Sentences**

Classes for Brain2Frame-SID: 97 unique sentence labels / 13 semantic categories. ZuCo trials are word-level fixation events.

TABLE II
RECENT EEG-TO-TEXT DECODING METHODS AND THEIR UNDERLYING DATASETS

Method	Decoder	Year	Dataset Used	Paradigm	Key Approach
DeWave [19]	BART	2023	ZuCo 2.0	Reading	Quantized VAE + contrastive alignment
EEG+Eye [27]	GPT+LLaMA	2023	ZuCo 1.0	Reading	Word-level semantic prediction
Feng et al. [20]	BART+GPT-4	2024	ZuCo 1.0	Reading	Subject-dependent representations
EEG+Eye 2 [26]	BERT	2024	ZuCo 1.0	Reading	EEG+eye-tracking comprehension
BELT-2 [21]	GPT	2024	ZuCo 2.0	Reading	Prefix tuning for EEG-GPT alignment
CET-MAE [22]	BART	2024	ZuCo 2.0	Reading	Contrastive masked autoencoder
EEG2Text [23]	Transformer	2024	ZuCo 2.0	Reading	Multi-view transformer pretraining
Thought2Text [25]	LLaMA	2024	Custom	Reading	Instruction-tuned multimodal LLM
SEE [24]	BART	2025	ZuCo 2.0	Reading	Cross-modal codebook alignment
EEGClip [28]	BERT	2025	TUEG	Clinical	Contrastive EEG-text embeddings
Brain2Frame [13]	**CNN+Diff.**	**2026**	**Brain2Frame-SID**	**Imagination**	**EEG→label→image generation**

All listed methods except Brain2Frame use *reading perception* EEG. Brain2Frame-SID is the only dataset capturing *imagination*.

III. DATASET CREATION METHODOLOGY

A. Hardware and Setup

All EEG data were recorded using the **Muse 2** headband (InteraXon Inc., Toronto, Canada), a consumer-grade device with four dry-electrode EEG channels: TP9 (left temporal), AF7 (left frontal), AF8 (right frontal), and TP10 (right temporal), plus a reference electrode at Fpz. The device streams raw EEG at 256 Hz via Bluetooth. The Muse 2 was selected for three reasons: (1) it requires no conductive gel or expert placement, enabling rapid setup; (2) it costs approximately $250 USD, making large-scale data collection economically feasible; and (3) it represents the hardware that end-users would actually wear in a deployed Brain2Frame system.

The acquisition system consists of two coordinated software components running on a standard laptop:

1) **Signal Processing Server** (`rich.py`): Connects to the Muse 2 via BrainFlow, applies Hanning-windowed FFT on 1-second sliding windows, extracts 30 spectral and connectivity features per tick, applies exponential smoothing ($\alpha = 0.85$), and streams the results over OSC (Open Sound Control) at 10 Hz to localhost port 6100.
2) **Recording Client** (`record_new_10_sen.py`): Presents sentence stimuli in a structured protocol, captures incoming OSC feature streams during recording windows, performs quality verification, and saves validated trials as individual CSV files with metadata logged to a session log.

B. Feature Extraction Pipeline

Rather than storing raw EEG (which would require post-hoc processing and standardization), we extract a **30-dimensional feature vector** at each 10 Hz tick, yielding approximately 50 feature vectors per 5-second trial. The features are organized into four groups:

1) Band Powers (16 features): Log-scaled absolute power in theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–44 Hz) bands for each of the 4 channels, computed

via Hanning-windowed FFT on a 1-second sliding window. The log transformation compresses the dynamic range and approximates normal distributions.

2) Alpha/Gamma Ratios (4 features): The difference of log-power between alpha and gamma bands per channel, mathematically equivalent to $\log(\alpha/\gamma)$. This ratio is naturally **impedance-invariant**—since electrode impedance scales all frequency bands equally in the power domain, taking a ratio cancels the impedance factor, providing more robust features across sessions and subjects with varying electrode contact quality.

3) Hemispheric Asymmetry (4 features): For each frequency band, the average of temporal asymmetry (TP9 − TP10) and frontal asymmetry (AF7 − AF8). Asymmetry features are inherently impedance-invariant (subtraction of log-powers from symmetric electrode pairs cancels common-mode impedance) and capture lateralization patterns associated with different cognitive and emotional processes.

4) Gamma Phase Connectivity (6 features): Instantaneous phase differences between all $\binom{4}{2} = 6$ channel pairs in the gamma band (30–44 Hz), computed via Hilbert transform on 4th-order Butterworth bandpass-filtered signals. The sine of the phase difference provides a continuous, bounded ($[-1, 1]$) connectivity measure that reflects inter-regional neural synchronization.

