

# Evaluating the Feasibility of Visual Imagery for an EEG-Based Brain–Computer Interface

Justin Kilmarx<sup>1</sup>, Ivan Tashev<sup>2</sup>, Fellow, IEEE, José del R. Millán<sup>3</sup>, Fellow, IEEE, James Sulzer<sup>4</sup>, and Jarrod Lewis-Peacock<sup>5</sup>

**Abstract**—Visual imagery, or the mental simulation of visual information from memory, could serve as an effective control paradigm for a brain-computer interface (BCI) due to its ability to directly convey the user's intention with many natural ways of envisioning an intended action. However, multiple initial investigations into using visual imagery as a BCI control strategies have been unable to fully evaluate the capabilities of true spontaneous visual mental imagery. One major limitation in these prior works is that the target image is typically displayed immediately preceding the imagery period. This paradigm does not capture spontaneous mental imagery as would be necessary in an actual BCI application but something more akin to short-term retention in visual working memory. Results from the present study show that short-term visual imagery following the presentation of a specific target image provides a stronger, more easily classifiable neural signature in EEG than spontaneous visual imagery from long-term memory following an auditory cue for the image. We also show that short-term visual imagery and visual perception share commonalities in the most predictive electrodes and spectral features. However, visual imagery received greater influence from frontal electrodes whereas perception was mostly confined to occipital electrodes. This suggests that visual perception is primarily driven by sensory informa-

tion whereas visual imagery has greater contributions from areas associated with memory and attention. This work provides the first direct comparison of short-term and long-term visual imagery tasks and provides greater insight into the feasibility of using visual imagery as a BCI control strategy.

**Index Terms**—Brain–computer interface, visual imagery, visual perception, electroencephalography.

## I. INTRODUCTION

THE concept of using brain signals recorded via electroencephalography (EEG) to control external devices has gained traction in recent years as a potential way to provide patients with severe neuromuscular disorders a way to communicate and interact with the world around them [1]. This technology, termed brain-computer interface (BCI), has since grown to cover applications such as robotic control [2], [3], [4], communication [5], [6], and even entertainment and gaming [7], [8], [9]. Various control strategies exist for BCI interaction, but each come with their own limitations that prevent BCIs from obtaining widespread use outside the lab [10]. The imagined movements of large body parts is perhaps the most common control paradigm for EEG BCI applications; however, it often suffers from lengthy training times [10], with inconsistent and unstable performance [11], and a restricted range of options for imagined movements [12]. Most attempts to address the limitations of this paradigm focus on advancing EEG signal processing and classification techniques [13]; however, an often overlooked solution is investigating other imagery-based control strategies [14], [15]. Perhaps the ideal approach would be to utilize the unlimited flexibility of visual imagery to provide a more ecological connection between mental imagery and the intended action. Therefore, the objective of this work is to investigate the efficacy of using visual imagery for EEG BCI control.

### A. Limitations of Current BCI Control Paradigms

When designing a BCI system, one major consideration is the means of interaction used to perform the task. These control paradigms can be divided into two methodologies: *exogenous* paradigms based on the brain's response to an external stimulus, and *endogenous* paradigms where

Manuscript received 18 January 2024; revised 21 May 2024; accepted 30 May 2024. Date of publication 6 June 2024; date of current version 19 June 2024. This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant DGE-1610403 and in part by the National Eye Institute of the National Institutes of Health under Award R01EY028746. (Corresponding author: Justin Kilmarx.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by The University of Texas at Austin Institutional Review Board under Application No. 2013100110-MOD06.

Justin Kilmarx was with the Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712 USA. He is now with DCS Corporation, Dayton, OH 45431 USA (e-mail: jkilmarx@dcscorp.com).

Ivan Tashev is with Microsoft Research, Redmond, WA 98052 USA (e-mail: ivantash@microsoft.com).

José del R. Millán is with the Department of Electrical and Computer Engineering and the Department of Neurology, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: jose.millan@austin.utexas.edu).

James Sulzer is with the Department of Physical Medicine and Rehabilitation, MetroHealth Medical Center and Case Western Reserve University, Cleveland, OH 44109 USA (e-mail: jss280@case.edu).

Jarrod Lewis-Peacock is with the Department of Psychology, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: jalewpea@utexas.edu).

Digital Object Identifier 10.1109/TNSRE.2024.3410870

participants learn to modulate their brain activity using mental imagery [10]. The most popular exogenous paradigms in EEG rely on measuring the brain's response to visual stimuli such as a flickering target [5], [16]. Most participants can learn to use these paradigms with high accuracy and minimal training [10]. However, these procedures can be time intensive and require a high level of sustained attention and visual focus which can cause fatigue [17] and would not be suitable for individuals with visual impairments or photosensitivity [18], [19]. Furthermore, there is often large variability in performance across individuals [20], [21], perhaps because the control strategy is not intuitive, and this confusion can take focus away from the desired application. For example, the user would have to remember which flickering target corresponds to the intended action instead of attending to the action directly.

Endogenous control paradigms with mental imagery can be used instead to overcome these challenges. Motor imagery of the movements of large body parts (e.g., right vs. left hand) is the most popular imagery-based paradigm [10]. Limitations for this approach include lengthy training times (weeks to months) [10], large inter- and intra-subject performance variations [11], non-intuitive control schemes for certain applications [22], and a limited variety of classes available for BCI control [12]. Furthermore, factors such as noise in the EEG signals, motivation, fatigue, and difficulty visualizing the intended action can greatly impact a user's ability to gain control of the BCI [23]. This leads to a challenge referred to as "BCI illiteracy" in which a substantial percentage of participants (approximately 15-30%) remain unable to achieve proper control of a BCI even after a standard training period [24]. Even for the participants who can attain some control, performance often falls short of the desired threshold rate for effective control (often set at 60% or 70% accuracy in classification of the intended action) [11], [13], [25]. These difficulties have been observed across all BCI paradigms regardless of the neural signal used [26]. Some studies have even seen that participants who are deemed "BCI illiterate" using one paradigm can reach proficiency with another that may be more matched to their specialized expertise [27], [28], [29]. For example, a recent study by Lee et al. [29] compared performance with a BCI when participants attempted to use a motor imagery, event related potential (ERP), and steady-state visual evoke potential (SSVEP) control paradigm. They found that 72.2% of the participants were deemed "BCI illiterate" on at least one of the paradigms with the imagery-based paradigm showing the highest rate at 53.7%. However, all participants were able to control at least one of the systems. This indicates that for these individuals that may have difficulty with one type of BCI, the availability of an alternative, more intuitive mental imagery paradigm such as visual imagery may be beneficial to achieve proper BCI control.

### B. Visual Imagery as an Alternative BCI Control Strategy

Visual imagery, or the spontaneous mental simulation of visual information from long-term memory, could be a useful BCI control strategy that has not yet been sufficiently

tested [30]. Several studies have shown that various categories of images (e.g., faces, animals, and inanimate objects) can be reliably distinguished using EEG when participants are observing an image [30], [31], [32]. However, very few studies have attempted to measure visual imagery using EEG, and those that do have shown mixed success [12], [30], [33]. Bobrov et al. [33] provides the first investigation into the use of visual imagery as a BCI control paradigm. In this study, they were able to reliably distinguish between visual imagery of faces, visual imagery of houses, and resting state with an average of 56% classification accuracy (chance 33%). However, this study was limited by the number of recruited subjects (N=7), amount of data collected (four sessions each approximately 5 min long), and the quality of data collected (first three sessions used the 16 channel Emotiv Systems Inc. EPOC headset). Lee et al. [12] was able to demonstrate a high average classification accuracy of around 40% (chance was 7.69%; N=22 participants) during an offline analysis of a single session of 13 visual imagery categories. This included words used for patient communication with concrete properties (e.g., ambulance, clock, or toilet) or abstract properties (e.g., hello, stop, or yes). In Kosmyna et al. [30], researchers performed offline classification between two classes of flower vs. hammer during visual observation and imagery. They were unable to achieve above chance accuracy between the two classes during visual imagery (average classification accuracy 52%, chance 50%), but they were able to distinguish trials when participants performed visual imagery vs rest (77% average classification accuracy; chance 50%) and between visual observation vs imagery (71% classification accuracy; chance 50%).

