

ACCEPTED MANUSCRIPT

## A review of rapid serial visual presentation-based brain-computer interfaces

To cite this article before publication: Stephanie Lees *et al* 2017 *J. Neural Eng.* in press <https://doi.org/10.1088/1741-2552/aa9817>

### Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2017 IOP Publishing Ltd.

During the embargo period (the 12 month period from the publication of the Version of Record of this article), the Accepted Manuscript is fully protected by copyright and cannot be reused or reposted elsewhere.

As the Version of Record of this article is going to be / has been published on a subscription basis, this Accepted Manuscript is available for reuse under a CC BY-NC-ND 3.0 licence after the 12 month embargo period.

After the embargo period, everyone is permitted to use copy and redistribute this article for non-commercial purposes only, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by-nc-nd/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions will likely be required. All third party content is fully copyright protected, unless specifically stated otherwise in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

# 1 A Review of Rapid Serial Visual Presentation-based Brain- 2 Computer Interfaces

3 **Stephanie Lees<sup>1</sup>, Natalie Dayan<sup>1</sup>, Hubert Cecotti<sup>1</sup>, Paul McCullagh<sup>1</sup>, Liam Maguire<sup>1</sup>, Fabien  
4 Lotte<sup>2</sup> and Damien Coyle<sup>1</sup>**

5  
6 <sup>1</sup> Faculty of Computing and Engineering, Ulster University, United Kingdom

7 <sup>2</sup> Inria Bordeaux Sud-Ouest/LaBRI/CNRS/Université de Bordeaux/IPB, Talence, France, 200  
8 Avenue de la Vieille Tour, 33405 Talence, France

9  
10 Email: dh.coyle@ulster.ac.uk

---

## 12 **Abstract**

13 Rapid serial visual presentation (RSVP) combined with the detection of event related brain responses  
14 facilitates the selection of relevant information contained in a stream of images presented rapidly to a human.  
15 Event related potentials (ERPs) measured non-invasively with electroencephalography (EEG) can be  
16 associated with infrequent targets amongst a stream of images. Human-machine symbiosis may be augmented  
17 by enabling human interaction with a computer, without overt movement, and/or enable optimization of  
18 image/information sorting processes involving humans. Features of the human visual system impact on the  
19 success of the RSVP paradigm, but pre-attentive processing supports the identification of target information  
20 post presentation of the information by assessing the co-occurrence or time-locked EEG potentials. This paper  
21 presents a comprehensive review and evaluation of the limited but significant literature on research in RSVP-  
22 based brain-computer interfaces (BCIs). Applications that use RSVP-based BCIs are categorized based on  
23 display mode and protocol design, whilst a range of factors influencing ERP evocation and detection are  
24 analyzed. Guidelines for using the RSVP-based BCI paradigms are recommended, with a view to further  
25 standardizing methods and enhancing the inter-reliability of experimental design to support future research  
26 and the use of RSVP-based BCIs in practice.

27  
28  
29 *Keywords— Rapid Serial Visual Presentation; Brain-Computer Interface; Event Related Potentials; Electroencephalography*

---

## 33 **1. Introduction**

34 Rapid Serial Visual Presentation (RSVP) is the process of sequentially displaying images at the  
35 same spatial location at high presentation rates with multiple images per second e.g., with a stimulus  
36 onset asynchrony no greater than 500ms but often lower than 100ms i.e., >10 stimuli presented per  
37 second. Brain-computer interfaces (BCI) are communication and control systems that enable a user  
38 to execute a task via the electrical activity of the user's brain alone (Vidal, 1973). RSVP-based BCIs  
39 are a specific type of BCI that is used to detect target stimuli, e.g. letters or images, presented  
40 sequentially in a stream, by detecting brain responses to such targets. RSVP-based BCIs are  
41 considered as a viable approach to enhance human-machine symbiosis and offers potential for  
42 human enhancement.

1 To date, the literature on RSVP-BCIs has not been comprehensively evaluated therefore it is timely  
2 to review the literature and provide guidelines for others considering research in this area. In this  
3 review we; 1) identify and contextualize key parameters of different RSVP-BCI applications to aid  
4 research development; 2) document the growth of RSVP-based BCI research; 3) provide an  
5 overview of key current advancements and challenges; 4) provide design recommendations for  
6 researchers interested in further developing the RSVP-BCI paradigm.

7  
8 This paper is organized as follows; Section 2", presents background information on the fundamental  
9 operating protocol of RSVP-BCIs. Section 3 details results of a bibliometric analysis of key terms  
10 "Rapid serial visual presentation", "RSVP", "Electroencephalography", "EEG", "Brain-Computer  
11 Interface", "BCI", "Event Related Potentials", "ERP and "Oddball" found within authoritative  
12 bibliographic resources. Section 4 provides an overview of performance measures. Section 5  
13 outlines existing RSVP-based BCI applications, presenting inter-application study comparisons and  
14 undertakes an analysis of the design parameters with inter-application study comparisons. Section 6  
15 provides a summary, discussion of findings and ongoing challenges.

## 16 2. Background

17 RSVP-based BCIs have been used to detect and recognize objects, scenes, people, pieces of relevant  
18 information and events in static images and videos. Many applications would benefit from an  
19 optimization of this paradigm, for instance counter intelligence, policing and health care, where  
20 large numbers of images/information are reviewed by professionals on a daily basis. Computers are  
21 unable to analyze and understand imagery as successfully as humans and manual analysis tools are  
22 slow (Mathan *et al.*, 2008; Gerson, Parra and Sajda, 2005). In studies carried out by Sajda *et al.*  
23 (2010), Poolman *et al.* (2008) and Bigdely-Shamlo *et al.* (2008), a trend of using RSVP-based BCIs  
24 for identifying targets within different image types has emerged. Research studies show the ability  
25 to use RSVP-based BCIs to drive a variety of visual search tasks including, in some circumstances,  
26 skills learned for visual recognition. Although the combination of RSVP and BCI has proven  
27 successful on several image sets, other research has attempted to establish whether or not greater  
28 efficiencies can be reached through the combination of RSVP-based BCIs and behavioural  
29 responses (Huang *et al.*, 2007).

### 30 2.1. Event Related Potentials and their use in RSVP-based BCIs

31  
32 Event-related potentials (ERPs) are EEG signals amplitude variations in the electroencephalogram  
33 (EEG) associated with the onset of a stimulus (usually auditory or visual) presented to a person.  
34 ERPs are typically smaller in amplitude ( $<10\mu\text{V}$ ) in comparison to the ongoing EEG activity ( $\sim 50$ -  
35  $100\mu\text{V}$ ) they are embedded within (Huang *et al.*, 2008; Acqualagna and Blankertz, 2011). As ERPs  
36 are locked in phase and time to specific events, they can be measured by averaging epochs over  
37 repeated trials (Huang *et al.*, 2011; Cecotti, Eckstein and Giesbrecht, 2012, 2014). Shared EEG  
38 signal features are accentuated and noise attenuated (Luck, 2005; M. X. Cohen, 2014). The outcome  
39 is represented by a temporal waveform with a sequence of positive and negative voltage deflections  
40 labeled as ERP components. ERPs are representative of summated cortical neural processing and  
41 behavioral counterparts, such as attentional orientation (Wolpaw and Wolpaw, 2012; M. X. Cohen,  
42 2014).

43  
44 The stream of images presented within a RSVP paradigm comprise frequent non-target images and  
45 infrequent target images; different ERP components are associated with target and non-target

1 stimuli (Bigdely-Shamlo *et al.*, 2008; M. Cohen, 2014; Sadjja *et al.*, 2014). BCI signal processing  
2 algorithms are used to recognise spatio-temporal electrophysiological responses and link them to  
3 target image identification, ideally on a single trial basis (Manor, Mishali and Geva, 2016).  
4

5 The most commonly exploited ERP in RSVP-based BCI applications is the P300. The P300 appears  
6 at approximately 250-750 ms post target stimulus (Polich and Donchin, 1988; Leutgeb, Schäfer and  
7 Schienle, 2009; Ming *et al.*, 2010; Zhang *et al.*, 2012). As specified by (Polich and Donchin, 1988)  
8 during the P300 experiment (commonly referred to as the 'Oddball' paradigm), participants must  
9 classify a series of stimuli which fall into one of two classes: targets and non-targets. Targets appear  
10 more infrequently than non-targets (typically ~5-10% of total stimuli in the RSVP paradigm) and  
11 should be recognizably different. It is known that P300 responses can be suppressed in an RSVP  
12 task if the time between two targets is <0.5 seconds; which is known as attentional blink  
13 (Raymond, Shapiro and Arnell, 1992; Kranczioch, Debener and Engel, 2003). The amplitude and  
14 the latency of the P300 are influenced by the target discriminability and the target-to-target interval  
15 in the sequence. The latency of the P300 is affected by stimulus complexity (McCarthy and  
16 Donchin, 1981; Luck, Woodman and Vogel, 2000). The P300 amplitude can vary as a result of  
17 multiple factors (Johnson, 1986), such as:  
18

- 19 • Subjective Probability – the expectedness of an event.
- 20 • Stimulus Meaning – comprised of: task complexity, stimulus complexity and stimulus  
21 value.
- 22 • Information Transmission – the amount of stimulus information a participant registers in  
23 relation to the information contained within a stimulus.  
24

## 25 2.2. RSVP-based BCI amongst the BCI Classes

26 BCI classes can be of three different types: active, reactive or passive (Zander *et al.*, 2010). An active BCI  
27 is purposefully controlled by the user through intentional modulation of neural activity, often  
28 independent of external events. Contrastingly, reactive BCIs generate outputs from neural activity  
29 evoked in response to external events, enabling indirect control by the user. Passive BCI makes use  
30 of implicit information and generate outputs from neural activity without purposeful control by the  
31 user. Active/reactive BCIs are commonly aimed at users with restricted movement abilities who  
32 intentionally try to control brain activity, whereas implicit or passive BCIs are more commonly  
33 targeted towards applications that are also of interest to able-bodied users (Zander and Kothe, 2011;  
34 Sasane and Schwabe, 2012).  
35  
36  
37

## 38 2.2. RSVP-based BCI Presentation Modes

39 RSVP-based BCIs have two presentation modes: static mode in which images appear and disappear  
40 without moving; and moving mode where targets within short moving clips have to be identified  
41 (Sajda *et al.*, 2010; Cecotti, Eckstein and Giesbrecht, 2012; Weiden, Khosla and Keegan, 2012).  
42 Both presentation modes can be used with or without a button press. With a button press, users  
43 indicate manually, by pressing a button, when they observe a target stimulus. A button press is used  
44 to establish baseline performance, reaction time and/or to enhance performance (discussed further in  
45 section 5.1).  
46

### 1 2.2.1. Static

2 In ‘static mode’, images displayed have identical entry and exit points; - the images are transiently  
3 presented on screen (typically for 100-500 ms) and then disappear. One benefit of static mode is that  
4 images occupy the majority of the display and therefore, identification of targets is likely even if  
5 they are only presented briefly. There are a number of different possible instructions a participant  
6 may be given:

- 7 • Prior to presentation, a target image may be shown to participants and participants are asked to  
8 identify this image in a sequence of proceeding images. Target recognition success rates can be  
9 achieved with presentation rates as high as 10/second (Cecotti, Eckstein and Giesbrecht, 2012).
- 10 • Participants may be asked to identify a *type of target* e.g., an animal within a collection of  
11 images. In this mode, the rate of presentation should be slowed down (4/second) (Wang *et al.*,  
12 2009).
- 13 • Immediately after image sequence presentation, the participant may be shown an image and  
14 asked: “did this image appear in the sequence you have just seen?” (Potter *et al.*, 2002).

### 15 2.2.2. Moving

16 There has been relatively little research regarding neural signatures of a target and/or anomalies in  
17 real world or simulated videos. In ‘moving mode’, short video clips are shown to participants, and  
18 within one video clip participants may be asked to identify one or more targets. It is important that  
19 these targets are temporally ‘spread out’ to avoid P300 suppression. There are different possible  
20 instructions a participant may be given:

- 21 • Prior to presentation, participants may be given a description of a target i.e., asked to  
22 identify, say a “person” or “vehicle” in a moving scene (Weiden, Khosla and Keegan,  
23 2012).
- 24 • Participants can be asked to identify a target event; in this case, the target is identified  
25 across space and time. The participant is required to integrate features from both motion and  
26 form to decide whether a behavior constitutes a target, for example, (Rosenthal *et al.*, 2014)  
27 defined the target as a person leaving a suspicious package in a train station.

### 28 2.3. Cognitive blindness

29 When designing an RSVP-based BCI, three different types of cognitive blindness should be  
30 considered namely, the attentional blink, change blindness and saccadic blindness. Generally,  
31 RSVP is a paradigm used to study the *attentional blink*, which is a phenomena that occurs when a  
32 participant’s attention is grabbed by an initial target image and a further target image may not be  
33 detectable for up to 500 ms after the first (Raymond, Shapiro and Arnell, 1992). Depending upon  
34 the duration of stimuli presentation the ration of target images/total images will change (e.g. if  
35 images are being presented at a duration of 100ms then there must be a minimum of 5 images  
36 between targets 1 and 2. In a sequence of 100 images there can be a maximum of 20 target images.  
37 Whereas if images are presented at 200ms this limits the maximum number of targets to 10/100  
38 images in total).

39 *Change blindness* occurs when a participant is viewing two images that vary in a non-trivial fashion,  
40 and has to identify the image differences. Change blindness can occur when confronted by images,  
41 motion pictures, and real world interactions. Humans have the capacity to get the gist of a scene  
42 quickly but are unable to identify particular within-scene features (Simons and Levin, 1997; Oliva,  
43  
44

1 2005). For example, when two images are presented for 100 ms each and participants are required to  
2 identify a non-trivial variation as the images are interchangeably presented, participants can take  
3 between 10-20s to identify the variation. This latency period in identifying non-trivial variations in  
4 imagery can be augmented through use of distractors or motion pictures (Rensink, 2000). In the  
5 context of designing an RSVP paradigm change blindness is of interest, as it will take longer for a  
6 user to identify a target within an image if it does not pop out from the rest of the image. Distractors  
7 within the image or cluttered images, will increase the time it takes a user to recognize a target,  
8 reducing the performance of the RSVP paradigm.

9 *Saccadic blindness* is a form of change blindness described by Chahine and Krekelberg (2009)  
10 where “humans move their eyes about three times each second. Those rapid eye movements called  
11 saccades help to increase our perceptual resolution by placing different parts of the world on the  
12 high-resolution fovea. As these eye movements are performed, the image is swept across the retina,  
13 yet we perceive a stable world with no apparent blurring or motion”. Saccadic blindness thus refers  
14 to the loss of image when a person saccades between two locations. Evidence shows that saccadic  
15 blindness can occur 50 ms before saccades and up to 50 ms after saccades (Diamond, Ross and  
16 Morrone, 2000). Thus, it is important that stimuli have a duration greater than 50 ms to bypass  
17 saccadic blindness, unless participants are instructed to attend a focus point and the task is gaze  
18 independent and thus does not demand saccades (such as during the canonical RSVP paradigm  
19 (section 5.4) ).

20 Having considered some of the factors influencing RSVP-based BCI designs, the remainder of the  
21 paper focuses on a bibliometric study of the RSVP literature highlighting the key methodological  
22 parameters and study trends. Studies are compared and contrasted on an intra- and inter-application  
23 basis. Later sections focus on study design parameters and provide contextualized recommendations  
24 for researchers in the field.