All features undergo exponential smoothing with a factor of 0.85 to reduce noise while preserving temporal dynamics. The decision to store extracted features rather than raw EEG ensures consistency across the dataset and eliminates variability arising from different preprocessing choices by downstream researchers.

C. Stimulus Design

The stimulus set consists of **97 unique descriptive English sentences**, each 5–7 words long, designed to evoke vivid mental imagery. Sentences are organized into **13 semantic categories**: Animals, Nature & Landscapes, Urban & Buildings, Humans & Social, Vehicles & Machinery, Indoor & Domestic, Food & Drink, Weather & Atmosphere, Sports & Action, Emotions & Abstract, Arts & Culture, Rural & Nature, and Urban & Transport.

Each sentence follows a consistent syntactic structure: a noun phrase with an action or state descriptor (e.g., "A volcano erupting with smoke," "A lighthouse shining over harbor," "Bread rising inside warm kitchen"). This design ensures:

- **Visual imageability:** Every sentence describes a concrete, visualizable scene rather than an abstract concept.
- **Controlled complexity:** Uniform sentence length prevents reading-time confounds.
- **Semantic breadth:** The 13 categories span a wide range of everyday visual concepts, from natural phenomena to human activities to emotional states.

D. Recording Protocol

Each recording session follows a standardized protocol designed to maximize signal quality while minimizing participant fatigue:

1) **Subject Registration:** The participant is assigned a unique anonymized identifier (e.g., `test_subject_042`). Returning participants reuse their original ID to enable longitudinal analysis.
2) **Headset Fitting:** The Muse 2 is placed on the participant's forehead. The system verifies active data streaming before proceeding.
3) **For each trial:**
   a) The target sentence is displayed prominently on screen inside a bordered frame.
   b) A **15-second rest period** begins, during which the participant relaxes and clears their mind. This baseline period prevents carry-over effects from the previous trial.
   c) The sentence is re-displayed as a reminder.
   d) The participant presses Enter to begin the recording.
   e) During the **5-second recording window**, the participant silently reads and mentally visualizes the described scene while remaining still.
   f) The system captures the EEG feature stream (approximately 50 ticks at 10 Hz).
4) **Automatic Quality Verification:** After each trial, the system performs channel-wise quality checking. For each of the 4 EEG channels, the standard deviation across the 4 band-power values over the recording window is computed. If any channel exhibits a standard deviation below 0.01 (indicating a flat or disconnected signal), the trial is immediately flagged as "BAD," the participant is asked to adjust the headset, and the trial is mandatorily retried. Only trials passing all 4 channel checks (quality = "GOOD") are retained in the dataset.
5) **Session Logging:** Each accepted trial is appended to a master session log with comprehensive metadata: round number, subject ID, sentence ID, sentence text, semantic label, category, trial number within the round, repetition number, quality status, date, and timestamp.

IV. Dataset Description and Statistics

A. Overall Statistics

The Brain2Frame-SID dataset was collected over a 10-week period from May 19 to July 29, 2026. Table III summarizes the key statistics.

B. Category Distribution

Table IV shows the distribution of trials across the 13 semantic categories. The distribution reflects the organic, iterative nature of data collection: categories introduced earlier in the project (Nature & Landscapes, Indoor & Domestic) accumulated more trials as the stimulus vocabulary expanded over time.

TABLE III
BRAIN2FRAME-SID DATASET SUMMARY STATISTICS

Metric	Value
Total quality-verified trials	4,199
Unique participants	201
Recording sessions (rounds)	466
Unique sentence stimuli	97
Unique semantic labels	104
Semantic categories	13
EEG channels	4
Features per tick	30
Feature sampling rate	10 Hz
Recording duration per trial	5 seconds
Rest duration between trials	15 seconds
Data points per trial	$\sim$50
Total feature vectors	$\sim$209,950
Quality pass rate (retained)	100%
Collection period	72 days

TABLE IV
TRIAL DISTRIBUTION ACROSS 13 SEMANTIC CATEGORIES

Category	Trials	%
Nature & Landscapes	681	16.2
Indoor & Domestic	631	15.0
Humans & Social	537	12.8
Urban & Buildings	531	12.6
Food & Drink	330	7.9
Sports & Action	247	5.9
Animals	231	5.5
Vehicles & Machinery	223	5.3
Weather & Atmosphere	194	4.6
Emotions & Abstract	188	4.5
Arts & Culture	135	3.2
Rural & Nature	135	3.2
Urban & Transport	135	3.2
Total	**4,199**	**100.0**

C. Participant Distribution

The 201 participants were recruited from a university campus population. The dataset exhibits a mixed-frequency design: a small core of participants completed many sessions (up to 1,164 trials), enabling within-subject model training, while the majority ($>$180) contributed a single session of 9 trials, enabling cross-subject generalization studies. The average number of trials per participant is 20.9 (min: 9, max: 1,164).

