



OPEN Using data from cue presentations results in grossly overestimating semantic BCI performance

Milan Rybář[✉], Riccardo Poli & Ian Daly[✉]

Neuroimaging studies have reported the possibility of semantic neural decoding to identify specific semantic concepts from neural activity. This offers promise for brain-computer interfaces (BCIs) for communication. However, translating these findings into a BCI paradigm has proven challenging. Existing EEG-based semantic decoding studies often rely on neural activity recorded when a cue is present, raising concerns about decoding reliability. To address this, we investigate the effects of cue presentation on EEG-based semantic decoding. In an experiment with a clear separation between cue presentation and mental task periods, we attempt to differentiate between semantic categories of animals and tools in four mental tasks. By using state-of-the-art decoding analyses, we demonstrate significant mean classification accuracies up to 71.3% during cue presentation but not during mental tasks, even with adapted analyses from previous studies. These findings highlight a potential issue when using neural activity recorded during cue presentation periods for semantic decoding. Additionally, our results show that semantic decoding without external cues may be more challenging than current state-of-the-art research suggests. By bringing attention to these issues, we aim to stimulate discussion and drive advancements in the field toward more effective semantic BCI applications.

Keywords Semantic decoding, Mental imagery, Cue presentation, Electroencephalography (EEG), Machine learning, Brain-computer interface (BCI), Functional magnetic resonance imaging (fMRI)

Neuroimaging research has shown the potential for semantic neural decoding, which aims to identify specific semantic concepts an individual is focused on, or thinking of, at a given moment in time from their neural activity^{1,2}. Most of these neuroimaging studies use experiments that are designed to cue participants to focus on instances of specific semantic concepts for short periods of time, for example by presenting visual or auditory stimuli. Neural activity is recorded while participants perform mental tasks involving the cued semantic concepts, for example, by silently naming the instance of the concept. This recorded activity is then processed to attempt to differentiate between distinct semantic concepts or categories.

This semantic neural decoding could be used by brain-computer interfaces (BCIs) for communication^{3,4}. BCIs enable a degree of direct control over external devices through signals originating from the human brain. There is a need to explore alternative control schemes for BCI-based communicators. Some options that have been explored are based on decoding language-based signals. For example, recent advances in speech decoding BCIs have shown exciting promise for restoring rapid communication^{5–10}. The most remarkable results have been achieved using invasive approaches, such as electrocorticography (ECoG). However, it is unclear whether it is possible to translate these results with a similar level of performance to electroencephalography (EEG) (or other non-invasive practical modalities)^{11–15}. Additionally, it has been shown that handwriting imagery can be decoded and used for typing purposes^{16,17}.

Semantic BCIs could aid communication by allowing direct communication of semantic concepts for people who experience difficulties communicating via other means^{18–20}. The majority of semantic decoding studies have used functional magnetic resonance imaging (fMRI) and only a few studies used EEG¹. Here, we focus on EEG which provides a better ecological validity in comparison with fMRI and is more suitable for non-invasive BCIs.

An important distinction between semantic BCIs and the majority of semantic decoding studies is that in semantic BCIs, users would freely choose and focus on a semantic concept of their choice (from a supported set of recognizable concepts). In other words, in a semantic BCI, there would not be an external cue to guide the

Brain-Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK. ✉email: contact@milanrybar.cz; i.daly@essex.ac.uk

participant's specific choice. Here, we begin with a preliminary step by using external cues but ensuring they are sufficiently separated from the mental tasks.

However, our efforts to translate these findings into a BCI paradigm have been met with significant challenges, even despite leveraging the same mental task as in the literature. Upon closer examination, there was a major difference between our approach and those approaches used in semantic decoding studies.

Because our objective revolves around advancing toward BCIs, we purposely avoid utilizing neural activity recorded during the cue presentation period for decoding purposes. This is because a BCI application allows users to freely communicate information from a choice of two or more options via decoding their brain activity. Consequently, it is not possible, within a BCI context, to know a priori which choice the user wishes to make and, therefore, it is not possible to rely on neural activity related to presentation of a specific cue. In contrast, the majority of semantic decoding studies in EEG have utilized such neural activity to differentiate between distinct semantic categories or concepts (see Table 1). This was typically done in one of two ways: (1) either the cue was present, for example as an image on the screen, while participants perform mental tasks, or (2) the cue presentation and mental task periods were explicitly separated but neural activity recorded during the cue presentation period together with the neural activity recorded during the mental task was used for decoding purposes. Additionally, this issue is mostly likely to be present in fMRI studies due to slow hemodynamic response unless cue presentation and mental task periods were sufficiently separated.

It is imperative to clarify that our intent is not to criticize the methodologies or analyses of prior studies. Rather, our objective is to push the boundaries further and endeavor to bridge the gap between neuroimaging research and practical BCI applications.

In this paper, we demonstrate radically different decoding performance during the cue presentation and mental task periods in two experiments in which we attempt to differentiate between the semantic categories of animals and tools. We use state-of-the-art decoding analyses and further adapt analyses from two EEG-based semantic decoding studies, one of which used the same mental task. While it is possible to decode the semantic category during the cue presentation period, decoders struggled to differentiate between the semantic category when using data recorded only during any of the tested mental tasks. We discuss potential reasons for these difficulties and their implications. By emphasizing this potential disparity between perceived and actual capabilities of semantic BCIs based on purely imaginary mental tasks without any external cue, we aim to stimulate discussion and drive advancements in the field toward more effective semantic BCI applications.

Background

Semantic neural decoding has been explored using a wide range of different neuroimaging methods, experimental designs, and decoding approaches (see a recent systematic literature review¹).

The majority of semantic decoding studies used fMRI, while a small number of studies used other methods such as EEG, magnetoencephalography (MEG), or invasive approaches. Due to its high spatial resolution, fMRI-based studies typically achieved the highest decoding performance when measured by standard machine learning metrics such as classification accuracy, rank accuracy, or leave-two-out pairwise comparison. However, these metrics do not take into account the time needed to make a decision. When the decoding performance is evaluated in terms of information transfer rate (ITR)^{30–32}, fMRI which is affected by a slow hemodynamic delay requires longer times to make a decision (with the lowest times around two seconds for fMRI versus less than half a second for EEG) and thus it achieves lower ITRs in comparison with electrophysiological neuroimaging methods (in a range of 0.02–9.08 for fMRI, 0.21–24.09 for EEG, and 0.09–149.83 for electrophysiological neuroimaging methods including invasive approaches)¹.

We identified EEG-based semantic decoding studies based on the recent systematic literature review of semantic neural decoding¹. Table 1 shows the studies that used either EEG alone or in combination with other neuroimaging modalities (primarily MEG). The majority of these EEG-based semantic decoding studies *used neural activity recorded when the cue was present for decoding purposes* (see the “Cue present” column in Table 1).

Temporal features	Spectral features	Task	Pres. Mod.	Cue present	Ref.
0-700ms	1-30Hz	out-of-category recognition	VAO	Yes	21
95-360ms	4.1-18.3Hz	silent naming task	V	Yes	22,23
200-700ms (category), 250-500ms (individual words)	1-30Hz	size judgment	AO	Yes	24
0-1.2s	optimized for each participant (0.5-55Hz)	passive listening, silent naming task	AO	Yes	25
0-1s	0.1-40Hz	remember all six elements presented in a sequence	O	Yes	26
0-1s	1-12Hz	out-of-category recognition	A	Yes	27
0-700ms	1-30Hz	in-category recognition	V	Yes	28
0-3500ms	2-100Hz	semantic judgment	V+O	No	29

Table 1. Semantic decoding studies using EEG from¹. The column ‘Pres. Mod.’ represents presentation modalities used: visual (V) (image), auditory (A) (spoken word), or orthographical (O) (written word). Multiple letters indicate that multiple modalities were tested separately, while V+O is an image with a corresponding written name. The column ‘Cue present’ refers to whether or not the cue presentation is present in the temporal window used by the semantic decoding model.

Even though it has been shown that semantic processing is independent of the stimulus presentation modality^{33,34}, several studies reported significantly different decoding accuracies for different stimulus modalities. For instance, an EEG study by Simanova et al.²¹ reported a higher mean classification accuracy when participants were cued by images in comparison with when they were cued by spoken or written words. A MEG-based study by Ghazaryan et al.³⁵ reported challenges of semantic decoding from orthographic stimuli in comparison with prior research based on visual stimuli.

Results

To investigate decoding performance during the cue presentation and mental task periods, we analyzed EEG data from our two experiments in which we attempted to differentiate between the semantic categories of animals and tools.

