Computers in Human Behavior Reports 18 (2025) 100677

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Electrophysiological evidence for differential semantic processing of words and objects presented in augmented reality

Tom Bullock [a,b,*,1], Emily Machniak [a,b,1], Joyce Passananti [c], Kangyou Yu [c], Radha Kumaran [c], You-Jin Kim [c,2], Jordan Garrett [a,b], Madhav Viswesvaran [a], Julia Ram [a], Tobias Höllerer [b,c], Barry Giesbrecht [a,b]

[a] Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, United States
[b] Institute for Collaborative Biotechnologies, University of California, Santa Barbara, Santa Barbara, CA, United States
[c] Department of Computer Science, University of California, Santa Barbara, Santa Barbara, CA, United States

ARTICLE INFO

Keywords:
Augmented reality
EEG
ERP
Classification
N400
3D models

ABSTRACT

Head-worn augmented reality (AR) offers exciting possibilities to help users learn new information. By blending digital content with the learner's real-world environment, AR can create engaging and enjoyable experiences that may improve knowledge retention. Electroencephalography (EEG) allows for discreet, continuous monitoring of brain activity at the scalp. This study examined whether the N400 event-related potential (ERP), a brain response linked to semantic processing, could be incorporated into a system combining AR and EEG. While the N400 is reliably elicited by a mismatch in meaning between two sequentially presented stimuli, there are two key outstanding questions. First, how do 3D objects presented in AR impact semantic processing as measured by the N400? Second, is there a reliable N400 to mismatches between an object and its name, in addition to mismatches in meaning? Twenty-four young adults viewed sequential pairs of stimuli through an AR headset while EEG was recorded. We manipulated whether the first stimulus was a 3D object or written word and whether the second stimulus matched or mismatched the first in terms of meaning or name. Participants' reaction times were slower for mismatching pairs when compared to matching pairs in all conditions. Averaged ERP and single-trial classification analyses showed robust differences in brain responses. Additionally, participants were more accurate and exhibited faster behavioral and brain responses for naming compared to meaning judgments. These results suggest the N400 could be used in a combined AR-EEG system to provide feedback on semantic understanding, potentially opening exciting new avenues for enhancing learning.

1. Introduction

Success in many areas of life—from acing an exam, to delivering a compelling presentation at work, to learning a new language—often depends on one's ability to learn and remember information effectively. Continuously testing oneself on new information throughout the learning process is a powerful learning strategy (Roediger & Butler, 2011), presumably because it requires active retrieval of information from memory, strengthening connections in the neural pathways associated with that information (Ye et al., 2020). AR offers several benefits for learning that may enhance these learning processes because it can blend educational content with relevant objects in a learner's everyday environment, creating a more engaging and enjoyable learning experience and leading to improved knowledge retention (Dunleavy et al., 2009; Ibrahim et al., 2018). As a first step, here we test whether there is evidence for semantic knowledge of objects presented in augmented reality (AR), using simultaneously measured brain activity via electroencephalography (EEG). EEG is an important tool that allows for continuous, noninvasive monitoring of rapid perturbations in human brain activity (Giesbrecht & Garrett, 2025). We propose that brain activity based recognition could be achieved by tracking a specific neural response known as the N400 which is linked to the processing of meaning (Kutas & Federmeier, 2011). The goal of this study is to investigate whether the N400 has the potential to be used in a combined

* Corresponding author. Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, United States.
E-mail address: tombullock@ucsb.edu (T. Bullock).
[1] Tom Bullock and Emily Machniak contributed equally to the article.
[2] You-Jin Kim is now at Texas A&M University, College Station, TX, United States.

https://doi.org/10.1016/j.chbr.2025.100677
Received 26 February 2025; Received in revised form 16 April 2025; Accepted 24 April 2025
Available online 24 April 2025

AR-EEG setup to provide feedback on object recognition, potentially making it a viable tool for assessing the presence of semantic knowledge about objects in the environment.

The N400 is a brain signal that occurs in response to conceptually meaningful stimuli (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980; Lau et al., 2008). It is often studied by presenting pairs of words one after the other, where the second word is either semantically congruent or incongruent with the first (Deacon et al., 2000; Gomes et al., 1997; Holcomb, 1988; Holcomb & Neville, 1990). When the word pair is semantically incongruent, a large negative deflection in the waveform is observed over central scalp regions, beginning approximately 400 ms after the onset of the second word. The N400 can be elicited by a wide range of stimulus types, including sounds, images, videos, objects and actions (Kutas & Federmeier, 2011) and can be observed across species (Boros et al., 2024). It is well characterized and has been used to study a wide range of phenomena, such as language comprehension, memory and attention (Bornkessel-Schlesewsky & Schlesewsky, 2019; Coronel & Federmeier, 2016; Fitz & Chang, 2019; Giesbrecht et al., 2007; Hodapp & Rabovsky, 2021; Kuperberg et al., 2020; Rabovsky et al., 2018; Sy et al., 2013). Although the N400 was originally labeled as a relative negativity peaking around 400 ms, the timing and shape of the response can vary as a function of experimental manipulation, so we use the term N400 here as a heuristic label for stimulus-related brain activity in the ~300–600 ms post-stimulus window with a characteristic waveform shape (Kutas & Federmeier, 2011). Because the N400 indexes the meaning of information relative to an individual's understanding of that information, it is a signal of interest for BCIs. Indeed, BCI researchers have used the N400 for a variety of applications (Dijkstra et al., 2020; Kaufmann et al., 2011, 2013; Lytaev, 2021; van Vliet et al., 2010).

Given that the N400 brainwave is triggered by unexpected or incongruent information, it could plausibly be used to detect whether a user has recognized (and thus learned) an object correctly. For example, imagine a scenario where a medical student is using AR glasses to help them study human anatomy. The student brings up a digital 3D model of the human musculoskeletal system in the glasses and then fixates on a specific muscle and attempts to recall its name. The glasses then show a text label next to the muscle with either the correct or incorrect name. If the label is incorrect and the student *knows* that it is incorrect, then their brain activity should reflect this mismatch and an N400 response should be observed. However, if they do not know the correct name and fail to recognize that the label is incorrect, then we would not expect an N400 to be observed. Thus, a combined AR-EEG system capable of detecting the presence of the N400 could be used to track user understanding and provide personal feedback. Another situation that might involve a user looking at objects in the physical world, rather than in a digital environment, is language learning - a scenario where AR can be a particularly powerful tool (Huynh et al., 2019a, Huynh et al., 2019b; Ibrahim et al., 2018). For example, imagine a student learning French. They look at a computer and try to remember the French word for it ("ordinateur"). Their AR glasses then show a word—it might be the correct translation ("ordinateur") or an incorrect one ("bouteille", meaning water bottle). If the incorrect word is shown and the student *knows* that it is incorrect, then an N400 mismatch response would be expected in their brain activity. But if they don't know the correct word, this signal would not appear, suggesting that they don't understand the word and require more practice. The AR system could then relay this information to the student and suggest words that need further revision.

To determine the viability of the N400 signal there are two knowledge gaps that need to be addressed. First, it is unclear how 3D objects presented in AR impact the N400 response. The N400 is typically studied using pairs of words, 2D images on a screen, sounds, or combinations of these stimuli (Barrett & Rugg, 1990; Calma-Roddin & Drury, 2020; Ganis et al., 1996; Holcomb, 1988; Lin et al., 2022; Nigam et al., 1992; Ortu et al., 2013). However, to our knowledge there have been no studies that have analyzed N400 responses to 3D stimuli presented via an AR headset. Considering that the dynamics of this endogenous signal are consistent across multiple stimulus presentation modalities (Federmeier & Laszlo, 2009), it stands to reason that the N400 will respond similarly to 3D object and word pairs. Second, the N400 has largely been studied in the context of a semantic mismatch between two stimuli presented in a sequential fashion. Here, the goal is to determine whether the N400 can provide a reliable signal to provide feedback on object recognition (i.e., naming), but the N400 is not well characterized for mismatches of this nature. Additionally, combining AR and EEG will present unique technical challenges, as it will require precise synchronization between the stimuli presented in the AR environment and the corresponding brain responses recorded in the EEG. A major challenge is that AR devices often lack the necessary ports (or access to the necessary ports) for sending event markers to EEG amplifiers, which is crucial for precise synchronization between the two systems.

