

Decoding Meaning: Cluster-Based Analysis of Semantic Processing Patterns in EEG

Joshua James Podmore¹, Moein Radman¹, Riccardo Poli¹, Silke Paulmann², Ian Daly¹

¹ Brain-Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, England, UK

² Department of Psychology, University of Essex, England, UK

E-mail: jp24194@essex.ac.uk

ABSTRACT: Conventional brain-computer interfaces (BCIs) typically rely on slow character-by-character selection mechanisms. Semantic BCIs (sBCIs) aim to overcome this limitation by decoding entire concepts directly from neural activity. Here, we introduce GRASP (Game of Retrieval and Semantic Processing), a novel auditory paradigm designed to isolate volitional semantic processing in EEG recordings. Participants ($n = 9$) identified concrete concepts during a gamified word-guessing task and provided response-locked markers (spacebar press). Importantly, 25% of trials consisted of phonologically matched pseudo-word sentences, serving as an active baseline condition. Time-frequency representations were computed using Morlet wavelets and statistically evaluated using threshold-free cluster enhancement permutation testing to characterise the spatio-spectro-temporal signatures of semantic processing within individual participants. As an initial proof-of-concept analysis, we focus on the body *vs.* dummy contrast within the GRASP paradigm. Our findings provide insight into the timing, scalp regions, and frequency bands associated with EEG-based semantic processing and may support the development of targeted preprocessing and machine-learning approaches for sBCIs.

INTRODUCTION

Progressive motor neurone diseases such as Amyotrophic Lateral Sclerosis (ALS) can ultimately produce a Locked-In State (LIS), in which cognitive function is preserved but volitional movement is lost [1]. Individuals living with LIS, as well as those affected by severe congenital motor impairments such as cerebral palsy, often require augmentative and alternative communication (AAC) systems to maintain any channel for expressing needs and intent [2]. Most existing BCI-based AAC solutions, such as P300 spellers and SSVEP [3] systems, share the fundamental constraint of iterative character-by-character selection. These place substantial demands on residual sensory and motor functions that may themselves be compromised in target user populations [4]. A single sentence can require dozens of sequential selections distributed across minutes of interaction, imposing severe limits on information transfer rate [5].

Semantic brain-computer interfaces (sBCI) offer a qualitatively different approach where a single decoding operation communicates a complete concept rather than

a single character, offering potential gains in communication efficiency and, in principle, independence from the intact sensory pathways on which conventional BCIs depend [6].

The process of recovering semantic content from distributed neural activity has been highlighted in several key papers [6]. Notably, Mitchell and colleagues [7] successfully predicted fMRI activation patterns for 60 concrete nouns from distributional semantic features, achieving cross-validated decoding accuracies of 0.68-0.85 and generalising to stimuli words unseen during classifier training. This is made possible by the distributed and category-specific organisation of semantic representations across the cortex, a principle formalised by Patterson and team [8].

Advances towards harnessing the mechanics of these distributions with more practical brain imaging were illustrated by Murphy et al [9], who demonstrated above-chance performance for binary classification (animals *vs.* tools) with Electroencephalography (EEG). While single-trial decoding yielded a mean accuracy of 72%, performance increased to 98% when predictions were aggregated across multiple stimulus repetitions. This is primarily attributable to the boost in data quality provided by anchoring EEG epochs to a spacebar press at the moment of semantic retrieval. This dramatically reduced the inclusion of uninformative samples and removed the need for computationally expensive sliding-window approaches. Despite this, the epoching was stimulus-locked to data post-image cue presentation. This undoubtedly introduced significant perceptual confounds and diminishes any claim of true semantic processing [10].

This form of direct cueing causes early visual cortical activations, attentional orienting, and lexical access processes to co-occur with the intended semantic signal [10]. Attempts to circumvent this via a delayed ideation paradigm in concert with fNIRS have shown promising results [11]. However, attempts to utilise these same stimuli and associated classification methods for EEG time-series data have been less fruitful [10]. This is likely due to weak task-engagement, low target-concept relevance, and the absence of trial-level behavioural validation.