One shortcoming in the aforementioned studies is that the two larger experiments by Kosmyna et al. [30] and Lee et al. [12] displayed the target category in each trial immediately before the imagery period. This could be considered more of a test of holding the object categories in working memory rather than spontaneous visual imagery [34]. This leaves the question open about whether spontaneous visual imagery can be decoded from EEG. To address this question, this current study provides participants with both visual and auditory cues of the intended mental imagery in separate experimental blocks.

Furthermore, the addition of the actual image during the cue period will allow a direct comparison between the neural signals elicited during observation and imagery. A study by Xie et al. [35] followed a similar procedure while looking for similarities between the mental activity during visual observation and imagery and found a correlation in the alpha band (8-13 Hz) between the two conditions. This is supported by the sensory recruitment hypothesis [36] which posits that the neural representations activated during perception can also be activated during short-term retention. However, the study by Lee et al. [12] found that activity in the higher gamma band (30-100 Hz) contained the most informative activity for visual imagery. This study seeks to add to this ongoing investigation for the most informative features for visual imagery decoding and the similarities between the neural activity during perception and imagery.

### C. Identifying Neural Mechanisms Contributing to BCI Performance

It is poorly understood why certain individuals are unable to control a BCI after a standard training protocol [11]. Previous literature from motor imagery has suggested many factors could play a role in performance variability including the user's basic demographics [37] (e.g., lifestyle, gender, or age), psychological traits [25], [38], [39] (motivation, confidence, or frustration), physiological traits [40], [41] (e.g., recruitment of motor imagery related brain networks), and anatomical structure [42] (e.g., structural integrity and myelination quality). Previous work from our group has indicated that difficulty learning to modulate desired brain activity in an fMRI neurofeedback task could be due to greater similarity in the brain activity patterns for each category [43] or overly rigid activity patterns (i.e., insufficient variability) for each category [44]. For this reason, we followed a similar approach as Kaneshiro et al. [45] to quantify the representational similarity between image categories using confusion matrices generated from multi-class classifications. We hypothesize that the classification of spontaneous visual imagery from long-term memory will reveal greater neural representation similarity between the image categories compared to short-term visual imagery from working memory. This work provides the first direct comparison of short-term and long-term visual imagery tasks measured by EEG in healthy adults and provides greater insight into the feasibility of using visual imagery as a BCI control paradigm.

We also administered a Vividness of Visual Imagery Questionnaire [46] (VVIQ) before the start of the experiment along with questionnaires of perceived psychological traits such as motivation, alertness, and frustration after each session. Participants' attention and engagement throughout the experiment were also monitored using eye tracking and pupillometry. Previous literature has shown that changes in the diameter of the pupil can occur in response to psychophysical and psychological stimuli [47]. Together, this information was collected to allow a more thorough exploration into the conditions contributing to successful decoding of visual imagery.

## II. METHODS

### A. Participants

A total of  $N=30$  healthy young adults between the ages of 18 and 40 years old were recruited from the Austin area for participation in this study. However, only  $N=26$  subjects (18 female, average age 22 years,  $SD=4.17$  years) were included in the analysis due to issues encountered during data collection. All methods were performed in accordance with the relevant guidelines and regulations of the University of Texas at Austin Institutional Review Board.

### B. Inclusion Criteria

The experiment conducted in this study involved participants performing visual imagery of a cued stimulus presented on a computer monitor. Each participant's neural activity was measured by EEG throughout the experiment. As such, all participants were required to meet the following inclusion

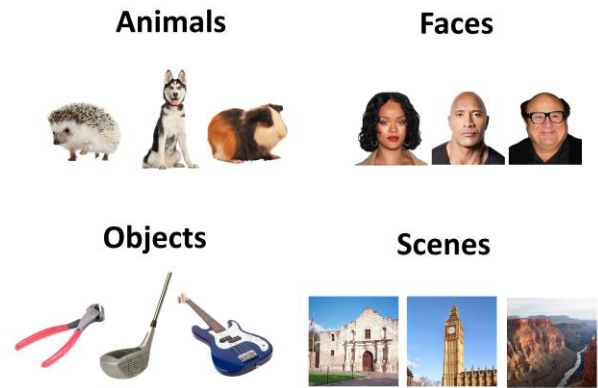


Fig. 1. Sample images used during experimental procedure. Participants selected one familiar image for each category of animals, famous faces, objects, or recognizable scenery and landmarks.

criteria: ability to provide informed consent, not be currently using any medication for psychiatric reasons, not be currently using any sedatives, no history of major psychotic disorders (including schizophrenia and bipolar disorder), no history of epilepsy or photosensitivity, no substance dependence, and good vision or minimal correction with contacts or eyeglasses. Participants were also asked to remove hair braids or any other tight hair styles and have clean hair (no oils, hair spray, or any other hair product) before participating in EEG recordings. This study also incorporated eye tracking, for which case individuals were excluded from participation in this study if they have glasses with more than one power (such as bifocals, trifocals, or progressive lenses), eye surgery (such as corneal, cataract, or intraocular implants), or eye movement or alignment abnormalities (such as amblyopia, strabismus, or nystagmus).

### C. Stimuli

Images for this experiment were obtained from an in-lab dataset of famous faces, animals, objects, and scenes that were chosen to be easily recognizable by the subject population (Fig. 1). The image categories were selected to be consistent with prior literature on representational similarity analysis and for their potential to provide distinct patterns of brain activity [45], [48]. Participants were instructed to select one image per category to use throughout the experiment that they were familiar with and could easily remember. All images were presented at a similar size (viewing angle  $\approx 3^\circ$ ) with a neutral gray background.

### D. Task Protocol

This study involved a single session of data collection. During the experiment, the participants were asked to perform mental visual imagery of four different categories of human faces, animals, objects, and scene images. The task included 5 blocks of visual imagery following either a visual observation or an auditory cue (Fig. 2a and 2b, respectively). During the observation cue blocks, the target image for each category was displayed with a small fixation cross at the center of the screen for 2.5 sec and the participant was instructed to

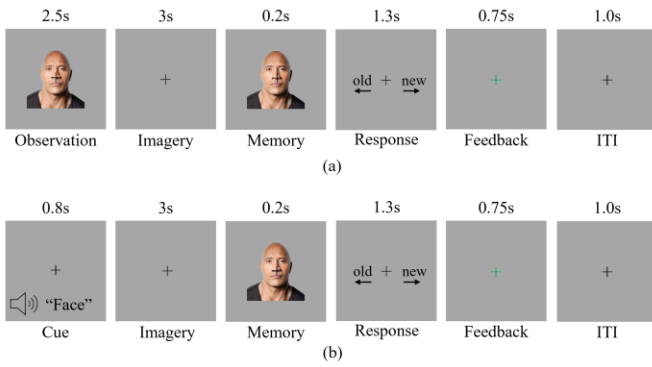


Fig. 2. Task protocol. (a) Observation cue block protocol. (b) Auditory cue block protocol.

passively view it while attempting to keep their gaze on the center cross. The image was then removed from the screen for 3 sec. and the participant was instructed to picture the image that they just saw in vivid detail while keeping their eyes open and fixated on the center cross. The auditory cues followed a similar procedure, except that a verbal cue of “face”, “scene”, “animal”, or “object” was played over a speaker and the participant was instructed to recall the associated image and picture it as vividly as possible in their mind. The timings for image presentation and visual imagery were chosen based on prior literature of decoding visual perception and imagery via EEG [12], [35] and to facilitate the cross-task analysis described below in section H.

After each imagery period in both conditions, an image was flashed on the screen for 200 ms. The displayed images were randomly presented with a 70% chance of being the target image and a 30% chance of being a non-target image. The participant was then instructed to quickly respond with a left or right button press if the presented image was the target category or a non-target category for that trial, respectively. If the response was correct, the fixation cross turned green. If the response was incorrect, the fixation cross would turn red. If the participant responded too slowly or did not provide a response for the trial, the fixation cross turned yellow. The primary purpose of this memory test was to ensure the participant was actively engaged throughout the experiment and to prevent mind wandering. This data could be used to remove trials where the participants may have missed the cue for that trial. Furthermore, this procedure was designed to mimic a typical BCI scenario where feedback of the predicted target is displayed at the end of each trial. The timing of the memory test was chosen based on feedback from initial pilot testing which found 200 ms made the task sufficiently challenging.