In the present study we address the previously outlined knowledge gaps and technical challenges. Participants completed variants of a pair-matching task, presented via an AR headset while brain activity was simultaneously recorded at the scalp with EEG. The task involved the presentation of a pair of stimuli on each trial. To investigate our first research question—the impact of 3D object stimuli on the N400—we manipulated the first stimulus in the pair so that it was either a written word (as is common in pair-matching tasks that are used in investigations of the N400) or a 3D object. To investigate our second research question—semantic associations versus naming associations—we manipulated the second stimulus so that it was either congruent or incongruent in meaning (i.e., semantics) or in terms of name (i.e., label). In order to ensure accurate synchronization across AR and EEG devices we also devised and validated a hardware solution that involved sending pulses via the headphone jack on the AR device at the same time as critical visual stimuli were presented via the headset. The two independent variables were manipulated in a full factorial design to create four conditions. Results showed that both behavioral and brain responses were different in all conditions, such that reaction times were slower to mismatching compared to matching stimulus pairs, and both ERP and classification analyses showed robust differences in neural responses to the second stimulus. Furthermore, participants responded more rapidly when they were required to make naming judgments compared to semantic association judgments. Together, the results represent a promising first step towards using the N400 in a system that combines AR and EEG to provide feedback on object recognition, potentially opening exciting new avenues for enhancing how we learn information.

2. Methods and materials

2.1. Participants

Twenty-four healthy, English language proficient adults aged 18–32 (15 females, 22 right handed, $M_{age} = 21.25$, $SD_{age} = 3.33$) volunteered to participate in the study. Participants reported normal or corrected-to-normal vision. Informed consent was provided at the beginning of each session. Participants received $15/hr for their participation (~2 h total per participant). All procedures were approved by the University's Human Subjects Committee.

The sample size was determined based on three factors. First, a power analysis ($\alpha = .05$, power $= .80$, large effect size) indicated a minimum of 15 participants. Second, a literature review showed that early N400 research typically used samples of 24 or fewer (Holcomb, 1988; Kutas & Van Petten, 1988), and N400 BCI studies carried out with non-clinical populations often have samples of 12–20 (Dijkstra et al., 2020). Third, full counterbalancing of our four conditions required a minimum of 24 participants. Considering these factors, we determined that a sample size of 24 was suitable.

2.2. Cognitive task

The task was presented to participants using a Magic Leap 1 AR headset (ML1, Magic Leap, Plantation, FL, USA). The task was implemented and controlled using custom scripts for Unity3D (2022.3.21f1 version, Unity Technologies, San Francisco, USA). Each trial of the task involved the sequential presentation of a pair of stimuli. The first stimulus could either be a word or a 3D object model, while the second stimulus was always a word. Each trial began with the presentation of an "X" at the center of the field of view (FOV) of users (1 s), acting as the fixation point. The first stimulus then appeared at the center of the FOV and remained on screen for 1 s. The first stimulus was either a word (e.g., "Sunglasses") or an object (e.g., sunglasses 3D model). Following a 1 s inter-stimulus interval, the second stimulus was then presented at the center of the FOV for 1 s. During the appearance of the second stimulus (e.g., "Eyes"), responses were made for that trial. Each trial took 6 s and this trial sequence repeated until the task reached either the end of a task block (i.e., block break) or the end of the task.

Two independent task manipulations were implemented to assess the impact of judgement type and stimulus modality on the N400 response. For the judgment manipulation, participants were required to make labeling or association judgments for each stimulus pair. For the modality manipulation, the first stimulus in the pair was either a 3D object or a written word. The "judgement" and "modality" variables were factorially combined to create four task conditions: model-word association (MWA), word-word association (WWA), model-word label (MWL), and word-word label (WWL). Conditions were presented in blocks of 100 trials.

In each condition, participants were required to indicate on each trial whether or not the stimulus pair was congruent (i.e., matching) or incongruent (i.e., non-matching). The nature of this congruence judgment was manipulated between task conditions. In the "association" conditions, participants were required to indicate if the second stimulus was semantically associated (i.e., congruent) or unassociated (i.e., incongruent) with the first stimulus. For example, in a MWA trial, if the model "clock" was followed by the word "time", the correct response would be congruent because clock and time are semantically associated. If, however, the model "clock" was followed by the word "syrup", the correct response would be incongruent. In a WWA trial, if the word "waffle" was followed by the word "syrup", the correct response would be congruent, again, because these words are semantically associated. If, "waffle" was followed by the word "time", the correct response would be incongruent. In the "label" conditions, participants were required to indicate if the second stimulus was the correct (i.e., congruent) or incorrect (i.e., incongruent) label for the first stimulus in the pair. For example, on a MWL trial, if the model "waffle" was followed by the word "waffle", the correct response would be congruent, but if the word was "western", the response would be incongruent. In a WWL trial, if the word "clock" was followed by the word "clock", the correct response would be congruent, but if it was followed by the word "syrup", the correct response would be incongruent. Participants registered their responses via a wireless keyboard (Arteck 2.4 GHz Compact Keyboard) by pressing the "m" key if they judged the pair to be congruent or the "z" key if they judged the pair to be incongruent.

The 3D models and words that were presented in the task were specifically selected so that they were unambiguous and relatively straightforward for participants to identify and respond to. The stimuli spanned a diverse set of categories representing familiar objects encountered in daily life, including household items (e.g., lamp, bed, dresser), office essentials (e.g., desk, monitor, chair), kitchenware (e.g., pan, silverware, appliances), sports equipment, and objects found in outdoor or public spaces. See Fig. 1a and b for a diagram of the trial sequence and examples of the stimulus pairs used in each condition.

2.3. General Procedure and EEG instrumentation

General Procedure. Upon arrival, the experimenter outlined the study and obtained informed consent from the participant. The participant was then fitted with an AR headset and completed practice tasks for WWL and MWA. After the participant completed the two practice tasks and confirmed no symptoms related to cybersickness, the AR headset was removed and EEG setup commenced.

While seated in a chair, the participant was fitted with an EEG cap (see *EEG Instrumentation*). The AR headset was then positioned over the EEG cap (see Fig. 1c). The experimenter provided task instructions and indicated the judgement type of the two response keys prior to each task condition. The participant was instructed to maintain fixation at center throughout each trial and to respond as quickly and accurately as possible. They were also instructed to relax and minimize head and body movements. Participants were given the opportunity to take a rest break halfway through each block and between blocks (~12 min total per condition/block). Condition order was counterbalanced between participants. Prior to each condition, participants completed a practice task (14 trials per condition) to ensure that they were familiar with the task and instructions. Upon completion of the full experiment the participant was thanked and paid for their time in the lab. Each session lasted ~2 h, including ~1 h for EEG setup. The full experimental procedure is outlined in Fig. 1d.

EEG Instrumentation and Recording. EEG data was recorded using a Brain Products ActiCHamp system (actiCHamp Plus, Brain Products GmbH, Gilching, Germany) consisting of 64 active electrodes arranged in an actiCAP elastic cap and placed in accordance with the 10–20 System. Electrodes TP9 and TP10 were adhered directly to the right and left mastoids. Connections were established between electrodes and the scalp using a viscous gel (SuperVisc, Brain Products). At the beginning of the investigation all impedances were reduced to below 15 kΩ. Data were sampled at 1000 Hz.

2.4. EEG and AR event synchronization

Precise time-synchronization between EEG recording and stimulus presentation within the AR headset is required to generate accurate ERPs. In traditional EEG/stimulus presentation setups, event codes are sent from a stimulus presentation computer to the EEG amplifier via a cable (TTL pulse). This is referred to as a "hardware trigger" approach. Hardware triggers, which are saved alongside EEG data at the exact time of their occurrence, offer the most precise way to synchronize EEG with stimulus presentation, which is critical when one is investigating event-related brain responses. Alternative synchronization methods are possible, such as sending event codes over a network from stimulus presentation device to EEG machine, or synchronizing event timestamps from the stimulus presentation and EEG machine data log files. However, while these approaches may eliminate the need for additional hardware, they sacrifice timing precision by introducing jitter, which can impact data quality. The Magic Leap 1 headset only has a single USB C slot which we were unable to access to send TTL pulses via Unity, so this standard approach was not viable. Instead, we developed an alternative custom hardware solution for sending event codes from the Magic Leap 1 headset to the EEG amplifier (Fig. 2). Specifically, a 3.5 mm audio cable was plugged into the headphone jack of the Magic Leap compute pack and fed into the output of a StimTrak device (Brain Products) which was connected to the EEG amplifier. During the task, at the onset of the second visual stimulus in each pair, a brief tone was played. This tone was not audible to the participant, but instead the sound was transmitted via the audio cable to the StimTrack, where it was converted to an electrical pulse which was precisely synchronized and stored alongside the recorded EEG data. Each tone created a large voltage increase from baseline, which could easily be detected and then converted into a timestamp. The timestamps were then synchronized with the trial structure and converted into meaningful event codes that could then be