Experiments

In Dataset 1, twelve participants performed four mental tasks: a silent naming task and three imagery sensory-based tasks using visual, auditory, and tactile perception. Participants were asked to visualize an object in their minds, imagine the sounds made by the object, and imagine the feeling of touching the object. Figure 1 shows an illustration of a concept trial. The concept was indicated by an image shown on the screen for 0.6 seconds. A sequence of mental tasks followed after a mask (an image of a checkerboard) which was shown for 0.6 seconds and a black screen which was shown for 0.5 seconds. Each mental task lasted for 3 seconds and was separated from the following one by a blank screen for 0.2 seconds. Each concept was presented five times, that is for a total of 90 trials per category (18 concepts, 5 repetitions each).

In Dataset 2, seven participants performed only the auditory imagery task (see Fig. 1). The cue presentation period of the concept image was extended to 1 second, the duration of the auditory imagery task was extended to 5 seconds, and the blank screen period before the mental task was changed to 0.5–1 second (uniformly distributed). Each concept was presented seven times, that is for a total of 126 trials per category (18 concepts, 7 repetitions each).

Analysis in the temporal domain

To explore the possibility of differentiating between the semantic categories of animals and tools, we used the most widely used analysis approach from the prior EEG-based semantic decoding studies shown in Table 1 which is an analysis in the temporal domain. This was conducted in the two most widely used frequency bands (see Table 1): 1–30 and 4–30 Hz. Preprocessed EEG data were filtered, downsampled to 128 Hz, and normalized. They were then classified by different classifiers in a stratified cross-validation scheme. We used a support vector machine (SVM) with the radial basis function kernel and $C = 1$, logistic regression (LR) with L1 or L2 norm, linear discriminant analysis (LDA), and a random forest (RF) classifier. Additionally, we used the SVM with a value for the C parameter chosen by a nested cross-validation, the LR with L1 norm with a regularization

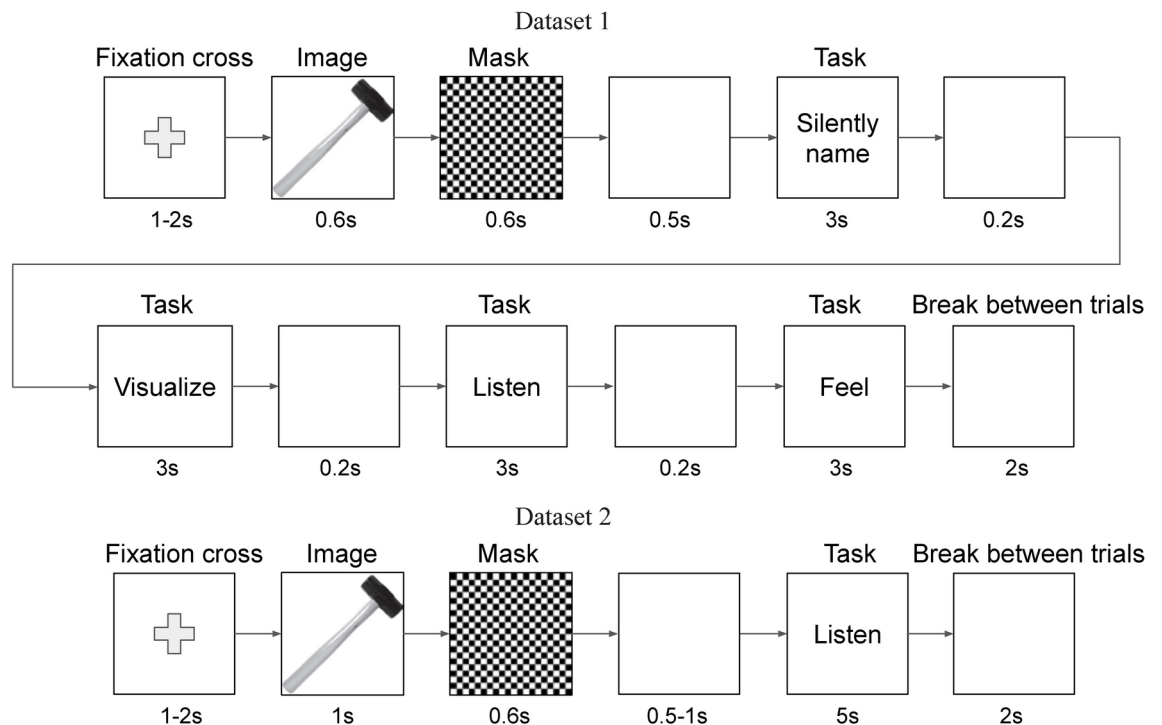


Fig. 1. Illustration of one concept trial in Datasets 1 and 2. The order of mental tasks is randomized across blocks in Dataset 1.

strength parameter value selected by a nested cross-validation, and the RF with a nested cross-validation search for a number of trees in the forest and a maximum depth of the tree.

Although, there were differences in decoding performance between the frequency bands of 1–30 and 4–30 Hz, the overall message was the same for both of them. We thus report here only the results using the frequency band of 1–30 Hz to simplify the presentation.

Figure 2 shows classification accuracies when the classifier can use information from all channels and all time points from the period of interest. It was possible to significantly differentiate between the semantic categories in the image presentation period with mean classification accuracies between 60.4% ($p < 0.01$, a bootstrapping simulation, see Methods) and 62.3% ($p < 0.001$) in Dataset 1 and between 66.9–69.9% ($p < 10^{-6}$) in Dataset 2 across all classifiers. On the other hand, mean classification accuracies were not significant in any mental task in both datasets.

A number of participants with individual significant classification accuracies ($p < 0.05$, a bootstrapping simulation) in the image presentation was 8.1 (out of ten participants) on average across all seven classifiers in Dataset 1 while all seven participants had significant accuracies in all classifiers in Dataset 2. These average numbers of participants for the mental tasks were 1.1, 1.0, 2.7, and 1.3 for the silent naming, visual imagery, auditory imagery, and tactile imagery task in Dataset 1, respectively, and 1.0 for the auditory imagery task in Dataset 2.

The percentages of channels with significant single-channel mean classification accuracies ($p < 0.05$) in the image presentation period were between 30.4–58.5% in Dataset 1 and between 51.8–78.9% in Dataset 2. In both datasets, the LDA had the lowest number of significant channels. On the other hand, these percentages for the mental tasks were below 21% in Dataset 1 and below 17.5% in Dataset 2. Figure 3 shows scalp maps with mean classification accuracies for each classifier.

To investigate the temporal evolution of semantic decoding, a sliding window approach was used. A temporal window of 109.375 ms was used (14 time points at a sampling frequency of 128 Hz) with a step of half the window size. A classifier was separately trained in each position of the temporal window by using only time points from that window. Figure 4 shows mean classification accuracies when the classifier can use information from all channels. It was possible to differentiate between the semantic categories during the image presentation period but not during any of the mental tasks in both datasets. The majority of classifiers had significant mean classification accuracies ($p < 0.05$) during the 54.7–429.7 ms period after the image onset (corresponding to the start of the first and the end of the last temporal window) in both datasets. The peak of the mean classification accuracies was in the temporal window of 164.1–265.6 ms after the image onset in both datasets with mean classification accuracies between 57.9% ($p < 0.05$) and 61.1% ($p < 0.01$) in Dataset 1 and between 64.4% ($p < 10^{-5}$) and 69.6% ($p < 10^{-6}$) in Dataset 2.

Figure 5 shows scalp maps indicating channels for which mean classification accuracies were statistically significant during the image presentation period using the SVM with C parameter selected by nested-cross-validation. The peak of the mean classification accuracies across all participants was again in the temporal window of 164.1–265.6 ms after the image onset in both datasets. Similarly, mean single-channel classification accuracies were not significant during any of the mental tasks in both datasets.

A comparison to state-of-the-art

To further highlight the need for separating the cue presentation period and the mental task period, we adapted analyses from two studies from Table 1 that used an image stimulus as a cue during the mental task period. It is imperative to clarify again that our intent is not to criticize these studies but to show the potential issue with attempting semantic category decoding using the neural data recorded during the cue presentation period.

