Biomedical Signal Processing and Control 70 (2021) 102996

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Deep learning helps EEG signals predict different stages of visual processing in the human brain

Nalin Mathur[a], Anubha Gupta[b,*], Snehlata Jaswal[c], Rohit Verma[d]

[a] Netaji Subhas University of Technology, New Delhi, India
[b] SBILab, Department of ECE, IIIT-Delhi, New Delhi, 110020, India
[c] Department of Psychology, Chaudhary Charan Singh University, Meerut, India
[d] Department of Psychology, All India Institute of Medical Sciences, New Delhi 110029, India

ARTICLE INFO

Keywords:
EEG
Visual feature binding
Visual working memory
Visual processing
Convolutional neural network

ABSTRACT

Analysis of electroencephalogram (EEG) signals to determine the nature of visual stimuli, being experienced by a person, is an active area of research. It is key to understand the link between human brain and behavior, especially for brain computer interface (BCI) applications and rehabilitation of patients suffering with neurological disorders. In this research, we conducted an experiment comparing two stages of visual processing, determined distinct EEG signals associated with them, and subsequently used a classifier to distinguish the two stages. EEG data was collected using a feature-binding experiment that required subjects to detect changes in color and shape binding after 100 ms and after 1500 ms. The two stages denoted by these study-test intervals were determined using features extracted from both time and frequency domains. These were used to separately train various machine learning classifiers. The time–frequency domain representation of the signal was used to train a convolutional neural network (CNN). Promising results were obtained. Thus, the contribution of the paper is two-fold. Firstly, we carry out EEG data analysis using deep learning to classify whether the EEG trial belongs to 100 ms class or 1500 ms class. Secondly, we connect these results to predict different stages of visual processing in human brain and visual feature binding. Thus, deep learning can help us predict the stages of visual processing and, hence, unlock important insights regarding the temporal dynamics of brain functioning. This can help in building relevant tools for BCI applications such as neuro-rehabilitation of subjects suffering impairments in visual feature binding.

1. Introduction

Visual feature binding is the brain process that integrates various features such as color, shape, location, size, orientation, etc. to form a coherent object [1–4]. Brain analyzes the incoming sensations and codes information about an object in different brain areas. Discrimination of individual features and subsequent integration of separate features into a unified representation of the object is basic to higher order information processing. Whether binding is automatic and instantaneous, or is a resource demanding process that takes place over a period of time, is still not a settled question. Inter alia, this question asks whether all features are bound together in the incipient stages of visual processing and then the object is maintained only in the visual working memory (VWM), or whether the binding process continues such that the representation of the object is steadily and gradually refined (and strengthened) in VWM. The answer to this question is practically important because patients with brain damage due to head injuries, stroke, etc. often show alterations in feature binding. Deficits in feature binding are also presumed to be an early cognitive marker of chronic disorders such as Alzheimer and Schizophrenia [5–8]. Designing rehabilitation tools and programs for such patients using brain computer interface (BCI) requires a precise delineation of the characteristics of different stages of feature binding. High temporal resolution of electroencephalography (EEG) signals makes them particularly useful in the study of information processing in human brain. Thus, it was of interest to study the electrophysiological correlates of binding at two distinct stages of visual processing in order to ascertain whether we can use this information to train a machine learning classifier.

Most studies of feature binding use a change detection task at the molar level of cognitive behavior. The task presents two visual displays to the participant who has to decide whether there is a change in the two displays. The first is the study display, which the participants have to memorize. In the intervening period, either a masking or blank display of varying study-test intervals is presented. Then, the

[*] Corresponding author.
E-mail addresses: xs2nalin@gmail.com (N. Mathur), anubha@iiitd.ac.in (A. Gupta), sneh.jaswal@gmail.com (S. Jaswal), rohit.aiims@gmail.com (R. Verma).

https://doi.org/10.1016/j.bspc.2021.102996
Received 12 April 2020; Received in revised form 6 July 2021; Accepted 17 July 2021
Available online 31 July 2021

second display, i.e., the test display is presented, which is either same as the first one or is slightly different. The difference is present in the target stimuli, while the rest are the distractors. Rensink [1] has extensively reviewed the varieties of these archetypal descriptions and their implications. Generally, the difference in the change detection task, if it occurs, is the addition of a new stimulus, deletion of an old one, or a swap in the already presented stimuli [1]. The present research uses only the last kind of change, a swap between the two stimuli.

The swap task was introduced by Wheeler and Treisman [2] specifically to study bindings. It is not possible to perform this task by remembering which features were presented, for all the features appear in the study as well as the test display. It is essential to remember how the features were combined to find which ones swapped and do the task successfully. Alvarez and Thompson [9] have used the term *feature switch detection* to describe this task. Their work has also shown that though this task under-estimates the binding capacity of VWM, it is an efficient paradigm for studying the factors affecting the fragile nature of bindings. The task is particularly suited to the present research, for it yields a single dependent measure of the differences in various stages of processing. By simply manipulating the study-test interval, one can change it from a test of iconic memory storage to a test of visual short-term memory (VSTM).

Our focus was to study the binding of two surface features, namely, color and shape. Since location serves as a powerful cue for binding, we decided to randomize locations from study to test. This variant of the *swap detection task* has been used earlier [3,4,10,11]. The length of the study test interval was manipulated to be either 100 ms or 1500 ms. These study-test intervals denote two different stages of visual processing: initial processing in iconic memory or visual sensory memory (VSM) and post perceptual processing in VWM [12–14]. After the stimulus vanishes, visual system retains almost all the information about the stimulus displayed as an icon until 100 ms (and a little while thereafter) as reported in seminal experiments by Sperling [14,15]. VSM, however, is not only about exhaustive storage. Rather, the spatiotopic representation is continuously worked upon and refined until the relevant information is transferred into the limited capacity VWM [16]. Researchers also opine that when testing at 1500 ms after the removal of the initial stimulus, we are delving into VSTM or VWM. Both the terms VSTM and VWM are often used interchangeably, although former implies limited capacity store, whereas latter implies both storage and processing. It is assumed that no icon would survive until 1500 ms because 300 ms or less is the usual estimated iconic memory [17–19].

In essence, this research was designed to test the effect of these two study-test intervals (representing two distinct stages of visual processing) on feature binding performance, extract the characteristic features of underlying EEG signals in these two experimental conditions, train a classifier using deep learning architecture, and test whether the classifier reliably predicts the stage of processing in human participants. In the first instance, it seems that performance should get better when the study test interval increases from 100 to 1500 ms. Among others, the time-based resource-sharing model of working memory suggests that increasing the study-test interval should improve performance as it allows more time for proper encoding and consolidation of stimuli [20,21]. However, it is important to remember that forgetting also occurs over time. Information in iconic memory decays rapidly and the limited capacity VSTM can hold only a limited number of items in the store. Hence, new learning quickly *knocks out* the old learning from the VSTM. To the extent that change detection is aided by iconic memory, performance will be better at 100 ms than at 1500 ms. Conversely, iconic memory may actually hamper change detection performance if the test display does not match what is being held in the iconic memory. Thus, it is of interest to compare the effect of the two study-test intervals at the behavioral level, besides studying the underlying EEG data associated with the two study-test intervals in order to train machines.

Olson et al. [22] tested the retention of single feature objects (object or location) and binding (of object and location) in controls and amnesiacs. They manipulated study-test intervals at two levels, one second and eight seconds. The main effect of study-test interval was not significant, although performance decreased in the longer interval. In the next experiment, they equated the memory load by intermixing the location and object identity trials. Now, the subject did not know what he would need to report in the test display. In this particular experiment, the main effect of study-test interval was significant manifesting poorer performance at longer interval.

Logie et al. [3] compared color–shape binding at regular study-intervals, 0, 500, 1000, 1500, 2000, and 2500 ms, keeping the locations of the stimuli either same or randomized from study to test. When locations were same, performance gradually decreased from 0 to 1500 ms and then stabilized. However, when locations were randomized, there was a slight but significant increase in performance from 0 to 1500 ms, after which performance was similar to that of unchanged locations. This pattern of results suggests that location is crucial for initial detection and encoding of feature bindings but that bound features might be stored independently of location after those representations are transferred to VSTM as also noted earlier [23].

The above result was replicated and substantiated in subsequent studies [10,11]. Using the same task, Jaswal and Logie [10] also tested the effect of presentation time and study-test intervals on sequentially presented stimuli for color shape binding. They manipulated the study-test interval at two levels, 0 and 2000 ms, and found better performance at 2000 ms than 0 ms [10]. The slight increase in performance from 0 to 1500 ms, when locations are randomized from study to test, most likely occurs because the spatiotopic iconic memory representation of the initial display hampers change detection in the test display at 0 ms. In contrast, at 1500 ms, there is no detrimental effect of the location based iconic representation as location has no special status in VSTM for bindings. Hence, performance is better.