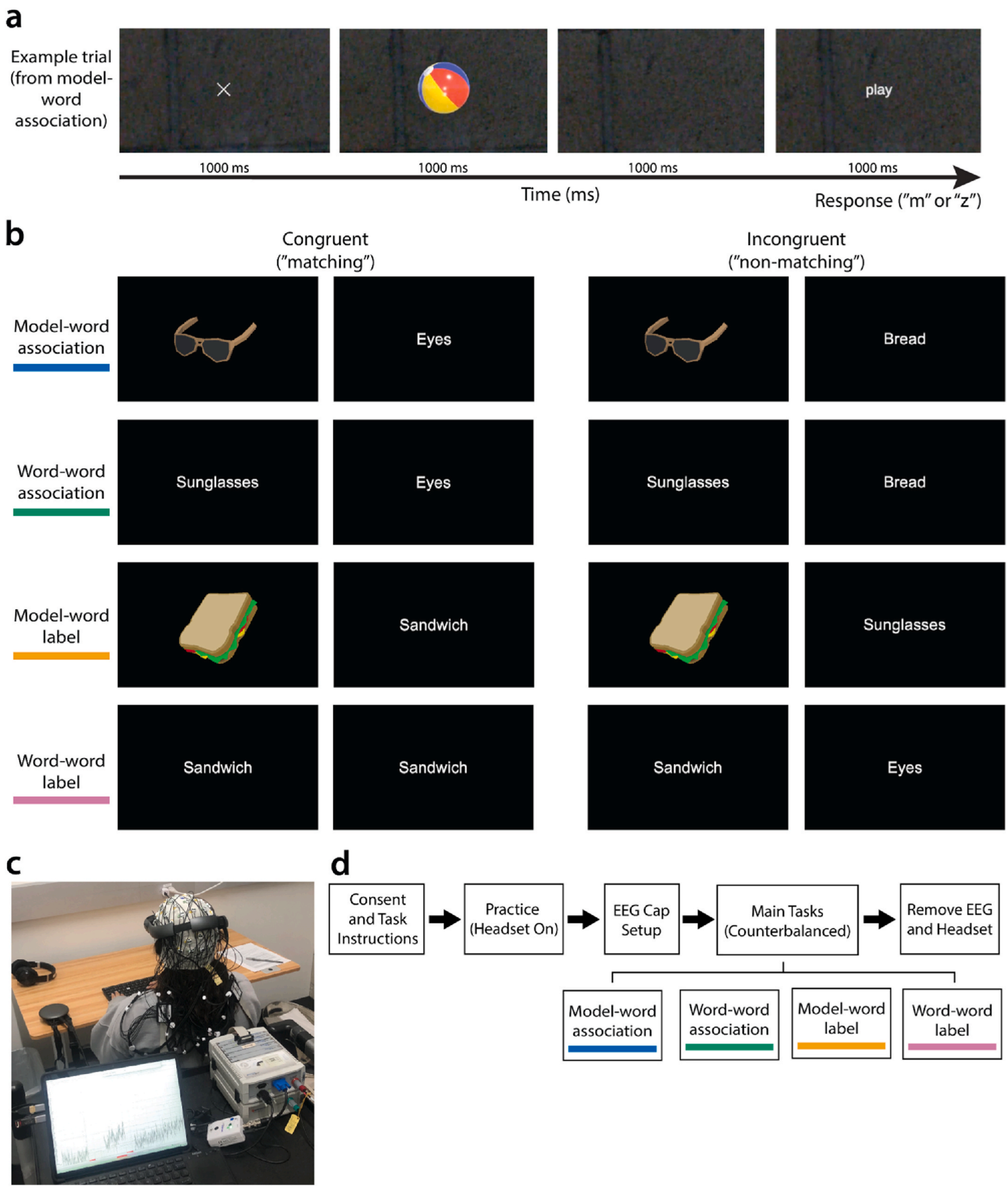

Fig. 1. ***Methods.*** (a) Trial sequence from the participant view within the AR headset (MWA trial example depicted above). Each trial began with a fixation cross (1000 ms), followed by the first stimulus (1000 ms), an inter-stimulus interval (1000 ms) and then the second stimulus (1000 ms). The participant had a brief time window (2000 ms) from the onset of the second stimulus in which to indicate with a keypress whether the first and second stimuli were congruent ("m") or incongruent ("z"). In this example the beach ball is associated with "play", so the correct response would be "m". (b) Schematic representations of stimulus pairs are shown for each of the four conditions for both congruent and incongruent trial examples. (c) A fully instrumented participant performing the task. (d) Overview of the complete experimental procedure.

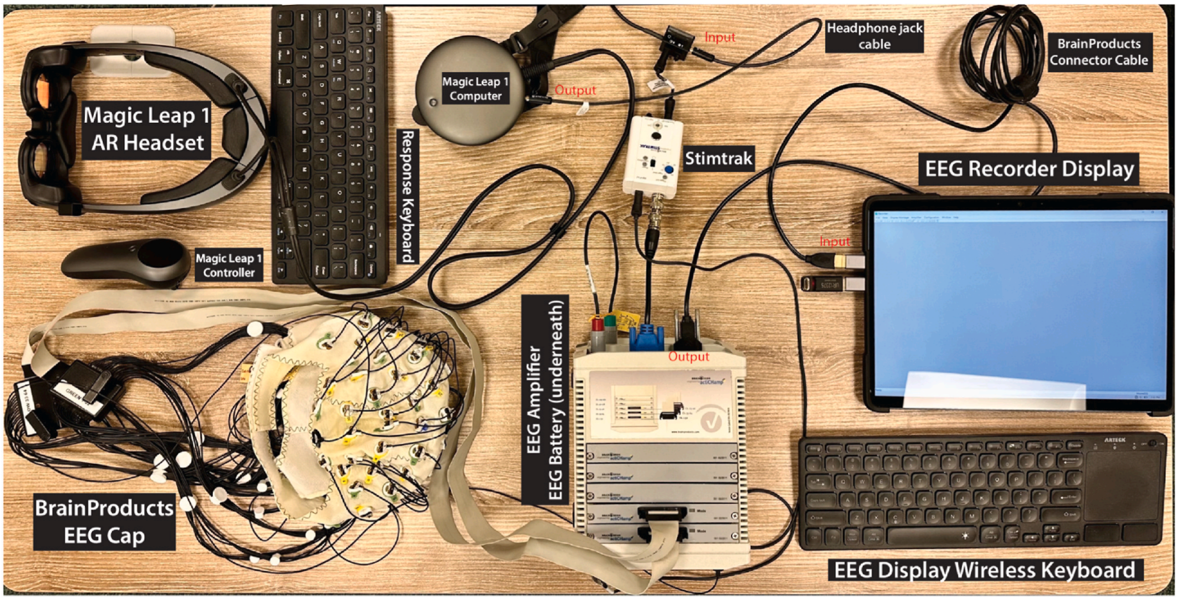

Fig. 2. ***Equipment.*** Participants were fitted with an EEG cap and AR headset and made responses via a keyboard. To precisely mark the second stimulus onset in our EEG recording, we also played a tone, which was not heard by the participant but rather sent to a separate device (StimTrack, Brain Products) via an audio cable connected to the headphone jack of the Magic Leap 1 headset. This tone was converted to a pulse, which was synchronized with the brain data, providing a precise timestamp.

used to parse the data into epochs in EEGLAB.

Next, we ran a test to validate the timing consistency of our hardware triggering approach. It was important to check whether there was any lag between the presentation of a stimulus on the Magic Leap screen and the onset of the tone, and if so, how consistent was this lag across trials. To determine the characteristics of any lag, we wrote a simple script in Unity that repeatedly rendered a high-contrast 3D model (a large white sphere) to the Magic Leap screen and simultaneously played a tone. The sphere and tone were presented 120 times at a rate of 1 Hz. To measure the precise onset of the visual stimulus (3D sphere), we attached a photodiode (Brain Products) to the Magic Leap screen and fed the photodiode into the EEG amplifier. When each sphere appeared, this caused a large voltage increase in the photodiode trace relative to baseline, which could easily be identified. The audio trace was converted to an electrical signal using the hardware approach described above, and the pulses generated by the tones were also easily identified. For each stimulus presentation, we were then able to compare the precise ground truth onsets for both visual and auditory stimuli and establish whether there was a lag between the two, and how consistent the lag was over many trials. We collected 120 trials of data and found that the mean lag between the presentation of the visual and auditory stimuli was 110 ms with 11 ms standard deviation. We then corrected the timing in all EEG analyses by the mean lag value.

2.5. EEG preprocessing

EEG Initial Preprocessing. To prepare the raw EEG data for analysis, we first converted the audio pulse triggers to meaningful event codes as described in Section 2.4 and next applied a 30 Hz low-pass filter to reduce muscle noise and a .1 Hz high-pass filter to remove slow drifts. The data were then downsampled from 1000 Hz to 250 Hz for faster processing. Noisy electrodes, identified based on low correlation with surrounding channels ($r < .85$), flatlining (>5 s), or excessive values (>4 standard deviations of the overall population), were interpolated using spherical interpolation. On average, 4.79 ± 2.62 (mean ± std) electrodes were interpolated per participant. All processing was performed offline using custom MATLAB scripts and EEGLAB toolbox functions (Delorme & Makeig, 2004).