Kivisaari and team [12] addressed prior paradigm limitations by having participants derive a target from three increasingly specific orthographic clue sentences, never presenting it directly. Their fMRI findings demonstrated that semantic representations emerge progressively as evidence accumulates across clues, paralleling natural language comprehension. Ghazaryan and collaborators (2024) [13] adapted this paradigm for MEG using 60 concrete Finnish nouns and zero-shot ridge regression to predict word2vec [14] semantic vectors from brain activity. Cross-subject averaged data reached significance after the third clue, however, participant-level decoding failed, attributed to MEG's low signal-to-noise ratio for written stimuli and declining engagement across the 12-18 repetitions required for signal averaging.

The present study introduces GRASP (Game of Retrieval and Semantic Processing). We aim to address the shortcomings of previous experimental paradigms by combining the best elements of each aforementioned approach. The GRASP paradigm is a word-guessing task. It features response-locked temporal indexing (spacebar press), functionally motivated and meta-analytically verified semantic categories (food/body parts/musical instruments) [15]. Additionally, GRASP employs an auditory stimulus presentation format to control for eye-movement artefacts and reading speed variation. Task engagement is boosted by ensuring all clues are unique, and the spacebar interrupt feature allows participants to conduct the task at their own pace. Further, GRASP introduces a novel within-session pseudo-word control condition in which one quarter of trials present acoustically matched sentences in which all content words are replaced by phonologically valid nonsense words. This provides an empirical acoustic baseline against which semantic responses can be contrasted.

This is key, as prior work in the EEG-based semantic decoding space has focused primarily on concept *vs.* concept binary decoding [6]. This assumes that the top-down labels assigned to any given set of samples are discrete. It should be plainly understood that the distance in semantic space between, e.g. foot and banana, is dramatically larger than, e.g. tongue and banana. Despite this, in a traditional semantic decoding experiment, the concepts of foot and tongue would typically be binned into a generic “body part” group. The same could be noted of cross-category comparisons between ear and piano, or fingers and violin. In other words, treating all these items as belonging to pure, orthogonally separable categories fails to accommodate the presence of a connected semantic topology [16]. Notably, these obstacles, which are inherent to concept *vs.* concept decoding, could underpin the low semantic-decoding performance reported in the corresponding literature.

Importantly, this cannot be overcome via the simplistic use of baseline or resting period data in a so-called concept *vs.* resting state setup. Since the resting brain is

intrinsically and continuously active in semantically structured ways [17, 18], any conventional rest period constitutes an uncontrolled, heterogeneous comparator, rendering these undirected experimental periods unsuitable for the characterisation of semantic decoding-related signals.

The utility of our “active-resting state” comes from the explicit positioning of each subject towards a specific low-semantic state across the semantic processing continuum. We hypothesise that this homogenises the active-rest class and allows us to investigate the semantic signals unique to specific concepts in a more discrete and structured manner with greater sensitivity and selectivity. Concretely, we believe that it is vital to first determine the typical spatio-spectro-temporal properties of semantic processing in EEG. We argue that this can only be done cleanly using a controlled, active-rest period. Any attempt to perform this task via concept *vs.* concept decoding is necessarily problematic given the continuous, non-discrete geometry of the cortical semantic topology.

The aims of this study include: (i) establishing whether the pseudo-word control condition can isolate semantic from acoustic processing contributions; and (ii) to characterise the spatial, spectral, and temporal profiles of EEG activity associated with the body *vs.* dummy contrast at the single-subject level as an initial proof-of-concept analysis within the GRASP paradigm. Additionally, we also intend to interpret these results to inform the development of targeted preprocessing stages and machine learning approaches that more precisely exploit the frequency bands, scalp regions, and response-locked time windows most discriminative of semantic processing.

MATERIALS AND METHODS

Participants: 9 neurologically healthy adults participated (avg: 23 years, 6 female). Recruitment was restricted to British English first-language speakers. Ethical approval was granted by the University of Essex Ethics Committee. Participants received £30 compensation irrespective of performance.

Stimulus Generation: Three semantic domains were employed: food (9 fruits/vegetables), musical instruments (9 items), and body parts (9 limbs/sense organs), yielding 27 concept items in total (see Figure 1). For each item, a three-part sentence-level definition was composed, with each sub-sentence increasing in specificity towards the ground truth term. Each definition was produced in two forms: an original sentence (real words) and a pseudo-sentence in which all content words were replaced by phonologically matched pseudo-words. Note, all function words such as stopwords, modals, conjunctions, and pronouns were left unchanged.