### 25 26 **3. Bibliometric study of the RSVP related literature**

27 A bibliometric review of the RSVP-based BCIs was conducted. The inclusion criteria for this  
28 review were studies that focused on EEG data being recorded while users were performing visual  
29 search tasks using an RSVP paradigm. The studies involved various stimulus types presented using  
30 the RSVP paradigm where participants had to identify target stimuli. All reported studies were not  
31 simply theoretical and had at least one participant. One or more of the keywords BCI, RSVP, EEG  
32 or ERP appeared in the title, abstract or keyword list. Only papers published in English were  
33 included. The literature was searched, evaluated and categorized up until August 2017. The  
34 databases searched were Web of Science, IEEE, Scopus, Google Scholar, and PubMed. The search  
35 terms used were: “Rapid serial visual presentation”, “RSVP”, “Electroencephalography”, “EEG”,  
36 “Brain-Computer Interface”, “BCI”, “Event Related Potentials”, “ERP and “Oddball”

37 Papers were excluded for the following reasons: 1. the research protocol had insufficient detail; 2.  
38 key aspects needed to draw conclusive results were missing; 3. the spectrum of BCI research  
39 reported was too wide (i.e. review papers not specific to RSVP), 4. A ‘possible’ research application  
40 was described but the study was not actually carried out; 5. The study was a repeated study by  
41 original authors with only minor changes. Due to the immaturity of RSVP-based BCI as research  
42 topic, conference papers were not excluded. Inclusion of conference papers was considered  
43 important in order to provide a comprehensive overview of the state-of-the-art and trends in the

1 field. Fifty-four papers passed initial abstract/title screening, these were then refined to the 45 most  
2 relevant papers through analysis of the entire paper contents. The date of the included publications  
3 ranged from 2003-2017.  
4

5 The relevant RSVP-based BCI papers are presented in Table 1 when a button press was required,  
6 and Table 2 when no button presses were conducted. RSVP based BCIs were evaluated in terms of  
7 the interface design. Table 1 and Table 2 show that there is considerable variation across the  
8 different studies in terms of the RSVP-BCI acquisition paradigm, including the total number of  
9 stimuli employed, percentage of target stimuli, size of on-screen stimuli, visual angle, stimulus  
10 presentation duration, and the number of study participants. Performance was measured using a  
11 number of metrics: the area under the Receiver Operating Characteristic (ROC) curve (Fawcett,  
12 2006), classification accuracy (%) and information transfer rate. ROC curves are used when  
13 applications have an unbalanced class distribution, which is typically the case with RSVP-BCI,  
14 where the number of target stimulus is much smaller than that of non-target stimuli. Many studies  
15 report different experimental parameters and some aspects of the studies have not been  
16 comprehensively reported. From Tables 1 and 2, it can be seen that the majority of applications  
17 using a button press as a baseline may be classified as surveillance applications while applications  
18 that do not use a button press are more varied. This may be because often surveillance applications  
19 have an industry focus, and quantified improvement relative to manual labelling alone is crucial for  
20 acceptance. In the majority of the applications where a button press was used, participants undertake  
21 trials with and without a button press and the difference in latency of response between the two is  
22 calculated to compare neural and behavioral response times. The results of the bibliometric analysis  
23 are further discussed in section 4, 5 and 6, following the analysis of key papers identified in the  
24 following section.  
25  
26

**Table 1. Design Parameters reviewed, Mode: Button press = Yes.** Table acronyms: SVM (Support Vector Machine), SFFS (Sequential Forward Feature Selection), N/A (Not available), BLDA (Bayesian Linear Discriminant Analysis), BCSP (Bilinear Common Spatial Pattern), CCSP (Composite CSP), CSP (Common Spatial Pattern), LDA (Linear Discriminant Analysis), C (EEG channel), FDA (Fisher Discriminant analysis), FDM (finite difference model), LLC (Linear Logistic Classifier), RBF SVM (Radial Basis Function SVM), PCA (Principle Component Analysis), LP (Laplacian classifier), LN (Linear Logistic regression), SP (Spectral Maximum Mutual Information Projection), FDA (Fisher Discriminant analysis), ACSP (Analytic CSP), HT (Human target), NHT (Non-human target), ST(Single Trial), DT (Dual Trial), BDA (Bilinear Discriminant Analysis), ABDA (Analytic BDA), DCA (Directed Components Analysis), HDCA (Hierarchical Discriminant Component Analysis), TO (Only background distractors), TN (Non-Target distractor stimuli & Background & Target stimuli), TvB (Target vs Background distractor), T v[B+NT] (Target vs both Background distractor and Non-Target).

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)	
1	(Healy and Smeaton, 2011)	Static	Categorization	4800	1.25	100	N/A	N/A	8	SVM linear kernel, SFFS	N/A	N/A	
2	(Cecotti, Sato-Reinhold, <i>et al.</i> , 2011)	Static	Categorization/face recognition	12000 trials	5 10 25 50	500	N/A	N/A	8	XDAWN + BLDA	0.768±0.074 0.821±0.063 0.815±0.068 0.789±0.070	78.7 76.4 77.0 71.5	
3	(Yu <i>et al.</i> , 2011)	Static	Categorization	>4000	~1.5	150	N/A	N/A	20	BCSP, SVM CCSP, SVM CSP, SVM	N/A	83.0±8.0 75.4±8.3 71.8±9.9	
4	(Ušćumlić, Chavarriaga and Millán, 2013)	Static	Categorization	1382	10	Eagles Tiger Train	250	N/A	Images occupy ~6*4 visual field	15	Gaussian EnsembleLDA(8C) EnsembleLDA(41C) Gaussian EnsembleLDA(8C) EnsembleLDA(41C) Gaussian EnsembleLDA(8C) EnsembleLDA(41C)	0.66 0.78 0.80 0.75 0.80 0.91 0.65 0.68 0.73	90.0 94.8 90.1
5	(Mohedano <i>et al.</i> , 2015)	Static	Categorization	3000	5	100 or 200	N/A	N/A	8	SVM	0.564 – 0.863	N/A	
6	(Acqualagnav <i>et al.</i> , 2010)	Static	RSVP Speller	30	User dependent	83 or 133	N/A	1	9	LDA	N/A	~70 ~85-90	



	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
7	(Touryan <i>et al.</i> , 2011)	Static	Face recognition	470-480	N/A	500	256*320	7 horizontally 9 vertically	22	PCA	0.868-0.991	60.4-92.0
8	(Sajda, Gerson and Parra, 2003)	Static	Surveillance	330	50	200 100 50	768*512	12.4 by 15.3	2	Spatial linear discriminator	0.79-0.96 0.74-0.80 0.84-0.79	N/A
9	(Gerson, Parra and Sajda, 2006)	Static	Surveillance	284	2	100	640* 426	33±3 * 25±3	5	Spatial linear discriminator	N/A	74-96
10	(Erdogmus, Mathan and Pavel, 2006)	Static	Surveillance	N/A	50	100 50	N/A	N/A	1	LP LN	0.90/0.95 (100/50ms) 0.37/0.66 (100/50ms)	N/A
									2	LP LN SP	0.87-0.83 0.87-0.82 0.89-0.86	
11	(Bigdely-Shamlo <i>et al.</i> , 2008)	Static	Surveillance	24394	40-60	~83	N/A	1.6 by 1.6	7	Bayes fusion of FDA	0.78-0.95	N/A
12	(Poolman, Frank, <i>et al.</i> , 2008)	Static	Surveillance	8300	4 or 1	100	500*500	2	3	DCA FDM	0.70-0.82	72-84
13	(Huang <i>et al.</i> , 2011)	Static	Surveillance	N/A	~1	60-150	500*500	22*22	33	RBF SVM Linear SVM LLC	0.848-0.941 0.846-0.927 0.753-0.834	N/A

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
									4	RBF SVM Linear SVM LLC	0.909-0.961 0.887-0.944 0.625-0.866	N/A
14	(Weiden, Khosla and Keegan, 2012)	Static/moving	Surveillance	2500	2	234	512*512	N/A	8	SVM	0.50-0.78 (static) 0.89-1.00 (video)	42 (static) 97 (video)
		Moving		7500	6 10 14				7		0.72-0.94 (video) 0.58-0.94 (video) 0.55-0.91 (video)	N/A
15	(Cecotti <i>et al.</i> , 2012)	Static/moving	Surveillance	30000	10	100	N/A	N/A	15	XDAWN, BLDA	~0.874-0.931 (static HT) ~0.675-0.937 (video NHT) ~0.875- 0.926 (video HT)	N/A
16	(Cecotti, Eckstein and Giesbrecht, 2012)	Static	Surveillance	300	10	200	683*384	≈ 13	10	XDAWN, BLDA	0.837 (ST) 0.838 (DT)	N/AN/A
17	(Yu <i>et al.</i> , 2014)	Static	Surveillance	> 4472	~1.61.6	150	400*400	N/A	22	CSP ACSP BDA ABDA	N/A	83.8±6 85.8±5 87.2±4 89.7±5
18	(Marathe, Ries and McDowell, 2014)	Moving	Surveillance	N/A	10	200	N/A	N/A	15	HDCA Sliding HDCA	0.8691 ± 0.0359 0.9494±0.9610	N/A
19	(Marathe <i>et al.</i> , 2015a)	Static	Surveillance	N/A	5 6	500	960*600	36.3 × 22.5	17	XDAWN, BLDA	~0.984 (TvB, TO) ~0.971 (TvB, TN) ~0.959 (Tv[B+NT], TN)	N/A

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
20	(Files and Marathe, 2016)	Static/moving	Surveillance	N/A	10	100	N/A	N/A	15	Linear classifiers	N/A	78.4-90.5
21	(Barngrover <i>et al.</i> , 2016)	Static	Surveillance	4384	4	200	100*50	N/A	19	SVM with Haar-like feature classifier	N/A	>70
22	(Marathe <i>et al.</i> , 2015b)	Static/moving	Intelligence	N/A	10	100 or 500	N/A	N/A	15	HDCA CSP XDAWN, BLDA	>0.9	>70

**Table 2. Design Parameters reviewed, Mode: Button press = No.** Table acronyms: FDA (Fisher Discriminant analysis), N/A (Not available), SWFP (Spatially Weighted Fisher Linear Discriminant – Principal Component Analysis), CNN (Convolutional Neural Network), HDPCA (Hierarchical Discriminant Principal Component Analysis Algorithm), HDCA (Hierarchical Discriminant Component Analysis), SVM (Support Vector Machine), RBF (Radial Basis Function) kernel, RDA (Regularized Discriminant Analysis), HMM (Hidden Markov Model), PCA (Principle Component Analysis), BDCA (Bilinear Discriminant Component Analysis), BFBD (Bilinear Feature Based Discriminants), BLDA (Bayesian Linear Discriminant Analysis), SWLDA (Step-wise Linear Discriminant Analysis), MLP (Multilayer Perceptron), LIS (Locked in syndrome), CV (Computer Vision), STIG (Spectral Transfer with Information Geometry), MSS (Max Subject-Specific Classifier), L1 ( $\ell_1$ -Regularized Cross-Validation), MV (Majority Vote), PMDRM (Pooled Riemannian Mean classification algorithm), AWE (Accuracy Weighted Ensemble), MT (Multi-Task Learning), CALIB (Within-Subject Calibration), RF (Random forest), BHCR (Bayesian Human Vision-Computer Vision Retrieval).

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
1	(Hope <i>et al.</i> , 2013)	Static	Medical	166	~1.1	100	189*189	N/A	2	FDA	0.75-0.78	N/A
2	(Galit <i>et al.</i> , 2014)	Static	Categorization	725	20	90-110	360*360	6.5 × 6.5	12	SWFP HDPCA HDCA	0.64-0.85 N/A N/A	66-82 66-81 57-70
				290						SWFP HDPCA	0.58-0.99 /0.99±0.55 0.99±0.67	91 N/A

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
										HDCA	0.87±0.05	N/A
3	(Mohedano <i>et al.</i> , 2014)	Static	Categorization	4224	15	200	N/A	N/A	5	SVM RBF	0.63-0.78	N/A
4	(Manor and Geva, 2015)	Static	Categorization	N/A	20	90-110	360*360	6.5 × 6.5	15	SWFP Deep CNN	0.652-0.850 0.692-0.858	70.0-83.1 66.2-82.5
5	(Huang <i>et al.</i> , 2017)	Static	Categorization	N/A	12.5	200	N/A	N/A	7	LDA + RF BHCR	0.873 0.987	N/A
6	(Orhan <i>et al.</i> , 2011a)	Static	RSVP Speller	26	~3.8	150	N/A	N/A	2	RDA	0.948-0.973	N/A
7	(Hild <i>et al.</i> , 2011)	Static	RSVP Speller	26	~3.6 (User dependent)	400	N/A	N/A	2 (1 LIS)	RDA	N/A	N/A
8	(Orhan <i>et al.</i> , 2012b)	Static	RSVP Speller	26	~3.8	150	N/A	N/A	2	RDA HMM	N/A	N/A
9	(Orhan <i>et al.</i> , 2012c)	Static	RSVP Speller	28	~3.6 (User dependent)	400 or 150	N/A	N/A	3 (1 LIS)	RDA PCA	N/A	Healthy controls=95 LIS=85
10	(Chennu <i>et al.</i> , 2013)	Static	RSVP Speller	25	4	133	N/A	N/A	11	SWLDA	0.82	86.02
			Matrix P300 Speller	25	4						0.84	88.58
11	(Orhan <i>et al.</i> , 2013d)	Static	RSVP Speller	28	~3.8	150	N/A	N/A	2	PCA RDA	0.812-0.998	N/A
12	(Oken <i>et al.</i> , 2014)	Static	RSVP Speller	28	~3.6 (Semi-user dependent)	400	N/A	3.8	15 (6 LIS)	PCA RDA	Healthy controls= 0.81-0.86 LIS= 0.73- 0.92	N/A
13	(Won <i>et al.</i> , 2017)	Static/ Moving	RSVP speller	36	N/A	N/A	N/A	near-central	8	Regularized LDA	N/A	88.9

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
14	(Cai <i>et al.</i> , 2013)	Static	Face recognition	160	6.25	500	400*400	N/A	8	SVM	0.802-0.921	90.3
15	(Sajda <i>et al.</i> , 2010)	Static	Surveillance	250	20	150	500*500	N/A	5	HDCA BDCA BFBD	N/A	0.76±0.07 0.83±0.91 0.91±0.07
16	(Rosenthal <i>et al.</i> , 2014)	Moving	Surveillance	250* 30s clips	30-50	5 times real-time	N/A	N/A	8	HDCA	>0.8	N/A
17	(Matran-Fernandez and Poli, 2014)	Static	Surveillance	2400	10	~83-200 (5, 6, 10,12 Hz)	640*640	59 Left Visual Field (LVF) target pictures and 85 Right Visual Field (RVF) target pictures	9	SVM	0.78 0.77 0.8 0.67	N/A
18	(Cecotti, Eckstein and Giesbrecht, 2014)	Static	Categorization	12,000	10	500	256*256	≈4.57	8	MLP BLDA Linear SVM	0.861±0.73 0.841±0.66 0.806±0.127	N/A
			Surveillance	900		100	683*384	≈26	10	XDAWN, MLP XDAWN, BLDA XDAWN, Linear SVM	0.845±0.63 0.850±0.61 0.847±0.63	
			Surveillance	4000		200	683*384	≈26	10	XDAWN, MLP XDAWN, BLDA XDAWN, Linear SVM	0.816±0.52 0.824±0.53 0.819±0.55	
19	(Manor, Mishali and Geva, 2016)	Static	Surveillance	N/A	~10	100 or 200	400*400	N/A	2	Supervised multimodal network  Semi-supervised multimodal network	N/A	88.1-93.9  81.4-90.3

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
20	(Waytowich <i>et al.</i> , 2016)	Moving/static	Surveillance	N/A	~11	100	N/A		32	Offline	STIG MSS LI MV PMDRM  STIG AWE MT CALIB	N/A
									17	Real-time feedback	STIG MSS LI MV PMDRM  STIG AWE MT CALIB	
21	(Yazdani <i>et al.</i> , 2010)	Static	Other	52	~2	500	N/A	N/A	5	SVM with radial basis function kernel	N/A	35±10.4 – 71.1±9.0 (F-measure range)
22	(Huang <i>et al.</i> , 2017)	Static	Categorization	96	12.5	200	N/A	N/A	7	Adaboost Bagging ANN RF SVM LR	0.873 0.987	0.887
23	(Lin Zhimin, Ying Zeng, Hui Gao, Li Tong, Chi Zhang, Xiaojuan Wang, Qunjian Wu, 2017)		Categorization	2000	10	250	N/A	N/A	7 8	SWLDA HDCA	0.7837-0.9148 0.9082-0.9522	N/A

#### 4. Validating inter-study comparison through performance measures

When comparing RSVP-studies it is important to acknowledge that researchers use different measures of performance. Before going into depth about signal processing techniques (section 5.7) it is important to discuss, firstly, the variations in approaches used to measure performance. To encourage valid inter-study comparison within and across RSVP application types, it is crucial to emphasize that we are, on the whole, reporting classification accuracy when it is calculated in terms of the number of correctly classified trials. Classification accuracy can be swayed by the imbalanced target and non-target classes, with targets being infrequently presented e.g. with a 10% target prevalence, if all trials are classed as non-targets, correct classification rate would be 90%. Hence, ROC values are also reported in this review where relevant information was provided in publications reviewed.