ERP and Classification Preprocessing. For both ERP and classification analyses we first re-referenced the EEG data and removed eye-movement artifacts using the AAR toolbox (Gomez-Herrero et al., 2006). Then, to isolate brain activity specifically related to the second word in each pair, we epoched the data around the onset of the second word (from −200 to 1000 ms) and baseline-corrected it using the 200ms pre-stimulus period. Trials with extreme voltage values [±100 μV] across a grouping of critical scalp channels ('CPz', 'CP3′, 'CP4′, 'C3′, 'Cz', 'C4′, 'FC3′, 'FCz', 'FC4′, 'F3′, 'Fz', 'F4′) were excluded (<2 % across all conditions) and incorrect responses were also removed. For ERP analysis the remaining trials were then averaged separately for congruent and incongruent word pairs to create averaged waveforms. For classification analysis, the trials were not averaged.

Quantifying ERPs. To analyze the differences in brain responses to congruent and incongruent word pairs, we created difference waves by subtracting the congruent ERP waveform from the incongruent waveform. This allowed us to clearly visualize and quantify how these brain responses diverged over time. We then identified the peak latency of each difference wave, which represents the time point with the maximum positive amplitude within the time window where the N400 would reasonably be expected to be observed (200–800 ms). If no clear peak was present for an individual participant and condition, we used the average peak latency for that condition. Finally, we calculated the mean amplitude of the difference wave by averaging the data points within a ±20 ms window around the peak latency. This approach allowed us to quantify the timing and magnitude of the brain's responses to our factorial combination of stimuli.

2.6. Machine learning approach

Classifying Stimulus Pair Congruency. A single logistic regression model was trained on voltage potentials recorded at each electrode across all trials to classify the congruency between presented stimulus pairs (i.e., incongruent versus congruent trials). The model was evaluated using a 10-fold cross-validation scheme, with each fold constrained to include an equal number of trials per condition and a comparable number of congruent and incongruent trials. Within each training fold, data were augmented by averaging voltage potentials over time using a

10 ms sliding window (step size = 1 sample), enabling the model to learn from the entire temporal profile of each trial. For testing, the trained model was applied to the held-out fold, and predictions were generated over time using average voltage activity within a 10 ms sliding window (step size = 1 sample). This approach enabled the assessment of when the trial congruency-related information was most discriminable. Model performance and generalizability were quantified using a balanced accuracy score. In short, this metric is computed by averaging the recall rate, or the number of true positives (TP) divided by total number of TP and false negatives, across class labels. The advantage of using this metric over a traditional un-balanced accuracy score is that it ensures above chance decoding is not driven by the model learning the prevalence of each class, but rather is due to differences in neural activity. Balanced accuracy scores across validation folds were averaged to yield a single measure of classification performance, which is referred to henceforth as “validation accuracy” for brevity. LASSO regularization was applied during model training to reduce overfitting and identify electrodes that contributed to congruency classification. A null comparison distribution was generated by repeating the 10-fold cross-validation procedure on randomly permuted trial congruency labels for 250 iterations. True validation accuracy was compared to this distribution (see Section 2.7 Hypothesis Testing) to determine whether it was significantly above chance.

2.7. Hypothesis Testing

Significance Testing. Statistical significance in all tests was assessed by using a non-parametric permutation-based resampling approach to empirically approximate null distributions for F and t statistics (Bullock et al., 2017, 2021, 2023a; Foster et al., 2016). This testing approach has the advantage of being robust to normality violations. Specifically, for each test, condition labels were shuffled within participants and 1000 iterations of the appropriate statistical tests were run (repeated-measures ANOVAs and/or paired-samples t tests) to generate null distributions of F and t statistics. Reliable differences were then tested for by computing the probability of obtaining F and t statistics from each null distribution that were greater than the observed F and t statistics. The F and t statistics are then reported along with the critical p-value (labeled p_{null}) which represents the probability of observing a value greater than this in the null distribution. Statistical test outcomes for non-continuous data (i.e., data that are averaged across testing blocks, such as accuracy and RT) are reported in the text. Here, to convey a more precise estimate of the observed statistic’s position in the null distribution, tests are reported as $p_{\text{null}}<.05$, $p_{\text{null}}<.01$ and $p_{\text{null}}<.001$. A test result of $p_{\text{null}}>.05$ indicates that the result was not statistically reliable. Effect sizes are reported as partial eta squared (η_p^2) for ANOVAs (small: $\eta_p^2 = .01$, moderate: $\eta_p^2 = .06$, large: $\eta_p^2 = .14$) and Cohen’s d for t-tests (small: $d = .02$; medium: $d = .05$; large: $d = .08$). Statistical test outcomes for continuous time-course data (i.e., ERP waveforms, classification) are reported visually in each figure to provide insight into the temporal dynamics of each measure. Here, ANOVA and t-test outcomes where $p_{\text{null}} < .05$ are represented by the presence of horizontal bars superimposed onto the plots, where the presence of a bar at a given time point indicates a significant difference. The time-course analyses presented here rely on repeated comparisons at multiple timepoints, which raises the possibility of increased inferential error, however, the effects that are described are present across multiple timepoints and no inference relies on a single comparison but rather a consistent pattern across time. Furthermore, a cluster-based correction procedure was performed for classification analyses to mitigate spurious statistical differences (Cohen, 2014).

Data Visualization. Data were visualized using functions from the Seaborn and Matplotlib Python Libraries (Hunter, 2007; Waskom, 2021) and MATLAB.

3. Results

First, we report participants’ behavioral performance on the task, to determine how our independent variables (judgement type: association, label; and modality: model, word) impacted participants’ accuracy and RT on each trial as they assessed whether each stimulus pair was congruent or incongruent. Second, we plotted ERPs to determine whether an N400 was present in each of the four conditions, and then ran analyses to characterize the differences in waveform morphology between conditions. Third, we present the results of a logistic regression classification analysis, to assess the viability of the N400 signal for use in a potential AR-BCI for learning.

3.1. Behavioral performance was consistently good in all conditions

Accuracy. Performance was generally good in all conditions (Fig. 3a). Accuracy data were submitted to a 2 [judgement type: association, label] x 2 [modality: model, word] x 2 [congruency: congruent, incongruent] repeated-measures ANOVA. Participants’ accuracy was higher in the label conditions compared to the association conditions, supported by a main effect of judgment type [$F(1,23) = 18.47$, $p_{\text{null}} < .001$, $\eta_p^2 = .45$]. There were no significant main effects of modality or congruency, and no interaction effects between any variables [all $p_{\text{null}} > .05$].

RT. Participants’ responses were faster in the label conditions compared to the association conditions, supported by a main effect of judgment type [$F(1,23) = 127.57$, $p_{\text{null}} < .001$, $\eta_p^2 = .84$]. Responses were also faster to congruent trials when compared to incongruent trials [$F(1,23) = 56.84$, $p_{\text{null}} < .001$, $\eta_p^2 = .72$] (Fig. 3b). Participants were slower at identifying word-word pairs compared to model-word pairs in the association conditions but not in the label conditions, supported by an interaction between judgment type and modality [$F(1,23) = 14.71$, $p_{\text{null}} < .001$, $\eta_p^2 = .39$] and pairwise comparisons [MWA vs WWA: $t(23) = 2.32$, $p_{\text{null}} <.05$, $d = .47$; MWL vs WWL: $t(23) = 1.68$, $p_{\text{null}} >.05$, $d = .34$; all other comparisons $p_{\text{null}} >.05$]. Participants were also slower at responding to incongruent trials when compared to congruent trials in the label conditions, supported by an interaction between judgement type and congruency [$F(1,23) = 12.61$, $p_{\text{null}} < .01$, $\eta_p^2 = .35$] and pairwise comparisons [MWA vs MWL: $t(23) = 3.72$, $p_{\text{null}} < .001$, $d = .76$; all other comparisons: $p_{\text{null}} > .05$] (Fig. 3c).

3.2. Robust N400 ERPs were observed in all conditions

ERPs and their corresponding difference waves (incongruent minus congruent), computed across fronto-central, central, and parietal regions, are displayed for each experimental condition in Fig. 4(a–d). All conditions showed robust statistical differences beginning between ~200 and 300 ms after the second stimulus appeared, confirming consistent N400 brain responses. However, the shape of the waveforms appeared to be visually different between conditions, suggesting our experimental manipulations (modality and judgment type) induce different patterns of brain activity. To better characterize these effects, we plotted the difference waves for each condition together (Fig. 4e). Mean amplitude (Fig. 4f) and peak latency (Fig. 4g) measures for the difference waves were submitted to a repeated-measures ANOVA. Mean amplitude was not different across conditions [all $F(1,23) < 4.50$, $p_{\text{null}} > .05$, $\eta_p^2 <.16$]. Peak latency was earlier in the label conditions compared to the association conditions [$F(1,23) = 49.61$, $p_{\text{null}} < .001$, $\eta_p^2 = .68$]. Stimulus modality had no effect [$F(1,23) = .001$, $p_{\text{null}} >.05$, $\eta_p^2 = .001$] and there was no interaction [$F(1,23) = .06$, $p_{\text{null}} > .05$, $\eta_p^2 = .003$].