Many theories and experiments confirm that location is a special feature. The Feature Integration Theory (FIT) suggests that various features are initially registered in parallel [24]. As attention is directed to a point in space, all information at that point is integrated. Thus, all identified features are mapped onto a master map of locations and hence, spatial attention precedes and guides attention to other features [25]. Their seminal experiments showed that participants were better at remembering locations than other object features and focusing attention on a particular spatial location then allows the features at that location to be bound together so that an item can be identified [2].

The Guided Search model also ascribes a special place to locations [26]. Experimental studies show that location is such an overwhelming cue for encoding stimuli and their features, that it is invariably used if present [27–32]. Studies from cognitive aging also postulate ‘location’ as a special feature linking impairments of binding when location is one of the features to be bound (e.g., location–shape) but being impervious to cognitive aging when binding does not involve location (e.g., shape–color) [33–36]. Thus, location plays a key role in the formation of bindings at the time of perception. It is reported that location aids retinotopic as well as spatiotopic representation of information in iconic memory, even if the stimuli vanishes [37–40]. Nevertheless, there is also little doubt that the importance of location as a feature diminishes over time in visual processing as shown in an early study [13]. This result has been replicated in several studies with different kinds of stimuli [12,27,41,42], and particularly, for bindings [3,10,11,23]. Our experimental task randomizes locations from study to test. Therefore, we expect that change detection performance will be lesser at 100 ms because the iconic memory hampers performance, whereas it will have no effect at 1500 ms where performance depends only on VWM capacity.

We also studied the electrophysiological activity in the brain, while the task is performed. Electroencephalography (EEG) is a non-invasive technique that uses multiple electrodes placed on the scalp to measure the electrical activity generated in the brain by the cerebral cortex nerve cells. EEG signals can broadly be classified into two categories: oscillatory signals and event related potentials (ERPs). Oscillatory signals are those which are related to the regular working of the human body such as digest ion, breathing, blood flow, etc. whereas an ERP is a measured brain response that is a direct result of a specific sensory, cognitive, or motor event [43]. Several previous studies have used EEG signals to identify and study ERPs related to various cognitive tasks in an effort to better understand the nuances of that particular cognitive task. EEG signals have also been combined with change detection tasks to better understand Parkinson's disease [42] and the working of VSTM [44].

EEG signals have been increasingly paired with deep learning and machine learning techniques in order to do various tasks such as predicting human response [45], representation of human visual features [46–49], learning representations [50], detecting emotions [51], and classifying motor imagery signals [52]. EEG signals are also used in the development of BCI systems, where external devices are controlled through thought-commands by real time automated analysis of EEG signals [53–58]. Hyperscanning is yet another interaction method that uses a combination of BCI and Virtual reality [59]. Interesting studies have been done exploring the use of single electrode for BCI applications [60]. Similarly, functional brain connectivity is learned and compared for population suffering with brain disorder against the healthy population using fewer electrode portable EEG machines [61]. All this research is being carried out with the aim to build solutions to support people with numerous physical or mental disabilities or to assist in living healthy life-styles.

These systems are majorly based on predictive analysis where human action is predicted through real time analysis of the EEG signals. Although BCI systems are not currently used to assist people with purely psychological disorders, their potential use in behavioral therapy and/or symptomatic relief by managing the patients' environment is not an empty dream. However, the first step in this process is the ability to accurately predict the environmental signals around a person by studying his/her EEG responses. Our major aim in this study is to explore whether the study test intervals in a change detection task, denoting the two stages of visual processing, could be accurately predicted by the analysis of EEG signals recorded during the experiment. Secondly, we have demonstrated the use of deep learning on the visual processing and working memory related EEG dataset. Although deep learning is being used increasingly in various applications, we have not encountered a similar use so far in cognitive research.

2. Materials

2.1. Apparatus

Subjects were seated in front of a 14″ desktop screen at a distance of approximately 1 m. EEG data was collected at a sampling rate of 256 Hz using the RMS Maximus portable EEG machine[1] keeping 21 electrodes on the scalp with Ag-Cl conductive paste in accordance with the international 10–20 system as shown in Fig. 1. The ground electrode was fixed at the nasion position while the reference electrode was placed 10% above the former.

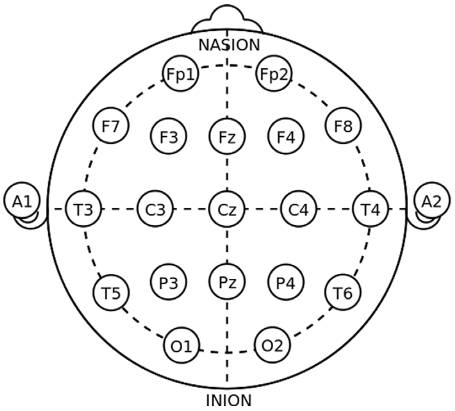


Fig. 1. International 10–20 system for application of scalp electrodes. *Source:* https://en.wikipedia.org/wiki/10-20_system_(EEG).

2.2. Experiment design and data collection

The experiment was conducted in accordance to previous experiments conducted in various studies [3,10,11]. The experimental protocol is shown in Fig. 2 and is described as follows. Subjects were shown a target and a test screen with a blank screen interspersed between them. The target and test screens consisted of four stimuli at random locations consisting of four shapes (triangle, parallelogram, plus and horseshoe) in four colors (blue, green, red and yellow). The experiment involved display of a target screen with four differently shaped objects in different colors at random locations on the screen, followed by a blank screen after which a test screen displayed the same objects at random locations having the same or randomly changed color from the target screen. The target screen was displayed for 200 ms, blank screen for a fixed duration of either 100 ms or 1500 ms and the test screen was response driven. Subjects responded by pressing keys 's' and 'd' if the test screen was similar or different from the target screen, respectively. Each subject was initially familiarized with the experiment by 25 practice trials after which they were tested on 96 trials. Location was randomized across all trials. The color–shape bindings in the target and test screens were same in 50% of the trials. Half of the trials showed a blank screen of 100 ms while the other half showed blank screen of 1500 ms. The two different lengths of the blank screen were randomly mixed within the 96 trials conducted for each subject. For each trial of a particular blank screen time, color–shape bindings were kept same in target and test screen for half of those trials. To give a brief estimate of the time duration of the experiment, we present the total time taken to complete 96 trials by a randomly chosen subject as follows: (0.1 s fixation time + 0.2 s target image time) x 96 trials + 0.1 s blank screen time x 48 trials + 1.5 s blank screen time x 48 trials + 117.5 s response time (of 96 trials, this response time varies from trial to trial) = 223.1 s = 3.7 min (approximately) or around 5.7 min overall including the time taken in the practice trials.

Subjects were asked to detect changes in color–shape binding between the target and test screens while inhibiting location as a feature. The experimental variable that was selected to be predicted during the test was the length of display of the blank screen between the target and the test screens. We also conducted analysis to confirm that the FIT was upheld during the experiment.

2.3. Preprocessing

EEG data was first filtered using a simple band-pass filter available in the EEGLab software [62] itself between 1 Hz and 60 Hz in order to remove undesired frequency band signals. Collection of EEG data is a non-invasive technique, as a result of which signals captured

[1] http://www.rmsindia.com/neurology.html.

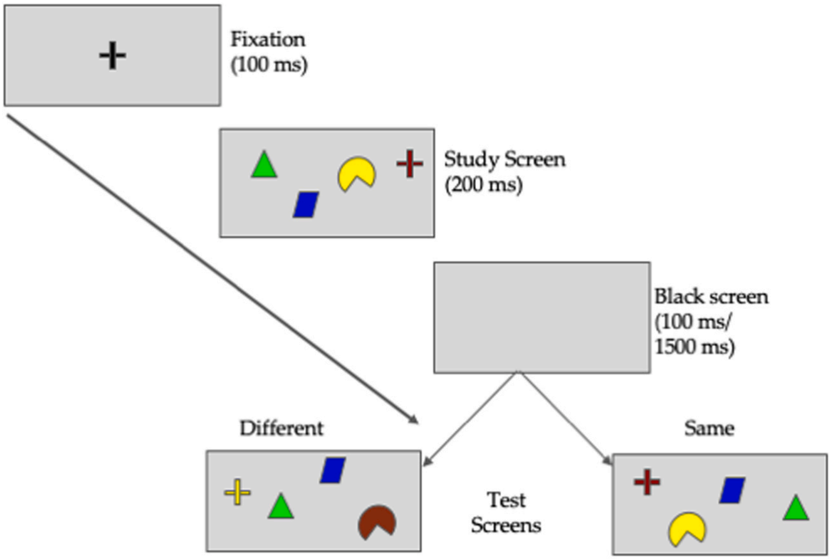


Fig. 2. Single trial of the experiment (Note: Stimuli are not drawn to scale).

by a particular electrode on the scalp are a combination of signals originating from different regions of the brain in the vicinity of the position of that electrode. Thus, there is a need to separate these components in order to better understand the captured EEG signal. Independent and uncorrelated temporal components were obtained using independent component analysis (ICA). Twenty one independent components were obtained from ICA as shown in Fig. 3. These 21 components also consisted of various brain related artifacts such as artifacts related to eye-ball movement, muscle movement, heartbeat, etc. For each subject, we studied the ICA component time-series, 2D and 3D scalp maps and observed the changes in the EEG spectrum to determine the presence of said artifacts. Electrooculogram (EOG) and Electromyogram (EMG), which are artifacts related to eye-ball movement and muscle movement, respectively (shown in Fig. 4), were removed to obtain 19 individual components. After ICA, experimental data of each subject was segmented into individual trials to obtain 96 different segments, each containing 19 individual components. Thus, we collected 96 different labeled segments for each subject which were further used to classify and predict the study-test intervals (implying different stages of processing).