In the literature, there are many variations on how performance is estimated and reported. The studies cited in the current section provide examples of performance measure variations from the literature. The intention of Files and Marathe (2016a) was to develop a regression-based method to predict hit rates and error rates whilst correcting for expected mistakes. There is a need for such methods, due to uncertainty and difficulty in correctly identifying target stimuli. The regression method developed by Files and Marathe, (2016a), had relatively high hit rates which spanned 78.4% to 90.5% across all participants. Contrastingly, as a measure of accuracy, Sajda *et al.* (2010) used hit rates expressed as a fraction of total targets detected per minute. Sajda *et al.* (2010) discuss an additional experiment that employed ROC values as an outcome measure. In Fuhrmann *et al.* (2014), where the RSVP application was categorization based, accuracy was defined as, the number of trials in which the classifier provided the correct response, divided by the total number of available trials, with regards to target/non-target classification. Yazdani *et al.* (2010) were concerned with surveillance applications of RSVP-based BCI and used the F-measure to evaluate the accuracy of the binary classifier in use. Precision (fraction of occurrences flagged that are of relevant) and recall (fraction of relevant occurrences flagged) were reported as the F-measure considers both these values.

Different variations in ROC value calculations were also discovered across the studies evaluated. Variability in the distribution of accuracy outcome measures is also founded upon whether the dataset is non-parametric e.g. median AUC is reported as opposed to the mean AUC (Matran-Fernandez and Poli, 2014). As a measure of accuracy, Rosenthal *et al.* (2014) conducted a bootstrap analysis, to show the sampled distribution of AUC values for HDCA classifiers where 1000 times over, labels were randomized, classifiers were trained and AUC values calculated through a “leaving one-out cross-validation” technique. Cecotti *et al.* (2012) presented a comparison of three class classifiers in a ‘one versus all’ strategy. The focus of Cecotti *et al.* (2012) was to compare the AUC to the volume under the ROC hyper-surface and the authors found a AUC of 0.878, which is suggestive of the possibility for discrimination between greater than two types of ERPs using single-trial detection. Huang *et al.* (2006) reported the AUC for session one of two experiments during button press trials. This paper demonstrates that with the three classifiers approach produces similar performance with AUC of >0.8 across the board (Huang *et al.*, 2006). Moreover, accuracy reportedly increases through collating evidence from two BCI users, and reportedly yielded a 7.7% increase in AUC compared to a single BCI user (Matran-Fernandez and Poli, 2014), using collaborative BCIs. This process was repeated 20 times to achieve an average accuracy

1 measurement that would not be relatable to other studies included in the bibliometric analysis that  
2 involved average performance over single trial test. Cecotti, Sato-Reinhold, *et al.* (2011) carried out  
3 a study where they compared varying target stimuli probability. Target probability has a significant  
4 effect on both behavioural performance and target detection. The best mean AUC is achieved with  
5 target probability of 0.10 AUC=0.82. The best target stimuli probability for optimal detection  
6 performance were 5% = 78.7%.

7  
8 This above review exemplifies how performance measures are used. The variability of accuracy  
9 analytics limits the extent to which inter-study comparability is feasible, nonetheless a high  
10 proportion of studies use AUC values and percentage accuracy as outcome measures therefore these  
11 measures provide the basis for comparisons in section 5. In the RSVP-based BCI application  
12 sections that follow, we provide additional information about the values reported in Tables 1 and 2.  
13 The intention being to validate why these performance metrics were selected when a number of  
14 different results are reported by the specified study, and to highlight inter-study idiosyncrasies that  
15 may need to be considered whilst comparing findings. In the next section, the different design  
16 parameters for the studies identified in Tables 1 and 2 are reviewed and a number of  
17 recommendations are suggested for the parameters that should be considered for RSVP-based BCI  
18 applications.

## 19 20 21 **5. Design parameters**

22 RSVP-based BCI applications to date can be grouped into surveillance, data categorization, RSVP  
23 speller, face recognition and medical image analysis applications. Often EEG-based RSVP-BCI  
24 system studies are multifactorial by design and report numerous results in the form of different  
25 outcome measures. In the RSVP-based BCI application section that follows, we provide examples  
26 of the different application types and examples of their design parameters.

27  
28 When designing an RSVP paradigm, there are eight criteria that we recommend be taken into  
29 consideration:

- 30 1) The type of target images and how rapidly these can be detected e.g., picture, number of  
31 words.
- 32 2) The differences between target and non-target images and how these influence the  
33 discrimination in RSVP paradigm
- 34 3) The display mode – static or moving stimuli and the background the images are presented  
35 on e.g., single color white, mixed, textured.
- 36 4) The response mode – consideration should be given as to whether a button press is used or  
37 not to confirm if person has identified a target.
- 38 5) The number of stimuli /the percentage of target stimuli – how many are presented  
39 throughout the duration of a session and the effect this could have on the ERP.
- 40 6) The rate at which stimuli are presented on screen throughout the duration of a session and  
41 the effect this has on the ERP.
- 42 7) The area (height × width), visual angle and the overt or covert attention requirement of the  
43 stimuli.



- 1 8) The signal processing pipeline - determine the features, channels, filters, and classifiers to  
2 use.

3 *5.1. Display and response modes*

4 A button press may be used in conjunction with either of the aforementioned presentation modes  
5 (section 2.2), and entails users having to click a button when they see a target. This mode is used as  
6 a baseline to estimate the behavioral performance and the difficulty of the task. In most research  
7 studies, participants undergo an experimental trial without a button press and a follow-on trial with  
8 a button press.

9 A button press can be used in RSVP-based BCI research in combination with the participant's EEG  
10 responses in order to monitor attention (Marathe *et al.*, 2014). The combination of EEG and button  
11 press can lead to increased performance in RSVP-based BCIs. Tasks that require sustained attention  
12 can cause participants to suffer from lapses in vigilance due to fatigue, workload or visual  
13 distractors (Boksem, Meijman and Lorist, 2005). The button press can be used to determine if there  
14 is a tipping point during the presentations when participants are unable to consciously detect target  
15 stimuli, while still identifying targets via EEG recordings (Potter *et al.*, 2014). However, the core  
16 advantage of the RSVP-based BCIs is the enhanced speed of using a neural signature instead of a  
17 behavioral response to determine if a user has detected an intended image of interest.

18  
19 Forty of the studies reported use static mode as a method of presentation, six of these papers used  
20 moving mode in conjunction with static mode while one study exclusively used moving mode.  
21 Moving mode is more complex than static mode as participants have to take in an entire scene rather  
22 than specific images. Moving mode uses motion onset in conjunction with the P300 for scenes in  
23 which the targets are moving, yielding a more realistic setting to validate RSVP-based BCIs  
24 (Weiden, Khosla and Keegan, 2012). All papers employing moving mode were found within the  
25 surveillance application category; this is unsurprising as the moving mode offers the opportunity to  
26 detect targets in realistic surveillance situations where movements of people or vehicles are of  
27 interest. For the other application areas i.e., medical, categorization etc. the static mode is likely to  
28 be the most appropriate.

29  
30 Won *et al.*, 2017 compared motion RSVP to standard RSVP, with the motion-type RSVP being the  
31 rapid presentation of letters of the alphabet, numbers 1-9 and a hyphen '-' used to separate words, in  
32 six different colour groups in one of six directions in line with the hands of a clock i.e. 2, 4, 6, 8, 10,  
33 12 whilst participants focused on a central point. An increase in performance accuracy with motion-  
34 type RSVP versus static-type was demonstrated, which could be accounted to the shorter latency  
35 and greater amplitudes of ERP components in the motion-type variation (Won *et al.*, 2017).

36  
37 Out of the studies found, 22 used a button press while 23 did not. 70% of surveillance applications  
38 used a button press. In categorization studies and face recognition studies the majority of  
39 applications used a button press. 89% of RSVP-speller applications did not use a button press.  
40 Typically, the BCI studies that involve spellers, focus on movement-free communication and high  
41 information transfer rates. Having a button press for confirmation of targets is not standard practice

1 in such applications (Umut Orhan *et al.*, 2012; Oken *et al.*, 2014). In many of the studies that did  
2 not utilize a button press, researchers are focused on different aspects of the RSVP paradigm other  
3 than reaction time. For example, researchers focused on the comparison of two classification  
4 methods, image durations etc. (Sajda *et al.*, 2010; Cecotti, Eckstein and Giesbrecht, 2014).  
5 Combining EEG responses with button press can improve accuracy although more signal processing  
6 is required in order to remove noise that occurs as a result of participant movement (Healy and  
7 Smeaton, 2011). Button press confirmation is unnecessary unless an assessment of physical reaction  
8 time is an important aspect of the study.  
9

10 Maguire and Howe (2016) instructed participants to use a button press following image blocks to  
11 indicate if a target was consciously perceived as present or absent. Such an approach is useful when  
12 studying RSVP based parameters and the limits of perception. However, button press responses  
13 might be less useful than EEG responses during RSVP for data labelling or image sorting, where  
14 the focus is to label individual images within the burst. Nonetheless, Bigdely-Shamlo *et al.* (2008)  
15 apply an image burst approach where a button press at the end of the image burst is used to  
16 determine if the participant saw a target image or not. The authors showed that airplanes could be  
17 detected in aerial shots with image bursts lasting 4100 ms and images presented at 12 Hz. The  
18 button press served well in determining correct and incorrect responses. In practice, however,  
19 button press may be superfluous or infeasible.

20 A body of researchers are of the opinion that RSVP-related EEG accuracy must surpass button press  
21 accuracy in order to be useful. However, this need not be the case as Gerson, Parra and Sajda (2006)  
22 report no significant differences in triage performance based on EEG recordings or button presses.  
23 Nevertheless button based triage performance is superior for participants that correctly respond to a  
24 high percentage of target images. Conversely, EEG-based triage alone is shown to be ideal for the  
25 subset of participants who respond correctly to fewer images Gerson, Parra and Sajda (2006).  
26 Hence, the most reliable strategy for image triaging in an RSVP based paradigm may be through  
27 reacting to the target image by real-time button presses in conjunction with an EEG based detection  
28 method. Target identification reflected in EEG responses can be confirmed by a button press, and  
29 through signal processing techniques both reported and missed targets can be identified.

30 Studies such as, Marathe *et al.*, (2014) propose methods for integrating button press information  
31 with EEG based RSVP classifiers to improve overall target detection performance. However,  
32 challenges arise when overlaying ERP and behavioural responses, such as issues concerning  
33 stimulation presentation speed and behavioural latency (Files and Marathe, 2016). Crucially, Files  
34 and Marathe, (2016) demonstrate that techniques for measuring real-time button press accuracy start  
35 to fail at higher presentation rates. Given evidence of human capacity for semantic processing  
36 during 20 Hz image streams (approximately 50 ms per image) and Response Times (RTs) often  
37 being an order of magnitude greater than EEG responses, button presses may be unsuitable for  
38 faster RSVP based image triaging.

1 Pending further studies investigating the reliability of fast detection of neural correlates, EEG based  
2 responses have the potential to exceed button press. However, it is not necessary for EEG based  
3 RSVP paradigms to surpass button press performance and evidence suggests that the complement of  
4 both modalities at comfortable lower presentation rates may indeed be the best approach.  
5 Nevertheless, ideally studies would contain an EEG only block and EEG plus button press block,  
6 where the button press follows the target and not the image burst. This would facilitate more  
7 accurate evaluation of differences and correlates between behavioural and neural response times.  
8 Interesting, (Bohannon *et al.*, 2017), present a heterogeneous multi-agent system comprising  
9 computer vision, human and BCI agents, and showed that heterogeneous multi-agent image systems  
10 may achieve human level accuracies in significantly less time than a single human agent by  
11 balancing the trade-off between time-cost and accuracy. In such cases a human-computer interaction  
12 may occur in the form of button press if the confidence in the response of other, more rapid agents  
13 such as RSVP-BCI agents or computer vision algorithm is low for a particular sequence of stimuli.

## 14 15 16 5.2. Type of stimuli

17 Surveillance is the largest RSVP BCI system application reported in this review, reflected as such  
18 by the discussion length of this subsection (Sajda, Gerson and Parra, 2003; Erdogmus, Mathan and  
19 Pavel, 2006; Gerson, Parra and Sajda, 2006; Poolman, P., Frank, R. M., Luu, P., Pederson, S. M.,  
20 and Tucker, 2008; Bigdely-Shamlo *et al.*, 2008; Sajda *et al.*, 2010; Huang *et al.*, 2011; Weiden,  
21 Khosla and Keegan, 2012; Cecotti, Eckstein and Giesbrecht, 2012; Matran-Fernandez and Poli,  
22 2014; Rosenthal *et al.*, 2014; Yu *et al.*, 2014; Marathe, Ries and McDowell, 2014; A. R. Marathe *et al.*,  
23 2015; Barngrover *et al.*, 2016; Cecotti, 2016; Files and Marathe, 2016).

24 In a surveillance application study carried out by (Huang *et al.*, 2011) targets were surface-to-air  
25 missile sites. Target and non-target images shared low-level features such as local textures, which  
26 enhances complexity. Nonetheless target images were set apart due to large-scale features like  
27 unambiguous road layouts. Another example of surveillance targets denoted by (Bigdely-Shamlo,  
28 Andrey Vankov, *et al.*, 2008) is where overlapping clips of London satellite images were  
29 superimposed with small target airplane images, which could vary in location and angle within an  
30 elliptical focal area. Correspondingly, in (Barngrover *et al.*, 2016), the prime goal was to correctly  
31 identify sonar images of mine-like objects on the sea bed. Accordingly, a three-stage BCI system  
32 was developed whereby the initial stages entail computer vision procedures e.g. Haar-like feature  
33 classification whereby pixel intensities of adjacent regions are summed and then the difference  
34 between regions is computed, in order to segregate images into image chips. These image chips  
35 were then fed into an RSVP type paradigm exposed to human judgment, followed by a final  
36 classification with Support Vector Machine (SVM).