The full experiment included 5 blocks of the observation cues and 5 blocks of the auditory cues. Each of the observation cue blocks consisted of 40 trials and lasted approximately 6 minutes. The auditory cue blocks had a shorter cue duration, so 48 trials were included for each block and also lasted approximately 6 minutes. After each block, the participants were required to take a minimum of 1 min break before continuing with the experiment, but they were allowed to take

additional time if needed. A 10 sec resting period preceded each block in which the participants were asked to fixate their gaze on the center cross and keep their mind blank. The entire session lasted approximately 1 hour and 15 minutes.

### E. Data Collection and Preprocessing

EEG data was collected from 32 channels in accordance with the 10-20 standard via the Brain Products actiCAP system with a sampling rate of 500 Hz. Eye movements and blinks were captured by placing four bipolar electrodes around the eyes with a reference placed on the mastoid. Data processing was carried out using the MNE Python package. Data from channels labeled as “bad” were removed and interpolated from the surrounding electrodes. Then, the signal was re-referenced to the common average to remove any background noise that is common across all channels. Eye movement and blink artifacts were removed from the signal using an independent component analysis (ICA) [49]. Artifact removal via ICA was carried out using the following procedure: implementing a high-pass filter at 1 Hz to remove signal drift, implementing notch filters at 60 Hz and its harmonics to remove powerline interference, annotating the task and break periods, running the ICA using MNE’s ICA algorithm with 32 components, manually selecting components that contained artifacts, removing these components, and applying the solution to the unfiltered data.

### F. Visual Perception Classification

Various feature extraction methods and classifiers were evaluated for the prediction of the four visual perception categories. We tested features in the delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-100 Hz) extracted via Morlet Wavelets, Fast Fourier Transform (FFT), and Common Spatial Patterns (CSP). The gamma band was also divided into low (30-60 Hz) and high (60-100 Hz) gamma for evaluation. For features extracted from the delta, theta, alpha, and beta bands, the data was first bandpass filtered between 1-40 Hz to remove low frequency signal drift and high frequency noise. For features including the gamma band, the data was bandpass filtered between 1-100 Hz. A notch filter at 60 Hz was also implemented to remove powerline interference. Only data from the 8 posterior EEG channels (O1, O2, Oz, P3, P4, P7, P8, and Pz) were used for feature selection.

Morlet Wavelets were employed in each desired frequency range equally spaced every 1 Hz for delta and theta bands, every 2 Hz for alpha and beta bands, and every 5 Hz for gamma bands. The mean power in each frequency bin was used as features for classification. For features extracted via the FFT, the mean power in each band was used for classification. CSP features were extracted using the CSP function from the MNE Python package with 8 components in the desired frequency range. After extraction, the features were then normalized using the MinMaxScaler from the scikit-learn Python package to scale the features between 0-1.

The classifiers tested in this analysis include Logistic Regression (LR) with a newton-cg solver, Linear Discriminant

analysis (LDA) with a shrinkage term of 0.1, and support vector machine (SVM) with a linear kernel. These classifiers are known to be robust in EEG classification including the decoding of visual imagery [12]. All classification approaches were cross-validated using a leave-one-block-out (5-fold) cross-validation approach.

### G. Visual Imagery Classification

A similar method as described above was also applied to the two visual imagery conditions to identify the optimal features and classifiers for the prediction of the four imagery categories. Due to the inherently low signal-to-noise ratio involved with visual imagery, an additional preprocessing stage of removing trials with signal amplitude exceeding 100 mV was implemented. This stage effectively removes trials where the participant may have been moving. This stage removed less than 5% of trials for each subject. In addition to the 8 EEG channels used during the visual perception classification, some frontal channels were found to carry information relevant to imagery classification. A grid search analysis of the best channels for imagery led us to include channels O1, O2, Oz, P3, P4, P7, P8, Pz, TP10, F7, F8, and FC6.

### H. Cross-Task Classification

Due to the similarities between the experimental conditions, we were interested in seeing if the inclusion of data from another condition can improve the classification accuracy. For example, can the inclusion of the perception periods improve the classification of the visual imagery periods? To test this approach, we concatenated the data from two conditions and performed a leave-one-block-out (10-fold) cross validation. For this analysis, we used the mean power in the 1-15 Hz band obtained by Morlet Wavelets over the 8 posterior channels used during the perception classification.

### I. Evaluation of Performance

The classifier's performance is evaluated based on the number of trials where the EEG classification output matches the target category for that trial beyond the level of chance. However, small sample sizes can lead to false positives, and Combrisson and Jerbi [50] have suggested to address this issue by adjusting the chance level as a function of sample size ( $n$ ) and number of classes ( $c$ ) using a binomial cumulative distribution. Using this method, the probability of a classification model to predict the correct label at least  $z$  times by chance is given by (1).

$$P(z) = \sum_{i=z}^n \binom{n}{i} \times \left(\frac{1}{c}\right)^i \times \left(\frac{c-1}{c}\right)^{n-i} \quad (1)$$

In this study, the statistically significant threshold level was calculated using the MATLAB (Mathworks Inc., MA, USA) function  $St(\alpha) = \text{binoinv}(1 - \alpha, n, 1/c) \times 100/n$ , where  $\alpha$  is the significance level given by  $\alpha = z/n$ . For this study, each session included a total sample size of  $n = 240$  observations for the observation period and visual imagery period following the observation cue. There were  $n = 288$  observations for the visual imagery period following the auditory cue. The

experiment consisted of  $c = 4$  classes which provided a significance threshold of 29.58% at  $p = 0.05$  for the observation and visual imagery periods following the observation cue. The imagery periods following the auditory cue had significance threshold of 29.17%. In other words, the classification model must achieve a classification accuracy above this threshold to be considered statistically significant.

### J. Pupillometry Data Analysis

Eye tracking data was captured throughout the experiment using a Tobii Pro Nano device. This data consisted of the  $x$  and  $y$  gaze positions along with the pupil diameters for each eye recorded at a sampling rate of 60 Hz. The pupillometry data was preprocessed using the methods outlined in Combrisson and Jerbi [50] and Winn et al. [51]. First, the data was segmented into trials starting from the onset of the cue to one second after the start of the memory test. Then, eye blink artifacts were corrected by identifying segments with *nan* values, removing 5 datapoints from the beginning and end of the *nan* segments, and interpolating the values from the surrounding data points. Trials where over 30% of the data was *nan* values were labeled as bad and were removed from the analysis. The data was then filtered with a second order Butterworth bandpass filter between 1 and 10 Hz to remove low frequency drift and high frequency noise and standardized using  $z$ -score. The trials were then baseline corrected by subtracting out the mean pupil dilating from the 1 sec long inter-trial period before the onset of each trial. Due to issues encountered during data collection, only  $N=22$  and  $N=15$  participants were included in the analysis of the eye tracking data for the observation and auditory blocks, respectively.

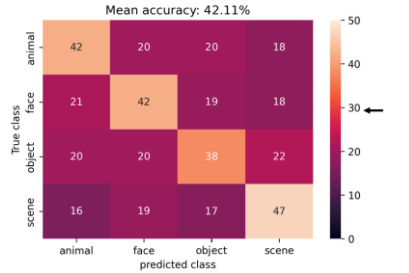
## III. RESULTS

Our first test was to observe the classification accuracy between the four imagery categories during the observation period. Table I presents the classification accuracies when using the Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) classifiers with various band power features extracted via Morlet Wavelets, Fast Fourier Transform (FFT), or Common Spatial Patterns (CSP). The right column presents the best combination of features combining the data in the Delta, Theta, and Alpha bands of brain activity. The highest accuracy was obtained by employing an LDA classifier trained on the mean Morlet wavelets in the 1-15 Hz range equally spaced every 2 Hz from the eight posterior EEG channels. This yielded a mean classification accuracy of 42.11% across all 26 subjects (Fig. 3a). This was significantly higher than the significance threshold of 29.58% ( $p < 0.01 \times 10^{-7}$ ). To localize the most predictive channels for the prediction of the observation trials, we also performed a searchlight analysis where only the data of a single EEG channel was used to perform the classification (Fig. 3b). As expected, this analysis revealed that the posterior electrodes directly over the primary visual cortex obtained the highest classification accuracy.