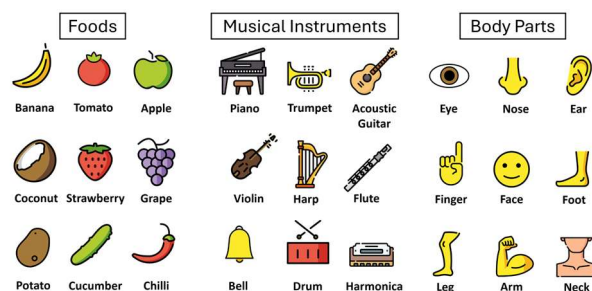


Figure 1: A pictographic of all 27 items for guessing in the GRASP clue paradigm. Each of the three categories comprises 9 items. Note, each block is constructed from 36 items total, with the remaining 9 filled by the so-called “dummy” trials to serve as a low-semantic comparator set.

Pseudo-word variants were generated in two stages. Firstly, automatic precaching used a constrained variant of the Wuggy pseudo-word generator. Each candidate was validated for syllable-count match, character-length, confirmed non-word (dictionary lookup), and text-to-speech (TTS) duration within 200ms of the original word.

All audio was synthesised using Microsoft Edge TTS to produce mono 22,050 Hz 16-bit WAV files. Word-level time boundaries were obtained for every utterance via the Montreal Forced Aligner (MFA) using a UK English pronunciation dictionary and acoustic model. The experimental structure comprised 12 blocks of 36 trials each, 27 original and 9 pseudo-trials per block. A fixed-subset randomisation strategy was applied, ensuring all participants received the same 144 pseudo-sentences with only trial order randomised.

All clue sentences were unique, and comprise 3 subcomponents increasing in specificity to the ground-truth item at each step, e.g. “This is known for its universal appeal, it is often served with butter, and is frequently baked with a cheese topping.”, answer: potato. Here is the same sentence transformed into a pseudo-sentence: “This is known for its unbizzal asteal, it is often sained with beller, and is frequently roked with a shoose flipping.”

Experimental Paradigm: The experiment was implemented in PsychoPy with audio playback via PyGame. Each session required approximately 3 hours (inc. setup & breaks). At the start of each block, a pictographic was presented of all 27 target items for 5 s (see Figure 1). Each trial commenced by displaying a central fixation cross, followed by the playback of the sentence WAV audio (see Figure 2). Participants were instructed to maintain fixation but could press the spacebar at any point to interrupt the audio and signal early identification. Following the audio, or its interruption, a text-entry box appeared instructing participants to type their best guess (30 s timeout). For pseudo-word trials, participants were also required to perform the spacebar interrupt upon identification and enter the term ‘dummy’ to verify attentive listening. If the audio played to completion, a 1.5 s dynamic Gaussian mask was displayed, followed by a 0.5 s pause.

Each trial concluded with a jittered inter-trial interval (ITI) of 0.5-1.25 seconds. Post-block responses were autocorrected using an item-prioritised strategy and a lemmatisation stage. The average trial-level semantic accuracy was then quantified as the mean cosine similarity between word2vec embeddings of the corrected responses and the ground-truth items. For dummy trials, the latency from the first pseudo-word onset to the spacebar response was also computed.

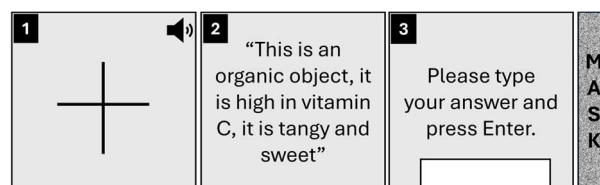


Figure 2: A representation of the stimulus presentation order. The “speaker” icon on the 1st tile is only displayed here to visually indicate the onset of the audio clue. Further, the 2nd tile is here only for demonstration purposes. Only the clue sentence audio is presented to participants. At no point are the clues provided orthographically. The 3rd tile is a faithful representation of the input screen users either see after the clue audio has completed, or immediately after the onset of a spacebar press. The final narrow panel illustrates the onset of the visual mask stage displayed at the end of each trial.