37 In the categorization application type images are sorted into different groups (Cecotti, Kasper, *et al.*,  
38 2011; Cecotti, Sato-Reinhold, *et al.*, 2011). Fuhrmann, Alpert *et al.* (2014), conducted a study

1  
2  
3  
4  
5  
6 1 whereby five image categories were presented: cars, painted eggs, faces, planes, and clock faces  
7 2 (Sadja *et al.*, 2014). A second study in Fuhrmann, Alpert *et al.*, (2014), containing target (cars) and  
8 3 non-target image (scrambled images of the same car) categories was conducted. In both RSVP  
9 4 experiments, the proposed Spatially Weighted Fisher Linear Discriminant – Principal Component  
10 5 Analysis (SWFP) classifier correctly classified a significantly higher number of images than the  
11 6 Hierarchical Discriminant Component Analysis (HDCA) algorithm. In terms of categorization,  
12 7 empirical grounds were provided for potential intuitive claims, stating that target categorization is  
13 8 more efficient when: there is only one target image type; or distractors are scrambled variations of  
14 9 the target image as opposed to different images all together (Sadja *et al.*, 2014).

15  
16  
17 10 Face recognition applications have been used to seek out whether a recognition response can be  
18 11 delineated from an uninterrupted stream of faces, whereby each face cannot be independently  
19 12 recognized (Touryan *et al.*, 2011). Two of the three studies evaluated utilized face recognition  
20 13 RSVP paradigm spin offs with celebrity/familiar faces as targets and novel, or other familiar or  
21 14 celebrity faces as distractors (Touryan *et al.*, 2011; Bangyu Cai *et al.*, 2013). Cecotti *et al.* 2011.,  
22 15 utilized novel faces as targets amongst cars with both stimuli types presented with and without  
23 16 noise. Utilizing the RSVP paradigm for face recognition applications is an unconventional  
24 17 approach, nonetheless the ERP itself has been used exhaustively to study neural correlates of  
25 18 recognition and declarative memory (Yovel and Paller, 2004; Guo, Voss and Paller, 2005;  
26 19 MacKenzie and Donaldson, 2007; Parra, Chiao and Paller, 2011). Specifically, with early and later  
27 20 components of the ERP having been associated with the psychological constructs of familiarity and  
28 21 recollection respectively (Smith, 1993; Rugg *et al.*, 1998). There is thus substantial potential for the  
29 22 utility of the RSVP based BCI paradigm for applications in facial recognition. In the future, RSVP-  
30 23 based BCI face recognition may be apposite in a real world setting in conjunction with security-  
31 24 based identity applications to recognize people of interest. Furthermore, Touryan *et al.*, (2011)  
32 25 claim that based on the success of their study, RSVP paradigm based EEG classification methods  
33 26 could potentially be applied to the neural substrates of memory. Indeed, some studies show  
34 27 augmentation in posterior positivity of ERP components for faces that are later remembered (Paller  
35 28 and Wagner, 2002; Yovel and Paller, 2004). That is to say, components of ERPs triggered by an  
36 29 initial stimulus may provide an indication of whether memory consolidation of said stimulus will  
37 30 take place, which provides an interesting avenue for utilizing RSVP based BCI systems for  
38 31 enhancing human performance. Based on these studies, it is clear that relatively novel face  
39 32 recognition paradigms have achieved success when used in RSVP-based BCIs.

40  
41  
42  
43 34 RSVP-based BCIs that assist with finding targets within images to support clinical diagnosis has  
44 35 received attention (Stoica *et al.*, 2013), for example, in the development of more efficient breast  
45 36 cancer screening methods (Hope *et al.*, 2013). Hope *et al.* (2013) is the only paper evaluated from  
46 37 the field of medical image analysis and hence described in detail. During an initial sub-study  
47 38 participants were shown mammogram images, where target lesions were present or absent. In a  
48 39 subsequent study, target red or green stimuli were displayed among a set of random non-target  
49 40 blobs. These studies facilitated comparison between ‘masses’ and ‘no masses’ in mammograms,  
50 41 and strong color based images versus random distractors. Images were presented against a grey  
51 42 background in three second bursts of 30 images (100 ms per image). A difference in the amplitude  
52 43 of the P300 potential was observed across studies, with a larger amplitude difference between target

1 and non-target images in the mammogram study. The researchers attributed this to the semantic  
2 association with mammogram images, in contrast to the lack thereof in the colored images-based  
3 study.

### 4 5 6 5.3. Total stimuli number and prevalence of target stimuli

7 The number of stimuli refers to the total number of stimuli i.e., the same stimulus can be shown  
8 several times. An exception to this is RSVP-speller studies where researchers only report on the  
9 number of symbols used i.e., 28 symbols - 26 letters of the alphabet, space and backspace (Hild *et*  
10 *al.*, 2011). In the RSVP-speller studies reviewed, the number of times each symbol is shown is not  
11 explicit. RSVP-speller applications are likely to have significantly fewer stimuli than the other  
12 aforementioned applications as participants are spelling out a specific word or sentence, which only  
13 has a small number of target letters/words. The integration of language models into RSVP-speller  
14 applications enables ERP classifiers to utilize the abundance of sequential dependencies embedded  
15 in language to minimize the number of trials required to classify letters as targets or non-targets  
16 (Orhan *et al.*, 2011; Kindermans *et al.*, 2014)). Some systems, such as the RSVP keyboard  
17 (described in Hild *et al.*, 2011; Orhan, Hild, *et al.*, 2012a; and Oken *et al.*, 2014) display only a  
18 subset of available characters in each sequence. This sequence length can be automatically defined  
19 or be a pre-defined parameter chosen by the researcher. The next letter in a sequence become highly  
20 predictable in specific contexts, therefore it is not necessary to display every character in the RSVP-  
21 speller. Studies show that target characters are generally displayed more than once before the  
22 character is selected. The length of a sequence and the ratio of target to non-target stimuli can have  
23 an effect on the typing rate/performance. In an online study by Acqualagna *et al.*, 2011, participants  
24 were shown 30 symbols that were randomly shuffled 10 times before a symbol was selected through  
25 classification and presented on screen. Orhan *et al.*, 2012, carried out an offline study whereby 2  
26 healthy participants were shown 3 sequences (consisting of 26 randomly ordered letters of the  
27 alphabet). Results of this study show that the number of correctly identified symbols more than  
28 doubled when using 3 sequences instead of 1 sequence to identify targets.

29  
30 Task complexity is enhanced by the multiplicity of target categories. In Poolman, *et al.*, 2008) there  
31 were two blocks of target presentations; a helipad block with a 4% target prevalence; and a surface-  
32 to-air missile and anti-aircraft artillery block with a 1% target prevalence. Additionally, in (Cecotti  
33 *et al.*, 2012) the targets were 50% vehicles, 50% people, with 50% being stationary and 50%  
34 moving. Further to this, (Weiden, Khosla and Keegan, 2012) demonstrate that presenting kinetic  
35 images during the RSVP paradigm as opposed to stationary images increases performance of EEG-  
36 based detection, and that this is negatively correlated with the cognitive load associated with the  
37 presented stimuli. In RSVP-speller applications task complexity varies based on what instructions  
38 participants are given e.g., (1) participants may be asked to “spell dog”; (2) “type a word related to  
39 weather”; (3) participants can be given a word bank containing 20 words and asked to “spell a word  
40 found within this word bank”. Half of the RSVP-speller-based BCI studies evaluated involved user  
41 defined sequence length (instruction 2 and 3) (Acqualagna *et al.*, 2010; Hild *et al.*, 2011; Umut  
42 Orhan *et al.*, 2012; Oken *et al.*, 2014), while the other half involved users been given a target  
43 word/sentence to spell (instruction 1). If a participant has to remember the sentence or how to spell

1 a long or unfamiliar word this can increase the complexity of a task (i.e., dog is much easier spelt  
2 than idiosyncrasy) (Primativo *et al.*, 2016). Note however that these different complexities in  
3 instructions are only present for evaluation/training tasks with the RSVP-BCI spellers. For their real  
4 use, participants choose themselves what they want to spell. The RSVP-based text application  
5 allows the number of stimuli before a target stimulus to be reduced (i.e. letters such as 'z' that are  
6 less commonly used can be shown less frequently).

7  
8 Excluding RSVP-speller applications, as it is already known that they do not require the same  
9 number of stimuli as the other applications, the number of stimuli used typically varied between  
10 studies from approximately 800 in the surveillance application study by (Sajda *et al.*, 2010) to  
11 26,100 in a categorization application study by (Sajda *et al.*, 2014). The most common target stimuli  
12 percentage range was 1-10% found in 61% of the studies reviewed, followed by 11-20% then  
13 >20%. There are a number of studies that focus specifically on the percentage of target stimuli. In a  
14 study by (Cecotti, Sato-Reinhold, *et al.*, 2011), researchers investigated the influence of target  
15 probability when categorizing face and car images. In this study, researchers use spatially filtered  
16 EEG signals as the input for a Bayesian classifier. Using eight healthy participants, this method was  
17 evaluated using four probability of target stimuli conditions i.e., 0.05, 0.10, 0.25, or 0.50. It was  
18 found that the target probability had an effect on participant's ability to detect targets and on  
19 behavioral performance. The best mean AUC (0.82) was achieved using the 0.1 probability  
20 condition. The results show that the percentage of targets shown in an RSVP paradigm has an  
21 effect on participants' performance. As number and percentage of target stimuli used can have an  
22 effect on the complexity of a task, it is important to keep the percentage of targets <10% to evoke  
23 the P300 and maximize detection rates. This was proposed to be in line with well-established P3  
24 measures, whereby bigger gaps between target trials reduce peak latency and increase amplitude  
25 (Gonsalvez and Polich, 2002).

#### 26 27 5.4. Duration of stimuli presentation

28 A key factor of the RSVP paradigm is the rate of presentation, as the focus of this paradigm is  
29 presenting data at a rapid rate so that large datasets can be analyzed in short periods. The duration  
30 for which stimuli were presented varied from 50 to 500 ms (Sajda, Gerson and Parra, 2003; Touryan  
31 *et al.*, 2011; B. Cai *et al.*, 2013). The upper limits for presentation time of stimuli during the RSVP-  
32 paradigm is ill-defined in the literature; however we found 500 ms per image to be the maximum  
33 RSVP duration used across all RSVP studies. The duration of stimuli typically differs between  
34 applications. Table 3 shows that the most common duration of stimuli was between 100-199 ms per  
35 image. The quickest duration of 50 ms per image was used in a study by (Sajda, Gerson and Parra,  
36 2003) where 2 participants were asked to identify scenes containing people in natural scenes. In  
37 each trial, the duration of the stimulus presentation was decreased from 200 to 100 to 50 ms per  
38 image. The results of this study showed that both participants had reduced performance for faster  
39 stimulus presentations i.e., 50 ms. This would suggest that the most suitable duration for RSVP-  
40 based BCI applications is 100-200 ms, to balance the trade-off between accuracy and speed.

41

1 Overall, these limited findings are suggestive of presentation rates >10Hz being infeasible for  
2 identification of neural correlates that allow successful identification of targets. Despite low a  
3 participant number in Sajda, Gerson and Parra, (2003), validation for this upper cut-off presentation  
4 rate may be provided by, Raymond, Shapiro and Arnell, 1992, where the attentional blink was first  
5 described. An RSVP paradigm was undertaken whereby the participant must register a target white  
6 letter in a stream of black letters and a second target 'X' amongst this stream. It was found that if  
7 the 'X' appeared within ~100-500ms of the initial target, errors in indicating whether the 'X' was  
8 present or not were likely to be made even when the first target was correctly identified (Raymond,  
9 Shapiro and Arnell, 1992). This is not to say that humans cannot correctly process information  
10 presented at >10Hz. Forster, (1970), has shown that participants can process words presented in a  
11 sentence at 16 Hz (16 words per second). However, the sentence structure may have influenced the  
12 correct detection rate, which has an average of four words per second for simple sentence structures  
13 and three words for complex sentences. Detection rates improve when presented at a slower pace  
14 e.g., four relevant words per second, with masks (not relevant words) presented between relevant  
15 words. Additionally, Fine and Peli, 1995, showed that humans can process words at 20 Hz in an  
16 RSVP paradigm.

17  
18 Potter *et al.*, (2014) assessed the minimum viewing time needed for visual comprehension, using  
19 RSVP of a series of 6 or 12 pictures presented at between 13 and 80 ms per picture, with no inter-  
20 stimulus interval. They found that observers could determine the presence or absence of a specific  
21 picture even when the pictures in the sequence were presented for just 13 ms each. The results  
22 suggest that humans are capable of detecting meaning in RSVP at 13 ms per picture. However, the  
23 finding challenges established feedback theories of visual perception. Specifically, research assert  
24 that neural activity needs to propagate from the primary visual cortex (V1) to higher cortical areas  
25 and back to the primary visual cortex before recognition can occur at the level of detail required for  
26 an individual picture to be detected, Maguire and Howe, (2016). Maguire and Howe, (2016) support  
27 Potter *et al.*, (2014) in that the duration of this feedback process is likely  $\geq 50$ ms, and suggest that  
28 this is feasible based on work done by Lamme and Roelfsema, (2000). Explicitly, Lamme and  
29 Roelfsema, (2000) estimated that response latencies at any hierarchical level of the visual system  
30 are ~10ms. Therefore, assuming that a minimum of five levels must be traversed as activity  
31 propagates from V1 to higher cortical areas and back again, this feedback process is unlikely to  
32 occur in <50ms. However, Maguire and Howe, (2016) suggested a potential confound of Potter *et*  
33 *al.*,(2014) was that pictures in the RSVP sequence, on occasion, contained areas with no high-  
34 contrast edges and hence may not have adequately masked preceding pictures. Consequently,  
35 Maguire and Howe, (2016) replicated the study rectifying the edges to ensure high-contrast  
36 covering the entire image. They were unable to find any evidence that meaning can be detected in  
37 an RSVP stream at 13 ms, or even 27 ms, per image but at 53 and 80 ms this is possible. Upon this  
38 basis, the limits of RSVP processing could be reduced to a minimum of ~20Hz. Nonetheless, further  
39 study is needed to investigate the limits of human capability to rapidly distinguish target from non-  
40 target information, in comparison to the limit in detecting target related ERPs versus non-target  
41 ERPs at 20Hz presentation rates.

42

1 In all three face recognition studies, each face image was displayed for 500ms (Cecotti, Sato-  
 2 Reinhold, *et al.*, 2011; Touryan *et al.*, 2011; B. Cai *et al.*, 2013). In two of the studies there was no  
 3 ISI (Cecotti, Sato-Reinhold, *et al.*, 2011; Touryan *et al.*, 2011), and in the other an ISI of 500ms was  
 4 given to ensure ample time for image processing (Bangyu Cai *et al.*, 2013). The speed at which face  
 5 images were shown is reduced in comparison to the other RSVP applications. RSVP spellers most  
 6 commonly use a duration of 400 ms, RSVP-spellers can benefit from slower stimulus duration with  
 7 the incorporation of a language model to enable the prediction of relevant letters. The estimation of  
 8 performance can be challenging in the RSVP paradigm when the ISI is small, as assigning a  
 9 behavioural response (i.e.; button press) to the correct image cannot be done with certainty. A  
 10 solution to this problem is to assign behavioral responses to each image, therefore researchers are  
 11 able to establish hits or false alarms (Touryan *et al.*, 2011). When two targets are temporally  
 12 adjacent with a SOA of 80 ms, participants are able to identify one of the two targets but not both.  
 13 SOA should be at least 400 ms and target images should not be shown straight after each other  
 14 (Raymond, Shapiro and Arnell, 1992). Acqualagna *et al.* 2010, had a four factorial design looking  
 15 at classification accuracy when the letters presented as no-colour or colour letters at either 83 or 133  
 16 ms with an ISI of 33ms (Acqualagna *et al.*, 2010). The number of sequence stimuli were presented  
 17 for enhanced accuracy rate in selecting letter of choice. After 10 sequences ~90% mean accuracy  
 18 was reached in 133ms colour presentation mode (100% for 6/9 participants). After 10 sequences in  
 19 133ms no colour presentation mode ~80% mean accuracy was reached (100% in 3/9 participants).  
 20 Whilst at presentation rates of 83ms mean accuracy rate was ~70% and there was no significant  
 21 effect of colour. This formulation is based on the chance rate of 3.33% (i.e. 1 in 30). This implies  
 22 that coloured letters enhances performance accuracy but not past a certain speed of stimulus  
 23 presentation.

