Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy

Stephanie Lees, Paul Mccullagh, Liam Maguire, Fabien Lotte, Damien Coyle

Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy

Stephanie Lees[1], Paul McCullagh[1], Liam Maguire[1], Fabien Lotte[2], Damien Coyle[1]

[1] intelligent Systems Research Centre, Faculty of Computing, Engineering and the Built Environment, Ulster University, UK
[2] Inria Bordeaux Sud-Quest/LaBRI (CNRS / Bordeaux INP / Univ. Bordeaux), Talence, France
Lees-s2@email.ulster.ac.uk

Abstract— Rapid serial visual presentation (RSVP) based brain-computer interfaces (BCIs) can detect target images among a continuous stream of rapidly presented images, by classifying a viewer's event related potentials (ERPs) associated with the target and non-targets images. Whilst the majority of RSVP-BCI studies to date have concentrated on the identification of a single type of image, namely *pictures*, here we study the capability of RSVP-BCI to detect three different target image types: *pictures*, *numbers* and *words*. The impact of presentation duration (speed) i.e., 100-200ms (5-10Hz), 200-300ms (3.3-5Hz) or 300-400ms (2.5-3.3Hz), is also investigated. 2-way repeated measure ANOVA on accuracies of detecting targets from non-target stimuli (ratio 1:9) measured via area under the receiver operator characteristics curve (AUC) for N=15 subjects revealed a significant effect of factor Stimulus-Type (*pictures, numbers, words*) ($F(2,28) = 7.243$, $p = 0.003$) and for Stimulus-Duration ($F(2,28) = 5.591$, $p = 0.011$). Furthermore, there is an interaction between stimulus type and duration: $F(4,56) = 4.419$, $p = 0.004$). The results indicate that when designing RSVP-BCI paradigms, the content of the images and the rate at which images are presented impact on the accuracy of detection and hence these parameters are key experimental variables in protocol design and applications, which apply RSVP for multimodal image datasets.

***Index Terms*— Rapid Serial Visual Presentation, Brain-Computer Interface, BCI, Event Related Potentials, Electroencephalography, EEG**

I. Introduction

Rapid serial visual presentation (RSVP) is characterized by sequentially displaying images at the same spatial location at high presentation rates [1][2][3]. Brain-computer interfaces (BCIs) are communication and control systems that enable users of this technology to send commands to a computer by using only their brain activity, generally measured using electroencephalography (EEG); and processed to extract relevant information [4]. The combination of RSVP and BCIs, has been used successfully in the detection of target stimuli. Many applications may benefit from optimized systems involving both humans and machines, for example, counter intelligence and policing, where large amounts of images need to be observed, classified and sorted by analysts searching for possible targets; and medical image screening, where target diseases may be identified [5][6]. Event-related potentials (ERPs) are measured or computed brain responses in EEG that occur in response to the onset or processing of a stimulus. ERPs are time-locked to specific events and are normally identified by averaging epochs over repeated trials [7]–[9]. In the RSVP paradigm, non-frequent target images are presented within streams of frequent non-target images. The ERP most commonly exploited with RSVP-BCI applications is the P300 component. The P300 refers to a positive component that appears from around 250ms to 500ms after the presentation of the target stimulus [10]–[13].

RSVP-BCIs have been used to detect and recognize target pictures of objects, scenes, people and events in static and motion images [14][15][16][17]. Computers are unable to analyze imagery as efficiently or successfully as people but manual analysis tools are slow [2][18]. In studies carried out by Sajda *et al*.[19], Poolman *et al.* [20] and Bigdely-Shamlo *et al.* [5], a trend of using RSVP-BCIs for rapidly identifying targets within different image types has emerged and the combination of RSVP and BCI has proven successful on several image sets. Other research has attempted to establish whether or not greater efficiencies can be reached through the combination of RSVP-BCIs and behavioral responses. Files and Marathe [21] showed that methods for measuring real-time button press can be combined with EEG in order to give better accuracy, assessed using the area under the receiver operating characteristic (ROC) curve. Whilst the combination of EEG and a button press can lead to increased performance in RSVP-BCIs, the core advantage of RSVP-BCIs is the enhanced speed of response of the ERPs in comparison to the reaction times normally associated with behavioural responses (e.g., overt motor responses such as tapping a button).

The majority of research focuses on a two-class problem i.e., detecting target images in sequences of non-target images that are completely different from each other. However, in real-life situations, non-target images are likely to share some of the

same characteristics of target images [22]. Within datasets used in industry, data are likely to be represented in one of three categories: *pictures*, *numbers* or *words*. Typical applications that incorporate these data types include database retrieval, computer games, surveillance, policing, and health care. The complexity of stimuli within RSVP studies varies depending upon the task the participant is required to carry out. Task complexity is boosted when the number of target categories is increased. It has been suggested that the percentage of targets should be lower than 10% to evoke the P300 and maximize correct detection rates [23]. In a study by Won *et al.* [24] researchers compared motion RSVP to static RSVP. Results showed an increase in performance accuracy with motion-image RSVP versus static-image, which could be attributed to the shorter latency and greater amplitudes of ERP components in the motion-image experiment [24].

In summary, although some studies have used pictures, individual letters and/or individual number digits in RSVP BCI paradigms, these image types were never systematically compared and response metrics evaluated. Most of the RSVP-BCI studies to date have focused on pictures of objects, places, people or animals; few have explored other types of data such as numbers or words. The RSVP-BCI paradigm could be useful for rapid word/number search and therefore a comparison with picture stimuli is required. Potential applications could include detecting missile silos in satellite images, searching for words or numbers contained within images in databases or in graphs; and analyzing info graphics that contain a mixture of pictures, numbers and words.

In addition, the optimal presentation speed is still unknown for all image types. To date there has been no systematic and formal comparison between the different data types, nor study to assess the best presentation rate for them. In this study, we address the following questions:

(i) Is there a difference in detection accuracy for each image type (*pictures*, *numbers and words*) and hence what are the implications of this for RSVP applications?

(ii) Is there an optimal presentation speed at which each image type (*pictures*, *numbers and words*) should be presented? In this work, we compare rates of 10-5Hz, 5-3.33Hz and 3.33-2.5Hz corresponding to presentation durations of 100-200ms, 200-300ms, and 300-400ms.

To answer these questions, we analyzed the performance of 15 subjects offline after an RSVP session. This paper is organized as follows; Section II presents methods and experimental parameters used to carry out the study. Section III details the performance of single trial classification. In Section IV, the findings are discussed and future work is suggested.

II. Methods

Volunteers were asked to participate in a single session to perform visual search tasks among 4500 images presented for different durations (100-200ms, 200-300ms or 300-400ms). Participants were directed to identify target images within a collection of non-target images. Fifteen participants (4 females, 11 males, age range 19-34 years) participated in the study at Ulster University, after giving their written consent. All had normal vision or corrected to normal vision. None had history of neurological disease or injury. The study was approved by Ulster University research and governance department after ethical review. EEG data were recorded non-invasively using a 16-channel g.USBamp (g.tec, Austria) with active gel-based electrodes and sampled at 256Hz. Electrodes were placed at the following locations based on the international 10-20 system: Fz, Cz, T7, T8, P7, P3, Pz, P4, P8, PO7, PO3, PO4, PO8, O1, Oz, O2. All electrodes were referenced to the left mastoid, and used a forehead ground at Fz (see Fig 1).