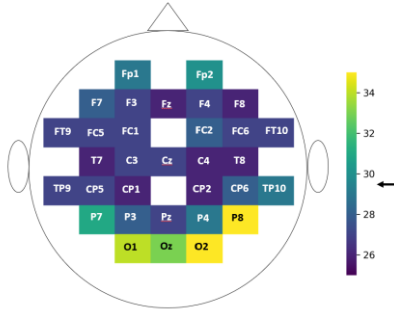
We next attempted to classify between the four visual imagery classes from short-term working memory following

**TABLE I**  
CLASSIFICATION ACCURACY (%) OF VARIOUS CLASSIFIERS AND FEATURES DURING VISUAL OBSERVATION

Feature Extraction	Classifier	Delta	Theta	Alpha	Beta	Low Gamma	High Gamma	Gamma	Delta, Theta, Alpha
Morlet	LR	34.53±6.32	39.24±6.00	37.85±4.91	32.98±4.48	28.76±3.70	26.85±2.75	28.42±3.75	41.08±6.48
	LDA	35.30±5.54	40.20±6.14	38.09±5.84	32.41±4.39	27.80±3.84	26.08±3.14	27.50±2.93	42.11±6.85
	SVM	34.21±5.29	38.29±6.48	37.69±5.61	32.04±4.28	28.81±3.77	27.45±2.72	28.09±3.28	40.43±7.54
FFT	LR	29.64±4.42	33.78±5.97	31.21±3.50	31.31±4.26	30.25±3.23	28.50±2.84	29.55±3.00	36.97±4.61
	LDA	30.12±4.64	33.49±6.29	31.26±3.51	30.96±4.63	30.44±3.80	28.47±2.78	30.36±3.36	36.82±5.15
	SVM	29.74±4.36	33.36±6.02	30.89±3.59	30.81±4.09	29.66±3.13	26.97±2.72	28.23±3.17	36.41±4.81
CSP	LR	31.89±3.83	38.38±6.75	32.83±4.50	28.92±4.03	25.19±3.03	25.78±2.97	25.84±2.32	32.65±5.22
	LDA	32.08±3.45	38.46±6.89	32.72±5.29	28.81±3.77	25.01±3.42	25.38±3.13	25.22±2.38	33.76±5.03
	SVM	32.13±3.73	37.90±6.81	32.59±4.45	28.38±3.04	25.46±2.88	24.93±2.19	25.25±2.03	33.75±4.41



(a)



(b)

**Fig. 3.** Classification of observation periods. (a) Mean confusion matrix obtained from classification of neural data during visual observation of the four image categories. (b) Heatmap of accuracy obtained during channel-wise classification of observation periods across all subjects. Black arrows on color bar demarcate significance threshold at 29.58% ( $p < 0.05$ ).

the observation cue. Similar preprocessing, feature extraction, and classification methods as for the observation periods were tested (Table II). We found that the LR classifier trained on the mean Morlet wavelets in the 1-15 Hz provided the best classification accuracy with a mean of 30.05% across all subjects which was significantly higher than the significance threshold of 29.58% ( $p < 0.05$ ) (Fig. 4a). Fig. 4b presents the average channel-wise heatmap across all subjects, which shows a trend similar to the observation period where posterior channels carry the most relevant information for classification. However, individual analysis of the data shows that there may be some greater contributions from some of the more frontal channels. Due to the low accuracy in decoding the visual imagery categories, no individual channels were found to exceed the significance threshold during the exploratory channel-wise searchlight analysis. Also included in Fig. 4c is the channel-wise heatmap of a subject with high classification

**TABLE II**  
CLASSIFICATION ACCURACY (%) OF VARIOUS CLASSIFIERS AND FEATURES DURING SHORT-TERM VISUAL IMAGERY

Feature Extraction	Classifier	Delta	Theta	Alpha	Beta	Low Gamma	High Gamma	Gamma	Delta, Theta, Alpha
Morlet	LR	27.14±3.72	28.77±3.73	28.79±4.24	26.53±4.14	24.96±2.92	25.14±2.88	25.04±3.80	30.05±3.58
	LDA	27.18±4.20	28.38±3.87	28.05±4.36	26.56±3.78	25.43±2.67	24.33±2.99	25.17±3.66	29.24±3.92
FFT	LR	26.39±3.67	28.41±3.27	28.87±4.09	27.09±4.70	25.41±2.88	25.67±2.70	25.19±3.45	28.58±4.02
	LDA	26.74±3.36	26.64±3.62	26.82±3.87	26.23±2.84	25.83±3.64	25.61±3.34	25.62±3.34	27.16±3.68
CSP	LR	26.63±3.64	26.81±4.38	26.52±3.77	25.74±2.45	25.83±3.12	26.21±3.09	25.81±3.56	27.27±3.86
	LDA	26.37±3.66	26.46±3.66	26.50±3.98	25.87±2.75	25.45±2.88	25.23±3.78	25.13±3.42	27.25±4.024
CSP	LR	27.48±3.29	26.03±4.00	26.34±3.92	27.03±4.12	24.88±2.89	24.47±2.81	24.53±2.62	28.01±4.28
	LDA	27.68±3.85	26.24±3.69	26.46±3.60	27.74±3.85	24.55±2.96	24.35±3.13	24.91±2.55	27.54±4.04
CSP	LR	27.46±3.19	26.47±3.73	25.56±3.95	26.51±3.94	24.70±2.79	24.76±2.58	24.71±2.26	27.40±4.08

**TABLE III**  
CLASSIFICATION ACCURACY (%) OF VARIOUS CLASSIFIERS AND FEATURES DURING LONG-TERM VISUAL IMAGERY

Feature Extraction	Classifier	Delta	Theta	Alpha	Beta	Low Gamma	High Gamma	Gamma	Delta, Theta, Alpha
Morlet	LR	25.25±3.36	26.05±2.52	25.51±2.89	25.79±3.20	25.40±2.47	26.10±3.13	26.18±2.93	26.74±3.06
	LDA	25.73±3.51	25.69±2.01	26.01±3.03	26.02±3.45	25.23±2.69	25.18±2.88	25.38±3.01	26.14±3.78
	SVM	25.38±2.98	26.45±2.73	24.75±3.24	26.25±3.05	25.16±2.68	25.07±2.67	25.99±2.72	25.76±3.63
FFT	LR	25.13±2.65	26.81±3.25	25.43±3.95	24.97±3.39	25.45±2.35	24.99±2.80	24.93±2.91	26.21±3.46
	LDA	24.95±2.81	26.85±3.04	24.72±3.86	25.16±3.51	25.72±2.69	25.62±3.08	25.48±2.52	26.25±3.50
	SVM	25.04±2.62	26.66±2.95	25.40±3.27	25.02±3.46	24.93±3.03	24.62±2.76	24.36±2.23	26.22±3.45
CSP	LR	26.50±3.48	25.87±3.10	26.05±3.15	25.31±2.81	24.88±2.74	25.00±2.37	25.33±2.79	25.61±2.73
	LDA	26.35±2.99	26.02±3.51	25.83±3.11	25.09±2.57	24.87±2.68	24.59±2.24	25.37±2.82	25.35±2.73
	SVM	26.48±3.06	26.23±3.09	25.91±3.39	25.32±2.36	25.12±2.53	25.57±2.49	24.76±2.63	25.74±2.84

accuracy to highlight the channels relevant for short-term visual imagery.

Finally, we tested the classification accuracy of the four visual imagery classes from long-term memory following the auditory cue (Table III). The best classification accuracy of 26.74% across all subjects was obtained by using 25 mean Morlet Wavelets logarithmically spaced in the 1-100 Hz band (Fig. 5a). However, this combination still did not pass the significance threshold of 29.17% for this section. Similar to the results from the short-term visual imagery analysis, no significant channels were revealed during the exploratory channel-wise searchlight analysis of the long-term imagery categories (Fig. 5b). However, individual analysis of subjects with higher accuracy also reveals significant contributions from frontal channels during visual imagery (Fig. 5c). Fig. 6 presents the classification accuracies for each subject across the tasks for comparison.