3.3. Classification

The main aim of this study was to assess whether the N400 is a viable signal for use in a combined AR-EEG system to assess semantic knowledge and potentially enhance retrieval learning. In the previous analysis

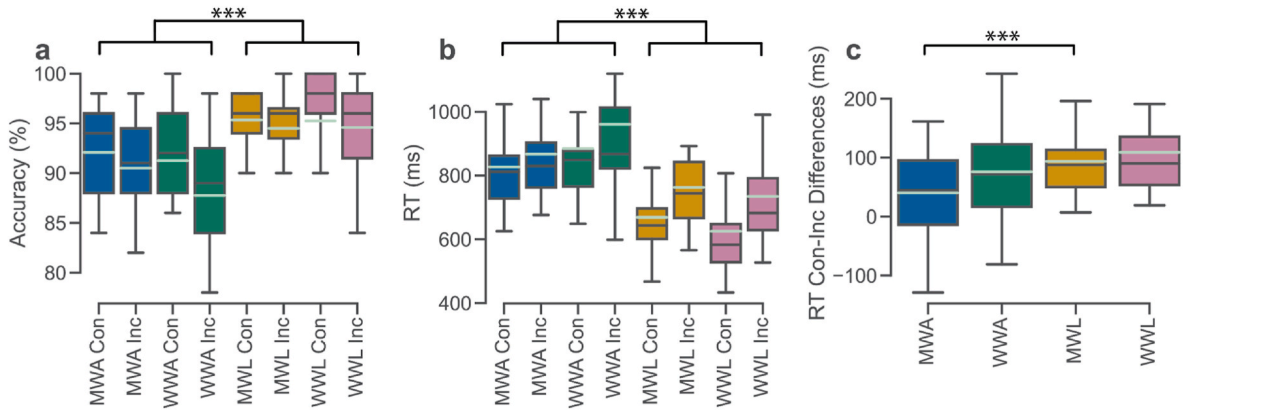

Fig. 3. ***Behavior***. Panels (a) and (b) depict accuracy and response times for stimulus pair judgments in each of the conditions, separated as a function of pair congruence (Con) or incongruence (Inc). In (c) we plot the RT differences between congruent and incongruent trials for each condition (difference = incongruent - congruent). Horizontal gray and white lines in boxplots represent group median and mean, respectively. ***$p_{\text{null}} < .001$.

we observed robust N400 ERPs in all conditions, confirming that we can successfully parse the brain's averaged responses to congruent and incongruent trials for all judgment and modality combinations. However, if the possible end goal is to take advantage of this signal for near real-time tracking of brain responses in a BCI system, then we need to determine how accurately brain responses to individual trials can be classified based on the congruence of each stimulus pair. A single logistic regression model trained and tested over time using a cross-validation approach revealed above-chance classification accuracy classifying congruent vs. incongruent trials in all four conditions relative to a permuted control. Successful decoding was observed to occur earlier in the label conditions (WWL: ~116ms; MWL: ~200ms) compared to the association conditions (MWA: ~304ms; WWA: ~304ms) (Fig. 5). There was a significant main effect of both judgement type (~192ms) and stimulus modality (~188ms) early on in the trial period, but no interactions.

4. General discussion

This study investigated the viability of the N400 ERP component as a candidate brain signal for assessing semantic knowledge within AR. We recorded 64-channel EEG while participants completed a pair-matching task presented via an AR headset. To examine the influence of stimulus modality on the N400, we manipulated whether the first stimulus in a pair was a 3D object or a written word. We also varied whether the second stimulus was congruent or incongruent with the first stimulus in terms of meaning or name, to compare semantic and naming associations. To ensure precise time synchronization between the AR and EEG systems, a novel hardware solution was implemented and validated, involving event-locked pulses sent through the headphone jack of the AR headset. Participants performed well at all variations of the task, responding faster to congruent pairs relative to incongruent pairs, and making faster judgments in label conditions compared to association conditions. Analysis of brain activity data over central and fronto-central scalp regions revealed patterns that complemented behavior, with robust differences between responses on congruent and incongruent trials and earlier differentiation in naming when compared to association conditions. Furthermore, machine-learning analyses revealed that the N400 could be used to classify congruent versus incongruent pairs on a single-trial basis, with successful classification occurring earlier in label compared to association conditions, hinting that the N400 might be a viable signal to use in an BCI context if classification accuracy were to be improved.

The N400 is typically studied by having participants judge the semantic congruency of a target stimulus (such as a word, image or sound) with a preceding stimulus. A congruency effect is then typically observed in their behavioral responses, such that responses are slower when the target stimulus is semantically incongruent (Gomes et al., 1997; Holcomb, 1988; Holcomb & Neville, 1990; Kutas & Federmeier, 2011). Here we show this congruency effect in all four conditions, indicating that the effect is present regardless of whether the first item in each stimulus pair is a 3D model or word, and if the task is to make an association or labeling judgment. However, there was a diminished congruency effect in MWA compared to MWL, which suggests that participants found it more challenging to make association judgments than labeling judgements when the first stimulus was a 3D model. In contrast, the congruency effect across WWA and WWL was not statistically different, implying that the smaller congruency effect was specific to MWA. It is not clear why the congruency effect is smaller in MWA, but one might presume this is because people are less familiar with our 3D models, making them harder to identify and process. Indeed, it is well-established that picture naming takes longer than word reading (Valente et al., 2016), arguably due to the "uncertainty factor", where a picture can often be named in several different ways but only a single response can be given for a written word (Ferrand, 1999). However, the data do not support this, as accuracy and RT in MWA is similar to WWA, where participants made association judgments on two words. We also observed notably faster responses to label conditions when compared to association conditions, indicating participants found naming judgments easier than label judgments. It was most critical that we observed a robust congruency effect in the MWL condition, as this condition most closely approximates a retrieval learning scenario, i.e., the user will view a physical or digital 3D object and then a text label will appear either correctly or incorrectly naming the object. This behavioral result indicates participants are making congruency judgments in MWL, which should elicit N400 responses.

Traditional ERP analyses of the brain data confirmed the presence of robust N400 responses in all experimental conditions across central and fronto-central scalp regions, with clear separation between congruent and incongruent trials ~200–600 ms post onset of the second stimulus regardless of the modality of the first stimulus or judgment type. These results further support the amodal nature of the N400, demonstrating that it can be reliably elicited by 3D models in AR, consistent with previous findings that the N400 can be elicited by words, 2D images presented in a screen, sounds, or combinations of these stimuli (Barrett & Rugg, 1990; Calma-Roddin & Drury, 2020; Ganis et al., 1996; Holcomb, 1988; Lin et al., 2022; Nigam et al., 1992; Ortu et al., 2013). The results also demonstrate that a reliable N400 can be elicited when participants make naming judgments, which expands on previous N400 studies where participants are required to make semantic association judgments (Kutas & Federmeier, 2011). Critically, for the purposes of our proposed AR-EEG learning system, it was most important that we