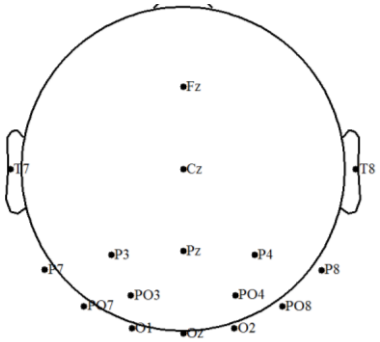

Fig. 1. Electrode site placement. EEG data were sampled at 256Hz from 16 channels setup in 10-20 system.

Stimuli were presented on a 17.3" UltraSharp FHD (1920*1080) wide view Anti-Glare LED-backlit monitor. The presentation sequence, recording and storage were controlled by scripts set up in Cogent [25] and MATLAB Simulink [26] packages. Additional data analysis and classification were performed using MATLAB (see section C below).

A. Experimental Design

Participants had to detect a target *picture*, *number* or *word* in a sequence of distractor pictures, numbers or words that were presented at the same location. 10% of images were target images and 90% were non-target images.

1) Stimuli

Pictures: *pictures* were selected from the 'morgueFile' database [27]. The target pictures are based on a study carried out by Wang *et al*. [28], namely: dalmatians, motorbikes, helicopters, starfish and candle sticks. The *pictures* that make up the non-targets are random images from within the same database that vary in type from animals to food (see Fig 2).

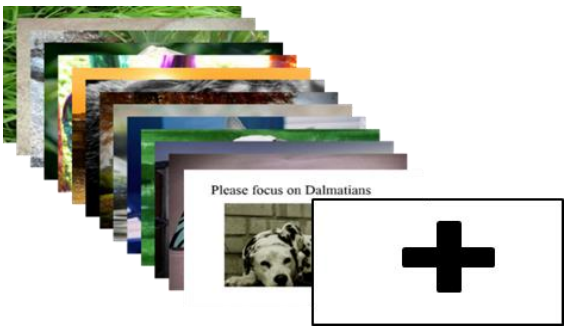

Fig. 2. Example of an RSVP *picture* series.

***Numbers*:** The *numbers* stimuli were randomly generated. The target/non-target numbers generated are in the range 100 to 500 (e.g. 101, 232, 357, 396, 157 etc.). The number images have a white background with the number presented in black font Times New Roman in the center, size 32 pixels. (see Fig 3).

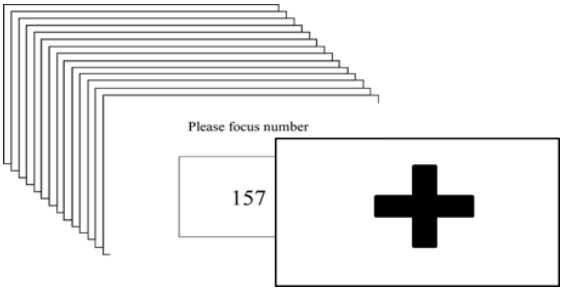

Fig. 3. Example of an RSVP number series.

***Words*:** Three-letter common *words* (tag, gum, him, any, pen etc...), were selected at random, also presented on a white background with black font Times New Roman, size 32 pixels (see Fig 4).

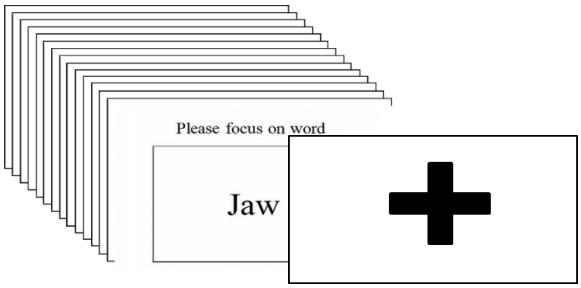

Fig. 4. Example of an RSVP word series.

All images were scaled to 560 × 360 pixels (width × height). The participants were seated 1 meter from the screen, which means that a stimulus covered 26° of the visual field of view.

2) Rate of presentation

Fig 5 details the granularity chosen to allow for varying presentation rate and image type. The duration of each image is either 100-200ms, 200-300ms or 300-400ms. The fundamental presentation unit is a Group of 100 images. A run comprises of 5 Groups (500 images) of either *numbers*, *words* or *pictures*. A Block is a set of the three Runs (1500 images with a Group of each type). The interval between each Run is 3 sec. Each session is made up of 3 Blocks, with different permutations of image types. The interval between each Block is 3 minutes. A session begins with fixation cross presented for 3 secs followed by a cue presented for 3 secs. Images were presented for randomly chosen durations between 100-200ms, 200-300ms or 300-400ms in each run illustrated in Fig 5. Table 1 shows one possible presentation order that a participant could receive. This was randomized, to prevent order effects. It has been shown that the length of the inter-stimulus interval (ISI), the temporal interval between the offset of one stimulus to the onset of another, as well as the variability, changes habituation in subjects and therefore is often randomized to prevent habituation and to minimize expectation effects [29]. In this study stimuli appear in rapid sequence one after the other and there is no ISI; therefore we introduced variability in the length of the stimulus presentation in each timing category (i.e., randomizing presentation times in each of the following intervals 100-200ms, 200-300ms and 300-400ms). Randomizing the timing of the inter-stimulus interval ensures that the alpha-wave activity of the participant does not become phase locked with the stimulus presentation rate [30].

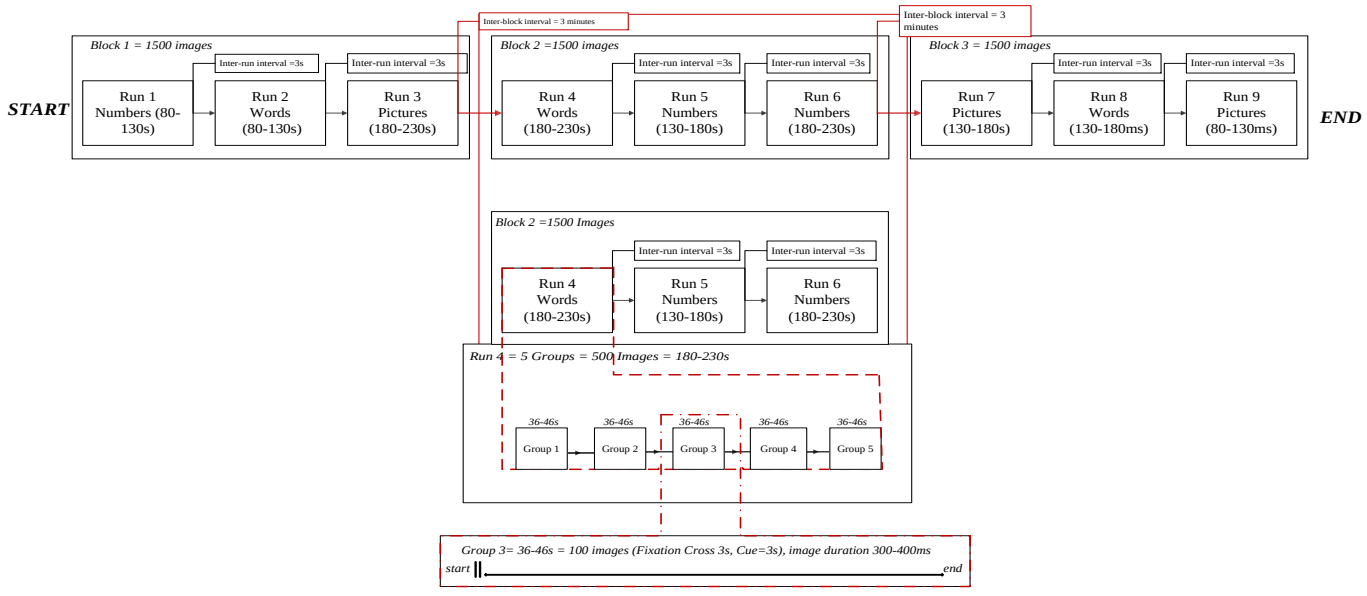

Fig. 5. A session comprises 3 blocks with an inter-block interval of 3 minutes. A block comprises 3 runs. Each run has 5 groups. A group is 100 images with 10% of the group comprising of target images.