EEG Recording and Preprocessing: EEG data were recorded using a 64-channel BioSemi ActiveTwo system at 2,048 Hz with a standard 10-20 montage and four external EOG channels (EXG1-4), with all subsequent preprocessing performed in MNE-Python [19]. Continuous data were preprocessed via a 1-80 Hz zero-phase bandpass finite-impulse response filter, downsampling to 512 Hz, and inter-block cropping to retain only task-relevant data. Following a 50 Hz notch filter, bad channels were detected via PyPrep [20] (RANSAC) and corrected using spherical spline interpolation.

Data were then re-referenced to an approximate infinity reference. To protect the fitting of the Extended Infomax ICA (32 components, 10,000 iterations), transient and muscle artefacts were annotated and temporarily excluded. Automated artefact rejection was then performed via z-score correlation ($z = 2.0$) with EOG and ECG channels. An additional behavioural filter was also applied to ensure the exclusive retention of exact matches (user input vs. ground truth item label) and trials within the typical range of the subject-specific typing speed.

Time-Frequency Analysis: EEG was analysed in a response-locked framework, with time zero defined as the spacebar press, examining spectral power differences between conditions over the 1.5 s pre-response window (−1.5 to 0 s). A response-locked alignment was preferred over stimulus-locking because the timing of conceptual identification varies across trials and participants. All analyses were conducted at the single-subject level, in recognition of the substantial inter-individual variability in EEG spectral profiles and scalp topographies.

The analyses were implemented in MNE-Python and SciPy [21]. Epochs were resampled to 256 Hz and restricted to the 64 EEG channels under the standard 10-20 montage. For each contrast, single-trial time-frequency representations (TFRs) were computed independently for each condition using a Morlet wavelet transform over a logarithmically spaced frequency range of 3-35 Hz (32 frequencies). Condition differences were tested simultaneously across the full scalp and time-frequency plane using a non-parametric permutation test with threshold-free cluster enhancement (TFCE; [22]). At each channel-frequency-time voxel, an independent-samples Student's t-statistic was computed.

A three-dimensional adjacency structure was defined jointly over channels (electrode neighbourhood graphs derived from the standard 10-20 montage), frequency bins (adjacent bins along the frequency axis), and time points (adjacent samples along the time axis). This ensured that cluster enhancement captures the natural smoothness of EEG spectral data in all three dimensions simultaneously rather than treating each dimension independently.

The permutation null distribution was constructed from 1,000 random permutations of condition labels. Statistical inference was performed using threshold-free cluster enhancement (TFCE) with a permutation-based cluster test. A TFCE statistic map was computed across the channel $\times$ frequency $\times$ time space. Significant voxels were defined as those belonging to clusters whose TFCE-based cluster-level p-values were below $\alpha = 0.05$, thereby controlling the family-wise error rate across the full spatio-spectro-temporal space.

The main aim of this analysis was to define the spatial, spectral, and temporal structure of semantic processing at the single-subject level rather than to perform binary classification. For each participant, the analysis yielded a three-dimensional TFCE statistic together with a voxel-level significance mask derived from TFCE-significant clusters, capturing the neural differentiation between body-part semantic trials and dummy trials.

RESULTS

In the present report, we focus on the body *vs.* dummy contrast, which provides a controlled comparison between semantically meaningful trials and pseudo-word trials lacking interpretable semantic content. This, therefore, offers a representative case for examining the 3-D spatio-spectro-temporal structure of semantic processing within the GRASP paradigm. Figure 3 illustrates the corresponding TFCE statistic map across all 3 dimensions for one subject (07) within a response-locked analysis window (−1.5s, 0s).

The topographic TFCE summary (Fig. 3A) reveals a broad left-lateralised spatial pattern, with the strongest TFCE statistics concentrated over fronto-central sites, extending into left temporal and centro-parietal electrode locations. The temporal TFCE profile (Fig. 3B) shows a

gradual increase in statistical strength beginning approximately 1.4 s prior to the spacebar press. TFCE values remain elevated across much of the pre-response interval and reach a maximum at around −0.6 s. Following this peak, TFCE values decrease but remain above early-window levels. The spectral TFCE summary (Fig. 3C) shows that elevated TFCE values are primarily concentrated in the low-frequency range, with a dominant peak around 7-9 Hz, with a smaller secondary elevation present in the 13-16 Hz range.