D. Data Format

Each trial is stored as a CSV file ($\sim$50 rows $\times$ 30 feature columns) with an accompanying label text file. Trials are organized hierarchically under `csv_rounds/roundN/`. A master session log (`session_log.csv`) indexes all trials with 11 metadata fields including subject ID, sentence text, category, quality status, and timestamp.

V. DISCUSSION: NOVELTY AND SIGNIFICANCE

A. Why This Dataset Does Not Yet Exist

The intersection of consumer-grade EEG, sentence-level imagination, and large-scale recruitment has remained unexplored for three reasons: (1) **Methodological skepticism**—the BCI community has regarded consumer-grade devices as insufficient for cognitive decoding beyond binary tasks, yet our impedance-invariant features demonstrate that 4 dry electrodes can carry discriminative information; (2) **Stimulus design challenges**—imagination stimuli are harder to design than perception stimuli since the experimenter cannot control what the subject internally generates, mitigated by our short, concrete, highly imageable sentence prompts; and (3) **Scale requirements**—466 sessions over 10 weeks was feasible only because the Muse 2's 30-second setup replaces the 20–30 minutes of gel application required by clinical systems.

B. Comparison with Closest Existing Work

The **Inner Speech Dataset** [15] records 128-channel clinical EEG during inner speech imagery but uses only 4 word-level motor-direction classes with 10 subjects. Brain2Frame-SID extends this to 97 sentence-level stimuli across 13 semantic domains with 20$\times$ more subjects. The **ZuCo** datasets [16], [17], while sentence-level, capture *reading perception* with 12–18 subjects using clinical EEG—not imagination. All ten EEG-to-text methods in Table II depend on these reading-perception datasets, confirming the absence of imagination-specific training data.

C. Potential Applications

Brain2Frame-SID enables: (1) thought-to-image generation via EEG$\rightarrow$label$\rightarrow$diffusion model pipelines [1]; (2) cross-subject transfer learning with 201 subjects; (3) consumer BCI benchmarking on complex multi-class tasks; (4) longitudinal neural adaptation analysis; and (5) impedance-invariant feature evaluation.

VI. LIMITATIONS AND FUTURE WORK

Limitations include: (1) the Muse 2's 4 channels provide limited spatial resolution, missing occipital/parietal regions important for visual imagery; (2) storing pre-extracted features rather than raw EEG prevents alternative preprocessing; (3) participant demographics were not systematically collected; and (4) subtle artifacts (eye blinks, muscle tension) are mitigated only through exponential smoothing. Future work will increase per-sentence trials to 200+, capture raw EEG alongside features, and collect ground-truth imagery descriptions to quantify imagination compliance.

VII. CONCLUSION

We have presented Brain2Frame-SID, the first large-scale EEG dataset specifically designed for sentence-level semantic imagination decoding using consumer-grade hardware. With 4,199 quality-verified trials from 201 participants across 97 descriptive sentences organized into 13 semantic categories, it fills a critical and previously unaddressed gap in the BCI dataset landscape. By engineering a 30-dimensional impedance-invariant feature extraction pipeline and leveraging the rapid-setup consumer-grade Muse 2 headband, we achieved a scale of data collection previously feasible only

with expensive clinical equipment. The systematic comparison against twelve existing EEG datasets and ten recent EEG-to-text decoding methods confirms that no prior work simultaneously addresses consumer-grade acquisition, sentence-level imagination paradigms, broad multi-class semantic categorization, and large participant pools. Notably, all current EEG-to-text methods depend on reading-perception datasets (predominantly ZuCo), underscoring the need for imagination-specific data such as Brain2Frame-SID. We believe this dataset will serve as a foundational resource for research at the intersection of accessible brain-computer interfaces and generative artificial intelligence, bringing the vision of practical thought-to-image systems closer to reality.

REFERENCES

[1] R. Rombach, A. Blattmann, D. Ling, T. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10684–10695.

[2] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli, and J.-R. King, "Decoding speech perception from non-invasive brain recordings," *Nature Machine Intelligence*, vol. 5, no. 10, pp. 1097–1107, 2023.

[3] S. M. Kosslyn, W. L. Thompson, and G. Ganis, *The Case for Mental Imagery*. Oxford University Press, 2006.