In a study by Simanova et al.²¹, participants were presented with either images (line drawings), written words, or spoken words of animals and tools. They were asked to respond, by pressing a mouse button, upon the

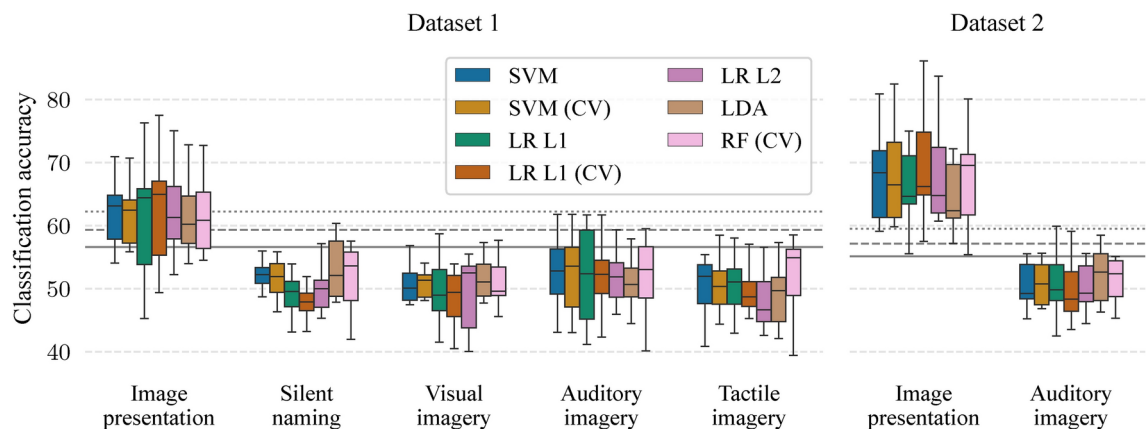


Fig. 2. Mean classification accuracies when classifiers can utilize information from all channels and all time points of corresponding periods. Horizontal lines represent significant borderlines for $p = 0.05$ (solid), $p = 0.01$ (dashed), and $p = 0.001$ (dotted).

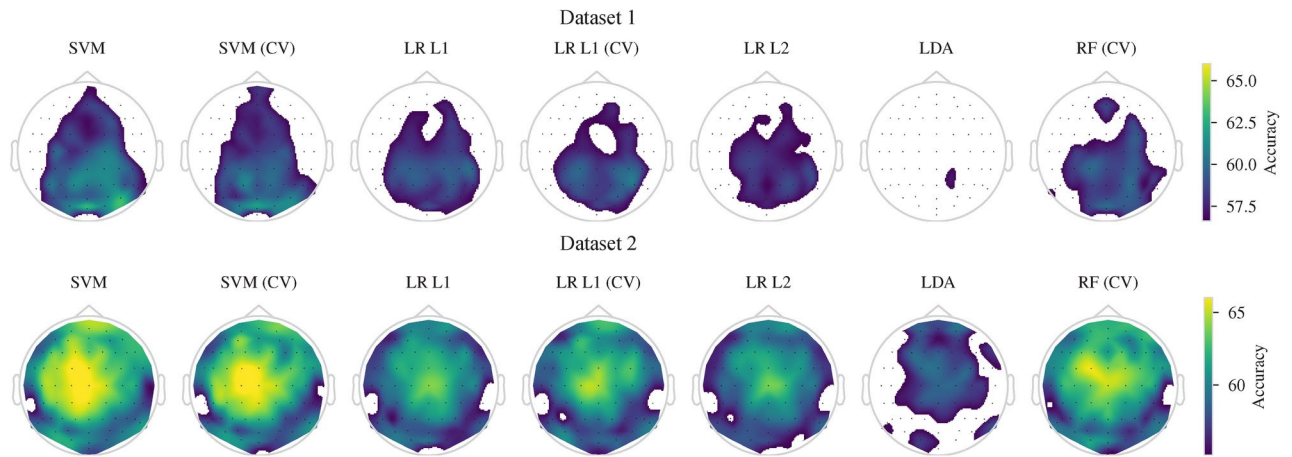


Fig. 3. Single-channel mean classification accuracies above the significance borderline ($p = 0.05$) for the image presentation when classifiers can utilize information from all time points of the image presentation period. White represents non-significant classification accuracies.

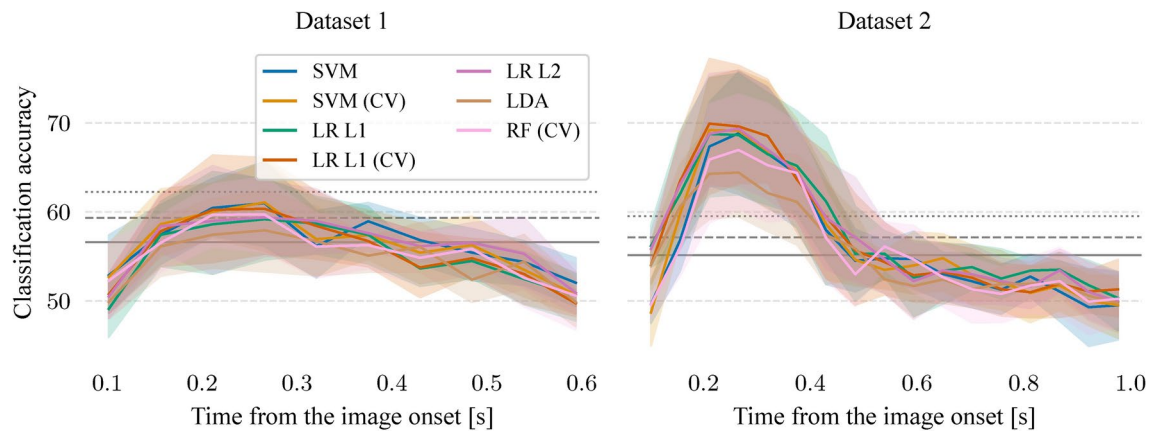


Fig. 4. Mean classification accuracies during the image presentation period when the classifier can use information from all channels. Classification accuracies are shown with mean and 95% confidence interval. Horizontal lines represent significant borderlines for $p = 0.05$ (solid), $p = 0.01$ (dashed), and $p = 0.001$ (dotted).

appearance of items from the non-target task categories (clothing or vegetables). In the case of image stimulus, the image was shown for 300 ms followed by a blank screen for 1000–1200 ms. However, their analysis used a temporal window of 0–700 ms after the stimulus onset which includes the cue presentation period. Filtered signals, in the frequency band 1–30 Hz, from all channels were classified by an LR (Bayesian logistic regression with a multivariate Laplace prior). They achieved a mean classification accuracy of 79% when participants were cued by images. We adapted their analysis by classifying the first 600 ms of the cue presentation period and the mental task periods in the same frequency band. Because our image presentation in Dataset 1 was only 600 ms, in comparison with the 700 ms used by Simanova et al., we used this temporal window of 600 ms for all classified periods.

Figure 6 shows the results of this adapted analysis, which used all channels and 600 ms periods. Mean classification accuracies for the image presentation were between 60.2–62.2% ($p < 0.01$) for tested classifiers in Dataset 1 and between 68.2–71.3% ($p < 10^{-6}$) in Dataset 2. Moreover, individual classification accuracies were significant for all participants in Dataset 2 ($p < 0.05$ for the LR with L1 norm, $p < 0.01$ for the LR with L1 norm and nested-cross-validation, and $p < 0.001$ for the rest). For any of the mental task periods, mean classification accuracies were not statistically significant in both datasets. The average number of participants with significant accuracies in the image presentation was 8.4 in Dataset 1 while all participants had significant accuracies in all classifiers in Dataset 2. These average numbers of participants for the mental tasks were 1.0, 1.1, 0.6, and 1.3 for the silent naming, visual imagery, auditory imagery, and tactile imagery task in Dataset 1, respectively, and 1.6 for the auditory imagery task in Dataset 2.