To improve the classification accuracy of the imagery periods, we were interested to see if the addition of data from the observation periods would improve the prediction ability. Contrary to our expectations, classification performance for all trial periods suffered when training included data from other periods. When the data from the observation periods were combined with the short-term visual imagery periods, the observation classification accuracy dropped to 38.47% and the imagery classification accuracy dropped to 27.30%. The observation periods and long-term imagery periods following the auditory cue concatenated together dropped to 36.50% and 26.42%, respectively. The classification accuracy of the imagery periods following the observation cue and auditory cue concatenated together dropped to 27.90% and 25.70%, respectively.

The time course of the Morlet Wavelet features for each of the tasks is presented in Fig. 7. All tasks exhibited primary

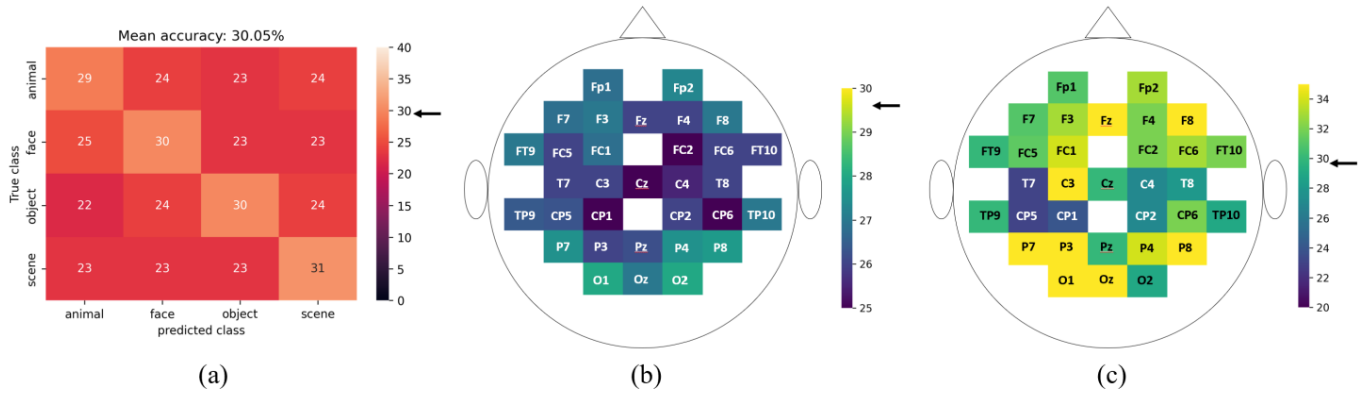


Fig. 4. Classification of short-term imagery periods following observation cue. (a) Mean confusion matrix obtained from classification of neural data during visual imagery of the four image categories following the observation cue demonstrates significantly above chance classification accuracy of 30.05%. (b) Heatmap of accuracy obtained during channel-wise classification of imagery periods following the observation cue across all subjects. (c) Heatmap of accuracy obtained during channel-wise classification of imagery periods following the observation cue for a subject with high classification accuracy. Black arrows on color bar demarcate significance threshold at 29.58% ( $p < 0.05$ ).

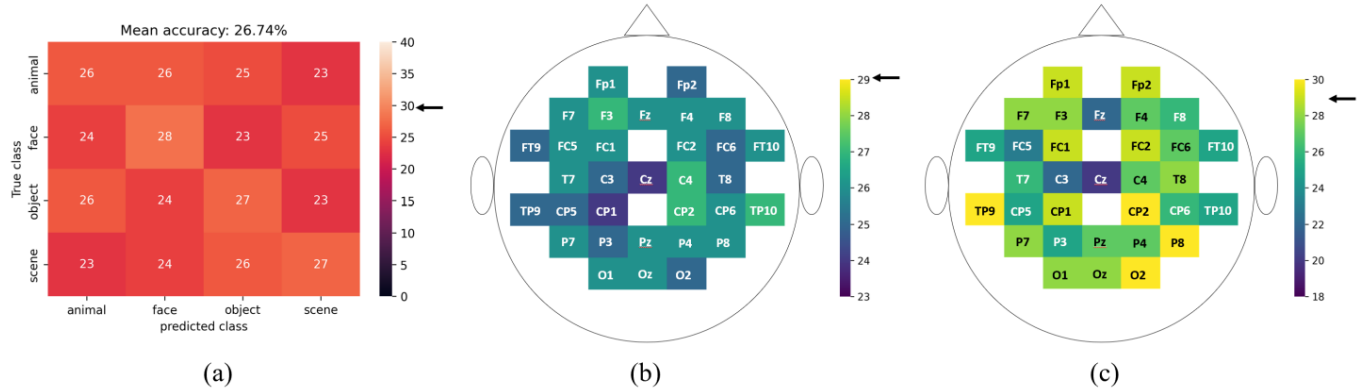


Fig. 5. Classification of long-term imagery periods following auditory cue. (a) Mean confusion matrix obtained from classification of neural data during visual imagery of the four image categories following the auditory cue provides an average classification accuracy of 26.74%. (b) Heatmap of accuracy obtained during channel-wise classification of imagery periods following the auditory cue across all subjects. (c) Heatmap of accuracy obtained during channel-wise classification of imagery periods following the auditory cue for a subject with high classification accuracy. Black arrows on color bar demarcate significance threshold at 29.17% ( $p < 0.05$ ).

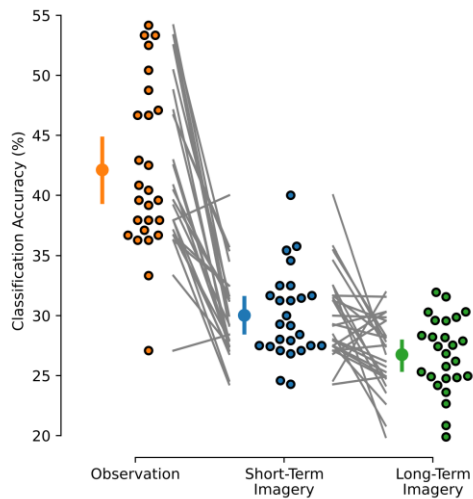
activity within the alpha band of brain activity; however, the timing for the appearance of the activity differed between each task. Observation of the image produced activity in the alpha band appearing almost immediately and dropping about 400 ms following the presentation of the image (Fig. 7a). The two imagery tasks produced a more sustained activity throughout the trial, which appeared approximately 500 ms after the start of the imagery period. However, the short-term visual imagery task peaks around 1500 ms after the start of the imagery period whereas the long-term visual imagery task peaks around 500 ms after the start and begins to decrease.

We also analyzed the pupillometry data to see if this could be used to identify when a participant was engaged in the visual imagery task. Fig. 8 presents the mean pupil dilation across participants during the observation cue blocks and the auditory cue blocks, respectively. During the observation cue blocks, the pupil contracts with the onset of the stimulus presentation. When the images are removed and the participant is instructed to perform visual imagery, the pupil dilates back to baseline. The pupil contracts again after the image is flashed

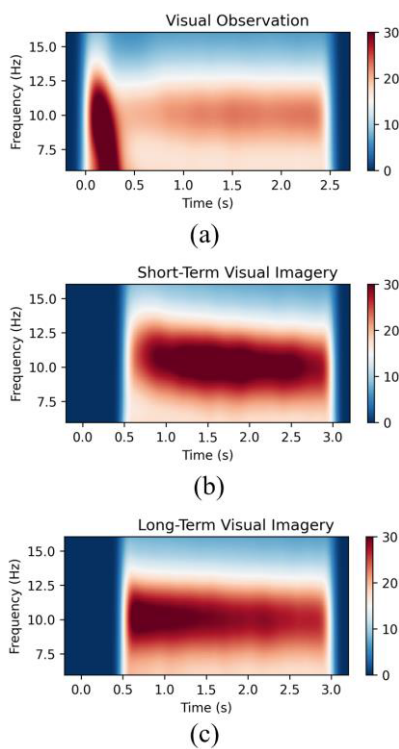
during the memory periods. During the auditory cue blocks, the pupil begins to dilate immediately following the auditory cue and peaks after approximately 1.5 sec before returning to baseline. The pupil contracts during the memory periods after the image is flashed. This dilation of the pupil during the imagery periods provides a good indication that participants were actively engaged in the task.