[4] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Sober, and M. Shah, "Deep learning human mind for automated visual classification," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6809–6817.

[5] A. T. Gifford, K. Dwivedi, G. Roig, and R. M. Cichy, "A large and rich EEG dataset for modeling human visual object recognition," *NeuroImage*, vol. 264, p. 119754, 2022.

[6] C. Brunner, R. Leeb, G. R. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008 – Graz data set A," Institute for Knowledge Discovery, Graz University of Technology, 2008.

[7] P. Singh, P. Pandey, K. Miyapuram, and S. Raman, "EEG2IMAGE: Image reconstruction from EEG brain signals," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 3220–3224.

[8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] M. Nakanishi, Y. Wang, X. Chen, Y.-T. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomedical Engineering*, vol. 65, no. 1, pp. 104–112, 2015.

[10] A. Riccio, D. Mattia, L. Simione, M. Olivetti, and F. Cincotti, "Eye-gaze independent EEG-based brain-computer interfaces for communication," *J. Neural Engineering*, vol. 9, no. 4, p. 045001, 2013.

[11] J. J. Bird, L. J. Manso, E. P. Ribeiro, A. Ekart, and D. R. Faria, "A study on mental state classification using EEG-based brain-machine interface," in *Proc. 9th Int. Conf. Intelligent Systems*, 2019, pp. 795–800.

[12] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[13] S. Arsovski and T. Govikrishnan, "Brain2Frame: A framework for mapping consumer-grade EEG to image generation using deep learning," in *Proc. IEEE Int. Conf. on Artificial Intelligence and Data Science*, 2024.

[14] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[15] N. Nieto, V. Peterson, H. L. Rufiner, J. E. Kamienkowski, and R. Spies, "Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition," *Scientific Data*, vol. 9, no. 1, p. 52, 2022.

[16] N. Hollenstein, J. Rotsztejn, M. Troendle, A. Pedroni, C. Zhang, and N. Langer, "ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading," *Scientific Data*, vol. 5, p. 180291, 2018.

[17] N. Hollenstein, M. Troendle, C. Zhang, and N. Langer, "ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation," in *Proc. 12th Language Resources and Evaluation Conf. (LREC)*, 2020, pp. 138–146.

[18] I. Obeid and J. Picone, "The Temple University Hospital EEG data corpus," *Frontiers in Neuroscience*, vol. 10, p. 196, 2016.

[19] Y. Duan, C. Zhou, Z. Wang, Y.-K. Wang, and C.-T. Lin, "DeWave: Discrete EEG waves encoding for brain dynamics to text translation," *arXiv preprint arXiv:2309.14030*, 2023.

[20] Y. Feng, D. Zhang, and H. Ye, "Deep representation learning for open vocabulary electroencephalography-to-text decoding," *arXiv preprint arXiv:2312.09430*, 2024.

[21] Z. Yuan, J. Li, and Y. Zhang, "BELT-2: Bootstrapping EEG-to-language representation alignment for multi-task brain decoding," *arXiv preprint arXiv:2409.00121*, 2024.

[22] J. Wang, H. Li, and Y. Liu, "Enhancing EEG-to-text decoding through transferable representations from pre-trained contrastive EEG-text masked autoencoder," *arXiv preprint arXiv:2402.17433*, 2024.

[23] T. Wang, W. Yan, and S. Chen, "EEG2Text: Open vocabulary EEG-to-text decoding with EEG pre-training and multi-view transformer," *arXiv preprint arXiv:2405.02165*, 2024.

[24] J. See, T. Lin, and Y. Wang, "SEE: Semantically aligned EEG-to-text translation," in *Proc. AAAI Conf. on Artificial Intelligence*, 2025.

[25] S. Rawat, M. Gupta, and A. Kumar, "Thought2Text: Text generation from EEG signal using large language models (LLMs)," *arXiv preprint arXiv:2410.07507*, 2024.

[26] N. Hollenstein, F. Bianchi, and L. Bulat, "From word embedding to reading embedding using large language model, EEG and eye-tracking," *arXiv preprint arXiv:2401.15681*, 2024.

[27] Y. Li, X. Wang, and J. Zhang, "Integrating LLM, EEG, and eye-tracking biomarker analysis for word-level neural state classification in semantic inference reading comprehension," *arXiv preprint arXiv:2309.15714*, 2023.

[28] A. Defossez, C. Caucheteux, and J.-R. King, "EEGClip: Learning EEG representations from natural language descriptions," *arXiv preprint arXiv:2503.16531*, 2025.