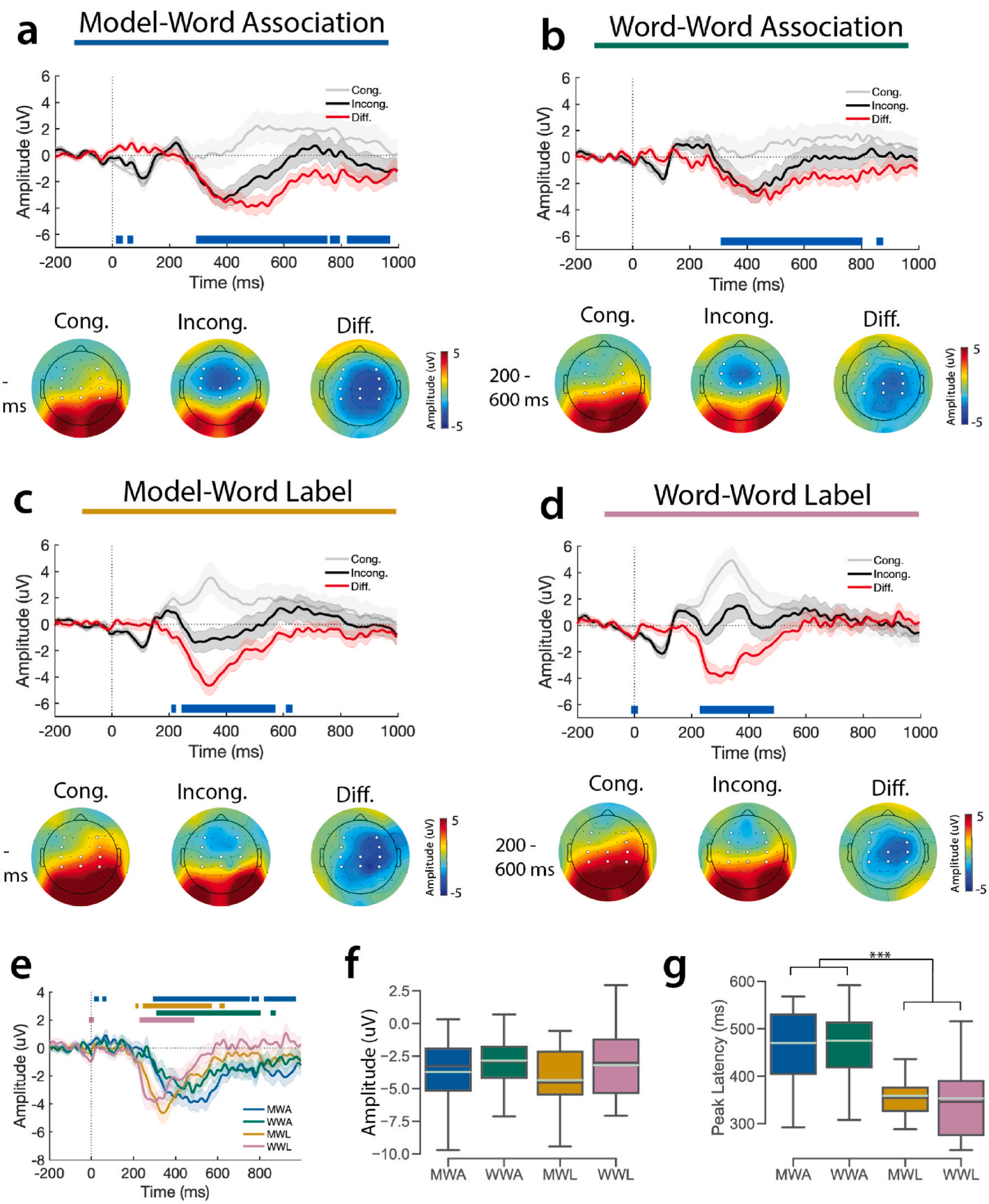

Fig. 4. ***Brain responses to stimulus pairs.*** Panels (a–d) show event-related potentials (ERPs) time-locked to the second word, with difference waves (incongruent minus congruent) highlighting the congruency effect. Horizontal blue bars at the base of each plot indicate significant differences between congruent and incongruent waveforms. Topographic maps depict scalp activity distribution averaged across a group of central and fronto-central electrodes (marked with white dots) and between 200 and 600 ms. Panel (e) overlays difference waves across conditions to allow for ease of comparison, while (f) and (g) show mean amplitude and peak latency of the difference waves, respectively. $^{***}p_{null} < .001$. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

observed an N400 response in the model-word label condition, and indeed this was the case. Interestingly, the shape of the waveforms was markedly different as a function of judgment type, with separation between congruent and incongruent waveforms occurring earlier in naming conditions when compared to association conditions, as evidenced by comparison of the timing of the ERP difference waves. Furthermore, when participants were required to make semantic judgments, the difference between congruent and incongruent trials was driven by the negative deflection to congruent trials (the typical N400 effect), but when participants were required to make naming judgments, the difference was mainly due to a positive deflection in congruent trials. This suggests that even though both types of judgments involve understanding the meaning of objects and words, the underlying brain processes are different. The consistent congruency effect across conditions and also the earlier separability when making label judgments compared to association judgments is consistent with the patterns observed in behavior, such that participants responded more rapidly to congruent than incongruent pairs in all conditions, and faster in label versus

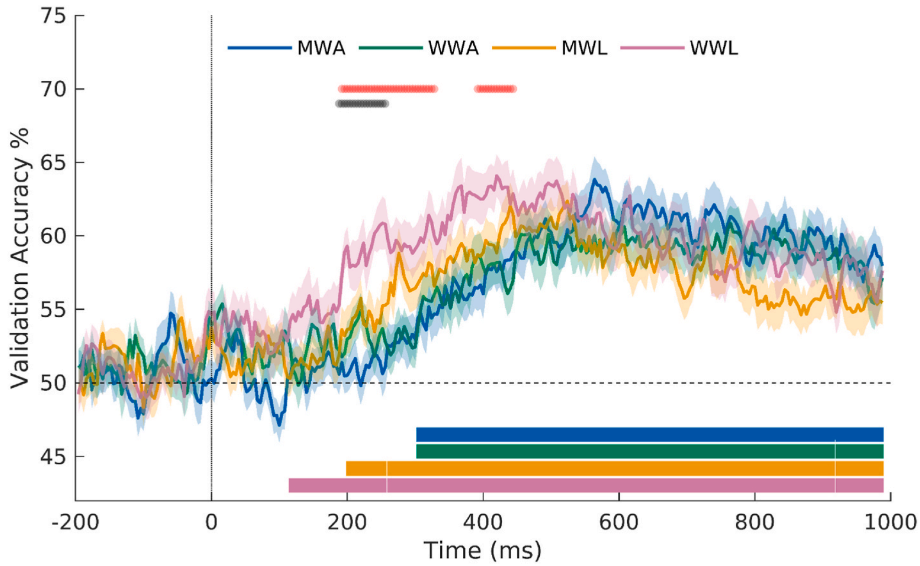

Fig. 5. ***Classification***. A logistic regression classifier was trained and tested across all trials and conditions using a 10-fold cross-validation approach. The model was evaluated at each time point using a sliding window approach, and its performance was quantified using the average balanced accuracy across validation folds. The horizontal bars at the base of the plot indicate timepoints where classification was significantly different from a permuted control in each condition ($p < .05$). The red and black horizontal lines at the top of the plot indicate timepoints where there was a main effect ($p < .05$) of judgement or stimulus type, respectively. Continuous error bars represent SEM. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

association conditions. A possible explanation for why naming judgments are faster than meaning judgments is due to differences in the directness and automaticity of the underlying neural processes. For both naming and meaning judgments, visual information must be processed up to the level of object recognition. This involves a hierarchical network of occipital and temporal regions, notably the inferotemporal cortex for object recognition (Cichy et al., 2014; DiCarlo et al., 2012) and left occipitotemporal cortex for word recognition (Mano et al., 2013; Xu et al., 2015). In addition, meaning judgments also require that the meanings of both items in the stimulus pair are accessed and compared, which may involve a more distributed network of left lateralized brain regions (Binder et al., 2009; Giesbrecht et al., 2004). These additional cognitive processing demands may contribute to the delayed RTs and N400 responses for meaning compared to naming judgments.

Having established the presence of an N400 response in all conditions when averaged across trials, the next step was to test whether neural responses to congruent and incongruent stimuli could be classified on a single-trial basis. Logistic regression classification confirmed that above-chance classification accuracy could be observed in each of the conditions ~304 ms post-stimulus onset. Successful decoding accuracy was observed earlier for the label relative to the association condition, which is consistent with the timing of ERP difference waves between these two experimental manipulations. The fact that classifier accuracy was above chance in all conditions (particularly MWL) suggests that the N400 is potentially a viable candidate for use in a BCI context. However, we must acknowledge that while classifier accuracy was at above chance levels in all conditions during the N400 window, it did not reach the level of performance required for a BCI. This is likely due to noise in EEG data and the relatively low number of trials presented to participants in each condition. Indeed, prior work has shown that increasing the number of stimuli per class dramatically boosts N400 detection rates (Dijkstra, Farquhar, & Desain, 2019; Dijkstra et al., 2020), suggesting this is a signal-to-noise issue. In the present study our main goal was to compare how different types of stimuli (3D objects vs. written words) and judgments (naming vs. association) affected the N400 brain response. To do this, we needed four different conditions in our experiment, limiting the amount of time spent on each condition, so we could only reasonably run 100 trials per condition. It is reasonable to predict that classification accuracy would be improved with a much larger number of trials. Nevertheless, despite the relatively low accuracy, the present data do represent an important first step in the direction toward identifying the neural markers of stimulus identification that can be utilized to create a BCI that facilitates user learning in virtual environments.

This study aimed to address two key questions about the N400, with the long-term goal of developing an integrated AR-BCI learning system. While we did not directly assess learning in this study, we can consider how our proposed system might function as a learning aid, and which types of learning task it might benefit. In the introduction we described two learning scenarios where the proposed system could be used: learning from digital content (e.g., a medical student studying anatomy) or from physical world objects (e.g., someone learning a new language). In both scenarios, learners would look at an object, try to remember its name, and then see a text label that is either correct or incorrect. An N400 response would be expected if the label is incorrect and the learner is aware of the error. However, if the learner doesn't know the correct name and doesn't recognize the error, an N400 response would not be expected. AR has shown strong potential to be particularly effective and enjoyable for language learning compared to traditional methods (Ibrahim et al., 2018). However, learning from physical world objects presents significant challenges. It would require rapid and accurate recognition of a wide range of objects from various viewpoints, correct segmentation from other items, and near-real-time labeling (see Huynh et al., 2019a, Huynh et al., 2019b). Physical objects can also be ambiguous or have multiple names. Moreover, language learning in real-world settings would most likely involve the learner moving around in the environment, implying eye, head and body movements, and potential shifting of the system on the head, all of which increases EEG noise (discussed in the next paragraph). In contrast, implementing AR-BCI learning from a digital model may be easier. Digital models could be pre-programmed with all required text labels, and the learning environment would likely be more controlled (e.g., indoors, consistent lighting/temperature), minimizing movement-related EEG artifacts. For both learning scenarios, the system would maintain a record of successfully and unsuccessfully recognized items, allowing for the provision of feedback to support learning. The optimal method for delivering this feedback is a topic for future research, but potential approaches include immersive feedback in AR. For example, in the language learning scenario, objects that the system identified as not having been correctly recognized by the user could be visually highlighted (e.g., by displaying a brightly colored frame around the object). In the medical student scenario, specific anatomical structures that the student has not recognized could be highlighted on the digital model. This immersive feedback could be presented to the user either in near real-time or during a separate dedicated session. Additional studies are required to determine the most effective way to provide feedback, with careful consideration given to avoiding information overload for the user.

