



**HAL**  
open science

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

Stephanie Lees, Natalie Dayan, Hubert Cecotti, Paul Mccullagh, Liam Maguire, Fabien Lotte, Damien Coyle

► **To cite this version:**

Stephanie Lees, Natalie Dayan, Hubert Cecotti, Paul Mccullagh, Liam Maguire, et al.. A Review of Rapid Serial Visual Presentation-based Brain-Computer Interfaces. *Journal of Neural Engineering*, 2017, pp.1-39. 10.1088/1741-2552/aa9817 . hal-01657643

**HAL Id: hal-01657643**

**<https://inria.hal.science/hal-01657643v1>**

Submitted on 7 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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  
11

---

## 12 **Abstract**

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

27  
28  
29 *Keywords— Rapid Serial Visual Presentation; Brain-Computer Interface; Event Related Potentials; Electroencephalography*  
30

---

## 31 32 33 **1. Introduction**

34 Rapid Serial Visual Presentation (RSVP) is the process of sequentially displaying images at the same  
35 spatial location at high presentation rates with multiple images per second e.g., with a stimulus onset  
36 asynchrony no greater than 500ms but often lower than 100ms i.e., >10 stimuli presented per second.  
37 Brain-computer interfaces (BCI) are communication and control systems that enable a user to execute  
38 a task via the electrical activity of the user's brain alone (Vidal, 1973). RSVP-based BCIs are a  
39 specific type of BCI that is used to detect target stimuli, e.g. letters or images, presented sequentially  
40 in a stream, by detecting brain responses to such targets. RSVP-based BCIs are considered as a viable  
41 approach to enhance human-machine symbiosis and offers potential for 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 overview  
5 of key current advancements and challenges; 4) provide design recommendations for researchers  
6 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 outlines  
13 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 large  
20 numbers of images/information are reviewed by professionals on a daily basis. Computers are unable  
21 to analyze and understand imagery as successfully as humans and manual analysis tools are slow  
22 (Mathan *et al.*, 2008; Gerson, Parra and Sajda, 2005). In studies carried out by Sajda *et al.* (2010),  
23 Poolman *et al.* (2008) and Bigdely-Shamlo *et al.* (2008), a trend of using RSVP-based BCIs for  
24 identifying targets within different image types has emerged. Research studies show the ability to use  
25 RSVP-based BCIs to drive a variety of visual search tasks including, in some circumstances, skills  
26 learned for visual recognition. Although the combination of RSVP and BCI has proven successful on  
27 several image sets, other research has attempted to establish whether or not greater efficiencies can  
28 be reached through the combination of RSVP-based BCIs and behavioural responses (Huang *et al.*,  
29 2007).

### 30 31 2.1. Event Related Potentials and their use in RSVP-based BCIs

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

42  
43 The stream of images presented within a RSVP paradigm comprise frequent non-target images and  
44 infrequent target images; different ERP components are associated with target and non-target stimuli  
45 (Bigdely-Shamlo *et al.*, 2008; M. Cohen, 2014; Sadjja *et al.*, 2014). BCI signal processing algorithms

1 are used to recognise spatio-temporal electrophysiological responses and link them to target image  
2 identification, ideally on a single trial basis (Manor, Mishali and Geva, 2016).

3  
4 The most commonly exploited ERP in RSVP-based BCI applications is the P300. The P300 appears  
5 at approximately 250-750 ms post target stimulus (Polich and Donchin, 1988; Leutgeb, Schäfer and  
6 Schienle, 2009; Ming *et al.*, 2010; Zhang *et al.*, 2012). As specified by (Polich and Donchin, 1988)  
7 during the P300 experiment (commonly referred to as the ‘Oddball’ paradigm), participants must  
8 classify a series of stimuli which fall into one of two classes: targets and non-targets. Targets appear  
9 more infrequently than non-targets (typically ~5-10% of total stimuli in the RSVP paradigm) and  
10 should be recognizably different. It is known that P300 responses can be suppressed in an RSVP task  
11 if the time between two targets is <0.5 seconds; which is known as attentional blink (Raymond,  
12 Shapiro and Arnell, 1992; Kranczioch, Debener and Engel, 2003). The amplitude and the latency of  
13 the P300 are influenced by the target discriminability and the target-to-target interval in the sequence.  
14 The latency of the P300 is affected by stimulus complexity (McCarthy and Donchin, 1981; Luck,  
15 Woodman and Vogel, 2000). The P300 amplitude can vary as a result of multiple factors (Johnson,  
16 1986), such as:

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

## 21 2.2. RSVP-based BCI amongst the BCI Classes

22  
23 BCI classes can be of three different types: active, reactive or passive (Zander *et al.*, 2010). An active BCI  
24 is purposefully controlled by the user through intentional modulation of neural activity, often  
25 independent of external events. Contrastingly, reactive BCIs generate outputs from neural activity  
26 evoked in response to external events, enabling indirect control by the user. Passive BCI makes use  
27 of implicit information and generate outputs from neural activity without purposeful control by the  
28 user. Active/reactive BCIs are commonly aimed at users with restricted movement abilities who  
29 intentionally try to control brain activity, whereas implicit or passive BCIs are more commonly  
30 targeted towards applications that are also of interest to able-bodied users (Zander and Kothe, 2011;  
31 Sasane and Schwabe, 2012).