Taken together, these results indicate that semantic pseudo-word differentiation in Subject 7 is characterised by a left-lateralised fronto-central network expressed predominantly in the theta and alpha range and emerging during the late stages of the pre-response interval.

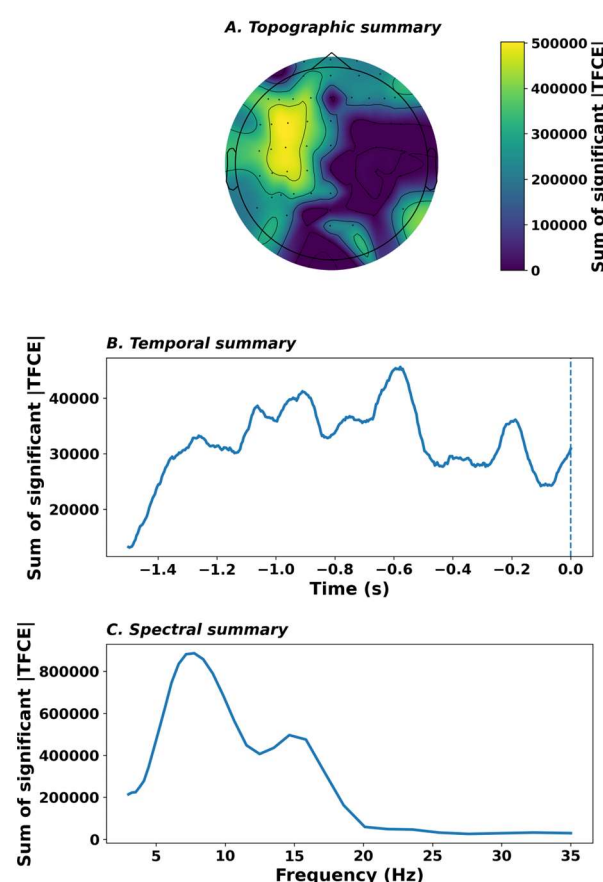


Figure 3: Spatio-spectro-temporal TFCE summary for Subject 07 in the body *vs.* dummy contrast. TFCE values are shown without normalisation to preserve single-subject interpretability.

Further, to illustrate the directionality of this effect at its dominant scalp locus, Figure 4 shows the mean log₁₀ Morlet-derived power difference between body and dummy trials at channel FC3, with all TFCE-significant clusters overlaid as black contours. The plot reveals a power increase in the theta-alpha range, extending all the way up to 20 Hz in some instances. Note also that the primary significant cluster spans most of the pre-response window. Additionally, isolated higher-frequency clusters are visible at approximately 20-25 Hz and 30-35 Hz, however, these should be understood as potentially subject-specific.

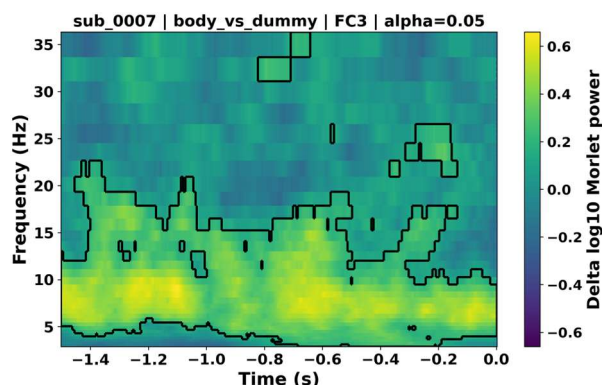


Figure 4: Mean log10 Morlet-derived power difference (body vs. dummy) at channel FC3 for Subject 07. FC3 was the highest-ranked channel in the single-subject TFCE summary (Fig. 3A). Black contours indicate TFCE-corrected significant clusters at $\alpha = 0.05$.

Further, to assess the consistency of the observed effects across participants, TFCE results were summarised descriptively across *all nine subjects*. Each participant exhibited at least one TFCE-significant cluster in the body vs. dummy contrast (see Figure 5).