TABLE I: EXAMPLE OF POSSIBLE RANDOMIZATION OF GROUP TYPES IN SESSION 1

	Run1	Run 2	Run 3
Block 1	Word (300-400ms)	Picture (300-400ms)	Number (300-400ms)
Block 2	Number (100-200ms	Word (200-300ms)	Picture (100-200ms)
Block 3	Picture (200-300ms)	Number (200-300ms)	Word (100-200ms)

B. Data preprocessing and feature extraction

Epochs were derived in association with the onset of each stimulus, beginning 200ms prior to the onset of the stimulus and lasting for 1000ms. Data were digitally filtered using a low-pass Butterworth filter (order 5, with cut-off at 10Hz) and subsequently resampled at 20Hz to reduce the number of features. Features comprising EEG signal amplitudes were extracted between 50ms to 450ms epoch (post stimulus). This produces nine features for each channel, irrespectively of the presentation durations. The features vector X is given in equation (1)

$$X = \{x_{11}, \dots x_{1n};\ \dots; x_{m1}, \dots x_{mn};\} \qquad (1)$$

where x is the down-sampled EEG signal for each trial within the 50ms to 450ms period post stimulus, n is the number of features taken from this period (n=9 i.e., every 50ms), and m is the number of best channels that are concatenated. For the channel ranking study as described below, m=1 (in this study n is not optimised). No trials were removed. Visual inspection of trials was performed and there were no obvious eye movement artefacts detected.

C. Calibration and testing

For each of the 9 runs there were 50 targets and 450 non-targets and a different classifier was setup and employed for each stimulus type and duration. For training and testing the data was split, 50% training and 50% testing randomly selected. In order to select channels, a Linear Discriminant Analysis (LDA) [31] classifier was trained to discriminate target vs. non target feature vectors extracted from single channels in a Leave One Out (LOO) cross validation on the 50% of the data used for training (50% excluded for testing). For each of the sixteen channels the average LOO classification accuracy (LOO-CA) was determined and channels were ranked by accuracy. The most commonly highest ranked channels across subjects were Pz, P3 and PO3. The top three ranked channels were concatenated to form a new feature vector (3 channel and 9 features, i.e., 27 features per vector). As there are nine non-target stimuli for each target stimulus, there are a number of options for selecting the non-target stimuli trials to form the second class for training a classifier with a balanced number of examples per class. One approach is to downsample the non-target stimuli data by randomly selecting from the non-target data an equal number of trials to that of the target data. However, as the number of target data trials was limited we up-sampled the target class data by repetition of target samples. This balances the target and non-target class and ensures that there are sufficient data to train the classifier. We used up-sampled targets vs non-target trials in the training data trials. A

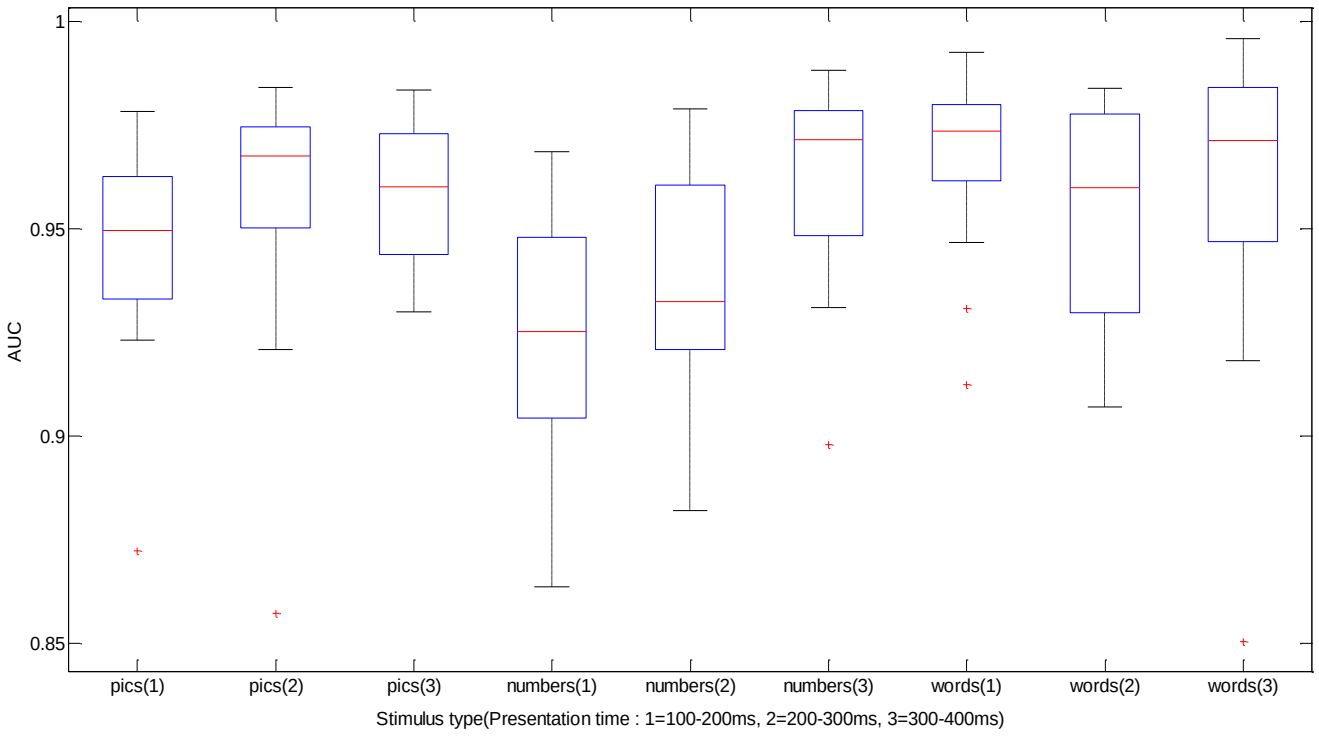

Fig. 6. Boxplots of AUC analysis across different data types at each of the presentation rates.

new LDA [31] classifier (using 50% of data) was produced to classify target vs non-target data on remaining 50% of unseen testing data.