## 32 2.2. RSVP-based BCI Presentation Modes

33  
34  
35  
36  
37 RSVP-based BCIs have two presentation modes: static mode in which images appear and disappear  
38 without moving; and moving mode where targets within short moving clips have to be identified  
39 (Sajda *et al.*, 2010; Cecotti, Eckstein and Giesbrecht, 2012; Weiden, Khosla and Keegan, 2012). Both  
40 presentation modes can be used with or without a button press. With a button press, users indicate  
41 manually, by pressing a button, when they observe a target stimulus. A button press is used to establish  
42 baseline performance, reaction time and/or to enhance performance (discussed further in section 5.1).

### 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 they  
5 are only presented briefly. There are a number of different possible instructions a participant may be  
6 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 preceding 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 images.  
11 In this mode, the rate of presentation should be slowed down (4/second) (Wang *et al.*, 2009).
- 12 • Immediately after image sequence presentation, the participant may be shown an image and  
13 asked: “did this image appear in the sequence you have just seen?” (Potter *et al.*, 2002).

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

### 27 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, RSVP  
31 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 the  
34 duration of stimuli presentation the ration of target images/total images will change (e.g. if images are  
35 being presented at a duration of 100ms then there must be a minimum of 5 images between targets 1  
36 and 2. In a sequence of 100 images there can be a maximum of 20 target images. Whereas if images  
37 are presented at 200ms this limits the maximum number of targets to 10/100 images in total).

38 *Change blindness* occurs when a participant is viewing two images that vary in a non-trivial fashion,  
39 and has to identify the image differences. Change blindness can occur when confronted by images,  
40 motion pictures, and real world interactions. Humans have the capacity to get the gist of a scene  
41 quickly but are unable to identify particular within-scene features (Simons and Levin, 1997; Oliva,  
42 2005). For example, when two images are presented for 100 ms each and participants are required to  
43 identify a non-trivial variation as the images are interchangeably presented, participants can take  
44 between 10-20s to identify the variation. This latency period in identifying non-trivial variations in

1 imagery can be augmented through use of distractors or motion pictures (Rensink, 2000). In the  
2 context of designing an RSVP paradigm change blindness is of interest, as it will take longer for a  
3 user to identify a target within an image if it does not pop out from the rest of the image. Distractors  
4 within the image or cluttered images, will increase the time it takes a user to recognize a target,  
5 reducing the performance of the RSVP paradigm.

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

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

### 21 22 **3. Bibliometric study of the RSVP related literature**

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

33 Papers were excluded for the following reasons: 1. the research protocol had insufficient detail; 2. key  
34 aspects needed to draw conclusive results were missing; 3. the spectrum of BCI research reported was  
35 too wide (i.e. review papers not specific to RSVP), 4. A ‘possible’ research application was described  
36 but the study was not actually carried out; 5. The study was a repeated study by original authors with  
37 only minor changes. Due to the immaturity of RSVP-based BCI as research topic, conference papers  
38 were not excluded. Inclusion of conference papers was considered important in order to provide a  
39 comprehensive overview of the state-of-the-art and trends in the field. Fifty-four papers passed initial  
40 abstract/title screening, these were then refined to the 45 most relevant papers through analysis of the  
41 entire paper contents. The date of the included publications ranged from 2003-2017.

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

20  
21

**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)	N/A

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
											~0.959 (Tv[B+NT], TN)	
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

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
				290					4	SWFP HDPCA HDCA	0.58-0.99 /0.99±0.55 0.99±0.67 0.87±0.05	91 N/A 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

	Reference	Mode	Application	Stimuli	Targets (%)	Duration (ms)	Size (px)	Visual angle (°)	Participants	Data analysis	ROC performance	Accuracy (%)
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
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