24  
 25 There is likely a significant interaction between the difficulty of target identification and  
 26 presentation rate. For example, the optimal presentation rate for a given stimulus set is highly  
 27 dependent on the difficulty of identifying targets within that set (Ward, Duncan and Shapiro, 1997).  
 28 Image sets with low clutter, high contrast, no occlusion, and large target size are likely amenable to  
 29 faster presentation rates; while image sets with high clutter, low contrast, high levels of occlusion,  
 30 with small target sizes will require slower presentation rates (Rousselet, Thorpe and Fabre-Thorpe,  
 31 2004; Serre, Oliva and Poggio, 2007; Hart *et al.*, 2013; Liu and Kwon, 2016). A more conclusive  
 32 analysis of the effect of stimulus presentation duration for each application type could be derived by  
 33 varying presentation rate durations between 100, 200, and 500 ms, whilst other parameters remain  
 34 fixed. With regards to temporal proximity of target images, 500ms should be taken to be the  
 35 minimum to maximize performance.

36  
 37 **Table 3. Variation of image duration in RSVP studies.**

Duration (ms)	Number of studies	Accuracy % range
<100	7	66-93
100-199	22	70-92
200-299	11	70-96
300-399	-	-
400-499	2	85-94



500+	8	78.4-90
------	---	---------

1

### 2 5.5. Image size/visual angle

3 Another RSVP design aspect to be considered is stimulus size. There is a large variation in image  
 4 sizes ranging from 256×256 pixels in a categorization application to 960×600 pixels in a  
 5 surveillance applications. In general, surveillance applications use larger images than the other  
 6 applications described. The most common image size used is 500×500 pixels. This is only used in  
 7 static surveillance applications and all surveillance studies using this image size achieved a high  
 8 accuracy (>80%). The other applications used smaller image sizes such as 360×360 pixels and  
 9 achieved high accuracies (i.e., 91% and 89.7%). Therefore, it can be concluded that for surveillance  
 10 studies, image sizes should be at least 500×500 pixels, although for all other applications the image  
 11 size may be smaller. A more complex task, where a target stimulus is presented in the background  
 12 of a larger image eliciting the N2 ERP. Early components such as the P1 and N2 are sensitive to the  
 13 spatial location of the stimuli (Saavedra and Bougrain, 2012).

14  
 15 One issue with reporting only image size is that it is always relevant to the distance viewed from  
 16 screen and its location on the screen with respect to the viewer i.e., the visual angle. The visual angle  
 17 is the angle an image subtends at the eye, reported in degrees of arc. In a study by (Dias and Parra,  
 18 2011) it was shown that participants performed best (90%) when the target stimulus was centered.  
 19 Performance consistently decreased to 50% in all participants as target stimulus were placed further  
 20 away from the center (4° of visual angle), this dropped further when target stimulus was placed at 8°  
 21 of visual angle. Although performance drops significantly participants are still able to detect target  
 22 stimulus shown in their peripheral visual field even at such rapid paces. Many papers report that the  
 23 visual angle of the stimuli can have an effect on performance. As a general principle, targets must  
 24 appear larger or be more distinct for detection at the outer edge of the visual field. The visual angle  
 25 can thus be deemed the most important measure as it accounts for distance from screen, image  
 26 location on screen and image size. Authors are therefore encouraged to report visual angle, as  
 27 reporting image size alone is not useful without the availability of distance from screen. For RSVP-  
 28 speller studies, none of the papers found reported on the size of the image or font, however some  
 29 reported the visual angle.

### 31 5.6 Target vs non-target Stimuli

32 Many different types of target images have been identified within this review. The majority of  
 33 research focuses on a two-class problem i.e., detecting target images in sequences of non-target  
 34 images that are completely different from each other. However, in real-life situations, non-target  
 35 images are likely to share some of the same characteristics as target images (A. R. Marathe *et al.*,  
 36 2015). These presentation sequences appear to be more like moving images than static images. In  
 37 (A. R. Marathe *et al.*, 2015) a more complex surveillance task was carried out where, in the first  
 38 task, participants were required to detect targets when targets are the only infrequent image whilst,  
 39 in the second task, targets were presented with non-targets (i.e. the target image could be found in  
 40 the background of a larger image). Participants were required to ignore everything else in the image,  
 41 a much more difficult task, and consequently the amplitude of the P300 was reduced. The results of

42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 this study found that the introduction of the infrequent non-target stimuli in the scene yielded a  
2 substantial slowing of the reaction time. Surveillance applications commonly use stimuli that are  
3 more complex where trained participants, such as intelligence analysts outperform novice  
4 participants, as they are able to give meaning to the stimuli. The RSVP-speller applications present  
5 their letters as images one at a time on screen (Hild *et al.*, 2011). Due to the nature of the RSVP  
6 paradigm, it is important that these letters are shown in a random order as participants pre-empting a  
7 target can have an effect on ERP responses (Oken *et al.*, 2014). Data categorization applications had  
8 the most variance between the different types of stimuli presented to a participant. However, these  
9 stimuli tend to be everyday items that participants can easily recognize.

### 10 11 5.7. Signal Processing

12 All applications have certain requirements in terms of speed and type of images displayed which,  
13 as outlined above, can influence the ERP and therefore also variations in performance as measured  
14 by detection accuracy. The signal processing framework plays an important role in being able to  
15 cope with variations in ERP and maximizing performance. There is a likely tradeoff between the  
16 design parameters used as described above and the levels sophistication build into the signal  
17 processing framework, which often varies across studies. Here we review some of the approaches  
18 applied.

#### 19 20 5.7.1. Pre-processing

21 To extract the relevant features, data is first pre-processed to improve the signal to noise ratio  
22 (SNR). The signal is pre-processed using varying band pass filters, depending on the application, in  
23 order to remove high frequency noise or artifacts (such as muscle activity). Generally, lower and  
24 upper cut-off frequencies of around 0.1 Hz and 30-40 Hz are used, respectively. The data is then  
25 often downsampled, and, for offline analyses, electrodes with substantial noise are removed through  
26 visual inspection of the EEG data or automated approaches based on thresholding or correlating  
27 artefacts in EEG channels with simultaneously recorded electrooculography (EOG) or  
28 electromyography (EMG). Data is then epoched into segments typically lasting ~600 ms, from 100  
29 ms prior to stimulus onset and the 500 ms post-stimulus onset. The starting point and duration of the  
30 epochs selected for further analysis vary from study to study.

#### 31 32 5.7.2. Feature extraction

33 Feature extraction is applied to the data for dimensionality reduction and to extract discriminant and  
34 non-redundant features. It can be difficult to carry out feature extraction due to the low SNR in  
35 single trial analysis. Conventionally averaging over multiple repeated trials is often used to  
36 overcome this. Many studies employ spatial filtering to extract ERPs from EEG. Some of the spatial  
37 filtering methods used include principal component analysis (PCA) (Sajda, Gerson and Parra, 2003;  
38 S *et al.*, 2014), independent component analysis (ICA) (Bigdely-Shamlo *et al.*, 2008; Blankertz *et*  
39 *al.*, 2011; Kumar and Sahin, 2013), or the xDAWN algorithm which maximizes the SNR between  
40 target and non-target stimuli classes (Rivet *et al.*, 2009; Rivet and Souloumiac, 2013; Cecotti,  
41 Eckstein and Giesbrecht, 2014). In the case of image triage where the intention is to classify single-

1 trial ERPs, spatial filters are used to enhance SNR and exploit spatial redundancy (e.g. Parra *et al.*,  
2 2005). Yu *et al.* 2011 went a step further by utilizing a methodology that considers spatial and  
3 temporal features to ensure augmented single-trial detection accuracy (Yu *et al.*, 2011). Bilinear  
4 common spatial pattern (BCSP) was suggested to outperform Common Spatial Patterns (CSP)  
5 filters (composite and common spatial pattern filters) (Yu *et al.*, 2011). It should be noted however  
6 that CSP spatial filters were not designed to classify ERP but to classify oscillatory EEG activity.  
7 CSP are indeed ignoring the EEG time course – i.e., the ERP – and are thus suboptimal for RSVP-  
8 BCI. We would recommend using spatial filters dedicated to ERP classification, such as xDAWN,  
9 which were used successfully in many RSVP-BCI. Spatial filters are normally only performed on  
10 high-density EEG data which might be impractical in certain real-life applications (Parra *et al.*,  
11 2005). High-density EEG data has been reported to increase accuracy (Ušćumlić, Chavarriaga and  
12 Millán, 2013). Table 4 shows the most common method used for different application types.

13  
14 Face recognition applications differ from other applications as face images evoke different ERPs, in  
15 addition to the P300. Faces typically evoke a N170 component that changes between targets and  
16 non-targets (Maurer, Rossion and McCandliss, 2008; Luo *et al.*, 2010). The vertex positive potential  
17 (VPP) is also associated with face recognition (Zhang *et al.*, 2012). The midfrontal FN400 and later  
18 parietal FP600 components have been associated with familiarity and recollection, respectively,  
19 (MacKenzie and Donaldson, 2007). Specifically, the amplitude of FP600 (a positive deflection  
20 >500 ms post-stimulus) was found to significantly correlate with the extent of face familiarity  
21 (Touryan *et al.*, 2011). The use of spatial filters that utilize spatial and temporal features may act as  
22 an advantage over conventional spatial filters that only exploit spatial redundancy e.g. (Yu *et al.*,  
23 2011). However, spatial filters can only be performed on high-density EEG data which might be  
24 impractical in certain real-life applications (Parra *et al.*, 2005).

### 25 26 27 5.7.3. Classification

28 This review found many different classification methods were used in the acknowledged studies,  
29 however some conclusions can be drawn. Linear classifiers are most populous within RSVP-based  
30 BCIs. Often EEG can contain information that enables classification of the stimuli correctly even  
31 when a participants behavioral response is incorrect (Sajda, Gerson and Parra, 2003; Bigdely-  
32 Shamlo *et al.*, 2008). The two most commonly used classifiers were Linear Discriminant Analysis  
33 (LDA) and Support Vector Machine (SVM), or variations of the two, such as Bayesian Linear  
34 Discriminant Analysis (BLDA) and Radial Basis Function Support Vector Machine (RBF SVM),  
35 respectively. Parra *et al.*, 2008 presented an RSVP framework that projects the EEG data matrix bi-  
36 linearly onto temporal and spatial axes (Parra *et al.*, 2008). This framework is versatile upon  
37 implementation, for example, it has been applied to classify target natural scenes and satellite  
38 missile images (Gerson, Parra and Sajda, 2006; Sajda *et al.*, 2010). Contrastingly, Alpert *et al.*, 2014  
39 presented a two-step linear classifier, which achieved classification accuracy suited to real-world  
40 applications (Sajda *et al.*, 2014). Whilst Sajda *et al.*, 2010 proposed a two-step system utilizing  
41 computer vision and EEG subsequently to optimize classification (Sajda *et al.*, 2010). The  
42 performance of an ensemble LDA classifier diminished when 8 centro-parietal EEG channels were  
43 utilized as opposed to the full 41 EEG channels (Ušćumlić, Chavarriaga and Millán, 2013).  
44 Contrastingly, (Healy and Smeaton, 2011) claimed that consideration of additional channels may

1 introduce noise as opposed to advancing categorical information, as indicated by results from one  
2 study participant.  
3

4 For the surveillance application, SVM achieved the highest percentage accuracies (Huang *et al.*,  
5 2011; Weiden, Khosla and Keegan, 2012). For the RSVP-speller application, the most common  
6 method of classification used was Regularized Discriminant Analysis (RDA). RDA achieved an  
7 AUC performance of 0.948-0.973(Orhan *et al.*, 2011). Step Wise Linear Discriminant Analysis  
8 (SWLDA) was also used in RSVP-speller applications with high AUC performance and accuracies  
9 (0.82, 0.84, 86%, 89%) (Hope *et al.*, 2013). In face recognition applications, the best AUC  
10 performance was produced using an SVM classifier (Cai *et al.*, 2013). Within this review, only one  
11 medical application was identified (Hope *et al.*, 2013) and researchers achieved high accuracy using  
12 a Fisher Discriminant Analysis. BLDA classifiers were also used, achieving high levels of accuracy  
13 (79%). The Spatially Weighted Fisher Discriminant (SWFP) algorithm outperformed the  
14 Hierarchical Discriminant Component Analysis (HDCA) algorithm by 10% in categorization  
15 applications. Touryan *et al.* 2011 demonstrated that EEG classification methods applied to  
16 categorization procedures can be adapted to rapid face recognition procedures (Touryan *et al.*,  
17 2011). Window sizes post stimulus onset of 128, 256 and 512 ms were fed into the classifiers. AUC  
18 values (average AUC = 0.945) are reported for the customized PCA models utilized to describe the  
19 changes in ERPs seen between familiar (famous and personal) and novel faces displayed for 500ms  
20 at a time. It is the customized version of these models i.e. the models developed for each participant  
21 using only that participant's data, which were shown to improve classification performance through  
22 the acknowledgment of discrete variability in the windowed ERP components.  
23  
24

25 Many of the BCI algorithms presented in tables 1 and 2 are linear, enabling simple/fast training  
26 with resilience to overfitting often caused by noise, implying suitability to single-trial EEG data  
27 classification. Nonetheless, linear methods can limit feature extraction and classification, and non-  
28 linear methods e.g. neural networks, are more versatile in modelling data of greater variability, also  
29 implying suitability to single-trial EEG data classification (Erdogmus, Mathan and Pavel, 2006;  
30 Yonghong Huang *et al.*, 2006; Lotte *et al.*, 2007). The use of neural networks, in particular deep  
31 neural network for RSVP-based BCI framework represents an attractive venture, and have shown  
32 promise over standard linear methods (Manor, Mishali and Geva, 2016; Huang *et al.*, 2017). A  
33 convolution neural network was shown to outperform a two-step linear classifier using the same  
34 dataset (Sadja *et al.*, 2014; Manor and Geva, 2015).  
35

36 The majority of studies reviewed investigate the effectivity of classifiers in identifying single-trial  
37 EEG correlates for target stimuli presented through an RSVP type paradigm. However, the spatial  
38 filtering technique as well as the type of classifier used has an impact on proficiency in detecting  
39 EEG of single trials (Bigdely-Shamlo *et al.*, 2008; Cecotti, Eckstein and Giesbrecht, 2014). For  
40 example, Independent Component Analysis reportedly identifies and divides multiple classes of  
41 non-brain response artefacts associated with eye and head movements, which would be useful for  
42 EEG de-noising during real-world applications when operators are mobile (Bigdely-Shamlo,  
43 Vankov *et al.*, 2008).  
44

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 Additionally, (Cecotti, Eckstein and Giesbrecht, 2014) evaluated three classifiers using three  
2 different spatial filtering methods, so all in all twelve techniques were compared for three different  
3 RSVP paradigms. Marathe *et al.*, 2015 utilized an Active Learning technique in a bid to reduce the  
4 training samples required to calibrate the classifier. Active Learning is a partially supervised  
5 iterative learning technique reducing the amount labeled data during required for training.  
6 Recalibration depends on parameters such as, human attentiveness, physical surroundings or task-  
7 specific factors. Looking at the real world applicability of RSVP based BCI systems, (Marathe *et*  
8 *al.*, 2015) build upon work addressing the issue of thorough recalibration required for real-time BCI  
9 system optimization.