The accuracies, achieved in detecting targets from non-target stimuli (ratio 1:9), are measured via area under the receiver operator characteristics curve (AUC) for each of the 15 participants. The effects of duration (3 different speeds i.e., 100-200ms (5-10Hz), 200-300ms (3.3-5Hz) or 300-400ms (2.5-3.3Hz)) and stimulus type (3 type of stimulus, i.e. pictures, numbers, and words) and the interaction effects between stimulus duration and type is investigated using a 2-way repeated measure (ANOVA) and pairwise post-hoc analyses with paired t-test and Bonferonni corrections for multiple comparisons.

III. Results

Results obtained on the testing set are shown in the boxplots for each of the different data types presented across all speeds in Fig 6. The ANOVA results reveal a significant difference in AUC between stimulus type ($F(2,28) = 7.243$, $p = 0.003$). Pairwise comparisons reveal that the AUC with *words* is significantly higher than *numbers* ($p<0.05$ Bonferroni corrected for multiple comparisons), but not significantly higher than *pictures* ($p > 0.05$). There are significant differences in AUC achieved with different stimulus presentation speeds ($F(2,28) = 5.591$, $p = 0.011$). Pairwise comparisons reveal that the AUC with 300-400ms is significantly higher than AUC for 100-200ms ($p < 0.05$, Bonferroni corrected for multiple comparisons), but is not higher than 200-300ms. Furthermore, there is an interaction effect between stimulus duration and stimulus type ($F(4,56) = 4.419$, $p = 0.004$). This interaction seems to be because the AUC for the shortest stimulus duration, 100-200ms, *words* stimuli produced higher AUC than *pictures* and *numbers*. This is illustrated in Fig 7, which shows the profile plots for effects stimulus time and stimulus type based on the estimated marginal means of AUC. In Fig 8 we show that performance significantly decreases when less than 6 channels are used. Thus, detection is best using Pz, P3, PO3, P4, P07, P08. However, even with one electrode an average of 75% AUC is achieved so if the real world setting limits the number of electrodes then there would be trade-off between accuracy and electrodes.

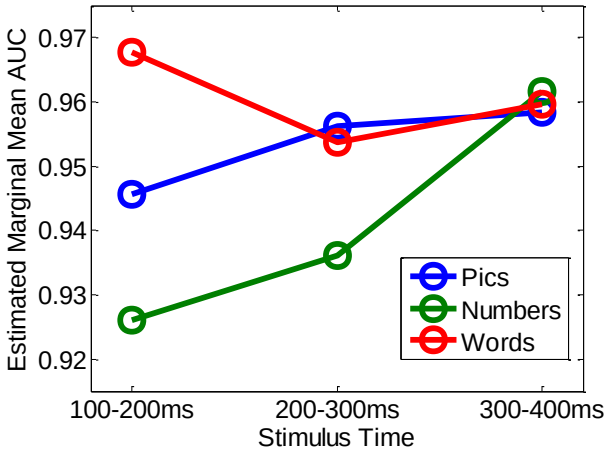

Fig. 7. Estimated marginal mean AUC showing interaction effects for Stimulus Time and Stimulus Type where stimulus time

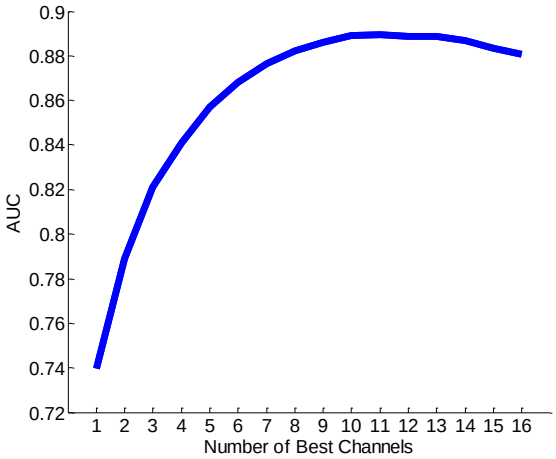

Fig. 8. AUC versus number of channels as channels increased, averaged across all participants. The channels are ranked in the following order across participants: Pz, P3, PO3, P4, PO7, PO8, PO4, P7, O1, P8, O2, Cz, T7, T8, Oz, Fz. The three highest ranking channels were calculated using LOOCV, as explained above.

The topographic map (Fig. 10) displays the comparative importance of each electrode channel based on AUC in single channel analysis for each channel for all three data types (*pictures*, *numbers* and *words*), shown at the three different durations (100-200ms, 200-300ms and 300-400ms). These topographical plots show the differences in brain area associated with each of the different stimulus/types and durations. With the topographical plots and number of best electrodes (Fig 10) it is possible to determine which electrodes have maximized AUC in the study. *Number*s show a distinctly different patterns of activation to *pictures* and *words* with the most rapid presentation rates showing maximum discrimination around inferior temporal gyrus, whilst for *pictures* and *words* maximum target detection appear around centroparietal area beginning at Cz. The channels are ranked in the following order across participants: Pz, P3, PO3, P4, PO7, PO8, PO4, P7, O1, P8, O2, Cz, T7, T8, Oz, Fz. This is based on the number of times channels are selected at each rank. The best number of channels did however differ across experimental conditions and we have now reported that in Fig 9. A repeated measures ANOVA showed a significant effect on number of channels used for stimulus type and duration with post hoc multiple comparison paired t-test revealing stimulus type Numbers presented at 100-200ms required significantly more ($p<0.05$) channels to achieve maximum performance than all other presentation durations/stimulus types except for pictures presented at 100-200ms. Pearson correlation coefficient to assess the correlation between AUC presented in the barchart in Fig 9 and number of electrodes used revealed there is a significant negative correlation ($R = -0.8371$, $p = 0.0049$)