3. Methods

EEG data were visualized and processed using EEGLab [62] that has been developed by the Swartz Center for Neuroscience in Matlab R2016.

3.1. Behavioral data analyses

The accuracy with which subjects detected the change in color–shape bindings was analyzed for both 100 ms and 1500 ms blank screen intervals using d-prime scores. The d-prime is a measure of accuracy employed on the basis of the signal detection theory which helped ensure that response bias was considered while calculating the accuracy with which change in binding was detected [63]. These scores were calculated for each subject individually in the two experimental conditions.

3.2. Feature extraction

The classification and prediction of study-test intervals (implying different stages of processing) was done using 14 state-of-the-art machine learning classifiers that were built on time-domain and frequency-domain features extracted from each of the remaining 19 components for each trial.

Table 1
EEG time domain features. ($z[n]$ is the analytical signal obtained using the Hilbert Transform of a real discrete time EEG signal $x[n]$, μ is the mean of $z[n]$, and σ is the standard deviation of $z[n]$.)

(1) Features based on statistical moments

- First and second moment of EEG signal. Aarabi et al. [67] and Löfhede et al. [64].

$F_{(t1)} = \mu = \frac{1}{N} \sum_{n=1}^{N} (|z[n] - \mu|)$

$F_{(t2)} = \sigma = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (\mu - |z[n]|)^2}$

- Normalized moments: Third and fourth moments of EEG signals [64,67].

$F_{(t3)} = \frac{1}{N\sigma^3} \sum_{n=1}^{N} (|z[n]| - \mu)^3$

$F_{(t4)} = \frac{1}{N\sigma^4} \sum_{n=1}^{N} (|z[n]| - \mu)^4$

(2) Features based on amplitude

- Median absolute deviation of EEG amplitude [64]

$F_{(t5)} = \frac{1}{N} \sum_{n=1}^{N} (|z[n] - \mu|)$

3.2.1. Time-domain features

Five discriminating features were extracted from the time-domain representation of the collected EEG and pre-processed EEG signal. Table 1 describes the relevant time-domain features that were also identified as being significant. The features selected are based upon amplitude such as median absolute deviation [64–66] and statistical moments such as mean, standard deviation, skewness, and kurtosis [64, 67] of the collected EEG signals. As these features were extracted from each of the 19 components of each trial, there were a total of 95 (19 × 5) time domain features that were extracted for each trial of each subject. All features were standardized after extraction by subtracting the mean and dividing by the standard deviation.

3.2.2. Frequency-domain features

A total of five frequency domain features were extracted from the frequency domain representation of the collected and pre-processed EEG signal. The frequency domain representation was obtained using Fourier transform. Table 2 described the discriminant and relevant frequency-domain features that have been identified. These features were based on spectral information of EEG signals such as average energy, spectral centroid, spectral flatness, spectral roll-off and spectral entropy [64,66,67]. As these features were extracted from each of the 19 components of each trial, there were a total of 95 (19 × 5) frequency domain features that were extracted for each trial of each subject. All

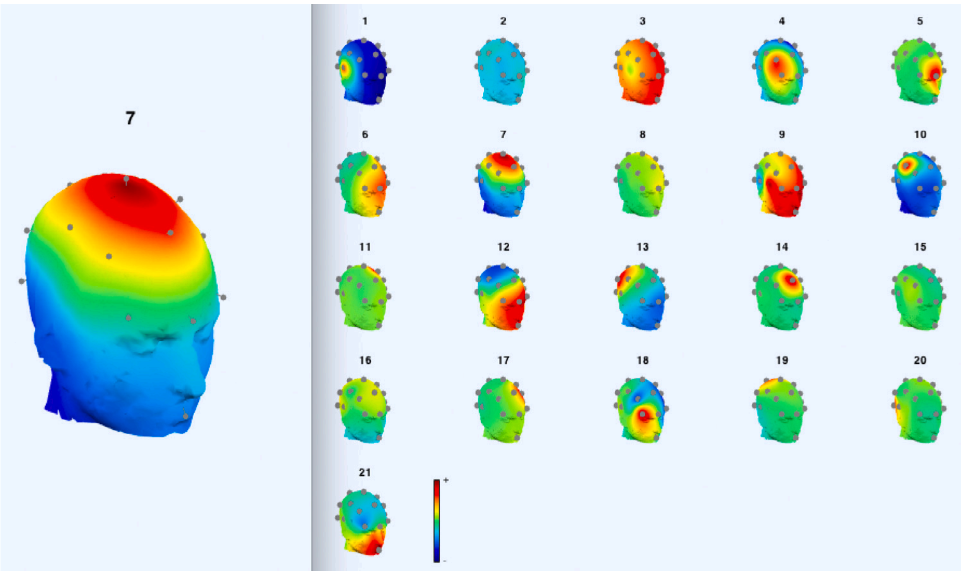


Fig. 3. 3-D scalp map of EEG Data.

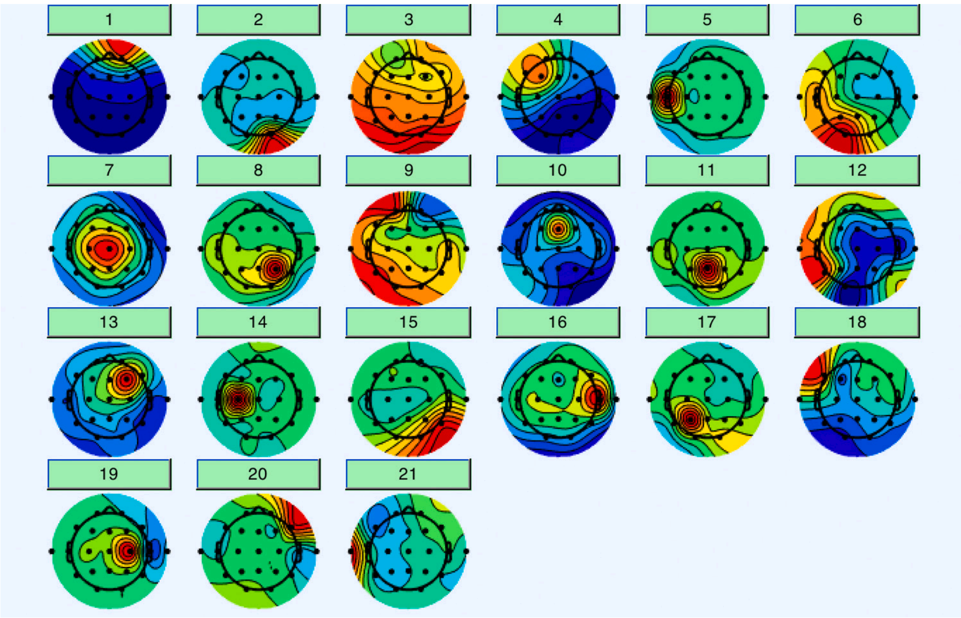


Fig. 4. Independent components of EEG data. Component 16 (EMG) and component 1 (EOG) were removed being artifacts.

features were standardized upon extraction by subtracting the mean and dividing by the standard deviation.

3.3. Prediction/estimation of study-test interval

The estimation of study-test intervals (implying different stages of processing) was carried forward by training state-of-the-art machine learning classification models using the features in Tables 1 and 2. The response of the subject and correctness of the response were added as two additional features. Response of the subject indicates whether the subject pointed out a change or a no change in the test and study screens. Correctness of the response indicates whether the subject's response was correct. Both these features were added to both the time and the frequency domain features. A total of 14 machine learning classifiers that are listed in Table 3 were used. They belonged to five different families of classifiers. The classifiers were used separately on time domain and frequency domain features for each subject. Thus, a total of 1512 simulations (14 classifiers x 54 subjects x 2 types of features) were conducted.

3.4. Deep learning: Convolutional Neural Network (CNN)

3.4.1. Data preparation

In order to train the proposed CNN, each trial of every single subject was converted into a three-dimensional matrix by taking the short-term Fourier transform (STFT) and stacking the different sections on top of each other as explained below. STFT is windowed Fourier transform used to determine the frequency content of signal in short duration. For this, a signal is divided into shorter segments of equal lengths and Fourier transform of each segment is computed. In our data, the length of the individual trial of a particular subject is different because it is equal to the time taken by the subject to press a response key that is variable from trial to trial. Thus, individual trials are represented by two-dimensional matrices of size (19, N) each, where 19 represents

Table 2
EEG frequency-domain features. $Z[k]$ is the Fourier transform of the analytical signal $z[n]$ of a real discrete time EEG signal $x[n]$. M corresponds to the maximum frequency of the signal.