10 There is growing interest in the use of transfer learning (TL) for calibration reduction or suppression  
11 to encourage the real-world applicability of BCIs (Wang et al., 2015). With TL, the EEG data or  
12 classifiers from a given domain is transformed in order to be applied to another domain, hence  
13 transferring data/classifier from one domain to another, possibly increasing the amount of data for  
14 the target domain (Wang et al., 2015). For RSVP-BCI, this typically consists in combining EEG  
15 data or classifiers from different participants, in order to classify EEG data from another participant,  
16 for which very little or even no calibration EEG data is available. An unsupervised transfer method,  
17 namely Spectral Transfer with Information Geometry (STIG), ranked and collated unlabelled  
18 predictions from a group of information geometry classifiers which was established through training  
19 on individual participants (Waytowich et al., 2016). Waytowich et al, 2016 showed that STIG can  
20 be used for single-trial detection in ERP-based BCIs, eliminating the requirement for taxing data  
21 collection for training. With access to limited data, STIG outperformed alternative zero-calibration  
22 and calibration reduction algorithms (Waytowich et al., 2016). Within the BCI community  
23 conventional TL approaches still necessitate training for each condition, however methodologies  
24 have been applied to eradicate the need for subject-specific data calibration, where large-scale data  
25 is leveraged from other participants (Wei et al., 2016). This demarcates the potential for single-trial  
26 classification via unsupervised TL and user-independent BCI technology deployment.

### 27 28 29 5.8. Suggested parameters

30 The parameters reviewed here have been selected as they have an effect on one or all of the  
31 following aspects of the RSVP paradigm; task complexity, stimulus complexity, stimulus saliency  
32 or information transmission. Performance within RSVP-based BCIs is measured as the participant's  
33 ability to correctly identify oddball images in a sequence. RSVP-based BCIs use two different  
34 measurements of performance such as accuracy (percentage of targets that are correctly identified  
35 using EEG) and ROC curves. 10% of papers assessed in this review did not report at least one out of  
36 these performance measures (ROC/ percentage accuracy). The accuracies of the different studies  
37 need to be put in context, as all the reviewed parameters and other observed parameters i.e. number  
38 of trials and participants will influence study accuracy. In Table 4 parameter recommendations are  
39 provided for designing RSVP-based BCIs within the different application types and these have been  
40 discussed thoroughly throughout section 5. In particular, Table 4 suggests parameters to use for  
41 each application, according to those leading to the best detection performances (accuracy or AUC)  
42 in studies comparatively. If no formal comparisons between parameters were available for a specific  
43 application or parameter, the most popular parameter values that yield good performances are  
44 mentioned.

1 **Table 4. Parameter and recommendations for RSVP-based BCIs**

2

Parameter	Surveillance	RSVP-speller	Face Rec	Categorization/ Medical
Stimuli No.	>5000	>5000	2000	>4000
% Targets	~5-10	≤5	~10	10-25
Stimulus presentation duration (ms)	100-200	500	500	100-200
Target examples	Helipads, planes, vehicles, people etc.	Letters	Faces	Animals, mammograms etc.
ERP component	P300	P300	N170	P300
Feature Extraction	XDAWN	-	XDAWN	BCSP/XDAWN
Classifier	BLDA, SVM, LP, SP	RDA/ SWLDA	SVM	BLDA

3  
4 Applying BCI systems commercially and outside the lab in real world scenario will ideally require  
5 the system to be robust during the execution of tasks of increasing difficulty. Section 5 summarized  
6 the five applications areas that have been studied to the greatest extent in the context of RSVP-  
7 based BCIs. Specifically, this section tackles intra application comparisons of various aspects of the  
8 papers that met the inclusion/exclusion criteria. A few of the papers found in this review carried out  
9 more than one study in different application types. The most common type of application found was  
10 surveillance applications, followed by RSVP-speller applications and categorization applications,  
11 after this were face recognition and lastly medical applications. Although there is a relatively  
12 limited number of studies, the design parameters and the focal points of different applications vary  
13 widely.

14  
15  
16 **6. Discussion and Conclusion**

17 With the increasing intensity in RSVP-based BCI research there is a need for further standardization  
18 of experimental protocols, to compare and contrast development of the different applications  
19 described in this review. This will aid the realization of a platform which researchers can use to  
20 develop RSVP paradigms and compare their results and determine the optimal RSVP based BCI  
21 paradigm for their application type. This paper presents a review of the available research, the  
22 defining elements of the research and a categorization approach that will facilitate coordination  
23 efforts among researchers in the field. Research has revealed that using a combination of RSVP

1 with BCI technology, allows the detection of targets at an expedited rate without detriment to  
2 accuracy.

3  
4 Understanding the neural correlates of visual information processing can create symbiotic  
5 interaction between human and machine through BCIs. Further development of RSVP-based BCIs  
6 will depend on both basic and applied research. Within the last five years, there have been  
7 advancements in how studies are reported and a sufficient body of evidence exists in support of the  
8 development and application of RSVP BCIs. However, there is a need for the research to be  
9 developed further and standardized protocols applied, so that comparative studies can be done for  
10 progressive research. Many ERP reviews have been carried out, however, this paper focuses on  
11 RSVP visual search tasks with high variability in targets and the parameters used. This paper gives  
12 guidelines on which parameters impact performance but also on which parameters should be  
13 reported so that studies can be compared. It is important that design aspects shown in Tables 1 and 2  
14 are reported and described within each research study. It has been shown that RSVP based BCIs can  
15 be used in processing target images in multiple application types with a low-target probability, but  
16 consistency of reporting method renders it difficult to truly compare one paradigm to another or one  
17 parameter-setup to another.

18 There is profuse reporting of percentage accuracy and area under the ROC curve values, nonetheless  
19 there is room for more studies to utilize this unofficial standardization across RSVP-based BCI  
20 research.

21  
22 To maximize relatability to pre-existing literature in terms of keeping one feature that contributes to  
23 cognitive load constant, it is recommended that studies utilizing greater than one category type as  
24 targets to conduct the same study with just one target category in the first instance.

25  
26 For all applications, it is of course necessary to choose an epoch for single trial ERP classification  
27 corresponding to the temporal evolution of the most robust ERP components that are, on the whole,  
28 pre-established in the literature as associated with the specified task at hand i.e., target stimuli  
29 identification due to their infrequency, recognisability, relevancy or contents. However, whether the  
30 duration of stimuli presentation must extend beyond the latency between ERP component  
31 appearances relative to stimuli presentation is questionable.

32  
33 This review found a single medical application. More research in applying the RSVP-based BCI  
34 paradigm to high throughput screening within medicine is highly encouraged upon the basis that  
35 similarly complex imagery has been categorized relatively successfully in other applications e.g.,  
36 side scan sonar imagery of mines or aircraft amongst birds eye view of maps in surveillance  
37 (Bigdely-Shamlo *et al.*, 2008; Barngrover *et al.*, 2016). The medical application of RSVP-based  
38 BCIs has immense potential in diagnostics and prognostics through recognition and tracking of  
39 established disease biomarkers, and accelerating high throughput health image screening.

40  
41 Studies utilized varying image sizes, visual angles and participant distance from screen. Researchers  
42 are encouraged to report visual angle as it accounts for both images size and distance of participant  
43 from screen. A potential way to facilitate uniformity of these variables is to utilize a head mounted  
44 display (HMD) or Virtual Reality (VR) headset such as an Oculus Rift (Foerster *et al.*, 2016). The  
45 rapid visual information processing capacity is heavily dependent on visual parameters and use of

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1 an HMD headset would enable standardization of viewing distance, room lighting and visual angle  
2 (Foerster *et al.*, 2016). Use of a VR headset could distort electrode positions, nonetheless this affect  
3 could be easily mitigated. BCIs employing motion-onset visual evoked potentials (mVEP) have  
4 been utilized with VR headsets in neurogaming, and shown to be feasible (Beveridge, Wilson and  
5 Coyle, 2016). The mVEP responses were evaluated in relation to mobile, complex and varying  
6 graphics within game distractors (Beveridge, Wilson and Coyle, 2016). (Foerster *et al.*, 2016) used  
7 the virtual reality device Oculus Rift for neuropsychological assessment of visual processing  
8 capabilities. This VR device is head-mounted and covers the entire visual field, thereby shielding  
9 and standardizing the visual stimulation and therefore may improve test-retest reliabilities.  
10 Compared to a CRT screen performances, visual processing speed, threshold of conscious  
11 perception and capacity of visual working memory did not differ significantly using the VR headset.  
12 VR headsets may therefore be applicable for standardized and reliable assessment and diagnosis of  
13 elementary cognitive functions in laboratory and clinical settings and maximise the opportunity to  
14 compare visual processing components between individuals and institutions and to establish  
15 statistical norm distributions. Recently, a new VR-EEG combined headset with electrodes  
16 embedded in occipital areas for ERP detection has been reported for neurogaming  
17 ([www.neurable.com](http://www.neurable.com)). RVSP-based BCI paradigms may therefore benefit from the head mounted  
18 visual displays however a vision obscuring headset may not be appropriate in some contexts as it  
19 could limit the ability of the users, e.g. a person with disabilities, to communicate with their peers  
20 and environment. Such a headset may prevent the expressive or receptive use of non-verbal  
21 communication skills such as eye movement and facial expressions that are vital for users with non-  
22 verbal communication skills.

23  
24 Advancements towards RSVP of targets during moving sequences have shown promising results,  
25 although it is more difficult to study movie clips since the stimulus start event is not as clear. A  
26 remaining challenge in this area is for researchers to design signal processing tools that can deal  
27 with imprecise stimulus beginning/end (Cecotti, 2015). However, an advantage of moving mode is  
28 that the target stimulus remains on the screen for longer than with static mode, allowing participants  
29 the opportunity to confirm a target stimulus. Moving stimuli studies to date have been limited to  
30 surveillance applications so there is a need for further investigation in this area. Just over half the  
31 papers used button press mode in conjunction with one of the other modes, as not all of the studies  
32 are concerned with comparing EEG responses to motor responses. It is important to develop a scale  
33 in order to rank the difficulty of tasks. This will enable the comparison of paradigms that are at the  
34 same level. The key outcomes of this study are shown in Table 4, provided as suggested guidelines.  
35 These are suggested parameters that may be useful to researchers when designing RSVP-based BCI  
36 paradigms within the different application types. From this review, we can conclude that using  
37 these parameters will enable more consistent performance for the different application types and  
38 will enable improved comparison with new studies.

39  
40 In acknowledgment of the need for standardization of parameters for RSVP-based BCI protocols,  
41 Cecotti, Satp-Reinhold *et al.*, 2011 raise an interesting proposal stating that other parameters could  
42 be automatically prescribed in accordance with the chosen target likelihood; such as the optimal ISI  
43 length, classifiers and spatial filters (Cecotti, Sato-Reinhold, *et al.*, 2011). Such an infrastructure for



1 parameter choices does not currently exist with studies focusing on the impact of different  
2 parameters.

3 Future studies would benefit from engaging with iterative changes in design parameters. This would  
4 allow for a comparative study of the different design parameters and enable the identification of  
5 parameters that most affect the experimental paradigm. A study involving increasing the rate of  
6 presentation until classification starts to deteriorate significantly for various types of stimulus  
7 categories may indicate the maximum possible speed of RSVP-BCI. Additionally, a future  
8 development for RSVP-based BCIs might be to use real life imagery with numerous distractor  
9 stimuli amongst the target stimuli. This is a more difficult task but it would enhance paradigm  
10 relatability to real-life applications. Hybridizing RSVP BCIs with other BCI paradigms has also  
11 started to receive more attention (Kumar and Sahin, 2013). Users of this system navigate using  
12 motor imagery movements (left, right, up and down). Search queries are spelt using the Hex-O-  
13 Speller and results retrieved from a web search engine may be fed back to the user using RSVP.  
14 This study shows the potential benefits of the RSVP paradigm and how it may be used in order to  
15 aid physically impaired users. Eye-tracking can be used as an outcome measure to assess and  
16 enhance RSVP stimuli and presentation modes. Specifically, using eye tracking researchers can  
17 establish where the participant's gaze is focused during erroneous trials and explore correlations  
18 between gaze variability and performance. With the RSVP-based BCI paradigm there is much scope  
19 to evaluate different data types/imagery. This is a fast growing field with a promising future. There  
20 are multiple opportunities and a large array of potential RSVP-BCI paradigm setups. Researchers in  
21 the field are therefore recommended to consider the literature to date and the comparative  
22 framework proposed in this paper.

## 23 24 **References**

- 25 Acqualagna, L. and Blankertz, B. (2011) 'A gaze independent spelling based on rapid serial visual  
26 presentation', in *2011 Annual International Conference of the IEEE Engineering in Medicine and  
27 Biology Society*. IEEE, pp. 4560–4563. doi: 10.1109/IEMBS.2011.6091129.
- 28 Acqualagnav, L. *et al.* (2010) 'A novel brain-computer interface based on the rapid serial visual  
29 presentation paradigm', in *2010 Annual International Conference of the IEEE Engineering in  
30 Medicine and Biology*. IEEE, pp. 2686–2689. doi: 10.1109/IEMBS.2010.5626548.
- 31 Barngrover, C. *et al.* (2016) 'A Brain-Computer Interface (BCI) for the Detection of Mine-Like  
32 Objects in Sidescan Sonar Imagery', *IEEE Journal of Oceanic Engineering*, 41(1), pp. 124–139.  
33 doi: 10.1109/JOE.2015.2408471.
- 34 Beveridge, R., Wilson, S. and Coyle, D. (2016) '3D graphics, virtual reality, and motion-onset  
35 visual evoked potentials in neurogaming', in Damien Coyle (ed.) *Brain-Computer Interfaces: Lab  
36 Experiments to Real-World Applications*. Vol 228. Elsevier, pp. 329–353. doi:  
37 10.1016/bs.pbr.2016.06.006.
- 38 Bigdely-Shamlo, N. *et al.* (2008) 'Brain Activity-Based Image Classification From Rapid Serial  
39 Visual Presentation', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5),  
40 pp. 432–441. doi: 10.1109/TNSRE.2008.2003381.
- 41 Blankertz, B. *et al.* (2011) 'Single-trial analysis and classification of ERP components — A  
42 tutorial', *NeuroImage*, 56(2), pp. 814–825. doi: 10.1016/j.neuroimage.2010.06.048.
- 43 Bohannon, A. W. *et al.* (2017) 'Collaborative image triage with humans and computer vision', *2016  
44 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference*