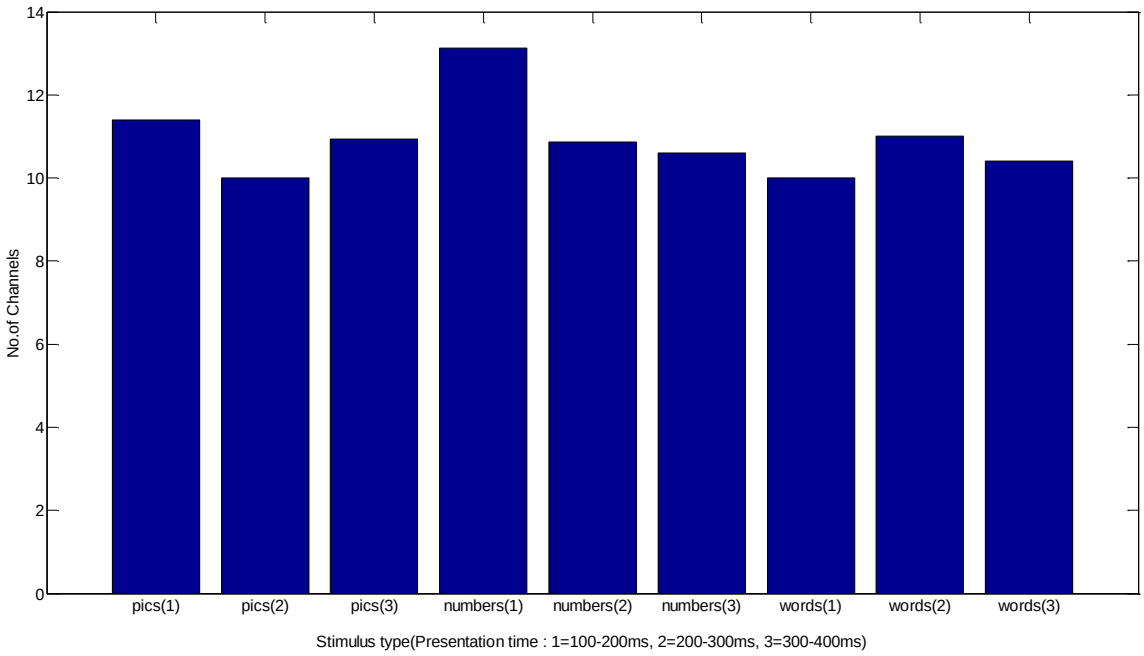

Fig. 9. The best number of channels across experimental conditions

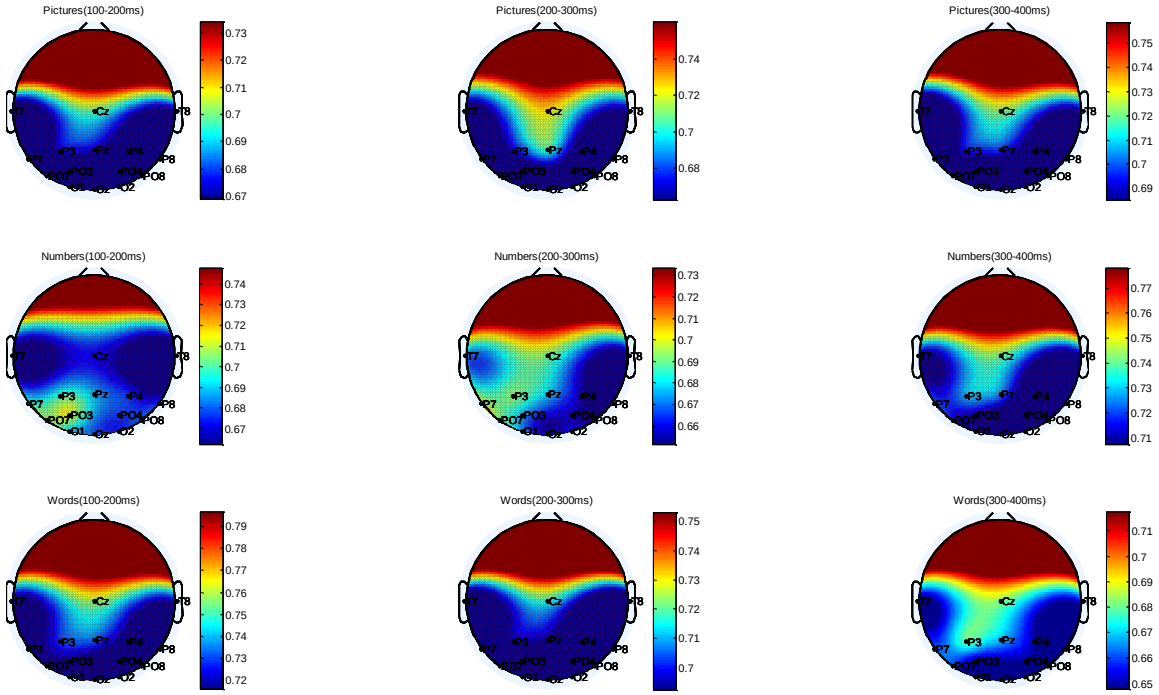

Fig. 10. Topographical plots based on AUC in single channel analysis for each channel. The graphs show all three data types (*pictures*, *numbers* and *words*), at the three different durations (100-200ms, 200-300ms and 300-400ms) and illustrate the differences in brain areas associated with each of the different stimulus/types and durations.

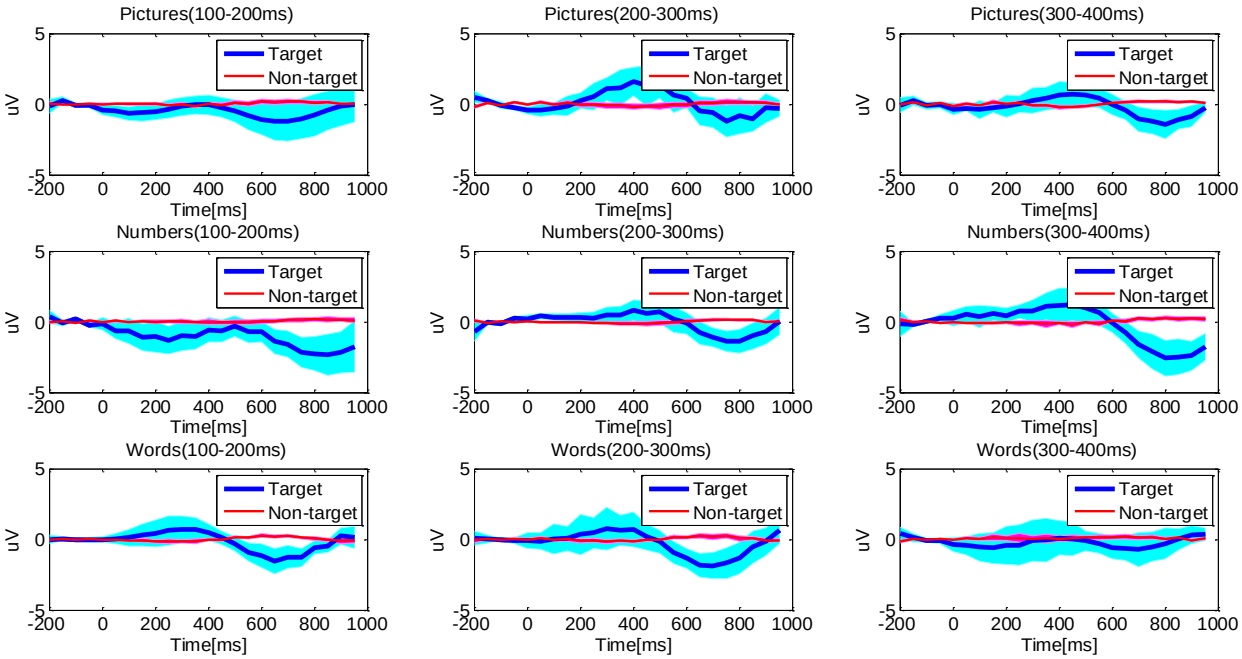

Fig. 11. Grand average ERPs for the 15 participants to Target and Non-Target stimuli (-200ms to 1000ms post stimulus) for each of the presentation speeds (100-200ms, 200-300ms and 300-400ms) and stimulus types.

Fig 11 shows target vs non-target ERP components on channel P3. Each graph shows the grand average ERPs for the 15 participants across the nine different experimental conditions with a time range of -200ms to 1000ms post stimulus onset. The amplitude of the P300 is lower for faster stimulus times. P300 delay increases when the stimulus duration increases. There are more oscillations in the longer stimulus durations, possibly due to ERPs from preceding stimuli. The continuous shaded region shows the standard deviation for target and non-target data. *pictures* and *numbers* at fast (100-200ms) presentation rates show minimal positive peaks around 300-400ms compared to all other stimuli at other speeds and presentation rates, which is correlated with lowest accuracy. Fig 11 reveals a strong late negativity to target stimuli occurring at approximately 700ms post-stimulus.