	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

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

2 When comparing RSVP-studies it is important to acknowledge that researchers use different  
3 measures of performance. Before going into depth about signal processing techniques (section 5.7) it  
4 is important to discuss, firstly, the variations in approaches used to measure performance. To  
5 encourage valid inter-study comparison within and across RSVP application types, it is crucial to  
6 emphasize that we are, on the whole, reporting classification accuracy when it is calculated in terms  
7 of the number of correctly classified trials. Classification accuracy can be swayed by the imbalanced  
8 target and non-target classes, with targets being infrequently presented e.g. with a 10% target  
9 prevalence, if all trials are classed as non-targets, correct classification rate would be 90%. Hence,  
10 ROC values are also reported in this review where relevant information was provided in publications  
11 reviewed.  
12

13 In the literature, there are many variations on how performance is estimated and reported. The studies  
14 cited in the current section provide examples of performance measure variations from the literature.  
15 The intention of Files and Marathe (2016a) was to develop a regression-based method to predict hit  
16 rates and error rates whilst correcting for expected mistakes. There is a need for such methods, due  
17 to uncertainty and difficulty in correctly identifying target stimuli. The regression method developed  
18 by Files and Marathe., (2016a), had relatively high hit rates which spanned 78.4% to 90.5% across  
19 all participants. Contrastingly, as a measure of accuracy, Sajda *et al.* (2010) used hit rates expressed  
20 as a fraction of total targets detected per minute. Sajda *et al.* (2010) discuss an additional experiment  
21 that employed ROC values as an outcome measure. In Fuhrmann *et al.* (2014), where the RSVP  
22 application was categorization based, accuracy was defined as, the number of trials in which the  
23 classifier provided the correct response, divided by the total number of available trials, with regards  
24 to target/non-target classification. Yazdani *et al.* (2010) were concerned with surveillance  
25 applications of RSVP-based BCI and used the F-measure to evaluate the accuracy of the binary  
26 classifier in use. Precision (fraction of occurrences flagged that are of relevant) and recall (fraction of  
27 relevant occurrences flagged) were reported as the F-measure considers both these values.  
28

29 Different variations in ROC value calculations were also discovered across the studies evaluated.  
30 Variability in the distribution of accuracy outcome measures is also founded upon whether the dataset  
31 is non-parametric e.g. median AUC is reported as opposed to the mean AUC (Matran-Fernandez and  
32 Poli, 2014). As a measure of accuracy, Rosenthal *et al.* (2014) conducted a bootstrap analysis, to  
33 show the sampled distribution of AUC values for HDCA classifiers where 1000 times over, labels  
34 were randomized, classifiers were trained and AUC values calculated through a “leaving one-out  
35 cross-validation” technique. Cecotti *et al.* (2012) presented a comparison of three class classifiers in  
36 a ‘one versus all’ strategy. The focus of Cecotti *et al.* (2012) was to compare the AUC to the volume  
37 under the ROC hyper-surface and the authors found a AUC of 0.878, which is suggestive of the  
38 possibility for discrimination between greater than two types of ERPs using single-trial detection.  
39 Huang *et al.* (2006) reported the AUC for session one of two experiments during button press trials.  
40 This paper demonstrates that with the three classifiers approach produces similar performance with  
41 AUC of >0.8 across the board (Huang *et al.*, 2006). Moreover, accuracy reportedly increases through  
42 collating evidence from two BCI users, and reportedly yielded a 7.7% increase in AUC compared to  
43 a single BCI user (Matran-Fernandez and Poli, 2014), using collaborative BCIs. This process was  
44 repeated 20 times to achieve an average accuracy measurement that would not be relatable to other  
45 studies included in the bibliometric analysis that involved average performance over single trial test.

1 Cecotti, Sato-Reinhold, *et al.* (2011) carried out a study where they compared varying target stimuli  
2 probability. Target probability has a significant effect on both behavioural performance and target  
3 detection. The best mean AUC is achieved with target probability of 0.10 AUC=0.82. The best target  
4 stimuli probability for optimal detection performance were 5% = 78.7%.

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

## 16 17 18 **5. Design parameters**

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

24  
25 When designing an RSVP paradigm, there are eight criteria that we recommend be taken into  
26 consideration:

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



## 1 5.1. Display and response modes

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

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

16  
17 Forty of the studies reported use static mode as a method of presentation, six of these papers used  
18 moving mode in conjunction with static mode while one study exclusively used moving mode.  
19 Moving mode is more complex than static mode as participants have to take in an entire scene rather  
20 than specific images. Moving mode uses motion onset in conjunction with the P300 for scenes in  
21 which the targets are moving, yielding a more realistic setting to validate RSVP-based BCIs (Weiden,  
22 Khosla and Keegan, 2012). All papers employing moving mode were found within the surveillance  
23 application category; this is unsurprising as the moving mode offers the opportunity to detect targets  
24 in realistic surveillance situations where movements of people or vehicles are of interest. For the other  
25 application areas i.e., medical, categorization etc. the static mode is likely to be the most appropriate.