It is important to acknowledge a number of limitations of this highly controlled proof-of-concept study. First, our sample was a relatively small convenience sample of college students, largely drawn from psychology and computer science undergraduate cohorts. These students may have more experience with cognitive tasks and/or AR devices. While this homogeneity could limit the generalizability of our results to other populations, our university has a relatively large undergraduate student body (>20,000) and is a minority-serving institution, meaning at least 25 % of the student body identifies as being from an underrepresented minority, indicating some heterogeneity in the sample. Second, our study purposefully selected stimuli that were easy to recognize and these were presented at fixation on a fixed time-schedule. We did not explicitly control our stimuli for visual complexity or word frequency, so it's possible that our findings may not generalize to more complex objects or less commonly used words. Participants were instructed to maintain fixation and minimize eye-movements. This approach differs from realistic retrieval-practice contexts, such as the anatomy or foreign language learning examples described in the introduction. In those situations, objects may be more ambiguous, and users would likely move their eyes freely to explore the physical and/or digital environment. The system would then need to detect when the user's gaze has dwelled on a critical object for a specified duration and initiate the presentation of a text label. Eye-movements disrupt patterns of brain activity (Bullock et al., 2023b; Irwin, 1996; Rolfs et al., 2011), so this will introduce additional noise into the system. Third, in the present study EEG was recorded using a research grade 64-channel gel electrode setup (ActiCHamp, Brain Products), which has a number of benefits such as low electrode impedances and complete scalp coverage, but would be impractical for actual real-world use in a BCI because it needs to be set up by a trained experimenter and the application process takes around 30–45 min. Instead, a BCI would be more likely to employ a more practical, lower-cost, consumer focused EEG system which will likely use fewer, dry electrodes, leading to higher impedances and reduced scalp coverage. Increased noise will reduce signal-to-noise ratio (SNR), potentially leading to less accurate classification. Furthermore, our hardware-trigger solution for AR-EEG synchronization requires custom hardware (StimTrak, Brain Products) and a headphone jack on the AR device, which may not be present on all AR devices. Fourth, our participants were seated in a temperature controlled room for the duration of the study. Under more realistic conditions a user might wear a BCI system while they are actively moving around the environment, meaning the system will have to contend with artifacts generated by the user's body movements, muscle noise and perspiration, which can impact classifier performance (Ding et al., 2019). Indeed, it is feasible that AR might eventually replace (or complement) smartphones and watches for continuous information consumption while the user is on-the-go (i.e., walking around in the environment; Kim et al., 2022; Kumaran et al., 2023), so any integrated BCI systems will also need to be robust to these large body movements. Furthermore, acute physical activity can impact cognitive function, with selective effects at specific stages of sensory and cognitive information processing in the human brain which could plausibly impact BCI performance (Bullock et al., 2015; Cao & Händel, 2019; Garrett et al., 2021, 2024; Giesbrecht et al., 2025; Giesbrecht & Garrett, 2025). In summary, translating these findings into a practical AR-EEG learning system will require testing larger, more diverse samples with a wider range of stimuli, while allowing for more natural eye, head, and body movements, and using a more practical EEG recording system.

5. Conclusion

In summary, this study represents an important first step in determining whether the N400 is a viable signal for use in a combined AR-EEG retrieval-learning system. We also validated a novel hardware-based solution for synching stimuli presented in AR with EEG recording. We demonstrated that the brain responds differently to congruent and incongruent pairs of stimuli, regardless of whether the first stimulus is a 3D model or a word, and regardless of whether participants are making naming or association judgments. This study lays the groundwork for future experiments aimed at improving classification accuracy of brain responses. These studies will increase the size of the training set and also test classification performance with a broader demographic of users in more realistic scenarios, such as those involving EEG systems with fewer, dry electrodes and unrestricted eye, head, and body movements.

CRediT authorship contribution statement

Tom Bullock: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Emily Machniak:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Conceptualization. **Joyce Passananti:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis. **Kangyou Yu:** Writing – review & editing, Software, Methodology, Investigation. **Radha Kumaran:** Writing – review & editing, Software, Methodology. **You-Jin Kim:** Writing – review & editing, Software, Methodology. **Jordan Garrett:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis. **Madhav Viswesvaran:** Writing – review & editing, Formal analysis. **Julia Ram:** Writing – review & editing, Investigation. **Tobias Höllerer:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Barry Giesbrecht:** Writing – review & editing, Writing – original draft, Resources, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Data availability

All data and custom analysis scripts are available via GitHub (https://github.com/attlab/AR_EEG_N400).

Funding statement

This work was generously supported by NSF award IIS-1911230.

Declaration of competing interest

None.

Acknowledgements

We thank Melissa Hernandez, Stina Johansson, Paola Belmontes, Catherine Wen and Elise Kormos for their help with data collection.

Data availability

Data will be made available on request.

References

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*(2), 201–212. https://doi.org/10.1016/0278-2626(90)90029-N

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00298

Boros, M., Magyari, L., Morvai, B., Hernández-Pérez, R., Dror, S., & Andics, A. (2024). Neural evidence for referential understanding of object words in dogs. *Current Biology, 34*(8), 1750–1754.e4. https://doi.org/10.1016/j.cub.2024.02.029

Bullock, T., Cecotti, H., & Giesbrecht, B. (2015). Multiple stages of information processing are modulated during acute bouts of exercise. *Neuroscience, 307*, 138–150. https://doi.org/10.1016/j.neuroscience.2015.08.046

Bullock, T., Elliott, J. C., Serences, J. T., & Giesbrecht, B. (2017). Acute exercise modulates feature-selective responses in human cortex. *Journal of Cognitive Neuroscience, 29*(4), 605–618. https://doi.org/10.1162/jocn_a_01082

Bullock, T., Giesbrecht, B., Beaudin, A. E., Goodyear, B. G., & Poulin, M. J. (2021). Effects of changes in end-tidal PO2 and PCO2 on neural responses during rest and sustained attention. *Physiological Reports, 9*(21), 1–25. https://doi.org/10.14814/phy2.15106

Bullock, T., MacLean, M. H., Santander, T., Boone, A. P., Babenko, V., Dundon, N. M., Stuber, A., Jimmons, L., Raymer, J., Okafor, G. N., Miller, M. B., Giesbrecht, B., & Grafton, S. T. (2023a). Habituation of the stress response multiplex to repeated cold pressor exposure. *Frontiers in Physiology, 13*, Article 752900. https://doi.org/10.3389/fphys.2022.752900

Bullock, T., Pickett, K., Salimian, A., Gregory, C., MacLean, M. H., & Giesbrecht, B. (2023b). Eye movements disrupt EEG alpha-band coding of behaviorally relevant and irrelevant spatial locations held in working memory. *Journal of Neurophysiology, 129*(5), 1191–1211. https://doi.org/10.1152/jn.00302.2021

Calma-Roddin, N., & Drury, J. E. (2020). Music, language, and the N400: ERP interference patterns across cognitive domains. *Scientific Reports, 10*(1), Article 11222. https://doi.org/10.1038/s41598-020-66732-0

Cao, L., & Händel, B. (2019). Walking enhances peripheral visual processing in humans. *PLoS Biology, 17*(10), 1–23. https://doi.org/10.1371/journal.pbio.3000511

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17*(3), 455–462. https://doi.org/10.1038/nn.3635

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. The MIT Press. https://doi.org/10.7551/mitpress/9609.001.0001

Coronel, J. C., & Federmeier, K. D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia, 84*, 36–43. https://doi.org/10.1016/j.neuropsychologia.2016.01.029

Deacon, D., Hewitt, S., Yang, C.-M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: Evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research, 9*(2), 137–146. https://doi.org/10.1016/S0926-6410(99)00050-6

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Dijkstra, K. V., Farquhar, J., & Desain, P. (2019). Electrophysiological responses of relatedness to consecutive word stimuli in relation to an actively recollected target word. *Scientific Reports, 9*(1), Article 14514. https://doi.org/10.1038/s41598-019-51011-4

Dijkstra, K. V., Farquhar, J. D. R., & Desain, P. W. M. (2020). The N400 for brain computer interfacing: Complexities and opportunities. *Journal of Neural Engineering, 17*(2), Article 022001. https://doi.org/10.1088/1741-2552/ab702e

Ding, Y., Huynh, B., Xu, A., Bullock, T., Cecotti, H., Turk, M., Giesbrecht, B., & Höllerer, T. (2019). Multimodal classification of EEG during physical activity. In *2019 international conference on multimodal interaction* (pp. 185–194). https://doi.org/10.1145/3340555.3353759

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*(1), 7–22. https://doi.org/10.1007/s10956-008-9119-1

Federmeier, K. D., & Laszlo, S. (2009). Chapter 1 time for meaning: Electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In *Psychology of learning and motivation* (Vol. 51, pp. 1–44). Elsevier. https://doi.org/10.1016/S0079-7421(09)51001-8.

Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychologica, 100*(3), 253–266. https://doi.org/10.1016/S0001-6918(98)00021-3

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology, 111*, 15–52. https://doi.org/10.1016/j.cogpsych.2019.03.002

Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2016). The topography of alpha-band activity tracks the content of spatial working memory. *Journal of Neurophysiology, 115*(1), 168–177. https://doi.org/10.1152/jn.00860.2015

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89–106. https://doi.org/10.1162/jocn.1996.8.2.89

Garrett, J., Bullock, T., & Giesbrecht, B. (2021). Tracking the contents of spatial working memory during an acute bout of aerobic exercise. *Journal of Cognitive Neuroscience, 33*(7), 1271–1286. https://doi.org/10.1162/jocn_a_01714

Garrett, J., Chak, C., Bullock, T., & Giesbrecht, B. (2024). A systematic review and Bayesian meta-analysis provide evidence for an effect of acute physical activity on cognition in young adults. *Communications Psychology, 2*(1), 1–21. https://doi.org/10.1038/s44271-024-00124-2

Giesbrecht, B., Bullock, T., & Garrett, J. (2025). Physically activated modes of attentional control. *Trends in Cognitive Sciences, 29*(3), 295–307. https://doi.org/10.1016/j.tics.2024.11.006

Giesbrecht, B., Camblin, C., & Swaab, T. Y. (2004). Separable effects of semantic priming and imageability on word processing in human cortex. *Cerebral Cortex, 14*(5), 521–529. https://doi.org/10.1093/cercor/bhh014

Giesbrecht, B., & Garrett, J. (2025). Electroencephalography. In *Encyclopedia of the human brain* (pp. 59–75). Elsevier. https://doi.org/10.1016/B978-0-12-820480-1.00007-3.

Giesbrecht, B., Sy, J. L., & Elliott, J. C. (2007). Electrophysiological evidence for both perceptual and postperceptual selection during the attentional blink. *Journal of Cognitive Neuroscience, 19*(12), 2005–2018. https://doi.org/10.1162/jocn.2007.19.12.2005

Gomes, H., Ritter, W., Tartter, V. C., Vaughan, H. G., & Rosen, J. J. (1997). Lexical processing of visually and auditorily presented nouns and verbs: Evidence from reaction time and N400 priming data. *Cognitive Brain Research, 6*(2), 121–134. https://doi.org/10.1016/S0926-6410(97)00023-2

Gomez-Herrero, G., Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Huffel, S., & Paesschen, W. (2006). Automatic removal of ocular artifacts in the EEG without an EOG reference channel. In *Proceedings of the 7th nordic signal processing symposium - norsig 2006* (pp. 130–133). https://doi.org/10.1109/NORSIG.2006.275210

Hodapp, A., & Rabovsky, M. (2021). The N400 ERP component reflects an error-based implicit learning signal during language comprehension. *European Journal of Neuroscience, 54*(9), 7125–7140. https://doi.org/10.1111/ejn.15462

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*(1), 66–85. https://doi.org/10.1016/0093-934X(88)90101-0

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language & Cognitive Processes, 5*(4), 281–312. https://doi.org/10.1080/01690969008407065

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering, 9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55. Computing in Science & Engineering.

Huynh, B., Orlosky, J., & Hollerer, T. (2019a). In-situ labeling for augmented reality language learning. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 1606–1611). https://doi.org/10.1109/VR.2019.8798358

Huynh, B., Orlosky, J., & Hollerer, T. (2019b). Semantic labeling and object registration for augmented reality language learning. In *2019 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 986–987). https://doi.org/10.1109/VR.2019.8797804

Ibrahim, A., Huynh, B., Downey, J., Hollerer, T., Chun, D., & O'donovan, J. (2018). ARbis pictus: A study of vocabulary learning with augmented reality. *IEEE Transactions on Visualization and Computer Graphics, 24*(11), 2867–2874. https://doi.org/10.1109/TVCG.2018.2868568

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*(3), 94–100. https://doi.org/10.1111/1467-8721.ep10772833

Kaufmann, T., Schulz, S. M., Grünzinger, C., & Kübler, A. (2011). Flashing characters with famous faces improves ERP-based brain–computer interface performance. *Journal of Neural Engineering, 8*(5), Article 056016. https://doi.org/10.1088/1741-2560/8/5/056016

Kaufmann, T., Schulz, S. M., Köblitz, A., Renner, G., Wessig, C., & Kübler, A. (2013). Face stimuli effectively prevent brain–computer interface inefficiency in patients with neurodegenerative disease. *Clinical Neurophysiology, 124*(5), 893–900. https://doi.org/10.1016/j.clinph.2012.11.006

Kim, Y.-J., Kumaran, R., Sayyad, E., Milner, A., Bullock, T., Giesbrecht, B., & Hollerer, T. (2022). Investigating search among physical and virtual objects under different lighting conditions. *IEEE Transactions on Visualization and Computer Graphics, 28*(11), 3788–3798. https://doi.org/10.1109/TVCG.2022.3203093

Kumaran, R., Kim, Y.-J., Milner, A. E., Bullock, T., Giesbrecht, B., & Höllerer, T. (2023). The impact of navigation aids on search performance and object recall in wide-area augmented reality. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–17). https://doi.org/10.1145/3544548.3581413

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in Psychophysiology, 3*, 139–187.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De) constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lin, Y., Liu, Z., & Gao, X. (2022). Sensitivity of N400 effect during speech comprehension under the uni- and bi-modality conditions. *Tsinghua Science and Technology, 27*(1), 141–149. https://doi.org/10.26599/TST.2021.9010008

Lytaev, S. (2021). Neurophysiological visual classification indicators in the brain-computer interface. In D. Harris, & W.-C. Li (Eds.), *Engineering psychology and cognitive ergonomics* (pp. 197–211). Springer International Publishing. https://doi.org/10.1007/978-3-030-77932-0.

Mano, Q. R., Humphries, C., Desai, R. H., Seidenberg, M. S., Osmon, D. C., Stengel, B. C., & Binder, J. R. (2013). The role of left occipitotemporal cortex in reading: Reconciling stimulus, task, and lexicality effects. *Cerebral Cortex, 23*(4), 988–1001. https://doi.org/10.1093/cercor/bhs093

Nigam, A., Hoffman, J. E., & Simons, R. F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience, 4*(1), 15–22. https://doi.org/10.1162/jocn.1992.4.1.15

Ortu, D., Allan, K., & Donaldson, D. I. (2013). Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia, 51*(9), 1742–1748. https://doi.org/10.1016/j.neuropsychologia.2013.05.003

Rabovsky, M., Hansen, S. S., & McClelland, J. L. (2018). Modelling the N400 brain potential as change in a probabilistic representation of meaning. *Nature Human Behaviour, 2*(9), 693–705. https://doi.org/10.1038/s41562-018-0406-4

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience, 14*(2), 252–258. https://doi.org/10.1038/nn.2711

Sy, J. L., Elliott, J. C., & Giesbrecht, B. (2013). Post-perceptual processing during the attentional blink is modulated by inter-trial task expectancies. *Frontiers in Human Neuroscience, 7*(October), 627. https://doi.org/10.3389/fnhum.2013.00627

Valente, A., Pinet, S., Alario, F.-X., & Laganaro, M. (2016). "When" does picture naming take longer than word reading? *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.00031

van Vliet, M., Mühl, C., Reuderink, B., & Poel, M. (2010). Guessing what's on your mind: Using the N400 in brain computer interfaces. In Y. Yao, R. Sun, T. Poggio, J. Liu, N. Zhong, & J. Huang (Eds.), *Brain informatics* (pp. 180–191). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15314-3_17.

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software, 6*(60), 3021. https://doi.org/10.21105/joss.03021

Xu, M., Wang, T., Chen, S., Fox, P. T., & Tan, L. H. (2015). Effective connectivity of brain regions related to visual word recognition: An fMRI study of C hinese reading. *Human Brain Mapping, 36*(7), 2580–2591. https://doi.org/10.1002/hbm.22792

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *Elife, 9*, Article e57023. https://doi.org/10.7554/eLife.57023