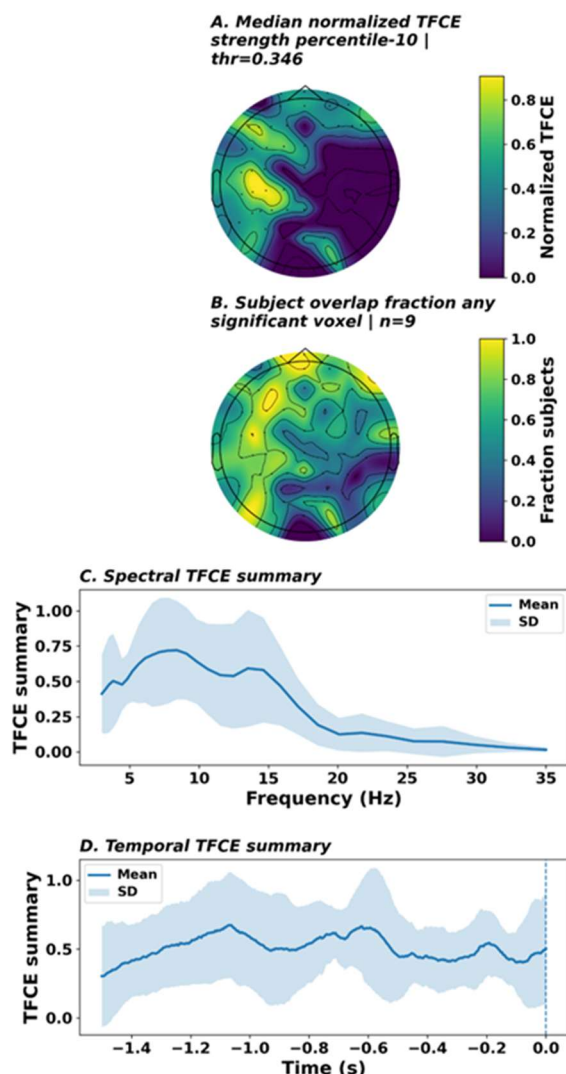


Figure 5: Cross-participant consistency of TFCE effects across spatial, spectral, and temporal domains (normalised within each participant prior to aggregation to account for inter-individual variation in absolute TFCE magnitude).

The spatial TFCE summary (Fig. 5A) shows the median normalised TFCE strength across participants projected onto the scalp topography, illustrating the cross-participant spatial distribution of the effect. The resulting map reveals a broadly left-lateralised pattern, with elevated TFCE values over fronto-central, central, and temporal electrodes of the left hemisphere. In contrast, right-hemisphere electrodes show comparatively lower TFCE values, indicating that the observed TFCE pattern is predominantly left-lateralised across participants.

The subject overlap map (Fig. 5B) shows the proportion of participants exhibiting TFCE-significant voxels at each scalp location. The highest overlap is observed over left fronto-central, central, and temporal electrodes, indicating that these regions represent the most consistent spatial locus of the effect across participants. The spectral TFCE summary (Fig. 5C) illustrates the distribution of TFCE values across frequencies summarised across participants. Elevated TFCE values are primarily concentrated within the theta to alpha range, with a smaller secondary contribution extending from the alpha band into the low-beta range. The temporal TFCE summary (Fig. 5D) illustrates the cross-participant temporal profile of TFCE values. TFCE values increase during the pre-response interval and exhibit two local maxima. One around -1.1 s and another around -0.6 s prior to the behavioural response. Following these peaks, TFCE values gradually decline.

DISCUSSION

This article provides an initial characterisation of the spatio-spectro-temporal signatures of semantic processing in EEG within the GRASP paradigm. Across participants, TFCE-based differentiation between semantic and pseudo-word trials was expressed as a predominantly left-lateralised spatial pattern spanning fronto-central, central, temporal, and centro-parietal scalp regions (see Figs. 3A, 5A, 5B). This spatial distribution broadly aligns with established models of semantic processing, which emphasise the role of left temporal cortices in semantic representation and frontal regions in semantic retrieval [16].

Spectrally, the strongest TFCE values were concentrated within the theta to alpha range, with smaller contributions extending into the low-beta band (see Figs. 3C & 5C). Oscillatory activity in this frequency range has previously been linked to semantic integration and lexical retrieval during language comprehension, suggesting that the signals identified here likely reflect neural processes involved in conceptual identification.