26  
27 Won *et al.*, 2017 compared motion RSVP to standard RSVP, with the motion-type RSVP being the  
28 rapid presentation of letters of the alphabet, numbers 1-9 and a hyphen '-' used to separate words, in  
29 six different colour groups in one of six directions in line with the hands of a clock i.e. 2, 4, 6, 8, 10,  
30 12 whilst participants focused on a central point. An increase in performance accuracy with motion-  
31 type RSVP versus static-type was demonstrated, which could be accounted to the shorter latency and  
32 greater amplitudes of ERP components in the motion-type variation (Won *et al.*, 2017).

33  
34 Out of the studies found, 22 used a button press while 23 did not. 70% of surveillance applications  
35 used a button press. In categorization studies and face recognition studies the majority of applications  
36 used a button press. 89% of RSVP-speller applications did not use a button press. Typically, the BCI  
37 studies that involve spellers, focus on movement-free communication and high information transfer  
38 rates. Having a button press for confirmation of targets is not standard practice in such applications  
39 (Umut Orhan *et al.*, 2012; Oken *et al.*, 2014). In many of the studies that did not utilize a button press,  
40 researchers are focused on different aspects of the RSVP paradigm other than reaction time. For  
41 example, researchers focused on the comparison of two classification methods, image durations etc.  
42 (Sajda *et al.*, 2010; Cecotti, Eckstein and Giesbrecht, 2014). Combining EEG responses with button

1 press can improve accuracy although more signal processing is required in order to remove noise that  
2 occurs as a result of participant movement (Healy and Smeaton, 2011). Button press confirmation is  
3 unnecessary unless an assessment of physical reaction time is an important aspect of the study.  
4

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

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

25 Studies such as, Marathe *et al.*, (2014) propose methods for integrating button press information with  
26 EEG based RSVP classifiers to improve overall target detection performance. However, challenges  
27 arise when overlaying ERP and behavioural responses, such as issues concerning stimulation  
28 presentation speed and behavioural latency (Files and Marathe, 2016). Crucially, Files and Marathe,  
29 (2016) demonstrate that techniques for measuring real-time button press accuracy start to fail at higher  
30 presentation rates. Given evidence of human capacity for semantic processing during 20 Hz image  
31 streams (approximately 50 ms per image) and Response Times (RTs) often being an order of  
32 magnitude greater than EEG responses, button presses may be unsuitable for faster RSVP based  
33 image triaging.

34 Pending further studies investigating the reliability of fast detection of neural correlates, EEG based  
35 responses have the potential to exceed button press. However, it is not necessary for EEG based RSVP  
36 paradigms to surpass button press performance and evidence suggests that the complement of both  
37 modalities at comfortable lower presentation rates may indeed be the best approach. Nevertheless,  
38 ideally studies would contain an EEG only block and EEG plus button press block, where the button  
39 press follows the target and not the image burst. This would facilitate more accurate evaluation of

1 differences and correlates between behavioural and neural response times. Interesting, (Bohannon *et*  
2 *al.*, 2017), present a heterogeneous multi-agent system comprising computer vision, human and BCI  
3 agents, and showed that heterogeneous multi-agent image systems may achieve human level  
4 accuracies in significantly less time than a single human agent by balancing the trade-off between  
5 time-cost and accuracy. In such cases a human-computer interaction may occur in the form of button  
6 press if the confidence in the response of other, more rapid agents such as RSVP-BCI agents or  
7 computer vision algorithm is low for a particular sequence of stimuli.

## 8 9 10 5.2. Type of stimuli

11 Surveillance is the largest RSVP BCI system application reported in this review, reflected as such by  
12 the discussion length of this subsection (Sajda, Gerson and Parra, 2003; Erdogmus, Mathan and Pavel,  
13 2006; Gerson, Parra and Sajda, 2006; Poolman, P., Frank, R. M., Luu, P., Pederson, S. M., and Tucker,  
14 2008; Bigdely-Shamlo *et al.*, 2008; Sajda *et al.*, 2010; Huang *et al.*, 2011; Weiden, Khosla and  
15 Keegan, 2012; Cecotti, Eckstein and Giesbrecht, 2012; Matran-Fernandez and Poli, 2014; Rosenthal  
16 *et al.*, 2014; Yu *et al.*, 2014; Marathe, Ries and McDowell, 2014; A. R. Marathe *et al.*, 2015;  
17 Barngrover *et al.*, 2016; Cecotti, 2016; Files and Marathe, 2016).