#### IV. DISCUSSION

This study demonstrates that decoding visual imagery from EEG is a challenging task. From the early work of Bobrov et al. [33], Kosmyna et al. [30], Lee et al. [12], [52], and Xie et al. [35], one of the major points of contention between the experimental procedures was the presentation of the target image during the task directly before the imagery period. It could be argued that this is not a true test of spontaneous visual imagery from long-term memory but rather holding the object in short-term working memory. Our study presents the first direct comparison between the ability to decode visual imagery following observation of the target

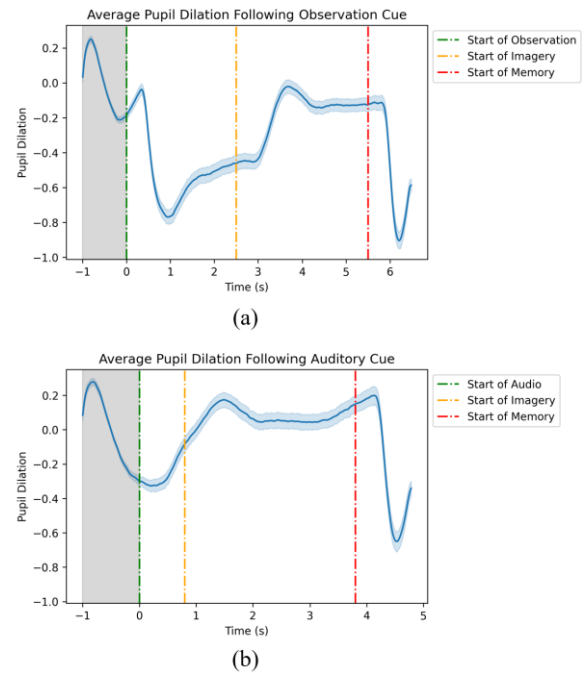


**Fig. 6.** Overall classification accuracy during observation (42.11%), short-term visual imagery (30.05%), and long-term visual imagery (26.74%) periods. Individual participants are presented beside the mean as circles with a black outline. Error bars indicate a 95% confidence interval.



**Fig. 7.** Mean frequency vs time plots of Morlet Wavelet features across the 8 posterior EEG channels for (a) visual observation, (b) short-term visual imagery, and (c) long-term visual imagery periods.

image and following an auditory cue for the target image. In accordance with our hypothesis, our classifier was able to achieve greater accuracy in predicting between the four image categories during the short-term visual imagery task following the observation cue compared to the long-term imagery task following the auditory cue. Also as expected, visual imagery produced a more nuanced pattern of activity that is more



**Fig. 8.** Pupil diameter changes averaged across all trials. (a) Mean pupil diameter during the observation cue blocks across all participants (b) Mean pupil diameter during the auditory cue blocks across all participants. The gray shaded region indicates the period used for baseline correction. The blue shaded region around the pupil diameter trace demarcates the 95% confidence interval.

difficult to untangle using multivariate decoding of EEG data compared to actual visual observation of the images.

In an early study by Lee et al. [53], researchers examined the differences in brain areas activated during visual perception and imagery. The results of this study showed a considerable overlap in activity between the two conditions in many areas of the brain; however, this overlap was neither uniform nor complete. They saw nearly complete overlap in frontal and parietal regions involved in various types of cognitive control processes such as the retrieval of episodic information, performing visual inspection, generation of visual images, attention, spatial working memory, and visuospatial processing. On the other hand, the activations in the occipital cortex were stronger and more diffuse during perception than during imagery. This exemplifies that the occipital regions are more strongly driven by sensory information rather than the information stored in memory. These regions are responsible for facilitating object detection and identification that are not necessarily required for the visualization of mental images [53]. Similar results were shown in this current study. The channel-wise searchlight results showed a large overlap in areas that were most informative for prediction during visual observation and imagery. The observation task was mostly driven by the occipital electrodes which were receiving the sensory information, but some of the frontal electrodes also seemed to carry relevant information. The imagery task showed much lower classification accuracies in the occipital channels and a greater influence of the frontal electrodes.

We also saw that spectral power below 15 Hz, and specifically alpha band power (8-13 Hz), provided the most



informative feature for classification during the observation periods and the short-term imagery periods. While this is consistent with the results found by Xie et al. [35], the results from Lee et al. [12] and Kosmyna et al. [30] indicates that higher gamma activity (30-100 Hz) may also carry information relevant for visual imagery. In our analysis, we also found that inclusion of the higher gamma range of brain activity may be beneficial for classification of the long-term visual imagery task. Unfortunately, the results were not significant, so a definitive conclusion cannot be made. Additionally, even though we found similar features and channels between the conditions, we were unable to utilize the data from the other conditions to improve the classification accuracy. It is possible that even though there is a significant overlap between the activity associated with the different experimental conditions, there is still too much variance or too little training data for the classifier to make use of the additional information. For example, the observation trials were mainly driven by activity in the occipital cortex related to the sensory perception of the presented image while the imagery trials received more contributions from the frontal areas associated with memory. Furthermore, our analyses revealed that the Morlet Wavelet features used in these analyses exhibited differences in the timing of activation across each of the tasks. It might also be that the two imagery conditions are confounded by a difference in the memory conditions used.

The imagery periods following the observation cue are more of a short-term working memory task while the imagery periods following the auditory cue are more of a long-term memory retrieval task. In a study by Ganis et al [54], researchers have shown a differential effect in the visual responses generated during visual imagery of famous faces generated from short-term memory and long-term memory. They found that both tasks activated similar areas of the brain, but the activity was greater during short-term memory when the subjects were asked to memorize specific pictures of celebrities. During the long-term task where they were asked to imagine the famous person without the presentation of an image, the neural activations were lower across all relevant brain regions. However, focusing on specific features of the imagined faces such as the eyes, lips or nose was shown to increase activation regardless of memory type. These results suggest that the type of cue and the instructions given to the participants can play a large role in the neural processes used and the activity evoked during mental visual imagery.

In the current experiment, retrieving a mental image from long-term memory following an auditory cue rather than retaining a visually presented image in working memory produced weaker or more variable neural activity that was more difficult to decode using common machine learning techniques. For mental imagery, the participants in this study were instructed to recreate the image in their mind and visualize the details as clearly as possible. However, in the post-experiment survey responses, multiple participants reported strategies of repeating the objects' name in their mind or thinking about the colors of the images, which may not be an optimal imagery strategy and introduces unsystematic variability in the neural signals during the imagery periods. Future work may benefit

from providing more detailed instructions for how to visualize the target category or providing feedback on the classifier's prediction in real-time to allow the participants to adapt their strategy throughout the experiment.

The results of our pupillometry analysis demonstrated an increase in pupil dilation during mental imagery in accordance with previous literature [47]. In the observation cue blocks, a decrease in the diameter when the stimulus was presented and an increase in diameter when the stimulus was removed is shown as expected. However, it is difficult to distinguish whether this dilation was due to effortful mental imagery or if the pupil was just returning to baseline after stimulus presentation. One of the limitations of this study with regards to pupillometry is that the trial periods were short (8.75 secs for the observation cue trial and 7.05 secs for the auditory cue trials) which may not be ideal for the analysis of pupillometry data [50], [51]. A more appropriate eye tracking experiment would provide longer resting periods between each task element to allow the pupil to fully return to baseline. Furthermore, our inter-trial period of 1 sec that was used for baseline correction may not be an appropriate time to allow the pupil diameter to fully return to baseline between trials. The large variability during this period may be due to eye blinks or movements which could interfere with the ability to interpret the data from the remainder of the trial.

## V. CONCLUSION

Visual imagery presents the possibility for an intuitive paradigm for BCI applications that can directly convey the user's intentions with many natural ways of envisioning an intended action. However, the work presented in this study reveals that true, spontaneous visual imagery from long-term memory is difficult to decode from EEG. This method of spontaneous visual imagery produces a more variable neural signal compared to short-term retention of a visual image in working memory. One potential limitation of this study is that participants were not given explicit instructions for how to perform the imagery and no feedback for successful imagery was provided throughout the experiment. Future work in decoding visual imagery from EEG may benefit from providing more explicit visualization instructions as well as multiple sessions with real-time feedback of visualization ability. This will allow the users to hone their strategies over time and provide more data to adapt to more advanced classification techniques.