IV. Discussion

In the study, we evaluated an RSVP-BCI with different data types/stimuli at various rates of presentation. Previous studies have shown that the difficulty of target identification has an effect on the P300 ERP [3][4]. Our results indicate that *pictures, numbers and words* could be detected at fast presentation rates, up to 10Hz with high accuracy. The highest variability between stimulus times is observed for *numbers* and performance at faster presentation rates for *numbers* differed significantly to that of the slower rates. To maximize performance, participants require longer rates of presentation when identifying *numbers*, with *numbers* having a significantly lower detection rate than *words* and *pictures* at the 10 Hz presented rate. A possible interpretation could be that there is more familiarity with the common *words* and *pictures* used. Studies show that if participants are familiar with a particular stimulus then they will identify it more readily within a group of stimuli [32][33]. It was also shown in [32] that a good reader is able to read five words per second and that the predictably of the next word in the text can have an effect on this. The perceived difficulty of a word can decrease accuracy, and hence a future study could be carried out using a RSVP-BCI on words of different complexities [34].

Neural activity propagates from the primary visual cortex (area VI) to higher cortical areas and back before recognition can occur at the level of detail required for an individual image to be detected. Lamme and Roelfsema [35] suggest response latencies at each hierarchical level of the visual system are approximately 10ms. Therefore supposing a minimum of five levels must be navigated as activity transmits from V1 to higher cortical areas along with reentrant loops, this is unlikely to occur in less than 50ms [36][37]. Thus, RSVP processing frequency has a theoretical maximum of approximately 20Hz. Nevertheless, more research is needed to explore the limits of human capability to rapidly detect target from non-target information. It is known that the P300 ERP can be suppressed if the time between two targets is less than 500ms [38][39] so some refractory period may be involved for repeated stimuli. The amplitude and the latency of the P300 are both influenced by the target discriminability and the target-to-target interval in the sequence. Moreover, the complexity of a stimulus affects the latency of the P300 [30][31]. Hence these factors can affect the reliability and accuracy of an RSVP-BCI [17]. A key feature of the RSVP-BCI paradigm is the rate of presentation; this is particularly relevant as the focus of this paradigm is presenting data to participants at a rapid rate, so that large amounts of data can be analyzed in short time periods. In the literature, the

reported duration of stimuli varied from 50 to 500ms [16][32][33], but the optimal presentation duration/rate is undetermined.

For this reason, this paper focused on investigating different stimulus presentation durations and therefore determining the optimal rate of presentation for each stimulus type. In a study by Sajda, *et al*. [14], two participants were asked to identify people in natural scenes. The duration of stimulus presentation was decreased from 200ms to 100ms and then to 50ms per image. The findings showed that the participants' performance reduced when stimuli were presented at the faster speed i.e., a duration of 50ms; the reasoning behind this may be due to the 'attentional blink' phenomenon [38]. In a study by Raymond *et al.* [38], it was found that if a second target appeared within ~100-500ms of the initial target, participants were less likely to identify the second target. This does not mean that participants cannot process information at rates higher than 10Hz, but suggests a masking process. Indeed Forster [34] showed that participants can process words presented in a sentence at up to 16 Hz and Fine and Peli [44] showed that participants can process words at 20 Hz in an RSVP paradigm. There is a direct interaction between target difficulty and presentation rate. The optimal presentation rate for a stimulus set is dependent on the difficulty of identifying targets [45]. Given that there are greater uncertainties with presentation rates in excess of 10Hz we deemed it appropriate to study at the maximum presentation rate of 10Hz down to 5Hz allowing for variability to enhance randomness of stimulus presentation, and compare this to presentation rates of between 5Hz to 3.33Hz and 3.33Hz to 2.5Hz. Further investigation for the different stimulus types should consider faster rates up to 20Hz.

A possible interpretation for the lower performance and higher variabilities with *numbers*, notably as compared to *words*, may be derived from the cognitive load theory (CLT) [46]. Indeed, when looking for a target 3-digit number, users have to maintain in their working memory 3 different digits, i.e., 3 items, to compare them with the 3-digits of each displayed number. For a 3 letter common word, the users do not have to maintain 3 different items in working memory, but only one: the word meaning/concept (which is represented as a single "scheme" according to the CLT). Indeed, a known common word is a not a random sequence of 3 letters, unlike a random 3-digit number. Thus maintaining a 3-letter common word in working memory should lead to lower cognitive load than maintaining a 3-digit random number in working memory. However, it is known that a higher cognitive load actually decreases the amplitude of the P300 (see [47][48]). This is confirmed by our data, which showed a lower P300 for the *numbers* category as compared to other categories, in particular as compared to the *words* category. To sum up, looking for target numbers may require more cognitive load than for other stimuli categories, which thus decreases the P300, which in turns reduces the RSVP-BCI accuracy. We had not anticipated this effect, therefore, in the future, it would be interesting to measure the users' cognitive load, e.g., using the NASA-Task Load Index questionnaire [49], or EEG markers of Workload [50], to assess the impact of cognitive load on RSVP-BCI performances. For a similar reason, it would be interesting to measure the influence of the users' working memory span on their RSVP-BCI performances with numbers.

Rate of presentation in an RSVP paradigm influences single trial detection performance. In ERP measurements, it is well known that the rate of presentation has a high importance when measuring P300. In a study by Potter *et. al.* [65] it was shown that even if pictures are shown at a rate of 6/s, participants are still able to identify target *pictures*, at least momentarily. Our analysis shows that an average of 0.9 AUC is observed for all image modalities and therefore it is anticipated that presentation rates could be increased, at least for some participants. Further study is required to investigate the limits of detection accuracy for image modality. It also remains to be explored whether participants that engage with a specific image modality more often perform better with that type of image e.g. a data analyst with numbers, a radiologist with *pictures* or a news presenter with *words*. *Numbers* have the most variation and have presented the lowest performance of all three image types at the fastest presentation rates.

Interestingly stimulus type *numbers* has the lowest accuracy and requires the maximum number of channels, significantly more than all other stimuli/durations, except stimulus type *pictures* at 100-200ms. The electrode utilization assessment indicates that different stimulus modalities and speeds require electrodes locations and the number of electrodes utilized to be specifically selected to maximize RSVP performance.

The topographical plots derived from single channel AUC (Fig 10) indicate the fast *numbers* processing occurs around inferior temporal gyrus, whilst for *pictures* it appears around C3 and *words* predominantly around Cz. For *pictures* this occurs in the opposite direction. Investigations using fMRI, electrophysiological recordings, and electrical stimulation methods have suggested that numerals may be visually processed differently than other stimuli, but many studies have not consistently identified a common brain region within the ventral visual stream [51]. Using intracranial electrophysiological recordings, Shum et al [51] observed a significantly higher response in the high-frequency broadband range (high gamma, 65–150Hz) to visually presented numerals, compared with morphologically similar (i.e., letters and false fonts) or semantically and phonologically similar stimuli (i.e., number words and non-number words). Anatomically, this preferential response was consistently localized in the inferior temporal gyrus (ITG) and anterior to the temporo-occipital incisure. Our results showing ERPs for *numbers* presented at 10Hz are classified maximally in left hemisphere closest to ITG are consistent with finding of ITG activation during processing numerals as shown in [51]. However, our result presents no evidence of lateralized activation being detectable and no right side activation at the faster presentation rates.