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

31 In the categorization application type images are sorted into different groups (Cecotti, Kasper, *et al.*,  
32 2011; Cecotti, Sato-Reinhold, *et al.*, 2011). Fuhrmann, Alpert *et al.* (2014), conducted a study  
33 whereby five image categories were presented: cars, painted eggs, faces, planes, and clock faces  
34 (Sadja *et al.*, 2014). A second study in Fuhrmann, Alpert *et al.* (2014), containing target (cars) and  
35 non-target image (scrambled images of the same car) categories was conducted. In both RSVP  
36 experiments, the proposed Spatially Weighted Fisher Linear Discriminant – Principal Component  
37 Analysis (SWFP) classifier correctly classified a significantly higher number of images than the  
38 Hierarchical Discriminant Component Analysis (HDCA) algorithm. In terms of categorization,

1 empirical grounds were provided for potential intuitive claims, stating that target categorization is  
2 more efficient when: there is only one target image type; or distractors are scrambled variations of  
3 the target image as opposed to different images all together (Sajda *et al.*, 2014).

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

26  
27 RSVP-based BCIs that assist with finding targets within images to support clinical diagnosis has  
28 received attention (Stoica *et al.*, 2013), for example, in the development of more efficient breast  
29 cancer screening methods (Hope *et al.*, 2013). Hope *et al.* (2013) is the only paper evaluated from the  
30 field of medical image analysis and hence described in detail. During an initial sub-study participants  
31 were shown mammogram images, where target lesions were present or absent. In a subsequent study,  
32 target red or green stimuli were displayed among a set of random non-target blobs. These studies  
33 facilitated comparison between ‘masses’ and ‘no masses’ in mammograms, and strong color based  
34 images versus random distractors. Images were presented against a grey background in three second  
35 bursts of 30 images (100 ms per image). A difference in the amplitude of the P300 potential was  
36 observed across studies, with a larger amplitude difference between target and non-target images in  
37 the mammogram study. The researchers attributed this to the semantic association with mammogram  
38 images, in contrast to the lack thereof in the colored images-based study.

### 1 5.3. Total stimuli number and prevalence of target stimuli

2 The number of stimuli refers to the total number of stimuli i.e., the same stimulus can be shown  
3 several times. An exception to this is RSVP-speller studies where researchers only report on the  
4 number of symbols used i.e., 28 symbols - 26 letters of the alphabet, space and backspace (Hild *et al.*,  
5 2011). In the RSVP-speller studies reviewed, the number of times each symbol is shown is not  
6 explicit. RSVP-speller applications are likely to have significantly fewer stimuli than the other  
7 aforementioned applications as participants are spelling out a specific word or sentence, which only  
8 has a small number of target letters/words. The integration of language models into RSVP-speller  
9 applications enables ERP classifiers to utilize the abundance of sequential dependencies embedded in  
10 language to minimize the number of trials required to classify letters as targets or non-targets (Orhan  
11 *et al.*, 2011; Kindermans *et al.*, 2014)). Some systems, such as the RSVP keyboard (described in Hild  
12 *et al.*, 2011; Orhan, Hild, *et al.*, 2012a; and Oken *et al.*, 2014) display only a subset of available  
13 characters in each sequence. This sequence length can be automatically defined or be a pre-defined  
14 parameter chosen by the researcher. The next letter in a sequence become highly predictable in  
15 specific contexts, therefore it is not necessary to display every character in the RSVP-speller. Studies  
16 show that target characters are generally displayed more than once before the character is selected.  
17 The length of a sequence and the ratio of target to non-target stimuli can have an effect on the typing  
18 rate/performance. In an online study by Acqualagna *et al.*, 2011, participants were shown 30 symbols  
19 that were randomly shuffled 10 times before a symbol was selected through classification and  
20 presented on screen. Orhan *et al.*, 2012, carried out an offline study whereby 2 healthy participants  
21 where shown 3 sequences (consisting of 26 randomly ordered letters of the alphabet). Results of this  
22 study show that the number of correctly identified symbols more than doubled when using 3  
23 sequences instead of 1 sequence to identify targets.