(1) Features based on power spectrum • Average energy [66,67]. $F_{(f1)} = \frac{1}{N} \sum_{k=1}^{M} \lvert Z[k] \rvert^2$
(2) Features based on spectral information • Spectral centroid: Average signal frequency weighed by magnitude of spectral centroid $F_{(f2)} = \frac{\sum_{k=1}^{M} k\lvert Z[k]\rvert}{\sum_{k=1}^{M} \lvert Z[k]\rvert}$ • Spectral flatness: Indicates smoothness of frequency distribution [64]. $F_{(f3)} = M(\prod_{k=1}^{M} Z[k])^{\frac{1}{M}} (\sum_{k=1}^{M} Z[k])^{-1}$ • Spectral roll-off: Spectral Concentration below threshold λ [64] $F_{(f4)} = \lambda \sum_{k=1}^{M} Z[k]$ The value of λ was equal to the frequency at which energy of the signal goes below 0.8 times of the total energy.
(3) Features based on entropy • Spectral entropy: measure of regularity of power signal [66] $F_{(f5)} = \frac{1}{log(M)} \sum_{k=1}^{M} P(Z[k])log(P(Z[k]))$

Table 3
Machine learning classifiers employed.

(I) Bayes classifiers 1. Naive Bayes 2. Bayes Net
(II) Traditional Classifiers 3. Binary SVM with Stochastic Gradient Descent (SGD) 4. Binary SVM with Sequential Minimum Optimization 5. Simple Logistic Regression
(III) Lazy classifiers 6. K-Nearest Neighbor
(IV) Rules based classifier 7. JRip
(V) Trees 8. Decision Stump 9. Hoeffding Tree 10. J-48 Tree 11. Logistic Model Tree (LMT) 12. Random Forest 13. Random Tree 14. REP Tree

Table 4
Convolution layers of the proposed architecture.

Layer name	Input size	Kernel size	Stride	#Filters
Conv1	19,51,L$_{max}$	3 × 3 × L$_{max}$	1	L$_{max}$
MaxPool1	16,48,L$_{max}$	4 × 4	1	–
Conv2	15,14,L$_{max}$	2 × 2 × L$_{max}$	1	50
MaxPool2	13,45,40	4 × 4	1	–
Conv3	12,44,50	2 × 2 × 50	1	80
MaxPool3	10,42,80	4 × 4	1	–
Conv4	9,41,80	2 × 2 × 80	1	100
MaxPool4	8,40,100	2z2	1	–

Table 5
Dense (fully connected)* layers of the proposed architecture.

Layer name	Input size	Output size	Dropout
Dense-1	27 300	50	0.3
Dense-2	50	20	0.3
Dense-3	20	10	0.3
Dense-4	10	1	0

* The output of the convolution layers was flattened before feeding to the dense layers.

the number of channels obtained after pre-processing (Section 2.3). N is dependent upon the length of the trial. For example, if the trial was 2.5 s long, the value of N would be $256 \times 2.5 = 640$ corresponding to the sampling frequency of 256 Hz. Some snapshots of STFT are shown in Fig. 5.

In order to take STFT of individual trials, each trial was divided into segments of 100 points each giving L number of matrices of size (19, 100) where the value of L depends on N (L is the maximum integer value of $N/100$). After dividing the trial into segments, Fourier transform of each segment was taken to yield L matrices of size (19, 51) that were stacked on top of each other to yield three-dimensional matrix of size (19, 51, L) for each trial. For each subject, there were a total of 96 trials. In order to ensure homogeneous input to CNN for a particular subject, the depth of the matrices of each trial was padded with zeros till the size of each trial became (19, 51, L_{max}) where L_{max} is the maximum value of L among all trials of that particular subject. This process of data preparation for CNN is shown in Fig. 6.

3.4.2. CNN architecture

We designed a CNN architecture with four convolutional layers followed by four densely connected layers as shown in Figs. 7 and 8 and explained in Tables 4 and 5. CNN is a class of multilayer feed forward neural networks that contain very few parameters compared to the conventional fully connected neural networks [68]. In a CNN architecture, we have three types of layers: (a) convolutional layers that have neurons with small visual field, e.g. say 3 × 3, where the visual field implies that the input to the current neuron is being received as the weighted linear combination of ($3 \times 3 = 9$) nine neurons of the previous layer. The weights of these 3 × 3 neuron connections are shared by all the neurons of that particular convolutional layer and is also called as a 3 × 3 filter or kernel. This weighted linear output is passed through the non-linear activation function of the neuron. Often, restricted linear activation (ReLU) is used as the activation function. (b) the max pool layer: this layer acts as the subsampling layer that is placed after the convolutional layer, and (c) fully connected layer: these layers are added just before the last softmax layer.

Same activation function was used for each layer of the neural network except for the last layer. The last (output) layer of the network had an activation of sigmoid. The rest of the network was tested on two different activation functions: ReLU and tan hyperbolic activation functions. Difference in performance achieved by the two functions was subsequently studied. The input image size and the number of filters of the first convolutional layer were different for each subject and was in accordance with the value of L_{max} for that subject.

3.4.3. CNN training

The initialization of weights during training was done using 'glorot uniform' initialization [69]. The subsequent training was conducted over 100 epochs at a learning rate of 0.001. Binary cross-entropy loss was chosen as the loss function and ADAM optimizer [70] as the optimizer of choice. The values of exponential decay rate for the first and second moment of ADAM optimizer were set as 0.9 and 0.999, respectively. The number of trainable parameters varied between 1,420,041 to 1,507,341 depending on the subject on whose data the CNN was trained.

4. Results and discussion

4.1. Behavioral data

Accuracy of change detection was the primary measure of interest in the behavioral data. A paired t-test showed significant difference in the average d-prime scores for 100 ms and 1500 ms study-test intervals ($t = 4.174, df = 53, p < 0.001$) with the average/mean score being

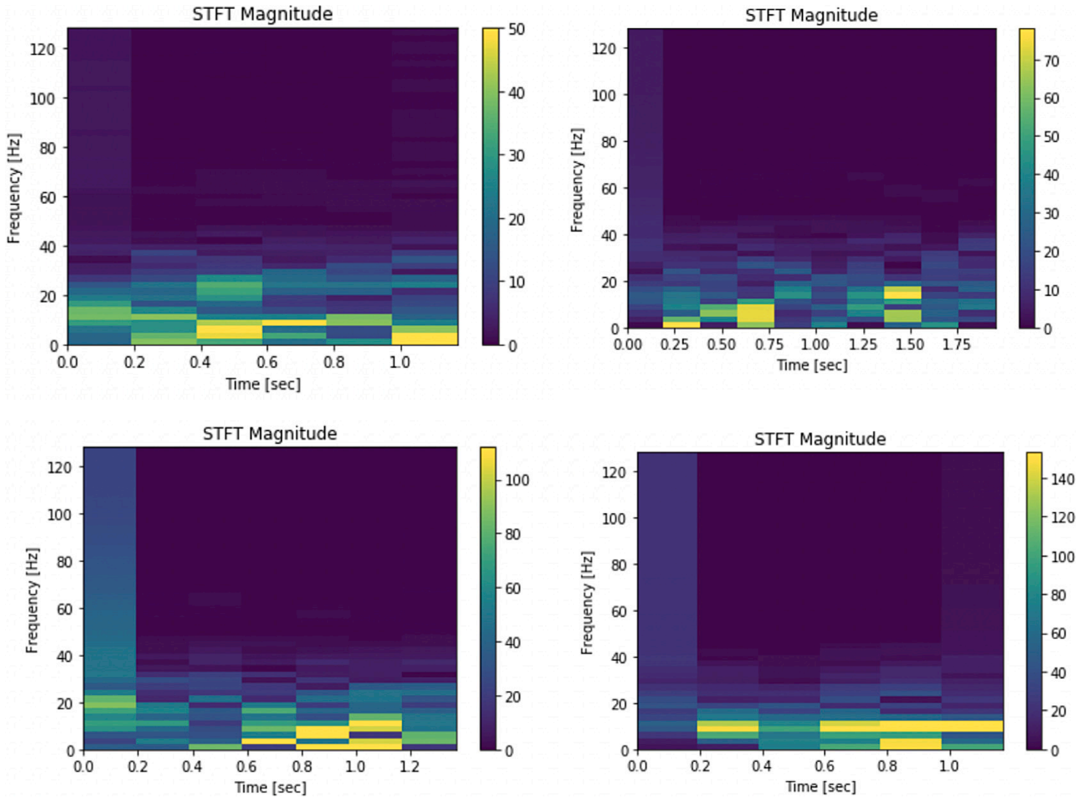


Fig. 5. Visual representation of EEG signals in the time–frequency domain. Images shown belong to different subjects.