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44
- 1 *Proceedings*, pp. 4046–4051. doi: 10.1109/SMC.2016.7844866.
- 2 Boksem, M. A. S., Meijman, T. F. and Lorist, M. M. (2005) ‘Effects of mental fatigue on attention:  
3 an ERP study.’, *Brain research. Cognitive brain research*, 25(1), pp. 107–16. doi:  
4 10.1016/j.cogbrainres.2005.04.011.
- 5 Cai, B. *et al.* (2013) ‘A rapid face recognition BCI system using single-trial ERP’, in *In Neural  
6 Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, p. (pp. 89–92).
- 7 Cai, B. *et al.* (2013) ‘A rapid face recognition BCI system using single-trial ERP’, in *2013 6th  
8 International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, pp. 89–92. doi:  
9 10.1109/NER.2013.6695878.
- 10 Cecotti, H., Sato-Reinhold, J., *et al.* (2011) ‘Impact of target probability on single-trial EEG target  
11 detection in a difficult rapid serial visual presentation task.’, *Conference proceedings : ... Annual  
12 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE  
13 Engineering in Medicine and Biology Society. Conference*, 2011, pp. 6381–6384.
- 14 Cecotti, H., Kasper, R. W., *et al.* (2011) ‘Multimodal target detection using single trial evoked EEG  
15 responses in single and dual-tasks.’, *Conference proceedings : ... Annual International Conference  
16 of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and  
17 Biology Society. Conference*, 2011, pp. 6311–4. doi: 10.1109/IEMBS.2011.6091557.
- 18 Cecotti, H. *et al.* (2012) ‘Multiclass classification of single-trial evoked EEG responses’, in *2012  
19 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE,  
20 pp. 1719–1722. doi: 10.1109/EMBC.2012.6346280.
- 21 Cecotti, H. (2015) ‘Toward shift invariant detection of event-related potentials in non-invasive  
22 brain-computer interface’, *Pattern Recognition Letters*. Elsevier Ltd., 66, pp. 127–134. doi:  
23 10.1016/j.patrec.2015.01.015.
- 24 Cecotti, H. (2016) ‘Single-Trial Detection With Magnetoencephalography During a Dual-Rapid  
25 Serial Visual Presentation Task’, *IEEE Transactions on Biomedical Engineering*, 63(1), pp. 220–  
26 227. doi: 10.1109/TBME.2015.2478695.
- 27 Cecotti, H., Eckstein, M. P. and Giesbrecht, B. (2012) ‘Effects of performing two visual tasks on  
28 single-trial detection of event-related potentials.’, *Conference proceedings : ... Annual International  
29 Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in  
30 Medicine and Biology Society. Conference*, 2012, pp. 1723–1726.
- 31 Cecotti, H., Eckstein, M. P. and Giesbrecht, B. (2014) ‘Single-trial classification of event-related  
32 potentials in rapid serial visual presentation tasks using supervised spatial filtering.’, *IEEE  
33 transactions on neural networks and learning systems*. Institute of Electrical and Electronics  
34 Engineers Inc., 25(11), pp. 2030–42. doi: 10.1109/TNNLS.2014.2302898.
- 35 Chahine, G. and Krekelberg, B. (2009) ‘Cortical contributions to saccadic suppression’, *PLoS ONE*,  
36 4(9).
- 37 Chennu, S. *et al.* (2013) ‘The cost of space independence in P300-BCI spellers.’, *Journal of  
38 neuroengineering and rehabilitation*. BioMed Central Ltd., 10(1), p. 82. doi: 10.1186/1743-0003-  
39 10-82.
- 40 Cohen, M. (2014) *Analyzing Neural Time Series Data: Theory and practice*. MIT Press.
- 41 Cohen, M. X. (2014) *Analyzing Neural Time Series Data: Theory and Practice (Issues in Clinical  
42 and Cognitive Neuropsychology)*. The MIT Press.
- 43 Diamond, M. R., Ross, J. and Morrone, M. C. (2000) ‘Extraretinal control of saccadic  
44 suppression.’, *The Journal of neuroscience : the official journal of the Society for Neuroscience*,

- 1 20(9), pp. 3449–55.
- 2 Dias, J. C. and Parra, L. C. (2011) ‘No EEG evidence for subconscious detection during Rapid
- 3 Serial Visual Presentation’, *2011 IEEE Signal Processing in Medicine and Biology Symposium*
- 4 *(SPMB)*. Ieee, pp. 1–4. doi: 10.1109/SPMB.2011.6120108.
- 5 Erdogmus, D., Mathan, S. and Pavel, M. (2006) ‘Comparison of Linear and Nonlinear Approaches
- 6 on Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks’, *The 2006 IEEE*
- 7 *International Joint Conference on Neural Network Proceedings*, pp. 1136–1142. doi:
- 8 10.1109/IJCNN.2006.1716229.
- 9 Fawcett, T. (2006) ‘An introduction to ROC analysis’, *Pattern Recognition Letters*, 27(8), pp. 861–
- 10 874.
- 11 Files, B. T. and Marathe, A. R. (2016) ‘A regression method for estimating performance in a rapid
- 12 serial visual presentation target detection task.’, *Journal of neuroscience methods*, 258, pp. 114–23.
- 13 doi: 10.1016/j.jneumeth.2015.11.003.
- 14 Fine, E. M. and Peli, E. (1995) ‘Scrolled and rapid serial visual presentation texts are read at similar
- 15 rates by the visually impaired.’, *Journal of the Optical Society of America. A, Optics, image*
- 16 *science, and vision*, 12(10), pp. 2286–92.
- 17 Foerster, R. M. *et al.* (2016) ‘Using the virtual reality device Oculus Rift for neuropsychological
- 18 assessment of visual processing capabilities’, *Scientific Reports*. Nature Publishing Group, 6(1), p.
- 19 37016. doi: 10.1038/srep37016.
- 20 Forster, K. I. (1970) ‘Visual perception of rapidly presented word sequences of varying
- 21 complexity’, *Perception & Psychophysics*, 8(4), pp. 215–221. doi: 10.3758/BF03210208.
- 22 Gerson, A. D., Parra, L. C. and Sajda, P. (2005) ‘Cortical origins of response time variability during
- 23 rapid discrimination of visual objects’, *NeuroImage*, 28(2), pp. 342–353.
- 24 Gerson, A. D., Parra, L. C. and Sajda, P. (2006) ‘Cortically coupled computer vision for rapid
- 25 image search.’, *IEEE transactions on neural systems and rehabilitation engineering: a publication*
- 26 *of the IEEE Engineering in Medicine and Biology Society*, 14(2), pp. 174–9. doi:
- 27 10.1109/TNSRE.2006.875550.
- 28 Gonsalvez, C. L. and Polich, J. (2002) ‘P300 amplitude is determined by target-to-target interval.’,
- 29 *Psychophysiology*, 39(3), pp. 388–96.
- 30 Guo, C., Voss, J. L. and Paller, K. A. (2005) ‘Electrophysiological correlates of forming memories
- 31 for faces, names, and face–name associations’, *Cognitive Brain Research*, 22(2), pp. 153–164. doi:
- 32 10.1016/j.cogbrainres.2004.08.009.
- 33 Hart, B. M. *et al.* (2013) ‘Attention in natural scenes: contrast affects rapid visual processing and
- 34 fixations alike.’, *Philosophical transactions of the Royal Society of London. Series B, Biological*
- 35 *sciences*. The Royal Society, 368(1628), p. 20130067. doi: 10.1098/rstb.2013.0067.
- 36 Healy, G. and Smeaton, A. F. (2011) ‘Optimising the number of channels in EEG-augmented image
- 37 search’, in *Proceedings of HCI 2011 - 25th BCS Conference on Human Computer Interaction*.
- 38 British Computer Society, pp. 157–162.
- 39 Hild, K. E. *et al.* (2011) ‘An ERP-based brain-computer interface for text entry using rapid serial
- 40 visual presentation and language modeling’, in *ACL HLT 2011 - 49th Annual Meeting of the*
- 41 *Association for Computational Linguistics: Human Language Technologies, Proceedings of Student*
- 42 *Session*, pp. 38–43.
- 43 Hope, C. *et al.* (2013) ‘High throughput screening for mammography using a human-computer
- 44 interface with rapid serial visual presentation (RSVP)’, in Abbey, C. K. and Mello-Thoms, C. R.
- 45 (eds) *Proceedings of SPIE - The International Society for Optical Engineering*, p. 867303. doi:

- 1 10.1117/12.2007557.
- 2 Huang, L. *et al.* (2017) 'BHCR: RSVP target retrieval BCI framework coupling with CNN by a  
3 Bayesian method', *Neurocomputing*, 238, pp. 255–268. doi: 10.1016/j.neucom.2017.01.061.
- 4 Huang, Y. *et al.* (2007) 'A fusion approach for image triage using single trial erp detection.', in  
5 *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, p. pages 473  
6 –476.
- 7 Huang, Y. *et al.* (2008) 'Large-scale image database triage via EEG evoked responses', *IEEE  
8 International Conference on Acoustics, Speech and Signal Processing*, pp. 429–432.
- 9 Huang, Y. *et al.* (2011) 'A framework for rapid visual image search using single-trial brain evoked  
10 responses', *Neurocomputing*, 74(12-13), pp. 2041–2051. doi: 10.1016/j.neucom.2010.12.025.
- 11 Johnson, R. (1986) 'A triarchic model of P300 amplitude.', *Psychophysiology*, pp. 367–384. doi:  
12 10.1111/j.1469-8986.1986.tb00649.x.
- 13 Kindermans, P.-J. *et al.* (2014) 'Integrating dynamic stopping, transfer learning and language  
14 models in an adaptive zero-training ERP speller', *Journal of Neural Engineering*, 11(3), p. 035005.  
15 doi: 10.1088/1741-2560/11/3/035005.
- 16 Kranczoch, C., Debener, S. and Engel, A. K. (2003) 'Event-related potential correlates of the  
17 attentional blink phenomenon', *Cognitive Brain Research*, 17(1), pp. 177–187. doi: 10.1016/S0926-  
18 6410(03)00092-2.
- 19 Kumar, S. and Sahin, F. (2013) 'Brain computer interface for interactive and intelligent image  
20 search and retrieval', in *2013 High Capacity Optical Networks and Emerging/Enabling  
21 Technologies*. IEEE, pp. 136–140. doi: 10.1109/HONET.2013.6729772.
- 22 Lamme, V. A. F. and Roelfsema, P. R. (2000) 'The distinct modes of vision offered by feedforward  
23 and recurrent processing', *Trends in Neurosciences*, 23(11), pp. 571–579. doi: 10.1016/S0166-  
24 2236(00)01657-X.
- 25 Leutgeb, V., Schäfer, A. and Schienle, A. (2009) 'An event-related potential study on exposure  
26 therapy for patients suffering from spider phobia.', *Biological psychology*, 82(3), pp. 293–300. doi:  
27 10.1016/j.biopsycho.2009.09.003.
- 28 Lin Zhimin, Ying Zeng, Hui Gao, Li Tong, Chi Zhang, Xiaojuan Wang, Qunjian Wu, and B. Y.  
29 (2017) 'Multi-Rapid Serial Visual Presentation Framework for EEG-based Target Detection',  
30 *BioMed Research International*.
- 31 Liu, R. and Kwon, M. (2016) 'Integrating oculomotor and perceptual training to induce a  
32 pseudofovea: A model system for studying central vision loss.', *Journal of vision*. Association for  
33 Research in Vision and Ophthalmology, 16(6), p. 10. doi: 10.1167/16.6.10.
- 34 Lotte, F. *et al.* (2007) 'A review of classification algorithms for EEG-based brain–computer  
35 interfaces', *Journal of Neural Engineering*. IOP Publishing, 4(2), pp. R1–R13. doi: 10.1088/1741-  
36 2560/4/2/R01.
- 37 Lucas, H. D., Chiao, J. Y. and Paller, K. A. (2011) 'Why Some Faces won't be Remembered: Brain  
38 Potentials Illuminate Successful Versus Unsuccessful Encoding for Same-Race and Other-Race  
39 Faces', *Frontiers in Human Neuroscience*, 5, p. 20. doi: 10.3389/fnhum.2011.00020.
- 40 Luck, S. J. (2005) *An introduction to the event-related potential technique*. MIT Press.
- 41 Luck, S., Woodman, G. and Vogel, E. (2000) 'Event-related potential studies of attention.', *Trends  
42 in cognitive sciences*, 4(11), pp. 432–440. doi: Doi 10.1016/S1364-6613(00)01545-X.
- 43 Luo, W. *et al.* (2010) 'Three stages of facial expression processing: ERP study with rapid serial  
44 visual presentation.', *NeuroImage*. NIH Public Access, 49(2), pp. 1857–67. doi:

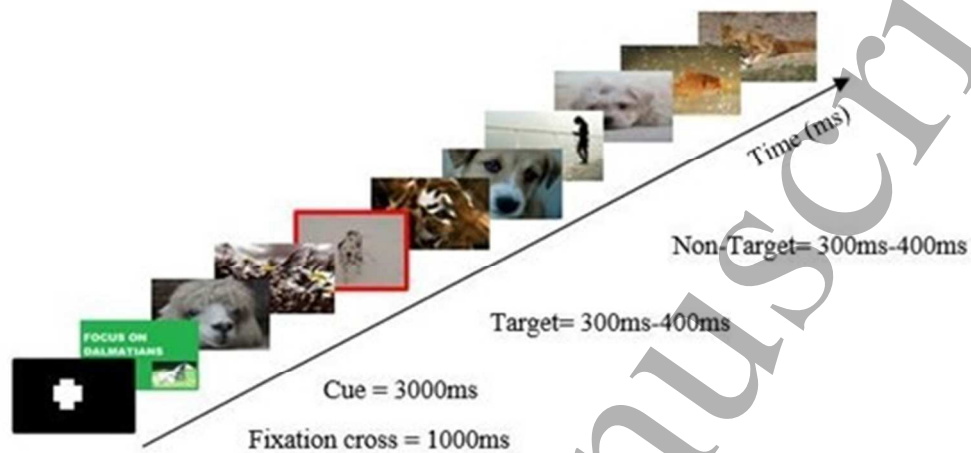
- 1 10.1016/j.neuroimage.2009.09.018.
- 2 MacKenzie, G. and Donaldson, D. I. (2007) 'Dissociating recollection from familiarity:  
3 Electrophysiological evidence that familiarity for faces is associated with a posterior old/new  
4 effect', *NeuroImage*, 36(2), pp. 454–463. doi: 10.1016/j.neuroimage.2006.12.005.
- 5 Maguire, J. F. and Howe, P. D. L. (2016) 'Failure to detect meaning in RSVP at 27 ms per picture',  
6 *Attention, Perception, & Psychophysics*. *Attention, Perception, & Psychophysics*, (April), pp. 1405–  
7 1413. doi: 10.3758/s13414-016-1096-5.
- 8 Manor, R. and Geva, A. B. (2015) 'Convolutional Neural Network for Multi-Category Rapid Serial  
9 Visual Presentation BCI', *Frontiers in Computational Neuroscience*, 9, p. 146. doi:  
10 10.3389/fncom.2015.00146.
- 11 Manor, R., Mishali, L. and Geva, A. B. (2016) 'Multimodal Neural Network for Rapid Serial Visual  
12 Presentation Brain Computer Interface.', *Frontiers in computational neuroscience*. *Frontiers Media*  
13 SA, 10, p. 130. doi: 10.3389/fncom.2016.00130.
- 14 Marathe, A. *et al.* (2015) 'Improved Neural Signal Classification in a Rapid Serial Visual  
15 Presentation Task using Active Learning', *IEEE Transactions on Neural Systems and*  
16 *Rehabilitation Engineering*, 4320(March 2016), pp. 1–1. doi: 10.1109/TNSRE.2015.2502323.
- 17 Marathe, A. R. *et al.* (2014) 'Confidence metrics improve human-autonomy integration',  
18 *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction -*  
19 *HRI'14*, pp. 240–241. doi: 10.1145/2559636.2563721.
- 20 Marathe, A. R. *et al.* (2015) 'The effect of target and non-target similarity on neural classification  
21 performance: a boost from confidence', *Frontiers in Neuroscience*, 9(August), pp. 1–11. doi:  
22 10.3389/fnins.2015.00270.
- 23 Marathe, A. R., Ries, A. J. and McDowell, K. (2014) 'Sliding HDCA: Single-trial eeg classification  
24 to overcome and quantify temporal variability', *IEEE Transactions on Neural Systems and*  
25 *Rehabilitation Engineering*, 22(March), pp. 201–211. doi: 10.1109/TNSRE.2014.2304884.
- 26 Mathan, S. *et al.* (2008) 'Rapid image analysis using neural signals', *Proceeding of the twenty-sixth*  
27 *annual CHI conference extended abstracts on Human factors in computing systems*. New York,  
28 New York, USA: ACM Press, p. 3309. doi: 10.1145/1358628.1358849.
- 29 Matran-Fernandez, A. and Poli, R. (2014) 'Collaborative Brain-Computer Interfaces for Target  
30 Localisation in Rapid Serial Visual Presentation', *6th Computer Science and Electronic*  
31 *Engineering Conference, CEEC 2014 - Conference Proceedings*, pp. 127–132.
- 32 Maurer, U., Rossion, B. and McCandliss, B. D. (2008) 'Category specificity in early perception:  
33 face and word n170 responses differ in both lateralization and habituation properties.', *Frontiers in*  
34 *human neuroscience*. *Frontiers Media SA*, 2, p. 18. doi: 10.3389/neuro.09.018.2008.
- 35 McCarthy, G. and Donchin, E. (1981) 'A metric for thought: a comparison of P300 latency and  
36 reaction time.', *Science (New York, N.Y.)*, 211(4477), pp. 77–80.
- 37 Ming, D. *et al.* (2010) 'Time-locked and phase-locked features of P300 event-related potentials  
38 (ERPs) for brain-computer interface speller', *Biomedical Signal Processing and Control*, 5(4), pp.  
39 243–251. doi: 10.1016/j.bspc.2010.08.001.
- 40 Mohedano, E. *et al.* (2014) 'Object Segmentation in Images using EEG Signals', in *Proceedings of*  
41 *the ACM International Conference on Multimedia - MM '14*. New York, New York, USA: ACM  
42 Press, pp. 417–426. doi: 10.1145/2647868.2654896.
- 43 Mohedano, E. *et al.* (2015) 'Exploring EEG for Object Detection and Retrieval', pp. 1–4. doi:  
44 10.1145/2671188.2749368.
- 45 Oken, B. S. *et al.* (2014) 'Brain-Computer Interface With Language Model–