24  
25 Task complexity is enhanced by the multiplicity of target categories. In Poolman, *et al.*, 2008) there  
26 were two blocks of target presentations; a helipad block with a 4% target prevalence; and a surface-  
27 to-air missile and anti-aircraft artillery block with a 1% target prevalence. Additionally, in (Cecotti *et*  
28 *al.*, 2012) the targets were 50% vehicles, 50% people, with 50% being stationary and 50% moving.  
29 Further to this, (Weiden, Khosla and Keegan, 2012) demonstrate that presenting kinetic images during  
30 the RSVP paradigm as opposed to stationary images increases performance of EEG-based detection,  
31 and that this is negatively correlated with the cognitive load associated with the presented stimuli. In  
32 RSVP-speller applications task complexity varies based on what instructions participants are given  
33 e.g., (1) participants may be asked to “spell dog”; (2) “type a word related to weather”; (3) participants  
34 can be given a word bank containing 20 words and asked to “spell a word found within this word  
35 bank”. Half of the RSVP-speller-based BCI studies evaluated involved user defined sequence length  
36 (instruction 2 and 3) (Acqualagnav *et al.*, 2010; Hild *et al.*, 2011; Umut Orhan *et al.*, 2012; Oken *et*  
37 *al.*, 2014), while the other half involved users been given a target word/sentence to spell (instruction  
38 1). If a participant has to remember the sentence or how to spell a long or unfamiliar word this can  
39 increase the complexity of a task (i.e., dog is much easier spelt than idiosyncrasy) (Primativo *et al.*,  
40 2016). Note however that these different complexities in instructions are only present for  
41 evaluation/training tasks with the RSVP-BCI spellers. For their real use, participants choose  
42 themselves what they want to spell. The RSVP-based text application allows the number of stimuli

1 before a target stimulus to be reduced (i.e. letters such as ‘z’ that are less commonly used can be  
2 shown less frequently).

3  
4 Excluding RSVP-speller applications, as it is already known that they do not require the same number  
5 of stimuli as the other applications, the number of stimuli used typically varied between studies from  
6 approximately 800 in the surveillance application study by (Sajda *et al.*, 2010) to 26,100 in a  
7 categorization application study by (Sajda *et al.*, 2014). The most common target stimuli percentage  
8 range was 1-10% found in 61% of the studies reviewed, followed by 11-20% then >20%. There are a  
9 number of studies that focus specifically on the percentage of target stimuli. In a study by (Cecotti,  
10 Sato-Reinhold, *et al.*, 2011), researchers investigated the influence of target probability when  
11 categorizing face and car images. In this study, researchers use spatially filtered EEG signals as the  
12 input for a Bayesian classifier. Using eight healthy participants, this method was evaluated using four  
13 probability of target stimuli conditions i.e., 0.05, 0.10, 0.25, or 0.50. It was found that the target  
14 probability had an effect on participant’s ability to detect targets and on behavioral performance. The  
15 best mean AUC (0.82) was achieved using the 0.1 probability condition. The results show that the  
16 percentage of targets shown in an RSVP paradigm has an effect on participants’ performance. As  
17 number and percentage of target stimuli used can have an effect on the complexity of a task, it is  
18 important to keep the percentage of targets <10% to evoke the P300 and maximize detection rates.  
19 This was proposed to be in line with well-established P3 measures, whereby bigger gaps between  
20 target trials reduce peak latency and increase amplitude (Gonsalvez and Polich, 2002).

#### 21 22 *5.4. Duration of stimuli presentation*

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

36  
37 Overall, these limited findings are suggestive of presentation rates >10Hz being infeasible for  
38 identification of neural correlates that allow successful identification of targets. Despite low a  
39 participant number in Sajda, Gerson and Parra, (2003), validation for this upper cut-off presentation  
40 rate may be provided by, Raymond, Shapiro and Arnell, 1992, where the attentional blink was first  
41 described. An RSVP paradigm was undertaken whereby the participant must register a target white

1 letter in a stream of black letters and a second target 'X' amongst this stream. It was found that if the  
2 'X' appeared within ~100-500ms of the initial target, errors in indicating whether the 'X' was present  
3 or not were likely to be made even when the first target was correctly identified (Raymond, Shapiro  
4 and Arnell, 1992). This is not to say that humans cannot correctly process information presented at  
5 >10Hz. Forster, (1970), has shown that participants can process words presented in a sentence at 16  
6 Hz (16 words per second). However, the sentence structure may have influenced the correct detection  
7 rate, which has an average of four words per second for simple sentence structures and three words  
8 for complex sentences. Detection rates improve when presented at a slower pace e.g., four relevant  
9 words per second, with masks (not relevant words) presented between relevant words. Additionally,  
10 Fine and Peli, 1995, showed that humans can process words at 20 Hz in an RSVP paradigm.