In a study by Murphy et al.²², participants were presented with images of mammals and tools. They were asked to silently name the shown object and to press a mouse button when finished (with a timeout of 3 seconds). The

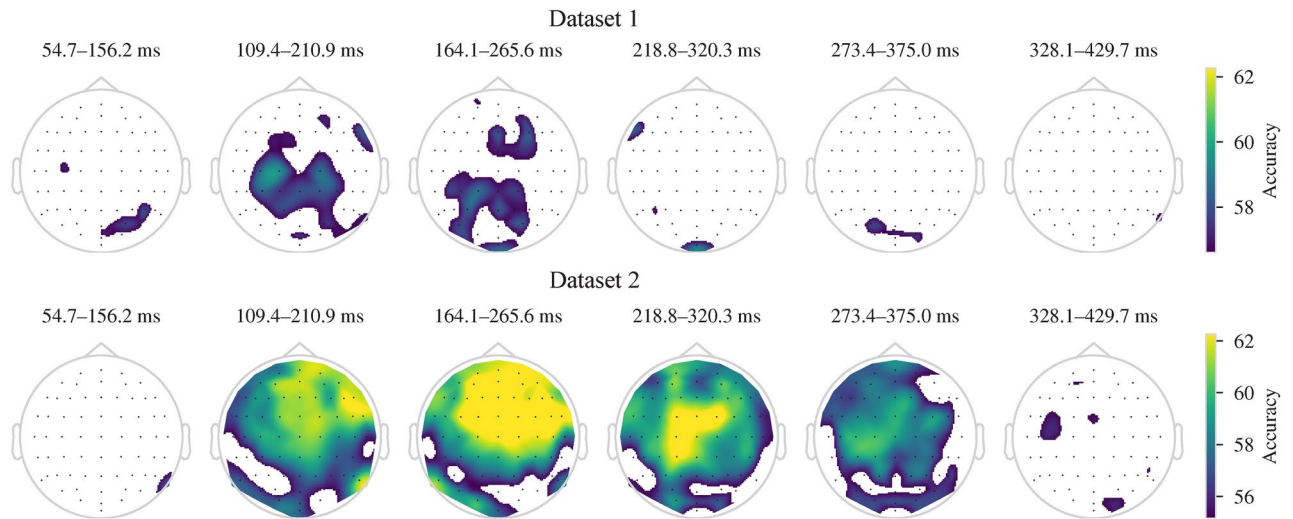


Fig. 5. Mean classification accuracies during the image presentation period for each channel using the SVM (CV). Times represent the start and end of each temporal window after the image onset. Scalp maps indicate performance above the significance borderline ($p = 0.05$). White represents non-significant classification accuracies.

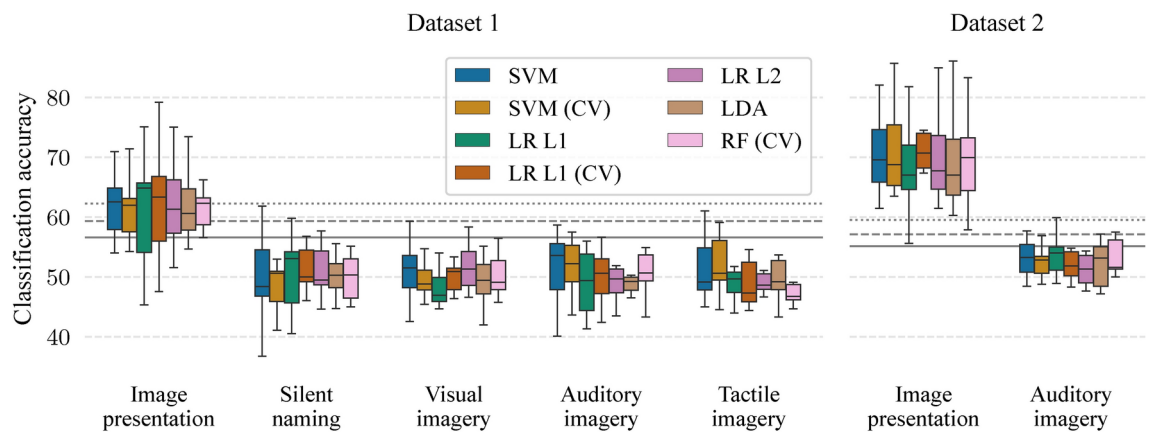


Fig. 6. Mean classification accuracies from adapted analysis²¹ in which classifiers can utilize information from all channels and the first 600 ms after the period onset. Horizontal lines represent significant borderlines for $p = 0.05$ (solid), $p = 0.01$ (dashed), and $p = 0.001$ (dotted).

image was shown for the whole task duration. Their analysis used a temporal window of 95–360 ms (after the task (or image) onset) and a frequency window of 4.1–18.3 Hz to extract epochs that were then spatially filtered by common spatial patterns (CSP)^{36,37}. A subset of N CSP components (in log variance) were classified by the SVM ($C = 1$). They were able to achieve a mean classification accuracy of 72% (with $N = 2$). We adapted their analysis by classifying $N \in \{2, 4, 6, 8, 10\}$ number of CSP components in the same temporal and frequency window.

Figure 7 shows the results of this adapted analysis using the SVM with the C parameter selected by nested-cross-validation. Mean classification accuracies for the image presentation were significant only for 8 and 10 CSP components ($p < 0.05$) in Dataset 1. In Dataset 2, mean accuracies were between 63.7% ($p < 10^{-5}$)–67.5% ($p < 10^{-6}$). For all of the mental task periods, mean classification accuracies were not statistically significant in both datasets. The average number of participants with significant accuracies across all classifiers in the image presentation was between 2.9–6.1 across different number of CSP components in Dataset 1 and between 4.1–7.0 in Dataset 2. In the mental tasks, these average number were between 0.0–2.0, 0.0–2.3, 1.0–1.7, and 0.3–2.0 for the silent naming, visual imagery, auditory imagery, and tactile imagery task in Dataset 1, respectively, and 0.0–1.0 for the auditory imagery task in Dataset 2.

Overall, both analyses revealed the possibility of differentiating between the semantic categories only during the image presentation period.

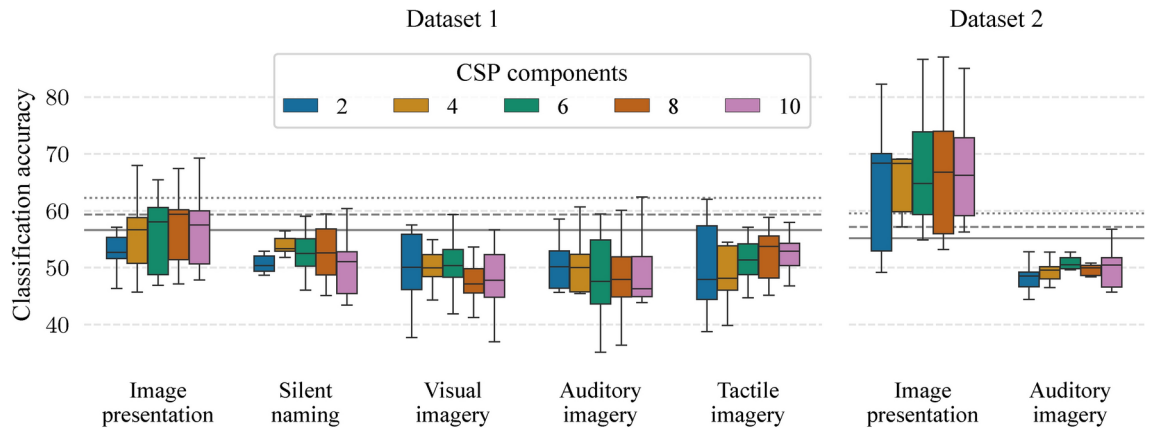


Fig. 7. Mean classification accuracies from adapted analysis²² in which the SVM (CV) can utilize a certain number of CSP components. Horizontal lines represent significant borderlines for $p = 0.05$ (solid), $p = 0.01$ (dashed), and $p = 0.001$ (dotted).

Discussion

Our results indicate a potential issue when semantic decoders are allowed to use neural activity recorded during the cue presentation period when moving toward semantic BCIs. We used state-of-the-art decoding analyses and further adapted analyses from two EEG-based semantic decoding studies to attempt to differentiate between the semantic categories of animals and tools during the image presentation and mental task periods.

Semantic decoding was possible during the image presentation period in all analyses. Mean classification accuracies were up to 69.9% when decoders could utilize information from all channels and all time points of the image presentation period. The sliding window approach revealed that the most informative period was between 54.7 and 429.7 ms after the image onset with the mean accuracy peak at the temporal window of 164.1–265.6 ms. Both of the two adapted analyses also allowed semantic decoding during the image presentation period with mean classification up to 71.3%.

However, *semantic decoding was not possible during any of the mental tasks across all participants in any of the analyses*. This is a more challenging problem for semantic BCIs, based on purely imaginary mental tasks without any external cue, than suggested by the current EEG-based semantic decoding studies. Furthermore, if the image presentation period had not been separated from the mental task period or if neural activity recorded during the image presentation period had been included with neural activity recorded during the mental task period, we could have claimed the possibility of semantic decoding with our presented analyses. A few participants had significant individual classification accuracies during some mental tasks and some classifiers, even though the semantic decoding was not possible during any of the mental tasks across all participants. In comparison, the majority of participants had significant individual accuracies during the image presentation period in all classifiers in both datasets.