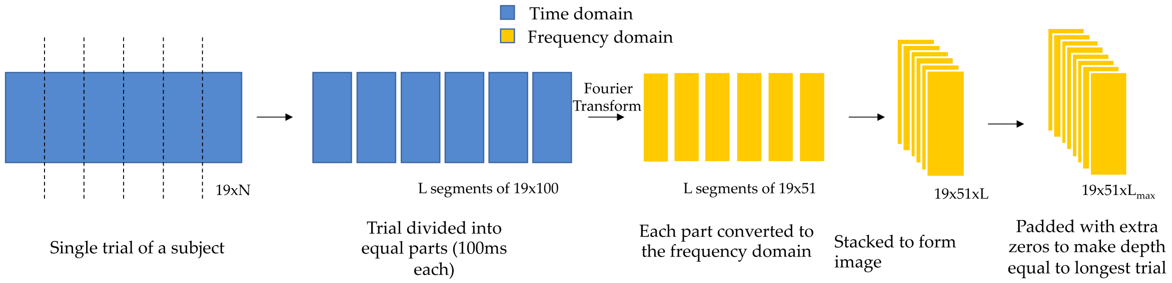


Fig. 6. Visual representation of changing a single trial into an STFT image to be fed to the CNN. L is the maximum integer value of $N/100$. L_{max} represents the depth of the longest trial.

lower for 100 ms (mean = 0.863, standard deviation = 0.497) than for 1500 ms (mean = 1.243, standard deviation = 0.497). The lower score for 100 ms as compared to 1500 ms has been observed in several previous studies which used this experimental task [3,10,11,23]. The presentation of multiple objects utilizes the powerful cue of location and allows configural encoding as shown by many studies of unifeature objects [27,71]. This relational encoding is preserved in iconic memory that was being tested at 100 ms, the icon being a spatiotopic representation of the stimuli that were seen [14,15].

The importance of location in binding has been emphasized in the FIT [24,25,72,73] as well as in the guided search model [26]. FIT suggested that binding is mediated by the links of separate features to a common location [24]. Treisman and Sato [25] propose that a "master map" of locations exists in our brain. Attention selects all the features associated with a particular location, and works as glue to bind those features [25]. Neuroscientists have found the evidence for such a master map. Several studies have also shown that activity in the retinotopically organized sub-regions of the visual and parietal cortex are critical for VSTM storage [74]. Studies also show that bindings are more vulnerable to location change and suggest that location plays a central role not only in encoding but also in maintenance and retrieval of bound objects [3,23,28,30].

As suggested by the feature integration theory [24,25] and the guided search model [26], location plays a key role in feature binding at the time of perception. In fact, other features are probably addressed through the master map of locations which exists in the brain. Features, and the object representations that they form, are inevitably encoded as a configuration. Thus, when the experimental task randomizes locations, the relational encoding of stimuli is disruptive of performance. Nevertheless, as the task is to remember only the binding between colors and shapes, gradually, locations, being irrelevant to the task, are deleted from VSTM. At 1500 ms, locations are no longer retained by participants in the object representations to have a disruptive effect on performance. The memory for color–shape bindings in VSTM at 1500 ms does not involve locations and thus performance is not as adversely affected by the randomization of location between study and test.

Results also suggest that although feature binding is a continuous process of consolidating relevant features and removing the irrelevant features from object representations, there are significant and reliable

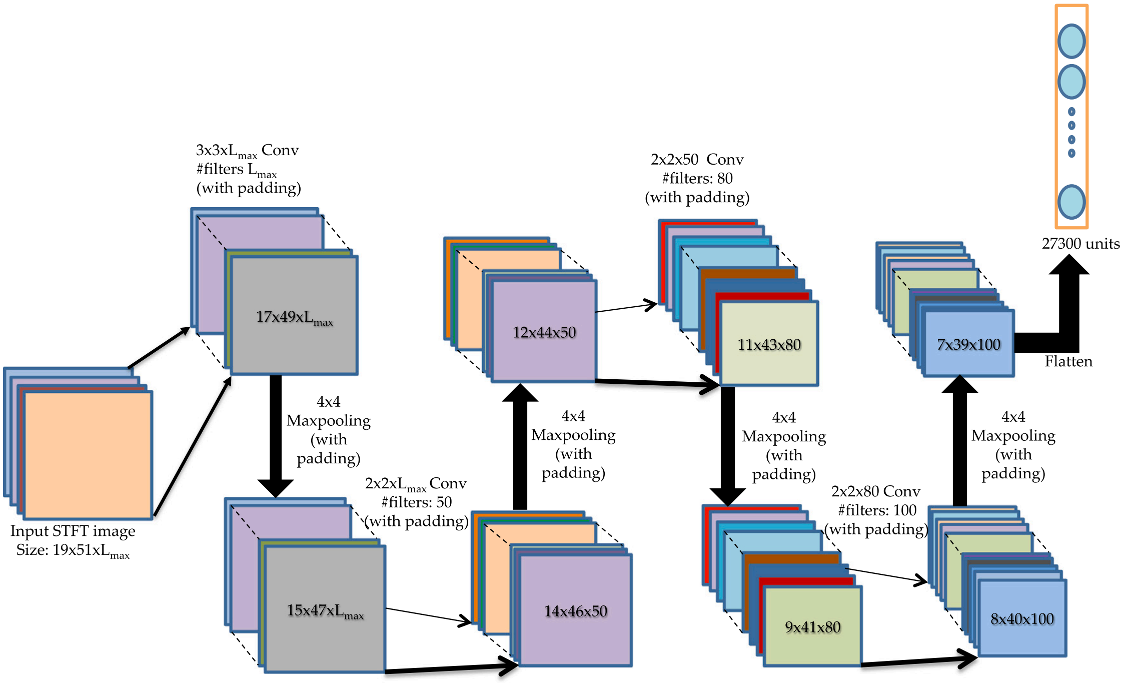


Fig. 7. Visual representation of the convolution layers of the proposed network architecture.

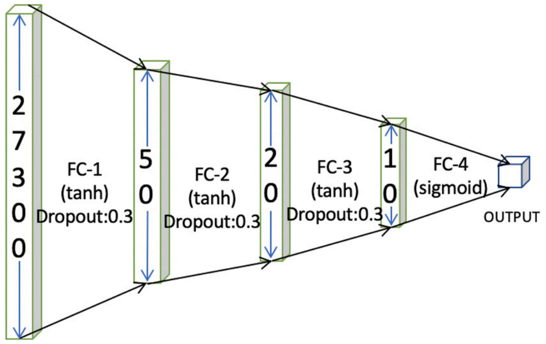


Fig. 8. Visual representation of the fully connected layers of the proposed network architecture.

differences in the state of brain functioning at 100 ms and 1500 ms. Different factors dominate brain at these different times, with associated behavioral as well as electrophysiological differences. Thus, it is possible to study the brain EEG signals to predict whether the brain is currently processing stimuli at 100 ms or 1500 ms.

4.2. Estimation of stages of processing using study-test intervals

Results for the estimation of study-test intervals (implying different stages of processing) were generated by averaging results obtained from 8-fold cross-validation for every subject in every classifier used. The best performance among the 14 machine learning classifiers using both frequency-domain and the time domain-features was noted and compared with the performance obtained using the proposed CNN architecture as shown in Table 6. For the CNN, the test accuracy is reported from each of the eight classifiers (corresponding to 8-fold cross validation) on the held-out fold on the 100th epoch. Results of CNN are calculated by considering the results of all the eight classifiers on their respective test fold. A summary of results showing the number of subjects on whom the best performance was obtained via the proposed CNN model vis-à-vis t-domain feature based or f-domain feature based models is shown in Table 7.

Machine learning classifiers showed better results when trained using the frequency domain features as compared to time domain features for all but one subject. This is observed due to the differences in α, β, and γ rhythms in EEG signals [75]. These rhythms contain a lot of information regarding the ongoing cognitive processes. The presence of these rhythms and subsequently the information contained in them, is better represented in the frequency domain than the time domain and hence, features extracted from the frequency domain show better results.

As seen in Tables 6 and 7, the proposed CNN architecture with tanh activation shows peak performance on 45 subjects and with ReLU activation shows peak performance on 11 subjects. Overall, the proposed CNN architecture shows peak performance for all but 3 subjects. The highest accuracy achieved with the proposed CNN is 100% for subject number 26. This shows extremely promising results using deep learning based architecture in EEG signal processing.

4.3. ReLU versus tan hyperbolic activation functions

The tan hyperbolic activation function (tanh) outperforms ReLU in 42 out of 54 subjects, while ReLU outperforms tanh activation on 6 out of 54 subjects. The remaining 6 subjects show equal performance with both tanh and ReLU activation functions. The ReLU activation function is a part of the family of non-saturated activation functions and is one of the most actively used activation functions in all of deep learning literature [76–78]. ReLU is a piecewise linear function that clips the negative part to zero and retains the positive part as it is. It has been noted in the literature to show high performance while working with image data.