## ACKNOWLEDGMENT

Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- [1] J. N. Mak and J. R. Wolpaw, "Clinical applications of brain-computer interfaces: Current state and future prospects," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 187–199, 2009, doi: [10.1109/RBME.2009.2035356](https://doi.org/10.1109/RBME.2009.2035356).
- [2] A. Frisoli et al., "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks," *IEEE Trans. Syst. Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1169–1179, Nov. 2012, doi: [10.1109/TSMCC.2012.2226444](https://doi.org/10.1109/TSMCC.2012.2226444).

- [3] J. Kilmarx, R. Abiri, S. Borhani, Y. Jiang, and X. Zhao, "Sequence-based manipulation of robotic arm control in brain machine interface," *Int. J. Intell. Robot. Appl.*, vol. 2, no. 2, pp. 149–160, Jun. 2018, doi: [10.1007/s41315-018-0049-7](https://doi.org/10.1007/s41315-018-0049-7).
- [4] M. Tariq, P. M. Trivailo, and M. Simic, "EEG-based BCI control schemes for lower-limb assistive-robots," *Frontiers Hum. Neurosci.*, vol. 12, p. 312, Aug. 2018, doi: [10.3389/fnhum.2018.00312](https://doi.org/10.3389/fnhum.2018.00312).
- [5] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J. Neurosci. Methods*, vol. 167, no. 1, pp. 15–21, Jan. 2008, doi: [10.1016/j.jneumeth.2007.07.017](https://doi.org/10.1016/j.jneumeth.2007.07.017).
- [6] N. Birbaumer, "Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, Nov. 2006, doi: [10.1111/j.1469-8986.2006.00456.x](https://doi.org/10.1111/j.1469-8986.2006.00456.x).
- [7] A. Y. Kaplan, S. L. Shishkin, I. P. Ganin, I. A. Basyul, and A. Y. Zhigalov, "Adapting the P300-based brain–computer interface for gaming: A review," *IEEE Trans. Comput. Intell. AI Games*, vol. 5, no. 2, pp. 141–149, Jun. 2013, doi: [10.1109/TCAIG.2012.2237517](https://doi.org/10.1109/TCAIG.2012.2237517).
- [8] M. Ahn, M. Lee, J. Choi, and S. Jun, "A review of brain–computer interface games and an opinion survey from researchers, developers and users," *Sensors*, vol. 14, no. 8, pp. 14601–14633, Aug. 2014, doi: [10.3390/s140814601](https://doi.org/10.3390/s140814601).
- [9] S. Borhani, J. Yu, J. Cate, J. Kilmarx, R. Abiri, and X. Zhao, "Clash of minds: A BCI car racing game in simulated virtual reality environment," in *Proc. Biomed. Eng. Soc. (BMES) Annu. Meeting*, 2018.
- [10] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, Feb. 2019, Art. no. 011001, doi: [10.1088/1741-2552/aaf12e](https://doi.org/10.1088/1741-2552/aaf12e).
- [11] M. Ahn and S. C. Jun, "Performance variation in motor imagery brain–computer interface: A brief review," *J. Neurosci. Methods*, vol. 243, pp. 103–110, Mar. 2015, doi: [10.1016/j.jneumeth.2015.01.033](https://doi.org/10.1016/j.jneumeth.2015.01.033).
- [12] S.-H. Lee, M. Lee, and S.-W. Lee, "Neural decoding of imagined speech and visual imagery as intuitive paradigms for BCI communication," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2647–2659, Dec. 2020, doi: [10.1109/TNSRE.2020.3040289](https://doi.org/10.1109/TNSRE.2020.3040289).
- [13] P. Wierzgała, D. Zapala, G. M. Wojcik, and J. Masiak, "Most popular signal processing methods in motor-imagery BCI: A review and meta-analysis," *Front. Neuroinform.*, vol. 12, p. 78, Nov. 2018, doi: [10.3389/fninf.2018.00078](https://doi.org/10.3389/fninf.2018.00078).
- [14] E. V. C. Friedrich, R. Scherer, and C. Neuper, "The effect of distinct mental strategies on classification performance for brain–computer interfaces," *Int. J. Psychophysiol.*, vol. 84, no. 1, pp. 86–94, Apr. 2012, doi: [10.1016/j.ijpsycho.2012.01.014](https://doi.org/10.1016/j.ijpsycho.2012.01.014).
- [15] B. Z. Allison and C. Neuper, "Could anyone use a BCI?" in *Brain–Computer Interfaces: Applying Our Minds to Human–Computer Interaction* (Human–Computer Interaction Series). London, U.K.: Springer, 2010, pp. 35–54, doi: [10.1007/978-1-84996-272-8\\_3](https://doi.org/10.1007/978-1-84996-272-8_3).
- [16] F.-B. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives," *Prog. Neurobiol.*, vol. 90, no. 4, pp. 418–438, Apr. 2010, doi: [10.1016/j.pneurobio.2009.11.005](https://doi.org/10.1016/j.pneurobio.2009.11.005).
- [17] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: Current challenges and emerging trends," *Frontiers Neuroeng.*, vol. 5, p. 14, Jul. 2012, doi: [10.3389/fneng.2012.00014](https://doi.org/10.3389/fneng.2012.00014).
- [18] L. M. McCane et al., "Brain–computer interface (BCI) evaluation in people with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis Frontotemporal Degeneration*, vol. 15, nos. 3–4, pp. 207–215, Jun. 2014, doi: [10.3109/21678421.2013.865750](https://doi.org/10.3109/21678421.2013.865750).
- [19] I. Volosyak, D. Valbuena, T. Luth, T. Malechka, and A. Graser, "BCI demographics II: How many (and what kinds of) people can use a high-frequency SSVEP BCI?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 3, pp. 232–239, Jun. 2011, doi: [10.1109/TNSRE.2011.2121919](https://doi.org/10.1109/TNSRE.2011.2121919).
- [20] F. Gemblér, P. Stawicki, and I. Volosyak, "Exploring the possibilities and limitations of multitarget SSVEP-based BCI applications," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 1488–1491, doi: [10.1109/EMBC.2016.7590991](https://doi.org/10.1109/EMBC.2016.7590991).
- [21] I. Volosyak, F. Gemblér, and P. Stawicki, "Age-related differences in SSVEP-based BCI performance," *Neurocomputing*, vol. 250, pp. 57–64, Aug. 2017, doi: [10.1016/j.neucom.2016.08.121](https://doi.org/10.1016/j.neucom.2016.08.121).
- [22] J. Pereira, P. Ofner, A. Schwarz, A. I. Sburlea, and G. R. Müller-Putz, "EEG neural correlates of goal-directed movement intention," *NeuroImage*, vol. 149, pp. 129–140, Apr. 2017.
- [23] R. Abiri, S. Borhani, J. Kilmarx, C. Esterwood, Y. Jiang, and X. Zhao, "A usability study of low-cost wireless brain–computer interface for cursor control using online linear model," *IEEE Trans. Human-Mach. Syst.*, vol. 50, no. 4, pp. 287–297, Aug. 2020, doi: [10.1109/THMS.2020.2983848](https://doi.org/10.1109/THMS.2020.2983848).
- [24] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain Topography*, vol. 23, no. 2, pp. 194–198, Jun. 2010.
- [25] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain–computer interface (BCI)?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 145–147, Jun. 2003, doi: [10.1109/TNSRE.2003.814481](https://doi.org/10.1109/TNSRE.2003.814481).
- [26] M. C. Thompson, "Critiquing the concept of BCI illiteracy," *Sci. Eng. Ethics*, vol. 25, no. 4, pp. 1217–1233, Aug. 2019, doi: [10.1007/s11948-018-0061-1](https://doi.org/10.1007/s11948-018-0061-1).
- [27] B. Allison, T. Luth, D. Valbuena, A. Teymourian, I. Volosyak, and A. Graser, "BCI demographics: How many (and what kinds of) people can use an SSVEP BCI?" *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 107–116, Apr. 2010, doi: [10.1109/TNSRE.2009.2039495](https://doi.org/10.1109/TNSRE.2009.2039495).
- [28] S. Becker, K. Dhindsa, L. Mousapour, and Y. Al Dabagh, "BCI illiteracy: It's us, not them. Optimizing BCIs for individual brains," in *Proc. 10th Int. Winter Conf. Brain-Comput. Interface (BCI)*, Feb. 2022, pp. 1–3.
- [29] M.-H. Lee et al., "EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, May 2019, Art. no. giz002.
- [30] N. Kosmyna, J. T. Lindgren, and A. Lécuyer, "Attending to visual stimuli versus performing visual imagery as a control strategy for EEG-based brain–computer interfaces," *Sci. Rep.*, vol. 8, no. 1, Sep. 2018, Art. no. 1, doi: [10.1038/s41598-018-31472-9](https://doi.org/10.1038/s41598-018-31472-9).
- [31] I. Simanova, M. van Gerven, R. Oostenveld, and P. Hagoort, "Identifying object categories from event-related EEG: Toward decoding of conceptual representations," *PLoS ONE*, vol. 5, no. 12, Dec. 2010, Art. no. e14465, doi: [10.1371/journal.pone.0014465](https://doi.org/10.1371/journal.pone.0014465).
- [32] P. Shenoy and D. S. Tan, "Human-aided computing: Utilizing implicit human processing to classify images," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2008, pp. 845–854.
- [33] P. Bobrov, A. Frolov, C. Cantor, I. Fedulova, M. Bakhnyan, and A. Zhavoronkov, "Brain–computer interface based on generation of visual images," *PLoS ONE*, vol. 6, no. 6, Jun. 2011, Art. no. e20674, doi: [10.1371/journal.pone.0020674](https://doi.org/10.1371/journal.pone.0020674).
- [34] R. Keogh and J. Pearson, "Mental imagery and visual working memory," *PLoS ONE*, vol. 6, no. 12, Dec. 2011, Art. no. e29221, doi: [10.1371/journal.pone.0029221](https://doi.org/10.1371/journal.pone.0029221).
- [35] S. Xie, D. Kaiser, and R. M. Cichy, "Visual imagery and perception share neural representations in the alpha frequency band," *Current Biol.*, vol. 30, no. 13, pp. 2621–2627, Jul. 2020, doi: [10.1016/j.cub.2020.04.074](https://doi.org/10.1016/j.cub.2020.04.074).
- [36] M. D'Esposito and B. R. Postle, "The cognitive neuroscience of working memory," *Annu. Rev. Psychol.*, vol. 66, no. 1, pp. 115–142, Jan. 2015, doi: [10.1146/annurev-psych-010814-015031](https://doi.org/10.1146/annurev-psych-010814-015031).
- [37] A. B. Randolph, "Not all created equal: Individual-technology fit of brain–computer interfaces," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, Jan. 2012, pp. 572–578, doi: [10.1109/HICSS.2012.451](https://doi.org/10.1109/HICSS.2012.451).
- [38] E. M. Hammer et al., "Psychological predictors of SMR-BCI performance," *Biol. Psychol.*, vol. 89, no. 1, pp. 80–86, Jan. 2012, doi: [10.1016/j.biopsycho.2011.09.006](https://doi.org/10.1016/j.biopsycho.2011.09.006).
- [39] W. Burde and B. Blankertz, "Is the locus of control of reinforcement a predictor of brain–computer interface performance?" in *Proc. 3rd Int. Brain-Comput. Interface Workshop Training Course*. Graz, Austria: Verlag der Technischen Universität Graz, 2006, pp. 76–77.
- [40] B. Blankertz et al., "Neurophysiological predictor of SMR-based BCI performance," *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, Jul. 2010, doi: [10.1016/j.neuroimage.2010.03.022](https://doi.org/10.1016/j.neuroimage.2010.03.022).
- [41] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, and J. Doyon, "Functional neuroanatomical networks associated with expertise in motor imagery," *NeuroImage*, vol. 41, no. 4, pp. 1471–1483, Jul. 2008, doi: [10.1016/j.neuroimage.2008.03.042](https://doi.org/10.1016/j.neuroimage.2008.03.042).
- [42] S. Halder et al., "Prediction of brain–computer interface aptitude from individual brain structure," *Frontiers Human Neurosci.*, vol. 7, p. 105, Apr. 2013, doi: [10.3389/fnhum.2013.00105](https://doi.org/10.3389/fnhum.2013.00105).
- [43] J. Kilmarx, E. Oblak, J. Sulzer, and J. Lewis-Peacock, "Towards a common template for neural reinforcement of finger individuation," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Jan. 2021.