The decoding performance differed between Datasets 1 and 2 when using the image presentation period. We believe there are several reasons for this. First, Dataset 2 had more repetitions per concept resulting in more trials per category to train the decoders (126 trials per category versus 90 trials in Dataset 1). Second, participants in Dataset 2 were members of our lab. Thus, there could be a difference in participants' motivation and engagement between the two datasets. Nevertheless, even these changes did not improve decoding performance for any of the mental tasks.

Perceiving an external cue and performing a mental task elicit complex neural activity that comprises, but is not limited to, perceptual processes, semantic activation caused by perception, and semantic activation caused by concept-related mental tasks. Thus, decoders might exploit the perceptual processing-related neural activity during the image presentation period. However, we do not believe that the semantic decoding performance is entirely based on perceptual processing while perceiving the image. If this were the case, we would expect to see mainly channels with significant accuracies from the brain areas related to perceptual processing. This is visual processing in our case and thus we would expect to see primarily channels from occipital areas. However, the decoders were able to utilize channels from all areas, not limited to occipital areas. Additionally, channels with maximal accuracies were not from occipital areas. The combination of motor, parietal, and frontal areas was heavily utilized by our decoders. This is especially clear on Dataset 2 (see Figures 3 and 5).

In the silent naming tasks, Murphy et al.²² reported that a wide range of occipital, parietal, and frontal areas played a role in separating the semantic categories of mammals and tools in EEG. Soto et al.³⁸ reported significant decoding of semantic categories in fMRI in the middle temporal gyrus, anterior and inferior temporal, inferior parietal lobe, and prefrontal regions. Nevertheless, prefrontal regions are typically believed to be involved in semantic control rather than in representing semantic knowledge^{39–41}.

Our imagery mental tasks differ from the mental tasks examined in previous EEG-based semantic decoding¹. Instead, our mental tasks should rather be compared with the results from mental imagery research^{42–47}. A review by McNorgan⁴³ found a general imagery network shared between different modalities. The review surveyed auditory, tactile, motor, gustatory, olfactory, and visual imagery in fMRI and PET studies. Shared activations

between modalities (regardless of mental tasks) were found in bilateral dorsal parietal, left inferior frontal, and anterior insula regions. Modality-specific imagery for most modalities was also associated with activations in corresponding sensorimotor regions (but not necessarily in primary sensorimotor areas), primarily left-lateralized.

Our imagery mental tasks did not attempt to constrain the imagery strategies used by the participants in any way. Although our participants were directed to use a particular sensory modality (e.g., tactile modality), we did not restrict participants to use only this modality. We allowed them to use anything that came naturally to them. Many participants reported using visual imagery in other imagery tasks. The current evidence suggests that the early visual cortex is involved in visual imagery when the imagery task requires high-fidelity representations^{43,48}.

Only a few semantic decoding studies attempted semantic decoding during a free recall paradigm with minimal cue support^{29,49,50}. During free recall, participants are often instructed to recall semantically related items consecutively. This is the most similar paradigm to our experimental design. For instance, a fMRI study⁴⁹ presented participants with semantic categories of animals and tools in four modalities: spoken words, written words, images, and natural sounds. Participants were asked to respond upon appearance to out-of-category exemplars. Additionally, the experiment ended with two free recall blocks in which participants were presented with a category name for 2 seconds and then instructed to recall all the entities from the cued category, seen during the experiment, in 40 seconds. A classifier was trained on the data from the actual stimulus presentation and then tested on the recall data. Mean classification accuracies across 14 participants were significant only when the classifier was trained on a combination of all modalities (67%) and partly when trained on a natural sounds modality (65%, $p = 0.01$, but below the p -value threshold for statistical significance used by the study). In general, studies that examined neural activity during free recall showed the rise and fall of the category-specific neural activity with different effects in different neuroimaging modalities^{29,50–55}.

No exploratory research can ever be completely comprehensive and we may have neglected to include some methods and approaches. We selected the current state-of-art analyses from prior EEG-based semantic decoding studies. It is important to note that the majority of our analyses considered primarily the temporal domain. This inherently assumes a time-locked neural activity to a certain degree. A high trial variability could explain the inability to achieve significant semantic decoding during the mental tasks. This would be especially the case for Dataset 1 in which participants were required to quickly switch between different mental tasks with only a 200 ms gap between the mental tasks. To avoid this issue, Dataset 2 used only one mental task with even longer periods for the image presentation and the mental task. However, this change in Dataset 2 did not improve the decoding performance. Our observations from other more complex analyses of these datasets⁵⁶, which are not reported here, suggest the need for (non-phase-locked) time-variant approaches due to considerable variability in temporal and spectral locations of statistically significant classification accuracies across participants and mental tasks. Examples of these methods are, for instance, recurrent neural networks such as long short-term memory networks^{57,58}. On the other hand, responses to stimuli often exhibit a relatively high signal-to-noise ratio (SNR) compared to responses to other cognitive events. For instance, event-related potentials (ERPs) typically have a higher SNR than event-related desynchronization/synchronization (ERD/S)⁵⁹. Furthermore, the cue-related neural activity is time-locked to the cue presentation, allowing for more effective averaging over trials compared to non-stimuli-locked events, which makes decoding easier.

We used a successful semantic decoding approach from prior studies to attempt to differentiate between the two most used semantic categories of animals and tools. Nevertheless, this broad characterization of the semantic categories could pose greater challenges when the decoding problem becomes more intricate, that is when the cue presentation and mental task periods are separated. A prediction of individual concepts might lead to better results. This is typically achieved by training a model that can predict coordinates in some feature space that can be then transformed into individual concepts. The idea is that semantically similar concepts would be located closer to each other in that feature space. There are many different approaches to model feature spaces^{1,60}. Additionally, this approach allows for generalization by zero-shot learning to predict items a model has not been trained on.

While we purposely avoided utilizing neural activity recorded during the cue presentation period for decoding purposes, experimental designs with separate blocks for only training semantic decoders could be used. In these experimental designs, neural activity evoked during the cue presentation period could be utilized to train semantic decoders, which might improve decoding performance during subsequent blocks when the cue period is no longer utilized. This approach would be based on the reinstatement of neural activity between encoding and retrieval from memory retrieval research^{29,61–63}. Despite this, recent findings indicate that even separating the cue presentation before the visual imagery task can influence decoding performance^{64,65}. Specifically, a study⁶⁴ demonstrated that visual imagery tasks preceded by either a visual cue (an image of the object) or an auditory cue (a spoken name of the object) affect outcomes differently. The authors argue that the paradigm when visual cues preceded the visual imagery tasks captures short-term retention in visual working memory^{66–68} rather than spontaneous mental imagery, which is necessary for an actual BCI application. They showed that while decoding the four image categories was possible during the visual imagery task following the visual cue, decoding was not possible during the visual imagery task following the auditory cue. Our study differs in two main aspects: their visual imagery task followed the image presentation immediately, while ours included a gap; and our visual imagery tasks were unconstrained, allowing participants to visualize their own representation of the presented concept, whereas their tasks required participants to vividly recall the specific image. Overall, these differences underscore the current challenges in developing effective imagery-based tasks for BCIs.

Methods

Participants

Dataset 1

Twelve right-handed native English speakers were recruited from the student and staff population of the University of Essex (3 males and 9 females, age range 20–57, mean 32.75, standard deviation 11.55). The recruitment search was limited to native English speakers to avoid potential differences in neural representations of semantic concepts by individuals who speak different languages^{69,70}. This restriction was done primarily for the silent naming task. All participants had normal or corrected-to-normal vision and gave written informed consent prior to the experiment. Participants received compensation of £16 for their time. The research protocol was approved by the Ethics Committee of the University of Essex on 25th October 2018. The experiment was performed in accordance with relevant guidelines and regulations. This experiment was also described in⁷¹.

Dataset 2

Seven participants were recruited from members of our lab (5 males and 2 females, age range 27–44, mean 34, standard deviation 5.98). All participants were right-handed and had normal or corrected-to-normal vision. None of the participants were native English speakers. This constraint was not required for this dataset because this experiment did not contain the silent naming task. All participants gave written informed consent before the experiment. The research protocol was approved by the Ethics Committee of the University of Essex on 20th February 2023 (ETH2223-0805). The experiment was performed in accordance with relevant guidelines and regulations.