The tanh activation function is a bounded function that is a part of the saturated activation functions family with value saturating to +1 for positive input and −1 for negative input. The overwhelming results obtained when using tanh activation function as compared to ReLU activation function can be attributed to the way in which both functions deal with the data. As seen from the STFT plots in Fig. 5, most of the brain activity is observed in lower frequencies. In fact, in

Table 6
Classification accuracy with different classifiers.

Subject	t-domain features (best performance)	f-domain features (best performance)	Proposed CNN (with ReLU activation)	Proposed CNN (with tanh activation)
1	65.27%	78.125%	**80.49%**	**80.49%**
2	67.021%	77.89%	78.98%	**81.91%**
3	80%	85.42%	**97.92%**	96.88%
4	72.63%	86.46%	84.32%	**88.54%**
5	70.53%	92.71%	92.71%	**94.79%**
6	74.74%	89.58%	91.67%	**93.75%**
7	65.26%	**72.92%**	57.29%	**72.92%**
8	65.26%	90.625%	74.53%	**94.79%**
9	66.315%	76.042%	85.42%	**88.54%**
10	70.21%	85.26%	**90.44%**	86.17%
11	63.16%	84.375%	82.18%	**90.62%**
12	73.68%	85.42%	**93.75%**	**93.75%**
13	67.37%	**81.25%**	76.04%	78.12%
14	74.74%	**80.21%**	68.75%	**80.21%**
15	78.95%	88.54%	**93.75%**	**93.75%**
16	66.32%	93.75%	95.83%	**97.92%**
17	72.63%	81.25	94.79%	**95.83%**
18	74.75%	89.58%	90.63%	**92.71%**
19	83.16%	89.58%	94.79%	**95.83%**
20	63.16%	83.34%	75.00%	**84.37%**
21	75.79%	85.42%	90.62%	**93.75%**
22	64.21%	71.875%	85.42%	**86.46%**
23	67.37%	88.54%	78.60%	**89.58%**
24	74.74%	85.42%	**95.83%**	93.75%
25	72.63%	88.54%	76.04%	**89.58%**
26	80%	95.83%	**100%**	**100%**
27	75.79%	86.46%	90.63%	**92.71%**
28	63.16%	72.92%	73.96%	**79.17%**
29	72.63%	81.25%	88.54%	**89.58%**
30	73.68%	89.58%	87.03%	**89.58%**
31	68.42%	72.92%	77.94%	**78.12%**
32	69.47%	86.46%	89.58%	**91.67%**
33	77.89%	89.58%	90.62%	**94.79%**
34	71.58%	84.375%	87.50%	**88.54%**
35	75.79%	83.34%	86.46%	**87.50%**
36	73.68%	79.17%	79.17%	**83.33%**
37	91.58%	91.67%	**95.83%**	**95.83%**
38	64.21%	79.17%	**89.58%**	**89.58%**
39	54.74%	93.75%	52.08%	**94.79%**
40	78.95%	80.21%	71.61%	**81.25%**
41	73.68%	80.21%	**87.50%**	83.33%
42	70.53%	83.34%	**94.79%**	90.62%
43	70.53%	87.5%	90.63%	**92.71%**
44	72.63%	86.46%	88.54%	**89.58%**
45	82.10%	92.713%	92.71%	**94.79%**
46	72.63%	80.21%	88.54%	**89.58%**
47	**73.68%**	69.79%	69.79%	61.46%
48	72.63%	86.46%	87.50%	**97.92%**
49	75.79%	86.46%	95.83%	**96.88%**
50	70.57%	80.21%	88.54%	**89.58%**
51	69.47%	76.04%	78.31%	**79.17%**
52	68.42%	**73.96%**	67.71%	65.63%
53	69.47%	**82.29%**	76.99%	80.21%
54	56.84%	64.58%	62.50%	**66.42%**
Average	71.49%	83.50%	84.37%	**87.95%**

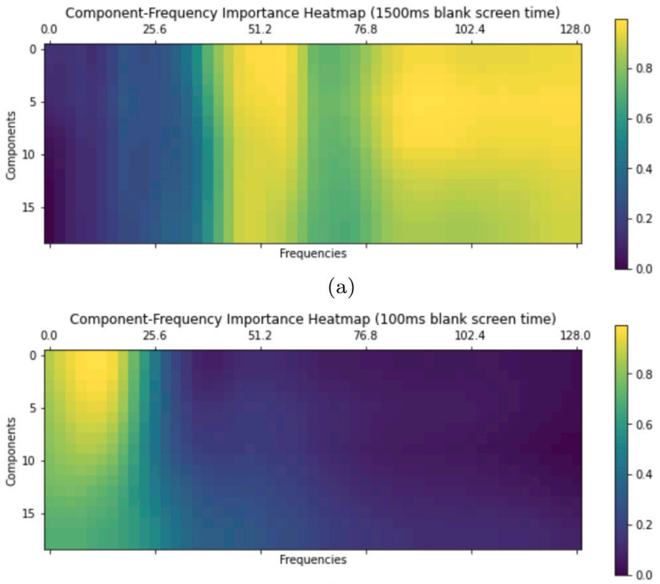
general, brain activity is also widely seen on lower frequencies. Hence, while certain high frequencies owing to noise may suddenly spike up, the lower frequencies are the ones that are being used by the CNN to classify the images. The main difference between tanh and ReLU is that ReLU is a linear function that starts from 'zero' and goes till infinity, while tanh is a function that tends to 'one' for higher input values. Hence, when we use ReLU with our data, it will equally pass lower and higher frequencies (spiky inputs) through the network which is not desirable as lower frequencies have more importance. This is where tanh is useful because it asymptotes towards 'one' for higher values, thus, giving more importance to data with lower frequencies which is desirable.

In order to ascertain whether the improvement with the CNN(tanh) classifier is statistically significant, we carried out paired t-test between the accuracy results of: (1) best classifier results with t-domain features (column 1, Table 6) versus those with CNN (tanh activation) (column 4, Table 6); and (2) best classifier results with t-domain features (column 2, Table 6) versus those with CNN (tanh activation) (column 4, Table 6). For both the comparisons, we first tested the normality assumption of the data in each column using the Jarque–Bera test. This was satisfied. Next, we compared the variance of both the comparison groups in (1) and (2). The comparison groups in (1) displayed unequal variances, while those in (2) displayed equal variances. Next, we carried out paired t-test with the corresponding variance setting (unequal in (1) and equal in (2)). The t-test rejected the null hypothesis of equal means in (1) with a p-value $= 8.6 \times 10^{-20}$ and (2) with the p-value of 0.003, i.e., the value in both the cases is less than 0.005. This shows that the improvement in accuracy is statistically significant with the CNN classifier.

Table 7
Summary of results obtained.

Features/model used	#peak performance (unique)	#peak performance (total)	Total subjects tested
t-domain	1	1	54
f-domain	3	5	54
CNN (relu activation)	5	11	54
CNN (tanh activation)	39	45	54

#peak performance (unique) indicates number of subjects that gave peak performance only on the given model (there can be multiple models giving peak performance for the same subject).
#peak performance (total) indicates number of subjects for whom peak performance was observed.



(a)

(b)

Fig. 9. Results of Grad-Cam algorithm on the STFT input images of one subject to ascertain the frequency regions where CNN is paying more attention to discriminate between the two visual binding tasks.

The ability of the deep learning method (CNN) used in the study to distinguish, with very high accuracy, between the two stages of visual processing indicates a distinction between the EEG signals captured in the two stages. It is, thus, important to assess the areas of the EEG signals that are leading our CNN to distinguish between the two stages as it can lead to more insights into the feature binding process.

In order to interpret the CNN and thus find areas of EEG signals that are relevant to either stage of visual processing, we used Gradient weighted Class Activation Maps (Grad-CAM) to generate visual explanations of our CNN. Grad-CAM [79] is extensively used in image classification tasks to generate visual explanations of where the CNN is paying most attention, while classifying a certain class. It leads to insights into the working of, otherwise very opaque, CNN models. In an RGB image, grad-CAM gives insights into the physical location of areas within the image that the CNN is paying attention to while classifying certain classes. In our CNN, the input is an STFT image (ICA component x Frequency x Time) of the EEG data, and thus grad-CAM will give insights into the ICA Component x Frequency areas our CNN is looking at while classifying certain classes. In Fig. 9, we present GRAD-CAM results on one of the subjects STFT map.

As seen above, the CNN is paying attention to higher frequencies while classifying for VWM (1500 ms) and lower frequencies while classifying for VSM (100 ms). This means that higher frequencies are more important while classifying the EEG data as that of VWM while lower frequencies are more important while interpreting data as that of VSM. This shift in frequencies indicates a steady and gradual refining of the binding process in the VWM.