- 1 Electroencephalography Fusion for Locked-In Syndrome', *Neurorehabilitation and Neural Repair*,  
2 28(4), pp. 387–394. doi: 10.1177/1545968313516867.
- 3 Oliva, A. (2005) 'Gist of the Scene', *Encyclopedia of Neurobiology of Attention*. Elsevier, San  
4 Diego, CA.
- 5 Orhan, U. *et al.* (2011) 'Fusion with language models improves spelling accuracy for ERP-based  
6 brain computer interface spellers.', *Conference proceedings : ... Annual International Conference of  
7 the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology  
8 Society. Conference*, 2011, pp. 5774–5777.
- 9 Orhan, U. *et al.* (2012) 'Improved accuracy using recursive bayesian estimation based language  
10 model fusion in ERP-based BCI typing systems.', *Conference proceedings : ... Annual International  
11 Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in  
12 Medicine and Biology Society. Conference*, 2012, pp. 2497–2500.
- 13 Orhan, U. *et al.* (2012) 'RSVP keyboard: An EEG based typing interface', *ICASSP, IEEE  
14 International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 645–648.  
15 doi: 10.1109/ICASSP.2012.6287966.
- 16 Orhan, U. *et al.* (2013) 'Offline analysis of context contribution to ERP-based typing BCI  
17 performance.', *Journal of neural engineering*, 10(6), p. 066003. doi: 10.1088/1741-  
18 2560/10/6/066003.
- 19 Paller, K. A. and Wagner, A. D. (2002) 'Observing the transformation of experience into memory.',  
20 *Trends in cognitive sciences*, 6(2), pp. 93–102.
- 21 Parra, L. *et al.* (2008) 'Spatiotemporal Linear Decoding of Brain State', *IEEE Signal Processing  
22 Magazine*, 25(1), pp. 107–115. doi: 10.1109/MSP.2008.4408447.
- 23 Parra, L. C. *et al.* (2005) 'Recipes for the linear analysis of EEG', *NeuroImage*, 28(2), pp. 326–341.  
24 doi: 10.1016/j.neuroimage.2005.05.032.
- 25 Polich, J. and Donchin, E. (1988) 'P300 and the word frequency effect', *Electroencephalography  
26 and Clinical Neurophysiology*, 70(1), pp. 33–45. doi: 10.1016/0013-4694(88)90192-7.
- 27 Poolman, P., Frank, R. M., Luu, P., Pederson, S. M., and Tucker, D. M. (2008) 'A single-trial  
28 analytic framework for eeg analysis and its application to target detection and classification.',  
29 *NeuroImage*, 42(2):787 – 798.
- 30 Potter, M. C. *et al.* (2002) 'Recognition memory for briefly presented pictures: The time course of  
31 rapid forgetting.', *Journal of Experimental Psychology: Human Perception and Performance*, 28,  
32 pp. 1163–1175. doi: 10.1037//0096-1523.28.5.1163.
- 33 Potter, M. C. *et al.* (2014) 'Detecting meaning in RSVP at 13 ms per picture.', *Attention, perception  
34 & psychophysics*, 76(2), pp. 270–9. doi: 10.3758/s13414-013-0605-z.
- 35 Primativo, S. *et al.* (2016) 'Perceptual and Cognitive Factors Imposing "Speed Limits" on Reading  
36 Rate: A Study with the Rapid Serial Visual Presentation', *PLOS ONE*. Edited by K. Paterson.  
37 Public Library of Science, 11(4), p. e0153786. doi: 10.1371/journal.pone.0153786.
- 38 Raymond, J. E., Shapiro, K. L. and Arnell, K. M. (1992) 'Temporary suppression of visual  
39 processing in an RSVP task: An attentional blink?', *Journal of Experimental Psychology: Human  
40 Perception and Performance*, 18, pp. 849–860. doi: 10.1037/0096-1523.18.3.849.
- 41 Rensink, R. A. (2000) 'When Good Observers Go Bad: Change Blindness, Inattentional Blindness,  
42 and Visual Experience', *Change*, 6(9), pp. 458–468.
- 43 Rivet, B. *et al.* (2009) 'xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–  
44 Computer Interface', *IEEE Transactions on Biomedical Engineering*, 56(8), pp. 2035–2043. doi:

- 1 10.1109/TBME.2009.2012869.
- 2 Rivet, B. and Souloumiac, A. (2013) 'Optimal linear spatial filters for event-related potentials based
- 3 on a spatio-temporal model: Asymptotical performance analysis', *Signal Processing*. Elsevier,
- 4 93(2), pp. 387–398.
- 5 Rosenthal, D. *et al.* (2014) 'Evoked neural responses to events in video', *IEEE Journal on Selected*
- 6 *Topics in Signal Processing*, 8(3), pp. 358–365. doi: 10.1109/JSTSP.2014.2313022.
- 7 Rousselet, G. A., Thorpe, S. J. and Fabre-Thorpe, M. (2004) 'How parallel is visual processing in
- 8 the ventral pathway?', *Trends in Cognitive Sciences*, 8(8), pp. 363–370. doi:
- 9 10.1016/j.tics.2004.06.003.
- 10 Rugg, M. D. *et al.* (1998) 'Dissociation of the neural correlates of implicit and explicit memory.',
- 11 *Nature*, 392(6676), pp. 595–598. doi: 10.1038/33396.
- 12 S, G. *et al.* (2014) 'Spatiotemporal representations of rapid visual target detection: a single-trial
- 13 EEG classification algorithm.', *IEEE transactions on bio-medical engineering*. IEEE Computer
- 14 Society, 61(8), pp. 2290–303. doi: 10.1109/TBME.2013.2289898.
- 15 Saavedra, C. and Bougrain, L. (2012) 'Processing Stages of Visual Stimuli and Event-Related
- 16 Potentials', *The NeuroComp/KEOpS'12 workshop*.
- 17 Sajda, P. *et al.* (2010) 'In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled
- 18 Computer Vision', *Proceedings of the IEEE*, 98(3), pp. 462–478. doi:
- 19 10.1109/JPROC.2009.2038406.
- 20 Sajda, P. *et al.* (2010) 'In a blink of an eye and a switch of a transistor: cortically coupled computer
- 21 vision', *Proceedings of the IEEE*, 98.
- 22 Sajda, P., Gerson, a. and Parra, L. (2003) 'High-throughput image search via single-trial event
- 23 detection in a rapid serial visual presentation task', *First International IEEE EMBS Conference on*
- 24 *Neural Engineering, 2003. Conference Proceedings.*, pp. 7–10. doi: 10.1109/CNE.2003.1196297.
- 25 Sasane, S. and Schwabe, L. (2012) 'Decoding of EEG Activity from Object Views: Active
- 26 Detection vs. Passive Visual Tasks'.
- 27 Serre, T., Oliva, A. and Poggio, T. (2007) 'A feedforward architecture accounts for rapid
- 28 categorization.', *Proceedings of the National Academy of Sciences of the United States of America*.
- 29 National Academy of Sciences, 104(15), pp. 6424–9. doi: 10.1073/pnas.0700622104.
- 30 Simons, D. J. and Levin, D. T. (1997) 'Change blindness', *Trends in Cognitive Sciences*, 1(7), pp.
- 31 261–267. doi: 10.1016/S1364-6613(97)01080-2.
- 32 Smith, M. E. (1993) 'Neurophysiological Manifestations of Recollective Experience during
- 33 Recognition Memory Judgments', *Journal of Cognitive Neuroscience*, 5(1), pp. 1–13. doi:
- 34 10.1162/jocn.1993.5.1.1.
- 35 Stoica, A. *et al.* (2013) 'Multi-brain fusion and applications to intelligence analysis', in *Proceedings*
- 36 *of SPIE - The International Society for Optical Engineering*, p. 87560N. doi: 10.1117/12.2016456.
- 37 Touryan, J. *et al.* (2011) 'Real-time measurement of face recognition in rapid serial visual
- 38 presentation', *Frontiers in Psychology*, 2(March), pp. 1–8. doi: 10.3389/fpsyg.2011.00042.
- 39 Ušćumlić, M., Chavarriaga, R. and Millán, J. D. R. (2013) 'An iterative framework for EEG-based
- 40 image search: robust retrieval with weak classifiers.', *PloS one*, 8(8), p. e72018. doi:
- 41 10.1371/journal.pone.0072018.
- 42 Vidal, J. J. (1973) 'Toward direct brain-computer communication.', *Annual review of biophysics*
- 43 *and bioengineering*, pp. 157–80. doi: 10.1146/annurev.bb.02.060173.001105.
- 44 Wang, J. *et al.* (2009) 'Brain state decoding for rapid image retrieval', *Proceedings of the seventeen*
- 45 *ACM international conference on Multimedia - MM '09*. New York, New York, USA: ACM Press,

- 1 p. 945. doi: 10.1145/1631272.1631463.
- 2 Ward, R., Duncan, J. and Shapiro, K. (1997) 'Effects of similarity, difficulty, and nontarget  
3 presentation on the time course of visual attention.', *Perception & psychophysics*, 59(4), pp. 593–  
4 600.
- 5 Waytowich, N. R. *et al.* (2016) 'Spectral Transfer Learning Using Information Geometry for a  
6 User-Independent Brain-Computer Interface', *Frontiers in Neuroscience*. Frontiers, 10, p. 430. doi:  
7 10.3389/fnins.2016.00430.
- 8 Weiden, M., Khosla, D. and Keegan, M. (2012) 'Electroencephalographic detection of visual  
9 saliency of motion towards a practical brain-computer interface for video analysis', in *Proceedings*  
10 *of the 14th ACM international conference on Multimodal interaction - ICMI '12*. New York, New  
11 York, USA: ACM Press, p. 601. doi: 10.1145/2388676.2388800.
- 12 Wolpaw, J. R. and Wolpaw, E. W. (2012) *Brain-computer interfaces: principles and practice*.  
13 Oxford University Press. doi: 10.1093/acprof:oso/9780195388855.001.0001.
- 14 Won, D.-O. *et al.* (2017) 'Shifting stimuli for brain computer interface based on rapid serial visual  
15 presentation', in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*.  
16 IEEE, pp. 40–41. doi: 10.1109/IWW-BCI.2017.7858152.
- 17 Yazdani, A. *et al.* (2010) 'The impact of expertise on brain computer interface based salient image  
18 retrieval', in *2010 Annual International Conference of the IEEE Engineering in Medicine and*  
19 *Biology*. IEEE, pp. 1646–1649. doi: 10.1109/IEMBS.2010.5626655.
- 20 Yonghong Huang *et al.* (2006) 'Comparison of Linear and Nonlinear Approaches on Single Trial  
21 ERP Detection in Rapid Serial Visual Presentation Tasks', in *The 2006 IEEE International Joint*  
22 *Conference on Neural Network Proceedings*. IEEE, pp. 1136–1142. doi:  
23 10.1109/IJCNN.2006.246818.
- 24 Yovel, G. and Paller, K. A. (2004) 'The neural basis of the butcher-on-the-bus phenomenon: when a  
25 face seems familiar but is not remembered', *NeuroImage*, 21(2), pp. 789–800. doi:  
26 10.1016/j.neuroimage.2003.09.034.
- 27 Yu, K. *et al.* (2011) 'A Bilinear Feature Extraction Method for Rapid Serial Visual Presentation  
28 Triage'. doi: 10.1109/IWBE.2011.6079025.
- 29 Yu, K. *et al.* (2014) 'The analytic bilinear discrimination of single-trial EEG signals in rapid image  
30 triage.', *PloS one*. Public Library of Science, 9(6), p. e100097. doi: 10.1371/journal.pone.0100097.
- 31 Zander, T. O. *et al.* (2010) 'Enhancing Human-Computer Interaction with Input from Active and  
32 Passive Brain-Computer Interfaces', in Tan, D. S. and Nijholt, A. (eds) *Brain -Computer Interfaces*.  
33 Springer London, pp. 181–199. doi: 10.1007/978-1-84996-272-8\_11.
- 34 Zander, T. O. and Kothe, C. (2011) 'Towards passive brain-computer interfaces: applying brain-  
35 computer interface technology to human-machine systems in general', *Journal of Neural*  
36 *Engineering*, 8(2), p. 025005. doi: 10.1088/1741-2560/8/2/025005.
- 37 Zhang, Y. *et al.* (2012) 'A novel BCI based on ERP components sensitive to configural processing  
38 of human faces.', *Journal of neural engineering*, 9(2), p. 026018. doi: 10.1088/1741-  
39 2560/9/2/026018.
- 40



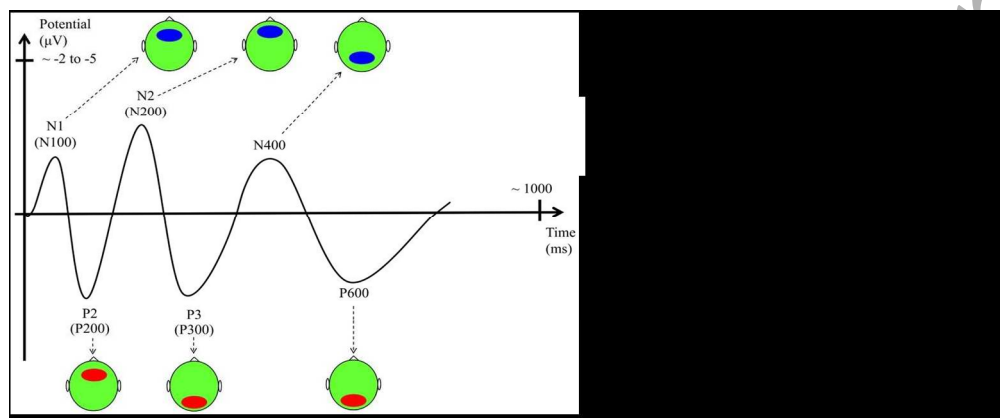


132x70mm (96 x 96 DPI)

Accepted Manuscript

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1036x427mm (96 x 96 DPI)

Accepted Manuscript