Tasks

Participants were instructed to perform a set of four individual mental tasks (silent naming, visual, auditory, and tactile imagery) after they were presented with images from two semantic categories of animals and tools. The order of mental tasks was randomized across blocks.

In the silent naming task, participants were instructed to name the presented object in their minds (in English). In the visual imagery task, participants were instructed to visualize the presented object, but using their representation of the concept instead of the particular image they had seen. In the auditory imagery task, they were instructed to imagine sounds made by the presented object when they were interacting with it. For instance, the sounds made by an animal (such as the meowing of a cat) or the sounds produced when using a tool (such as the banging of a hammer). Lastly, in the tactile imagery task, participants were instructed to imagine the feeling of touching the presented object. For instance, when petting an animal or touching different parts of a tool. In addition to the above-mentioned descriptions and examples of the mental tasks, participants were instructed to use the imagery strategy that came most naturally to them. For all imagery tasks, they were instructed to be engaged for the whole mental task duration (3 seconds).

Dataset 2 differences

Only the auditory imagery task was used (for 5 seconds).

Stimuli

Images of 18 animals and 18 tools were sourced from the Internet. We selected the pair of semantic categories and the visual modality (images) to cue participants because they were the most used in the semantic decoding studies¹. Images were converted to gray-scale, cropped, resized to 400 × 400 pixels, and contrast stretched. The object was presented on a white background. The selected concepts are listed below.

Animals: bear, cat, cow, crab, crow, dog, donkey, duck, elephant, frog, lion, monkey, owl, pig, rooster, sheep, snake, and tiger.

Tools: axe, bottle-opener, broom, chain saw, computer keyboard, computer mouse, corkscrew, hammer, hand saw, Hoover, kettle, knife, microwave, pen, phone, scissors, shovel, and toothbrush.

Experimental design

Figure 1 shows an illustration of a concept trial. Each concept trial started with a black fixation cross on a white background for 1–2 seconds (uniformly distributed). The image of a concept was then presented for 0.6 seconds. A mask (the image of a checkerboard) followed for 0.6 seconds to reduce visual persistence and thus to get rid of potential effects of perceptual processing related neural activity from the image presentation after this mask presentation⁷². A blank white screen was shown for 0.5 seconds before a sequence of all four mental tasks. Each mental task lasted for 3 seconds and was separated from the following one by a blank white screen for 0.2 seconds. The type of the mental task was indicated by text presented on the screen for the whole mental task duration: “Silently name”, “Visualize”, “Listen”, or “Feel”. After the last task, a short break for 2 seconds was given to participants in which a blank screen was presented, which changed color over time from white to black and back. In total, the presentation of one concept took 17.3–18.3 seconds.

Each concept was presented five times, that is for a total of 90 trials per category (18 concepts, 5 repetitions each). The experiment contained 15 blocks with 12 concepts per block (207.6–219.6 seconds). Each block was followed by a break of at least 30 seconds. There was one longer break of at least 3 minutes in the middle of the experiment. Additionally, the experiment started with two short blocks (86.5–91.5 seconds, a random subset of 5 concepts, each repeated two times) for familiarization with the experiment, that are excluded from the analysis.

The order of concepts and mental tasks was pseudo-randomized. No concept was repeated twice in succession. All mental tasks in one block had the same order. Different blocks had different orders of mental tasks while no order in a given block was repeated in the following block.

Dataset 2 differences

The experiment used to record Dataset 2 was a simplified variant of the experiment used to record Experiment 1 and used only the auditory imagery task (see Fig. 1). The cue presentation period of the concept image was extended to 1 second, the duration of the auditory imagery task was extended to 5 seconds, and the blank screen period before the mental task was changed to 0.5–1 second (uniformly distributed). In total, the presentation of one concept took 10.1–11.6 seconds. Each concept was presented seven times and, thus, the total number of trials per category was increased to 126 (18 concepts, 7 repetitions each). The experiment contained 14 blocks with 18 concepts per block (181.8–208.8 seconds). The breaks between blocks were the same as in the first experiment. The order of concepts was pseudo-randomized as in the first experiment, so as not to repeat any concept twice in succession, with an additional constraint that the semantic categories are balanced in each block.

Data acquisition

EEG data were acquired with a BioSemi ActiveTwo system with 64 electrodes positioned according to the international 10–20 system, plus one electrode on each earlobe as references, with a sampling rate of 2048 Hz.

EEG data preprocessing

The EEG signals, from both datasets, were referenced to the mean of electrodes placed on the left and right earlobe. Channels with bad signal quality were manually identified and removed. The EEG data were high-pass filtered to remove slow drift artifacts by an IIR 4th-order Butterworth filter with a 1 Hz cutoff (forward-backward filtering, `sosfiltfilt` method from SciPy⁷³). Epochs representing concept trials (a presentation of a concept image with a sequence of all mental tasks) were extracted from the last second of the fixation cross, before the image presentation, until the end of the last task (14.3 seconds after the image onset in Dataset 1 and 7.1 seconds in Dataset 2). We inspected the concept trials for artifacts in any channel and marked bad concept trials that were then excluded from further analysis. Participant 8 in Dataset 1 was excluded from further analysis due to overall bad signal quality and artifacts in most concept trials.

Artifacts from eye blinks were suppressed by independent component analysis (ICA)^{74,75}. Spatial ICA (FastICA from SciPy⁷³ with default settings except with a maximum number of iterations set to 10^4 and with a number of components corresponding to a number of valid channels) was fitted on the valid concept trials for each participant. One component from the ICA that represented eye blinks was removed from the high-pass filtered data.

Analysis in the temporal domain

The analysis in the temporal domain was conducted in the two most widely used frequency bands (see Table 1): 1–30 and 4–30 Hz. In both cases, the preprocessed EEG data were first high-pass filtered by an IIR 4th-order Butterworth filter (1 or 4 Hz cutoff, forward-backward filtering, `sosfiltfilt` method) and then convolved with a FIR low-pass filter (designed by the `firwin` method from SciPy⁷³, 0.661 s length, 30 Hz cutoff, 5 Hz bandwidth)⁷⁶. Lastly, the filtered data were downsampled to 128 Hz and restructured into epochs of interest (excluding data from the excluded concept trials).

Two approaches for the classification were investigated: a sliding window approach and the classification of the whole period of interest. The sliding window approach was used to investigate the temporal evolution of semantic decoding. A temporal window of 109.375 ms was used (14 time points at a sampling frequency of 128 Hz) with a step of half the window size. A classifier was separately trained in each position of the temporal window by using only time points from that window. In the classification of the whole period, a classifier could use information from all time points in the period of interest.

In both approaches, data in each temporal window were normalized (z-scored) for each channel separately (based on training folds during cross-validation) and a classifier was trained in a stratified 15-fold cross-validation. We used these analyses on each channel separately and on all channels.

We used the following classifiers (all from `scikit-learn`⁷⁷ with default settings unless specified): the SVM with the radial basis function kernel and $C = 1$, LR with L1 or L2 norm (with a maximum number of iterations set to 10^5), LDA, and the RF classifier. Additionally, we used the SVM with a value for the C parameter chosen by a nested cross-validation, the LR with L1 norm with a regularization strength parameter value selected by a nested cross-validation, and the RF with a nested cross-validation search for a number of trees in the forest from {10, 100, 200, 500} and a maximum depth of the tree from {5, 10} or unlimited. In all instances, a stratified 10-fold inner cross-validation was used.

Decoding performance was measured by classification accuracy. Statistical significance for the mean classification accuracy across participants was calculated using a bootstrapping simulation (based on 10^6 simulations using a classifier that randomly chooses a class). Significance borderlines were then 56.6% for $p = 0.05$, 59.33% for $p = 0.01$, and 62.25% for $p = 0.001$ in Dataset 1 and 55.16%, 57.14%, and 59.52% respectively in Dataset 2.

A comparison to state-of-the-art

We adapted the analysis from a study by Simanova et al.²¹ in the following pipeline. The preprocessed EEG data were filtered in the frequency band of 1–30 Hz and the first 600 ms were extracted from periods of interest. Each channel was separately normalized (z-scored). A classifier used all channels for the classification. We tested all classifiers as in the previous section. Note that the LR with L1 norm is similar to the approach reported in²¹ (but without explicitly using a Bayesian approach). Because our image presentation in Dataset 1 was only 600 ms, in comparison with their 700 ms, we used a temporal window of 600 ms for all classified periods.