5. Conclusions

The current study shows the possibility of determining visual stimuli by analysis of EEG signals using deep learning and machine learning methods. The study also sheds light on how activation functions that show best performance for data like EEG may be different from mainstream activation functions that have been widely accepted in the deep learning community. A new method to use STFT of a signal in order to generate input to a convolutional neural network (a type of deep learning architecture) has also been introduced.

The high accuracy with which machine learning and deep learning algorithms trained on EEG data could identify the study test intervals confirms the difference in cognitive processes involved when dealing with visual stimuli in visual processing and in VSTM.

In keeping with some classic studies in visual memory [13,80], one may surmise that at 100 ms, a spatiotopic representation influences the performance of the participants. At this time the participants have almost all items in their iconic memory. This spatiotopic representation results in a lower performance at 100 ms because there is a mismatch between the representation of the study display and the incoming test display, making the identification of binding swaps more difficult, as locations of the original stimuli are still very much part of their representation. However, at 1500 ms, the performance of the participants is contingent only on the items that are maintained in the limited capacity VSTM. Locations of the stimuli have already been discarded from the representation. The identification of swaps is better because stimuli are maintained as only color shape bindings, and a visual search of the test display in a serial fashion matches the stored representations with the stimuli in the test display to find the mismatched ones leading to better performance. There is no interference from the irrelevant feature of locations. Indeed, some researchers have also proposed a 'fragile' visual short term store between iconic and visual working memory [40,81]. The current EEG evidence, however, shows no sharp distinctions to denote these two (or three) stages of memory and suggests only a gradual strengthening of working memory representations. The STFT inputs being analyzed by the GRAD-CAM (as shown by an example of one subject's data in Fig. 9) carried out on all channels of subjects clearly demonstrated that only lower frequencies played significant roles (show higher amplitudes) in the blank period of 1500 ms. Several studies show that visual working memory processing of objects is associated with CDA which is a low frequency event related potential [82,83]. Thus, it can be inferred that although all features are bound together initially, gradually the binding process continues such that the representation of the object is steadily and gradually refined (and strengthened) in VWM, such that at 1500 ms, only the relevant features define the object representation in memory.

The current work focuses on determining the stages of visual processing of a subject, where models are built separately for individual subjects. This can be termed as intra-subject training. It is well known by now that biomedical signals can be used as biometric data as well. In other words, brain signals of each subject also have a characteristic mark. In order to train a CNN classifier that can train on only the task for all the subjects, irrespective of the subject characteristics, requires a larger dataset. This is the limitation of the current study. The aim of future studies will be to conduct inter-subject training of models where models trained on some particular subjects can be used to determine the stages of visual processing in another subject.

There is also a scope of re-engineering the model that has been trained in order to determine the basic logic with which the model does its calculation. This can unlock key insights into the working of visual processing in the human brain.

CRediT authorship contribution statement

Nalin Mathur: Methodology, Formal analysis, Software, Investigation, Writing - review & editing. **Anubha Gupta:** Conceptualization, Data collection and curation, Project administration, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Snehlata Jaswal:** Conceptualization, Experiment design, Data collection and curation, Software for data collection, Validation, Writing – original draft, Writing – review & editing. **Rohit Verma:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all the anonymous subjects who participated in the study and contributed the data. We also thank Dr. Vinayak Naik, BITS Pilani, Goa Campus, India and Dilnashin Anwar for support in data collection.

References

[1] R.A. Rensink, Change detection, Annu. Rev. Psychol. 53 (1) (2002) 245–277.
[2] M.E. Wheeler, A.M. Treisman, Binding in short-term visual memory, J. Exp. Psychol. [Gen.] 131 (1) (2002) 48.
[3] R.H. Logie, J.R. Brockmole, S. Jaswal, Feature binding in visual short-term memory is unaffected by task-irrelevant changes of location, shape, and color, Mem. Cogn. 39 (1) (2011) 24–36.
[4] S. Jaswal, The importance of being relevant, Front. Psychol. 3 (2012) 309.
[5] M.A. Parra, S. Abrahams, R.H. Logie, L.G. Mendez, F. Lopera, S. Della Sala, Visual short-term memory binding deficits in familial Alzheimer's disease, Brain 133 (9) (2010) 2702–2713.
[6] M. Pietto, M.A. Parra, N. Trujillo, F. Flores, A.M. García, J. Bustin, P. Richly, F. Manes, F. Lopera, A. Ibáñez, et al., Behavioral and electrophysiological correlates of memory binding deficits in patients at different risk levels for Alzheimer's disease, J. Alzheimer's Dis. 53 (4) (2016) 1325–1340.
[7] P. Bob, O. Pec, A.L. Mishara, T. Touskova, P.H. Lysaker, Conscious brain, metacognition and schizophrenia, Int. J. Psychophysiol. 105 (2016) 1–8.
[8] J.M. Gold, C.M. Wilk, R.P. McMahon, R.W. Buchanan, S.J. Luck, Working memory for visual features and conjunctions in schizophrenia., J. Abnorm. Psychol. 112 (1) (2003) 61.
[9] G.A. Alvarez, T.W. Thompson, Overwriting and rebinding: Why feature-switch detection tasks underestimate the binding capacity of visual working memory, Vis. Cogn. 17 (1–2) (2009) 141–159.
[10] S. Jaswal, R. Logie, Configural encoding in visual feature binding, J. Cogn. Psychol. 23 (2011) 586–603, http://dx.doi.org/10.1080/20445911.2011.570256.
[11] S. Jaswal, R.H. Logie, The contextual interference effect in visual feature binding: What does it say about the role of attention in binding? Q. J. Exp. Psychol. 66 (4) (2013) 687–704.
[12] D.E. Irwin, Information integration across saccadic eye movements, Cogn. Psychol. 23 (3) (1991) 420–456.
[13] W. Phillips, On the distinction between sensory storage and short-term visual memory, Percept. Psychophys. 16 (2) (1974) 283–290.
[14] G. Sperling, The information available in brief visual presentations, Psychol. Monogr.: Gen. Appl. 74 (11) (1960) 1.
[15] G. Sperling, A model for visual memory tasks, Hum. Factors 5 (1) (1963) 19–31.
[16] D.E. Erwin, The extraction of information from visual persistence, Am. J. Psychol. (1976) 659–667.
[17] M. Coltheart, The persistences of vision, Philos. Trans. R. Soc. Lond. B 290 (1038) (1980) 57–69.
[18] G.R. Loftus, C.A. Johnson, A.P. Shimamura, How much is an icon worth? J. Exp. Psychol.: Hum. Percept. Perform. 11 (1) (1985) 1.
[19] G.R. Loftus, J. Duncan, P. Gehrig, On the time course of perceptual information that results from a brief visual presentation., J. Exp. Psychol.: Hum. Percept. Perform. 18 (2) (1992) 530.
[20] P. Barrouillet, S. Bernardin, V. Camos, Time constraints and resource sharing in adults' working memory spans, J. Exp. Psychol. [Gen.] 133 (1) (2004) 83.
[21] P. Barrouillet, V. Camos, The time-based resource-sharing model of working memory, in: The Cognitive Neuroscience of Working Memory, Vol. 455, 2007, pp. 59–80.
[22] I.R. Olson, K. Page, K.S. Moore, A. Chatterjee, M. Verfaellie, Working memory for conjunctions relies on the medial temporal lobe, J. Neurosci. 26 (17) (2006) 4596–4601.
[23] A. Treisman, W. Zhang, Location and binding in visual working memory, Mem. Cogn. 34 (8) (2006) 1704–1719.
[24] A.M. Treisman, G. Gelade, A feature-integration theory of attention, Cogn. Psychol. 12 (1) (1980) 97–136.
[25] A. Treisman, S. Sato, Conjunction search revisited, J. Exp. Psychol.: Hum. Percept. Perform. 16 (3) (1990) 459.
[26] J.M. Wolfe, Guided search 2.0 a revised model of visual search, Psychon. Bull. Rev. 1 (2) (1994) 202–238.
[27] Y. Jiang, I.R. Olson, M.M. Chun, Organization of visual short-term memory, J. Exp. Psychol: Learn. Mem. Cogn. 26 (3) (2000) 683.
[28] A. Hollingworth, Object-position binding in visual memory for natural scenes and object arrays, J. Exp. Psychol.: Hum. Percept. Perform. 33 (1) (2007) 31.
[29] S.R. Mitroff, G.A. Alvarez, Space and time, not surface features, guide object persistence, Psychon. Bull. Rev. 14 (6) (2007) 1199–1204.
[30] A.M. Richard, S.J. Luck, A. Hollingworth, Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information, Cognition 109 (1) (2008) 66–88.
[31] S. Van der Stigchel, H. Merten, M. Meeter, J. Theeuwes, The effects of a task-irrelevant visual event on spatial working memory, Psychon. Bull. Rev. 14 (6) (2007) 1066–1071.
[32] B. Wyble, H. Bowman, M.C. Potter, Categorically defined targets trigger spatiotemporal visual attention, J. Exp. Psychol.: Hum. Percept. Perform. 35 (2) (2009) 324.
[33] J.R. Brockmole, M.A. Parra, S. Della Sala, R.H. Logie, Do binding deficits account for age-related decline in visual working memory? Psychon. Bull. Rev. 15 (3) (2008) 543–547.
[34] L.A. Brown, J.R. Brockmole, The role of attention in binding visual features in working memory: Evidence from cognitive ageing, Q. J. Exp. Psychol. 63 (10) (2010) 2067–2079.
[35] I.R. Olson, J.X. Zhang, K.J. Mitchell, M.K. Johnson, S.M. Bloise, J.A. Higgins, Preserved spatial memory over brief intervals in older adults, Psychol. Aging 19 (2) (2004) 310.
[36] M.A. Parra, S. Abrahams, R.H. Logie, S. Della Sala, Age and binding within-dimension features in visual short-term memory, Neurosci. Lett. 449 (1) (2009) 1–5.
[37] B.G. Breitmeyer, W. Kropfl, B. Julesz, The existence and role of retinotopic and spatiotopic forms of visual persistence, Acta Psychol. 52 (3) (1982) 175–196.
[38] J.A. Feldman, Four frames suffice: A provisional model of vision and space, Behav. Brain Sci. 8 (2) (1985) 265–289.
[39] K. McRae, B.E. Butler, S.J. Popiel, Spatiotopic and retinotopic components of iconic memory, Psychol. Res. 49 (4) (1987) 221–227.
[40] I.G. Sligte, H.S. Scholte, V.A. Lamme, Are there multiple visual short-term memory stores? PLoS One 3 (2) (2008) e1699.
[41] G. Alvarez, A. Oliva, The role of global layout in visual short-term memory, Vis. Cogn. 15 (1) (2007).
[42] D.E. Irwin, Memory for position and identity across eye movements, J. Exp. Psychol: Learn. Mem. Cogn. 18 (2) (1992) 307.
[43] S.J. Luck, An Introduction to the Event-Related Potential Technique, MIT press, 2014.
[44] E.-Y. Lee, N. Cowan, E.K. Vogel, T. Rolan, F. Valle-Inclan, S.A. Hackley, Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information, Brain 133 (9) (2010) 2677–2689.
[45] S. Girdher, A. Gupta, S. Jaswal, V. Naik, Predicting human response in feature binding experiment using EEG data, in: 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), IEEE, 2020, pp. 24–28.
[46] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, M. Shah, Deep learning human mind for automated visual classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6809–6817.
[47] K. Qiao, J. Chen, L. Wang, C. Zhang, L. Zeng, L. Tong, B. Yan, Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices, Front. Neurosci. 13 (2019) 692.
[48] J. Jiang, A. Fares, S.-H. Zhong, A context-supported deep learning framework for multimodal brain imaging classification, IEEE Trans. Hum.-Mach. Syst. 49 (6) (2019) 611–622.
[49] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, J. Schmidt, M. Shah, Decoding brain representations by multimodal learning of neural activity and visual features, IEEE Trans. Pattern Anal. Mach. Intell. (2020).
[50] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from EEG with deep recurrent-convolutional neural networks, 2015, arXiv preprint arXiv:1511.06448.
[51] S. Jirayucharoensak, S. Pan-Ngum, P. Israsena, EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation, Sci. World J. 2014 (2014).
[52] Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, J. Neural Eng. 14 (1) (2016) 016003.
[53] R.H. Abiyev, N. Akkaya, E. Aytac, I. Günsel, A. Çağman, Brain-computer interface for control of wheelchair using fuzzy neural networks, BioMed Res. Int. 2016 (2016).