This RSVP BCI work will be further developed to gain (i) a better understanding of image types, (ii) enable minimization of reaction time, (iii) determine the earliest ERPs for detection, (iv) enable the most reliable detection of time-locked ERPs and (v) identify which user factors impact their performances with a given stimulus type or duration, to provide the best RSVP-

BCI to each user. This may enable the development of tools to assess a person's predisposition to types of quantitative representations and their associated subjective qualitative statements. This may also aid optimization of the delivery of visual information to ensure best decisions in situations where fast-paced decision are necessary.

A limitation of our study is that we did not control for types of target images used. Variation in ERP responses (timing and amplitude) occur depending on whether or not the stimulus has meaning (e.g., in this study if participants own a Dalmatian then they are likely to have a quicker response with a higher amplitude) [52]. The peaks and troughs of a stimulus-locked ERP waveform allow us to visualize cognitive processing as it unfolds during a trial. The P300 is elicited by a class of task related events. Its amplitude has been shown to be directly proportional to the participants expectancy of a stimulus [53][54]. In future studies we intend to control the stimulus type ensuring that subjects are not predisposed to any particular stimulus. Another limitation of our study is that the presentation rates were almost overlapping at the limits of each band of presentation duration, e.g., around 200ms in 100-200ms vs 200-300ms, thus some stimulus presentation times could differ by only as much as a few milliseconds, even though they are different groups in the analysis. In future studies times between groups will be separated by a minimum of 40ms e.g., 100-180ms vs 220-300ms.

V. Conclusion

The human brain is considered the most powerful visual information processing system as it can evaluate a scene in a few hundred milliseconds [55]. Humans exploit this capability all the time, however there is a bottleneck for humans responding via the normal muscular channels to the fast interpretation and detection of information in image scenes. This poses problems when the response is required rapidly or if there is a requirement to process large data volumes efficiently and/or monitor data rapidly. To learn how to exploit this capability, research has focused on understanding the neural correlates of visual information processing to create symbiotic interaction between humans and machines through BCIs. Our research questions were: "are there differences in accuracy of detecting ERPs for different image/stimulus types and different presentation rates and are there interaction effects between stimulus type and stimulus duration?" Our results revealed a significant effects of factor Stimulus-Type *(pictures, , numbers, words*) and of factor Stimulus-Duration as well as an interaction between stimulus type and duration. Such interaction notably suggested that at the shortest stimulus duration, *words* stimuli produced higher AUC than other stimuli, in particular compared to *numbers*.

The words stimuli can be detected at higher speeds (equivalent to 100-200ms duration) with similar detection accuracy. *Pictures* can also be detected at higher presentation rates but with detriment to the accuracy. With *numbers* data type there was a significant decrease in accuracy from 200-300ms vs 100-200ms, therefore in this case a tradeoff between speed vs accuracy is not beneficial.

This study contributes to the knowledge relating to RSVP BCI paradigms, especially as the study of the optimal setup for RSVP-BCI is ongoing and remains open [17]. It shows (1) the feasibility of using RSVP-BCI to identify targets in multiple image-types and (2) for the first time, the differences and similarities between different stimuli presented at varying rates. A major focus is a comparative analysis of performance achieved used different modalities of stimuli, determining which stimulus types, if any, are more difficult to detect in RSVP at different rates of presentation. This is the first step in developing RSVP interface for image triage using different modalities. A real world detection scenario may require that a subject is searching for images containing *words*, *numbers* and/or *words*. Our work implies that the accuracy of detection for each category can be >90%, but that the individual categories depend on presentation rate and the scalp topography is dependent on stimulus type (Fig 9).

Many applications would benefit from optimized RSVP-BCI systems, for example, counter intelligence, policing and clinical diagnosis where large amounts of images or information need to be observed, analyzed, understood and classified on a daily basis by analysts. Follow up work will explore the use of RSVP-BCI with multimodal presentations in conjunction with visual analytics tools to assess the differences in performance for targets containing mixture of *pictures*, *numbers* and *words*.

References

[1] J. R. Wolpaw; and E. W. Wolpaw;, *Brain Computer Interface : Principles and practice*. Oxford ; New York : Oxford University Press, 2012.

[2] S. Mathan *et al.*, "Rapid image analysis using neural signals," *Proceeding twenty-sixth Annu. CHI Conf. Ext. Abstr. Hum. factors Comput. Syst.*, p. 3309, 2008.

[3] R. Spence and M. Witkowski, *Rapid Serial Visual Presentation*. 2013.

[4] J. Farquhar and N. J. Hill, "Interactions between pre-processing and classification methods for event-related-potential classification: best-practice guidelines for brain-computer interfacing.," *Neuroinformatics*, vol. 11, no. 2, pp. 175–92, Apr. 2013.

[5] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain Activity-Based Image Classification From Rapid Serial Visual Presentation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp. 432–441, Oct. 2008.

[6] J. M. Aquino and K. M. Arnell, "Attention and the processing of emotional words: Dissociating effects of arousal," *Psychon. Bull. Rev.*, vol. 14, no. 3, pp. 430–435, 2007.

[7] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Effects of performing two visual tasks on single-trial detection of event-related potentials.," *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, pp. 1723–1726, 2012.

[8] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering.," *IEEE Trans. neural networks Learn. Syst.*, vol. 25, no. 11, pp. 2030–42, Nov. 2014.

[9] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, and K. E. Hild, "A framework for rapid visual image search using single-trial brain evoked responses," *Neurocomputing*, vol. 74, no. 12–13, pp. 2041–2051, Jun. 2011.

[10] J. Polich and E. Donchin, "P300 and the word frequency effect," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 1, pp. 33–45, Jul. 1988.

[11] Y. Zhang, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "A novel BCI based on ERP components sensitive to configural processing of human faces.," *J. Neural Eng.*, vol. 9, no. 2, p. 026018, Apr. 2012.

[12] D. Ming *et al.*, "Time-locked and phase-locked features of P300 event-related potentials (ERPs) for brain–computer interface speller," *Biomed. Signal Process. Control*, vol. 5, no. 4, pp. 243–