We adapted the analysis from a study by Murphy et al.²² in the following pipeline. The preprocessed EEG data were filtered in the frequency band of 4.1–18.3 Hz and epochs were extracted at the temporal window of

95–360 ms. We performed dimensionality reduction by principal component analysis (PCA) before the CSP step to avoid the issue of decreased dimensionality of EEG signals after ICA for eye blinks suppression during the preprocessing phase⁷⁸. The number of dimensions was decreased by one because one IC was removed. Then data were spatially filtered by CSP and a subset of N CSP components were extracted. The first half of N selected CSP components had the largest eigenvalues and the second half had the smallest eigenvalues. The log variance of the selected CSP components was classified by all considered classifiers. We tested $N \in \{2, 4, 6, 8, 10\}$.

Data availability

The data that support the findings of this study are available on request from the corresponding authors [M.R. and I.D.]. The data will be publicly available after the end of the ongoing project.

Code availability

The code that supports the findings of this study is available at <https://github.com/milan-rybar/semantic-neural-decoding-cue>.

Received: 27 March 2024; Accepted: 7 November 2024

Published online: 14 November 2024

References

- Rybář, M. & Daly, I. Neural decoding of semantic concepts: A systematic literature review. *J. Neural Eng.* **19**, 021002. <https://doi.org/10.1088/1741-2552/ac619a> (2022).
- Akama, H. & Murphy, B. Emerging methods for conceptual modelling in neuroimaging. *Behaviormetrika* **44**, 117–133. <https://doi.org/10.1007/s41237-016-0009-1> (2017).
- Graimann, B., Allison, B. & Pfurtscheller, G. Brain-Computer Interfaces: A Gentle Introduction. In *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*, (Springer (ed. Graimann, B.)) 1–27 (Berlin Heidelberg, 2009). https://doi.org/10.1007/978-3-642-02091-9_1.
- Nicolas-Alonso, L. F. & Gomez-Gil, J. Brain computer interfaces, a review. *Sensors* **12**, 1211–1279. <https://doi.org/10.3390/s120201211> (2012).
- Angrick, M. et al. Online speech synthesis using a chronically implanted brain-computer interface in an individual with ALS. *Sci. Rep.* **14**, 9617. <https://doi.org/10.1038/s41598-024-60277-2> (2024).
- Willett, F. R. et al. A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036. <https://doi.org/10.1038/s41586-023-06377-x> (2023).
- Metzger, S. L. et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046. <https://doi.org/10.1038/s41586-023-06443-4> (2023).
- Luo, S. et al. Stable decoding from a speech BCI enables control for an individual with ALS without recalibration for 3 months. *Adv. Sci.* **10**, 2304853. <https://doi.org/10.1002/advs.202304853> (2023).
- Metzger, S. L. et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. *Nat. Commun.* **13**, 6510. <https://doi.org/10.1038/s41467-022-33611-3> (2022).
- Moses, D. A. et al. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. *N. Engl. J. Med.* **385**, 217–227. <https://doi.org/10.1056/NEJMoa2027540> (2021).
- Zhang, L., Zhou, Y., Gong, P. & Zhang, D. Speech imagery decoding using EEG signals and deep learning: A survey. *IEEE Transactions on Cognitive and Developmental Systems* 1–18. <https://doi.org/10.1109/TCDS.2024.3431224> (2024).
- Tang, J., LeBel, A., Jain, S. & Huth, A. G. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat. Neurosci.* **26**, 858–866. <https://doi.org/10.1038/s41593-023-01304-9> (2023).
- Dash, D., Ferrari, P. & Wang, J. Decoding imagined and spoken phrases from non-invasive neural (MEG) signals. *Front. Neurosci.* **14**, 290. <https://doi.org/10.3389/fnins.2020.00290> (2020).
- Cooney, C., Folli, R. & Coyle, D. Neurolinguistics research advancing development of a direct-speech brain-computer interface. *iScience* **8**, 103–125. <https://doi.org/10.1016/j.isci.2018.09.016> (2018).
- Herff, C. & Schultz, T. Automatic speech recognition from neural signals: A focused review. *Front. Neurosci.* **10**, 204946. <https://doi.org/10.3389/fnins.2016.00429> (2016).
- Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M. & Shenoy, K. V. High-performance brain-to-text communication via handwriting. *Nature* **593**, 249–254. <https://doi.org/10.1038/s41586-021-03506-2> (2021).
- Pei, L. & Ouyang, G. Online recognition of handwritten characters from scalp-recorded brain activities during handwriting. *J. Neural Eng.* **18**, 046070. <https://doi.org/10.1088/1741-2552/ac10a0> (2021).
- Kübler, A. et al. A brain-computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. *Ann. N. Y. Acad. Sci.* **1157**, 90–100. <https://doi.org/10.1111/j.1749-6632.2008.04122.x> (2009).
- Hill, N. J. et al. A practical, intuitive brain-computer interface for communicating ‘yes’ or ‘no’ by listening. *J. Neural Eng.* **11**, 035003. <https://doi.org/10.1088/1741-2560/11/3/035003> (2014).
- Kleih, S. C. et al. The WIN-speller: a new intuitive auditory brain-computer interface spelling application. *Front. Neurosci.* **9**, 346. <https://doi.org/10.3389/fnins.2015.00346> (2015).
- Simanova, I., van Gerven, M., Oostenveld, R. & Hagoort, P. Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS ONE* **5**, e14465. <https://doi.org/10.1371/journal.pone.0014465> (2010).
- Murphy, B. et al. EEG decoding of semantic category reveals distributed representations for single concepts. *Brain Lang.* **117**, 12–22. <https://doi.org/10.1016/j.bandl.2010.09.013> (2011).
- Murphy, B. & Poesio, M. Detecting Semantic Category in Simultaneous EEG/MEG Recordings. In *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics*, CN '10, 36–44 (Association for Computational Linguistics, USA, 2010).
- Chan, A. M., Halgren, E., Marinkovic, K. & Cash, S. S. Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage* **54**, 3028–3039. <https://doi.org/10.1016/j.neuroimage.2010.10.073> (2011).
- Suppes, P., Lu, Z.-L. & Han, B. Brain wave recognition of words. *Proc. Natl. Acad. Sci.* **94**, 14965–14969. <https://doi.org/10.1073/pnas.94.26.14965> (1997).
- Alizadeh, S., Jamalabadi, H., Schönauer, M., Leibold, C. & Gais, S. Decoding cognitive concepts from neuroimaging data using multivariate pattern analysis. *Neuroimage* **159**, 449–458. <https://doi.org/10.1016/j.neuroimage.2017.07.058> (2017).
- Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S. & Bonte, M. EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Front. Psychol.* **6**, 71. <https://doi.org/10.3389/fpsyg.2015.00071> (2015).
- Behroozi, M., Daliri, M. R. & Shekarchi, B. EEG phase patterns reflect the representation of semantic categories of objects. *Med. Biol. Eng. Comput.* **54**, 205–221. <https://doi.org/10.1007/s11517-015-1391-7> (2016).