[54] Z. Bahri, S. Abdulaal, M. Buallay, Sub-band-power-based efficient brain computer interface for wheelchair control, in: 2014 World Symposium on Computer Applications & Research (WSCAR), IEEE, 2014, pp. 1–7.

[55] R. Chai, S.H. Ling, G.P. Hunter, Y. Tran, H.T. Nguyen, Brain–computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization, IEEE J. Biomed. Health Inf. 18 (5) (2014) 1614–1624, http://dx.doi.org/10.1109/JBHI.2013.2295006.

[56] E.A. Curran, M.J. Stokes, Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems, Brain Cogn. 51 (3) (2003) 326–336.

[57] G.E. Fabiani, D.J. McFarland, J.R. Wolpaw, G. Pfurtscheller, Conversion of EEG activity into cursor movement by a brain-computer interface (BCI), IEEE Trans. Neural Syst. Rehabil. Eng. 12 (3) (2004) 331–338.

[58] X. Gao, D. Xu, M. Cheng, S. Gao, A BCI-based environmental controller for the motion-disabled, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 137–140.

[59] I. Gumilar, E. Sareen, R. Bell, A. Stone, A. Hayati, J. Mao, A. Barde, A. Gupta, A. Dey, G. Lee, et al., A comparative study on inter-brain synchrony in real and virtual environments using hyperscanning, Comput. Graph. 94 (2021) 62–75.

[60] D. Anwar, P. Garg, V. Naik, A. Gupta, A. Kumar, Use of portable EEG sensors to detect meditation, in: 2018 10th International Conference on Communication Systems & Networks (COMSNETS), IEEE, 2018, pp. 705–710.

[61] E. Sareen, L. Singh, A. Gupta, R. Verma, G.K. Achary, B. Varkey, Functional brain connectivity analysis in intellectual developmental disorder during music perception, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (11) (2020) 2420–2430.

[62] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (1) (2004) 9–21.

[63] H. Stanislaw, N. Todorov, Calculation of signal detection theory measures, Behav. Res. Methods Instrum. Comput. 31 (1) (1999) 137–149.

[64] J. Löfhede, M. Thordstein, N. Löfgren, A. Flisberg, M. Rosa-Zurera, I. Kjellmer, K. Lindecrantz, Automatic classification of background EEG activity in healthy and sick neonates, J. Neural Eng. 7 (1) (2010) 016007.

[65] J. Mitra, J.R. Glover, P.Y. Ktonas, A.T. Kumar, A. Mukherjee, N.B. Karayiannis, J.D. Frost Jr, R.A. Hrachovy, E.M. Mizrahi, A multi-stage system for the automated detection of epileptic seizures in neonatal EEG, J. Clin. Neurophysiol.: Off. Publ. Am. Electroencephalogr. Soc. 26 (4) (2009) 218.

[66] B. Greene, S. Faul, W. Marnane, G. Lightbody, I. Korotchikova, G. Boylan, A comparison of quantitative EEG features for neonatal seizure detection, Clin. Neurophysiol. 119 (6) (2008) 1248–1261.

[67] A. Aarabi, F. Wallois, R. Grebe, Automated neonatal seizure detection: a multistage classification system through feature selection based on relevance and redundancy analysis, Clin. Neurophysiol. 117 (2) (2006) 328–340.

[68] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[69] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–256.

[70] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.

[71] L.D. Blalock, B.A. Clegg, Encoding and representation of simultaneous and sequential arrays in visuospatial working memory, Q. J. Exp. Psychol. 63 (5) (2010) 856–862.

[72] L. Huang, A. Treisman, H. Pashler, Characterizing the limits of human visual awareness, Science 317 (5839) (2007) 823–825.

[73] A. Treisman, How the deployment of attention determines what we see, Vis. Cogn. 14 (4–8) (2006) 411–443.

[74] Y. Xu, Reevaluating the sensory account of visual working memory storage, Trends Cogn. Sci. 21 (10) (2017) 794–815.

[75] C. Babiloni, C. Del Percio, F. Vecchio, F. Sebastiano, G. Di Gennaro, P.P. Quarato, R. Morace, L. Pavone, A. Soricelli, G. Noce, et al., Alpha, beta and gamma electrocorticographic rhythms in somatosensory, motor, premotor and prefrontal cortical areas differ in movement execution and observation in humans, Clin. Neurophysiol. 127 (1) (2016) 641–654.

[76] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.

[77] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (9) (2015) 1904–1916.

[78] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.

[79] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[80] H. Smithson, J. Mollon, Do masks terminate the icon? Q. J. Exp. Psychol. 59 (1) (2006) 150–160.

[81] A.R. Vandenbroucke, I.G. Sligte, J.G. de Vries, M.X. Cohen, V.A. Lamme, Neural correlates of visual short-term memory dissociate between fragile and working memory representations, J. Cogn. Neurosci. 27 (12) (2015) 2477–2490.

[82] T.F. Brady, V.S. Störmer, G.A. Alvarez, Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli, Proc. Natl. Acad. Sci. 113 (27) (2016) 7459–7464.

[83] R. Luria, H. Balaban, E. Awh, E.K. Vogel, The contralateral delay activity as a neural measure of visual working memory, Neurosci. Biobehav. Rev. 62 (2016) 100–108.