251, Oct. 2010.
[13] V. Leutgeb, A. Schäfer, and A. Schienle, “An event-related potential study on exposure therapy for patients suffering from spider phobia.,” *Biol. Psychol.*, vol. 82, no. 3, pp. 293–300, Dec. 2009.
[14] P. Sajda, a. Gerson, and L. Parra, “High-throughput image search via single-trial event detection in a rapid serial visual presentation task,” *First Int. IEEE EMBS Conf. Neural Eng. 2003. Conf. Proceedings.*, pp. 7–10, 2003.
[15] A. Matran-Fernandez and R. Poli, “Collaborative Brain-Computer Interfaces for Target Localisation in Rapid Serial Visual Presentation,” *6th Comput. Sci. Electron. Eng. Conf. CEEC 2014 - Conf. Proc.*, pp. 127–132, 2014.
[16] D. Rosenthal, P. Deguzman, L. C. Parra, and P. Sajda, “Evoked neural responses to events in video,” *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 3, pp. 358–365, 2014.
[17] S. Lees *et al.*, “A review of rapid serial visual presentation-based brain-computer interfaces,” *J. Neural Eng.*, vol. 15, no. 2, 2018.
[18] A. D. Gerson, L. C. Parra, and P. Sajda, “Cortical Origins of Response Time Variability during Rapid Discrimination of Visual Objects,” *Neuroimage*, vol. 28, no. 2, pp. 326–341, 2005.
[19] P. Sajda *et al.*, “In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled Computer Vision,” *Proc. IEEE*, vol. 98, no. 3, pp. 462–478, Mar. 2010.
[20] D. M. Poolman, P., Frank, R. M., Luu, P., Pederson, S. M., and Tucker, “A single-trial analytic framework for eeg analysis and its application to target detection and classification.,” *NeuroImage, 42(2)787 – 798.*, 2008.
[21] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, “A fusion approach for image triage using single trial erp detection.,” in *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, 2007, p. pages 473 –476.
[22] A. R. Marathe *et al.*, “The effect of target and non-target similarity on neural classification performance: a boost from confidence,” *Front. Neurosci.*, vol. 9, no. August, pp. 1–11, 2015.
[23] H. Cecotti, J. Sato-Reinhold, J. L. Sy, J. C. Elliott, M. P. Eckstein, and B. Giesbrecht, “Impact of target probability on single-trial EEG target detection in a difficult rapid serial visual presentation task.,” in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 2011, 2011, pp. 6381–4.
[24] D.-O. Won, H.-J. Hwang, K.-R. Muller, and S.-W. Lee, “Shifting stimuli for brain computer interface based on rapid serial visual presentation,” in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*, 2017, pp. 40–41.
[25] C. Hutton, E. Featherstone, and O. Josephs, “Cogent 2000 v125 User Manual – 14 / 04 / 03,” pp. 1–51, 2000.
[26] D. Houcque, “Introduction to MATLAB for Engineering Students,” *Northwest. Univ. version*, no. August, 2005.
[27] “http://www.morguefile.com.” .
[28] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, “Brain state decoding for rapid image retrieval,” *Proc. seventeen ACM Int. Conf. Multimed. - MM '09*, p. 945, 2009.
[29] D. V Buonomano, J. Bramen, and M. Khodadadifar, “Influence of the interstimulus interval on temporal processing and learning : testing the state-dependent network model,” pp. 1865–1873, 2009.
[30] W. Geoffrey, “A brief introduction to the use of event-related in studies of perception and attention,” *Atten. Percept. Psychophys.*, vol. 72, no. 8, pp. 2031–2046, 2010.
[31] F. Lotte, M. Congedo, L. Anatole, F. Lamarche, and B. A. A, “A review of classification algorithms for EEG-based brain – computer interfaces To cite this version : A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces,” 2007.
[32] G. Ôquist, “Adaptive Rapid Serial Visual Presentation,” *Context*, no. December, 2001.
[33] Marvin M. Chun and M. C. Potter, “A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation,” *J. Exp. Psychol. Hum. Percept. Perform.*, 1995.
[34] K. I. Forster, “Visual perception of rapidly presented word sequences of varying complexity,” *Percept. Psychophys.*, vol. 8, no. 4, pp. 215–221, 1970.
[35] V. A. F. Lamme and P. R. Roelfsema, “The distinct modes of vision offered by feedforward and recurrent processing,” *Trends Neurosci.*, vol. 23, no. 11, pp. 571–579, Nov. 2000.
[36] J. F. Maguire and P. D. L. Howe, “Failure to detect meaning in RSVP at 27 ms per picture,” *Attention, Perception, Psychophys.*, no. April, pp. 1405–1413, 2016.
[37] M. C. Potter, B. Wyble, C. E. Hagmann, and E. S. McCourt, “Detecting meaning in RSVP at 13 ms per picture.,” *Atten. Percept. Psychophys.*, vol. 76, no. 2, pp. 270–9, 2014.
[38] J. E. Raymond, K. L. Shapiro, and K. M. Arnell, “Temporary suppression of visual processing in an RSVP task: An attentional blink?,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 18, pp. 849–860, 1992.
[39] C. Kranczioch, S. Debener, and A. K. Engel, “Event-related potential correlates of the attentional blink phenomenon,” *Cogn. Brain Res.*, vol. 17, no. 1, pp. 177–187, Jun. 2003.
[40] G. McCarthy and E. Donchin, “A metric for thought: a comparison of P300 latency and reaction time.,” *Science*, vol. 211, no. 4477, pp. 77–80, 1981.
[41] S. Luck, G. Woodman, and E. Vogel, “Event-related potential studies of attention.,” *Trends Cogn. Sci.*, vol. 4, no. 11, pp. 432–440, 2000.
[42] J. Touryan, L. Gibson, J. H. Horne, and P. Weber, “Real-time measurement of face recognition in rapid serial visual presentation,” *Front. Psychol.*, vol. 2, no. March, pp. 1–8, 2011.
[43] B. Cai, S. Xiao, L. Jiang, Y. Wang, and X. Zheng, “A rapid face recognition BCI system using single-trial ERP,” in *In Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, 2013, p. (pp. 89-92).
[44] E. M. Fine and E. Peli, “Scrolled and rapid serial visual presentation texts are read at similar rates by the visually impaired.,” *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.*, vol. 12, no. 10, pp. 2286–92, Oct. 1995.
[45] R. Ward, J. Duncan, and K. Shapiro, “Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention.,” *Percept. Psychophys.*, vol. 59, no. 4, pp. 593–600, May 1997.
[46] J. Swezller, “Cognitive Load Theory, Learning Difficulty, and Instructional Design,” vol. 4, pp. 295–312, 1994.
[47] H. K. Gomarus, M. Althaus, A. A. Wijers, and R. B. Minderaa, “The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task : An ERP and ERD / ERS study,” vol. 117, pp. 871–884, 2006.
[48] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, “Efficient mental workload estimation using task-independent EEG features,” vol. 13, pp. 1–10, 2016.
[49] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,” *Adv. Psychol.*, vol. 52, no. C, pp. 139–183, 1988.
[50] C. Mühl, C. Jeunet, F. Lotte, and M. A. Hogervorst, “EEG-based workload estimation across affective contexts,” vol. 8, no. June, pp. 1–15, 2014.
[51] J. Shum *et al.*, “A Brain Area for Visual Numerals,” *J. Neurosci.*, 2013.
[52] S. H. Patel and P. N. Azzam, “Characterization of N200 and P300: Selected studies of the Event-Related Potential,” *International Journal of Medical Sciences*. 2005.
[53] and B. Y. Lin Zhimin, Ying Zeng, Hui Gao, Li Tong, Chi Zhang, Xiaojuan Wang, Qunjian Wu, “Multi-Rapid Serial Visual Presentation Framework for EEG-based Target Detection,” *Biomed Res. Int.*, 2017.
[54] J. Polich, “Updating P300: An Integrative Theory of P3a and P3b,” *Clin Neurophysiol*, vol. 118, no. 10, pp. 2128–2148, 2007.
[55] L. Gibson, J. Touryan, A. Ries, K. Mcdowell, H. Cecotti, and B. Giesbrecht, “Adaptive Integration and Optimization of Automated and Neural Processing Systems ― Establishing Neural and Behavioral Benchmarks of Optimized Performance,” no. July, 2012.