- [44] E. Oblak, J. Lewis-Peacock, and J. Sulzer, "Differential neural plasticity of individual fingers revealed by fMRI neurofeedback," *J. Neurophysiol.*, vol. 125, no. 5, pp. 1720–1734, May 2021.
- [45] B. Kaneshiro, M. Perreau Guimaraes, H.-S. Kim, A. M. Norcia, and P. Suppes, "A representational similarity analysis of the dynamics of object processing using single-trial EEG classification," *PLoS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0135697, doi: [10.1371/journal.pone.0135697](https://doi.org/10.1371/journal.pone.0135697).
- [46] D. F. Marks, "Visual imagery differences in the recall of pictures," *Brit. J. Psychol.*, vol. 64, no. 1, pp. 17–24, Feb. 1973, doi: [10.1111/j.2044-8295.1973.tb01322.x](https://doi.org/10.1111/j.2044-8295.1973.tb01322.x).
- [47] E. Granholm and S. R. Steinhauer, "Pupillometric measures of cognitive and emotional processes," *Int. J. Psychophysiol.*, vol. 52, no. 1, pp. 1–6, Mar. 2004, doi: [10.1016/j.ijpsycho.2003.12.001](https://doi.org/10.1016/j.ijpsycho.2003.12.001).
- [48] N. Kriegeskorte, "Representational similarity analysis—Connecting the branches of systems neuroscience," *Frontiers Syst. Neurosci.*, vol. 2, p. 4, Nov. 2008, doi: [10.3389/neuro.06.004.2008](https://doi.org/10.3389/neuro.06.004.2008).
- [49] K.-S. Hong and M. J. Khan, "Hybrid brain–computer interface techniques for improved classification accuracy and increased number of commands: A review," *Frontiers Neurobotics*, vol. 11, p. 35, Jul. 2017, doi: [10.3389/fnbot.2017.00035](https://doi.org/10.3389/fnbot.2017.00035).
- [50] E. Combrisson and K. Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *J. Neurosci. Methods*, vol. 250, pp. 126–136, Jul. 2015, doi: [10.1016/j.jneumeth.2015.01.010](https://doi.org/10.1016/j.jneumeth.2015.01.010).
- [51] M. B. Winn, D. Wendt, T. Koelewijn, and S. E. Kuchinsky, "Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started," *Trends Hearing*, vol. 22, Jan. 2018, Art. no. 233121651880086.
- [52] S. Mathôt and A. Vilotijević, "Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis," *Behav. Res. Methods*, vol. 55, no. 6, pp. 3055–3077, Aug. 2022.
- [53] S.-H. Lee, M. Lee, J.-H. Jeong, and S.-W. Lee, "Towards an EEG-based intuitive BCI communication system using imagined speech and visual imagery," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 4409–4414, doi: [10.1109/SMC.2019.8914645](https://doi.org/10.1109/SMC.2019.8914645).
- [54] G. Ganis, W. L. Thompson, and S. M. Kosslyn, "Brain areas underlying visual mental imagery and visual perception: An fMRI study," *Cogn. Brain Res.*, vol. 20, no. 2, pp. 226–241, Jul. 2004, doi: [10.1016/j.cogbrainres.2004.02.012](https://doi.org/10.1016/j.cogbrainres.2004.02.012).
- [55] A. Ishai, "Visual imagery of famous faces: Effects of memory and attention revealed by fMRI," *NeuroImage*, vol. 17, no. 4, pp. 1729–1741, Dec. 2002, doi: [10.1006/nimg.2002.1330](https://doi.org/10.1006/nimg.2002.1330).