29. Morton, N. W. et al. Category-specific neural oscillations predict recall organization during memory search. *Cereb. Cortex* **23**, 2407–2422. <https://doi.org/10.1093/cercor/bhs229> (2013).
30. McFarland, D. J., Sarnacki, W. A. & Wolpaw, J. R. Brain-computer interface (BCI) operation: optimizing information transfer rates. *Biol. Psychol.* **63**, 237–251. [https://doi.org/10.1016/S0301-0511\(03\)00073-5](https://doi.org/10.1016/S0301-0511(03)00073-5) (2003).
31. Wolpaw, J. et al. Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **8**, 164–173. <https://doi.org/10.1109/TRE.2000.847807> (2000).
32. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> (1948).
33. Patterson, K., Nestor, P. J. & Rogers, T. T. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* **8**, 976–987. <https://doi.org/10.1038/nrn2277> (2007).
34. Warrington, E. K. & Shallice, T. Category specific semantic impairments. *Brain* **107**, 829–853. <https://doi.org/10.1093/brain/107.3.829> (1984).
35. Ghazaryan, G. et al. Trials and tribulations when attempting to decode semantic representations from MEG responses to written text. *Lang. Cogn. Neurosci.*[SPACE]<https://doi.org/10.1080/23273798.2023.2219353> (2023).
36. Müller-Gerking, J., Pfurtscheller, G. & Flyvbjerg, H. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.* **110**, 787–798. [https://doi.org/10.1016/S1388-2457\(98\)00038-8](https://doi.org/10.1016/S1388-2457(98)00038-8) (1999).
37. Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. & Müller, K.-R. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* **25**, 41–56. <https://doi.org/10.1109/MSP.2008.4408441> (2008).
38. Soto, D., Sheikh, U. A., Mei, N. & Santana, R. Decoding and encoding models reveal the role of mental simulation in the brain representation of meaning. *R. Soc. Open Sci.* **7**, 192043. <https://doi.org/10.1098/rsos.192043> (2020).
39. Wagner, A. D., Paré-Blagojev, E., Clark, J. & Poldrack, R. A. Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron* **31**, 329–338. [https://doi.org/10.1016/S0896-6273\(01\)00359-2](https://doi.org/10.1016/S0896-6273(01)00359-2) (2001).
40. Poldrack, R. A. et al. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* **10**, 15–35. <https://doi.org/10.1006/nimg.1999.0441> (1999).
41. Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A. & Jefferies, E. The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex* **21**, 1066–1075. <https://doi.org/10.1093/cercor/bhq180> (2011).
42. Pearson, J. The human imagination: the cognitive neuroscience of visual mental imagery. *Nat. Rev. Neurosci.* **20**, 624–634. <https://doi.org/10.1038/s41583-019-0202-9> (2019).
43. McNorgan, C. A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Front. Hum. Neurosci.* **6**, 285. <https://doi.org/10.3389/fnhum.2012.00285> (2012).
44. Kosslyn, S. M., Thompson, W. L. & Ganis, G. *The Case for Mental Imagery* (Oxford University Press, 2006).
45. Kosslyn, S. M., Ganis, G. & Thompson, W. L. Neural foundations of imagery. *Nat. Rev. Neurosci.* **2**, 635–642. <https://doi.org/10.1038/35090055> (2001).
46. Nanay, B. Multimodal mental imagery. *Cortex* **105**, 125–134. <https://doi.org/10.1016/j.cortex.2017.07.006> (2018).
47. Lacey, S. & Sathian, K. Multisensory object representation: Insights from studies of vision and touch. *Progress Brain Res.* **191**, 165–176. <https://doi.org/10.1016/B978-0-444-53752-2.00006-0> (2011).
48. Kosslyn, S. M. & Thompson, W. L. When is early visual cortex activated during visual mental imagery?. *Psychol. Bull.* **129**, 723–746. <https://doi.org/10.1037/0033-2909.129.5.723> (2003).
49. Simanova, I., Hagoort, P., Oostenveld, R. & van Gerven, M. A. J. Modality-independent decoding of semantic information from the human brain. *Cereb. Cortex* **24**, 426–434. <https://doi.org/10.1093/cercor/bhs324> (2014).
50. Polyn, S. M., Natu, V. S., Cohen, J. D. & Norman, K. A. Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963–1966. <https://doi.org/10.1126/science.1117645> (2005).
51. Sederberg, P. B. et al. Gamma oscillations distinguish true from false memories. *Psychol. Sci.* **18**, 927–932. <https://doi.org/10.1111/j.1467-9280.2007.02003.x> (2007).
52. Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R. & Fried, I. Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* **322**, 96–101. <https://doi.org/10.1126/science.1164685> (2008).
53. Long, N. M., Oztekin, I. & Badre, D. Separable prefrontal cortex contributions to free recall. *J. Neurosci.* **30**, 10967–10976. <https://doi.org/10.1523/JNEUROSCI.2611-10.2010> (2010).
54. Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B. & Kahana, M. J. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proc. Natl. Acad. Sci.* **108**, 12893–12897. <https://doi.org/10.1073/pnas.1015174108> (2011).
55. Polyn, S. M., Kragel, J. E., Morton, N. W., McCluey, J. D. & Cohen, Z. D. The neural dynamics of task context in free recall. *Neuropsychologia* **50**, 447–457. <https://doi.org/10.1016/j.neuropsychologia.2011.08.025> (2012).
56. Rybář, M. *Towards EEG/fNIRS-Based Semantic Brain-Computer Interfacing* (University of Essex, 2023).
57. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444. <https://doi.org/10.1038/nature14539> (2015).
58. Craik, A., He, Y. & Contreras-Vidal, J. L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **16**, 031001. <https://doi.org/10.1088/1741-2552/ab0ab5> (2019).
59. Pfurtscheller, G. & Lopes da Silva, F. Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clin. Neurophysiol.* **110**, 1842–1857. [https://doi.org/10.1016/S1388-2457\(99\)00141-8](https://doi.org/10.1016/S1388-2457(99)00141-8) (1999).
60. Bruffaerts, R. et al. Redefining the resolution of semantic knowledge in the brain: Advances made by the introduction of models of semantics in neuroimaging. *Neurosci. Biobehav. Rev.* **103**, 3–13. <https://doi.org/10.1016/j.neubiorev.2019.05.015> (2019).
61. Wing, E. A., Ritchey, M. & Cabeza, R. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *J. Cogn. Neurosci.* **27**, 679–691. https://doi.org/10.1162/jocn_a_00740 (2015).
62. Yaffe, R. B. et al. Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci.* **111**, 18727–18732. <https://doi.org/10.1073/pnas.1417017112> (2014).
63. Yaffe, R. B., Shaikhouni, A., Arai, J., Inati, S. K. & Zaghoul, K. A. Cued memory retrieval exhibits reinstatement of high gamma power on a faster timescale in the left temporal lobe and prefrontal cortex. *J. Neurosci.* **37**, 4472–4480. <https://doi.org/10.1523/JNEUROSCI.3810-16.2017> (2017).
64. Kilmarx, J., Tashev, I., Millán, J. R., Sulzer, J. & Lewis-Peacock, J. Evaluating the feasibility of visual imagery for an EEG-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 2209–2219. <https://doi.org/10.1109/TNSRE.2024.3410870> (2024).
65. Kosmyna, N., Lindgren, J. T. & Lécuyer, A. Attending to visual stimuli versus performing visual imagery as a control strategy for EEG-based brain-computer interfaces. *Sci. Rep.* **8**, 13222. <https://doi.org/10.1038/s41598-018-31472-9> (2018).
66. Keogh, R. & Pearson, J. Mental imagery and visual working memory. *PLoS ONE* **6**, e29221. <https://doi.org/10.1371/journal.pone.0029221> (2011).
67. D'Esposito, M. & Postle, B. R. The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* **66**, 115–142. <https://doi.org/10.1146/annurev-psych-010814-015031> (2015).
68. Xie, S., Kaiser, D. & Cichy, R. M. Visual imagery and perception share neural representations in the alpha frequency band. *Curr. Biol.* **30**, 2621–2627.e5. <https://doi.org/10.1016/j.cub.2020.04.074> (2020).
69. Yang, Y., Wang, J., Bailer, C., Cherkassky, V. & Just, M. A. Commonalities and differences in the neural representations of English, Portuguese, and Mandarin sentences: When knowledge of the brain-language mappings for two languages is better than one. *Brain Lang.* **175**, 77–85. <https://doi.org/10.1016/j.bandl.2017.09.007> (2017).

70. Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T. & Raizada, R. D. S. Semantic structural alignment of neural representational spaces enables translation between English and Chinese words. *J. Cogn. Neurosci.* **28**, 1749–1759. https://doi.org/10.1162/jocn_a_01000 (2016).
71. Rybář, M., Poli, R. & Daly, I. Decoding of semantic categories of imagined concepts of animals and tools in fNIRS. *J. Neural Eng.* **18**, 046035. <https://doi.org/10.1088/1741-2552/abf2e5> (2021).
72. Rolls, E. T. & Tovee, M. J. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. R. Soc. Lond. B* **257**, 9–15. <https://doi.org/10.1098/rspb.1994.0087> (1994).
73. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2> (2020).
74. Jung, T.-P. et al. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**, 163–178. <https://doi.org/10.1111/1469-8986.3720163> (2000).
75. Islam, M. K., Rastegarnia, A. & Yang, Z. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology* **46**, 287–305. <https://doi.org/10.1016/j.neucli.2016.07.002> (2016).
76. Widmann, A., Schröger, E. & Maess, B. Digital filter design for electrophysiological data - a practical approach. *J. Neurosci. Methods* **250**, 34–46. <https://doi.org/10.1016/j.jneumeth.2014.08.002> (2015).
77. Pedregosa, F. et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
78. Rybář, M., Daly, I. & Poli, R. Potential pitfalls of widely used implementations of common spatial patterns. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 196–199. <https://doi.org/10.1109/EMBC44109.2020.9176314> (2020).

Author contributions

All authors conceived the experiments, M.R. conducted the experiments, and M.R. analyzed the results. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.R. or I.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024