11  
12 Potter *et al.*, (2014) assessed the minimum viewing time needed for visual comprehension, using  
13 RSVP of a series of 6 or 12 pictures presented at between 13 and 80 ms per picture, with no inter-  
14 stimulus interval. They found that observers could determine the presence or absence of a specific  
15 picture even when the pictures in the sequence were presented for just 13 ms each. The results suggest  
16 that humans are capable of detecting meaning in RSVP at 13 ms per picture. However, the finding  
17 challenges established feedback theories of visual perception. Specifically, research assert that neural  
18 activity needs to propagate from the primary visual cortex (VI) to higher cortical areas and back to  
19 the primary visual cortex before recognition can occur at the level of detail required for an individual  
20 picture to be detected, Maguire and Howe, (2016). Maguire and Howe, (2016) support Potter *et al.*,  
21 (2014) in that the duration of this feedback process is likely  $\geq 50$ ms, and suggest that this is feasible  
22 based on work done by Lamme and Roelfsema, (2000). Explicitly, Lamme and Roelfsema, (2000)  
23 estimated that response latencies at any hierarchical level of the visual system are ~10ms. Therefore,  
24 assuming that a minimum of five levels must be traversed as activity propagates from V1 to higher  
25 cortical areas and back again, this feedback process is unlikely to occur in <50ms. However, Maguire  
26 and Howe, (2016) suggested a potential confound of Potter *et al.*, (2014) was that pictures in the RSVP  
27 sequence, on occasion, contained areas with no high-contrast edges and hence may not have  
28 adequately masked preceding pictures. Consequently, Maguire and Howe, (2016) replicated the  
29 study rectifying the edges to ensure high-contrast covering the entire image. They were unable to find  
30 any evidence that meaning can be detected in an RSVP stream at 13 ms, or even 27 ms, per image but  
31 at 53 and 80 ms this is possible. Upon this basis, the limits of RSVP processing could be reduced to  
32 a minimum of ~20Hz. Nonetheless, further study is needed to investigate the limits of human  
33 capability to rapidly distinguish target from non-target information, in comparison to the limit in  
34 detecting target related ERPs versus non-target ERPs at 20Hz presentation rates.

35  
36 In all three face recognition studies, each face image was displayed for 500ms (Cecotti, Sato-  
37 Reinhold, *et al.*, 2011; Touryan *et al.*, 2011; B. Cai *et al.*, 2013). In two of the studies there was no  
38 ISI (Cecotti, Sato-Reinhold, *et al.*, 2011; Touryan *et al.*, 2011), and in the other an ISI of 500ms  
39 given to ensure ample time for image processing (Bangyu Cai *et al.*, 2013). The speed at which face  
40 images were shown is reduced in comparison to the other RSVP applications. RSVP spellers most  
41 commonly use a duration of 400 ms, RSVP-spellers can benefit from slower stimulus duration with  
42 the incorporation of a language model to enable the prediction of relevant letters. The estimation of

1 performance can be challenging in the RSVP paradigm when the ISI is small, as assigning a  
 2 behavioural response (i.e.; button press) to the correct image cannot be done with certainty. A solution  
 3 to this problem is to assign behavioral responses to each image, therefore researchers are able to  
 4 establish hits or false alarms (Touryan *et al.*, 2011). When two targets are temporally adjacent with a  
 5 SOA of 80 ms, participants are able to identify one of the two targets but not both. SOA should be at  
 6 least 400 ms and target images should not be shown straight after each other (Raymond, Shapiro and  
 7 Arnell, 1992). Acqualagna *et al.* 2010, had a four factorial design looking at classification accuracy  
 8 when the letters presented as no-colour or colour letters at either 83 or 133 ms with an ISI of 33ms  
 9 (Acqualagna *et al.*, 2010). The number of sequence stimuli were presented for enhanced accuracy rate  
 10 in selecting letter of choice. After 10 sequences ~90% mean accuracy was reached in 133ms colour  
 11 presentation mode (100% for 6/9 participants). After 10 sequences in 133ms no colour presentation  
 12 mode ~80% mean accuracy was reached (100% in 3/9 participants). Whilst at presentation rates of  
 13 83ms mean accuracy rate was ~70% and the there was no significant effect of colour. This formulation  
 14 is based on the chance rate of 3.33% (i.e. 1 in 30). This implies that coloured letters enhances  
 15 performance accuracy but not past a certain speed of stimulus presentation.