Temporally, the TFCE profiles indicated that these differentiating signals emerged well before the overt behavioural response, with local maxima occurring approximately 1 s and 0.6 s prior to the spacebar press (see Figs. 3B & 5D). This is consistent with cognitive processes related to semantic retrieval rather than the motor execution of the response.

As seen in Figure 4, the primary TFCE-significant cluster, situated in the theta-alpha range, spans most of the pre-response window. The fact that it does not appear to resolve within this window is of interest, however, the epoching onset could not be extended further back to observe potential propagation events without introducing overlap with preceding trials and compromising trial quality. We believe that this represents a potentially fruitful avenue of investigation worth pursuing in future work.

Taken together, these findings suggest that the GRASP paradigm can reliably elicit EEG signatures of semantic processing. As an initial proof-of-concept, the present results highlight candidate scalp regions, frequency bands, and response-locked temporal windows that may be exploited in future machine-learning approaches for semantic brain-computer interfaces.

CONCLUSION

Beyond characterising these effects, the present findings demonstrate the value of response-locked analysis for isolating semantic processing from motor execution. Here, we highlight the spatial, spectral, and temporal features that may be informative for future decoding or machine learning applications. As a proof of concept, this work lays the foundation for developing sBCIs and motivates further studies with larger samples, richer stimulus sets, and cross-participant analyses.

ACKNOWLEDGEMENTS

This work was supported by funding from the Leverhulme Trust [Grant Number: RPG-2023-207].

REFERENCES

- 1] Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. *N. Engl. J. Med.* 2017;377:162–172
- 2] Linse K, Aust E, Joos M, Hermann A. Communication matters: pitfalls and promise of high-tech communication devices in palliative care of severely physically disabled patients with amyotrophic lateral sclerosis. *Front. Neurol.* 2018;9:603
- 3] Abiri R, Borhani S, Sellers EW, Jiang Y, Zhao X. A comprehensive review of EEG-based brain-computer interface paradigms. *J. Neural Eng.* 2019;16:011001
- 4] Wolpaw JR, Wolpaw EW. *Brain-Computer Interfaces: Principles and Practice.* Oxford University Press, Oxford, UK (2012)
- 5] Waller A. Telling tales: unlocking the potential of AAC technologies. *Int. J. Lang. Commun. Disord.* 2019;54:159–169
- 6] Rybář M, Daly I. Neural decoding of semantic concepts: a systematic literature review. *J. Neural Eng.* 2022;19:021002
- 7] Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, et al. Predicting human brain activity associated with the meanings of nouns. *Science* 2008;320:1191–1195
- 8] Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? *Nat. Rev. Neurosci.* 2007;8:976–987
- 9] Murphy B, Talukdar P, Mitchell T. EEG decoding of semantic category reveals distributed representations for single concepts. *Brain Lang.* 2011;117:12–22
- 10] Rybář M, Poli R, Daly I. Using data from cue presentations results in grossly overestimating semantic BCI performance. *Sci. Rep.* 2024;14:28003
- 11] Rybář M, Poli R, Daly I. Decoding of semantic categories of imagined concepts of animals and tools in fNIRS. *J. Neural Eng.* 2021;18:046035
- 12] Kivisaari SL, van Vliet M, Hultén A, Lindh-Knuuttila T, Faisal A, Salmelin R. Reconstructing meaning from bits of information. *Nat. Commun.* 2019;10:927
- 13] Ghazaryan G, van Vliet M, Saranpää A, Lammi L, Lindh-Knuuttila T, Hultén A, et al. Trials and tribulations when attempting to decode semantic representations from MEG responses to written text. *Lang. Cogn. Neurosci.* 2024;39:1149–1160
- 14] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* 2013;26:3111–3119
- 15] Radman M, Podmore JJ, Poli R, Paulmann S, Daly I. Decoding semantic categories: insights from an fMRI ALE meta analysis. *J. Neural Eng.* 2025;22:061006
- 16] Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 2016;532:453–458
- 17] Raichle ME. The restless brain. *Brain Connect.* 2011;1:3–12
- 18] Jackson RL, Hoffman P, Pobric G, Lambon Ralph MA. The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. *J. Neurosci.* 2016;36:1490–1501
- 19] Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 2013;7:267
- 20] Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* 2015;9:16
- 21] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 2020;17:261–272
- 22] Smith SM, Nichols TE. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 2009;44:83–98