16  
 17 There is likely a significant interaction between the difficulty of target identification and presentation  
 18 rate. For example, the optimal presentation rate for a given stimulus set is highly dependent on the  
 19 difficulty of identifying targets within that set (Ward, Duncan and Shapiro, 1997). Image sets with  
 20 low clutter, high contrast, no occlusion, and large target size are likely amenable to faster presentation  
 21 rates; while image sets with high clutter, low contrast, high levels of occlusion, with small target sizes  
 22 will require slower presentation rates (Rousselet, Thorpe and Fabre-Thorpe, 2004; Serre, Oliva and  
 23 Poggio, 2007; Hart *et al.*, 2013; Liu and Kwon, 2016). A more conclusive analysis of the effect of  
 24 stimulus presentation duration for each application type could be derived by varying presentation rate  
 25 durations between 100, 200, and 500 ms, whilst other parameters remain fixed. With regards to  
 26 temporal proximity of target images, 500ms should be taken to be the minimum to maximize  
 27 performance.  
 28

29 **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

30

31 *5.5. Image size/visual angle*

32 Another RSVP design aspect to be considered is stimulus size. There is a large variation in image  
 33 sizes ranging from 256×256 pixels in a categorization application to 960×600 pixels in a surveillance  
 34 applications. In general, surveillance applications use larger images than the other applications



1 described. The most common image size used is 500×500 pixels. This is only used in static  
2 surveillance applications and all surveillance studies using this image size achieved a high accuracy  
3 (>80%). The other applications used smaller image sizes such as 360×360 pixels and achieved high  
4 accuracies (i.e., 91% and 89.7%). Therefore, it can be concluded that for surveillance studies, image  
5 sizes should be at least 500×500 pixels, although for all other applications the image size may be  
6 smaller. A more complex task, where a target stimulus is presented in the background of a larger  
7 image eliciting the N2 ERP. Early components such as the P1 and N2 are sensitive to the spatial  
8 location of the stimuli (Saavedra and Bougrain, 2012).

9  
10 One issue with reporting only image size is that it is always relevant to the distance viewed from  
11 screen and its location on the screen with respect to the viewer i.e., the visual angle. The visual angle  
12 is the angle an image subtends at the eye, reported in degrees of arc. In a study by (Dias and Parra,  
13 2011) it was shown that participants performed best (90%) when the target stimulus was centered.  
14 Performance consistently decreased to 50% in all participants as target stimulus were placed further  
15 away from the center (4° of visual angle), this dropped further when target stimulus was placed at 8°  
16 of visual angle. Although performance drops significantly participants are still able to detect target  
17 stimulus shown in their peripheral visual field even at such rapid paces. Many papers report that the  
18 visual angle of the stimuli can have an effect on performance. As a general principle, targets must  
19 appear larger or be more distinct for detection at the outer edge of the visual field. The visual angle  
20 can thus be deemed the most important measure as it accounts for distance from screen, image location  
21 on screen and image size. Authors are therefore encouraged to report visual angle, as reporting image  
22 size alone is not useful without the availability of distance from screen. For RSVP-speller studies,  
23 none of the papers found reported on the size of the image or font, however some reported the visual  
24 angle.

## 25 26 *5.6 Target vs non-target Stimuli*

27 Many different types of target images have been identified within this review. The majority of  
28 research focuses on a two-class problem i.e., detecting target images in sequences of non-target  
29 images that are completely different from each other. However, in real-life situations, non-target  
30 images are likely to share some of the same characteristics as target images (A. R. Marathe *et al.*,  
31 2015). These presentation sequences appear to be more like moving images than static images. In (A.  
32 R. Marathe *et al.*, 2015) a more complex surveillance task was carried out where, in the first task,  
33 participants were required to detect targets when targets are the only infrequent image whilst, in the  
34 second task, targets were presented with non-targets (i.e. the target image could be found in the  
35 background of a larger image). Participants were required to ignore everything else in the image, a  
36 much more difficult task, and consequently the amplitude of the P300 was reduced. The results of this  
37 study found that the introduction of the infrequent non-target stimuli in the scene yielded a substantial  
38 slowing of the reaction time. Surveillance applications commonly use stimuli that are more complex  
39 where trained participants, such as intelligence analysts outperform novice participants, as they are  
40 able to give meaning to the stimuli. The RSVP-speller applications present their letters as images one  
41 at a time on screen (Hild *et al.*, 2011). Due to the nature of the RSVP paradigm, it is important that  
42 these letters are shown in a random order as participants pre-empting a target can have an effect on  
43 ERP responses (Oken *et al.*, 2014). Data categorization applications had the most variance between  
44 the different types of stimuli presented to a participant. However, these stimuli tend to be everyday  
45 items that participants can easily recognize.

1

## 2 5.7. Signal Processing

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

9

### 10 5.7.1. Pre-processing

11

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

22

### 23 5.7.2. Feature extraction

24

25 Feature extraction is applied to the data for dimensionality reduction and to extract discriminant and  
26 non-redundant features. It can be difficult to carry out feature extraction due to the low SNR in single  
27 trial analysis. Conventionally averaging over multiple repeated trials is often used to overcome this.  
28 Many studies employ spatial filtering to extract ERPs from EEG. Some of the spatial filtering methods  
29 used include principal component analysis (PCA) (Sajda, Gerson and Parra, 2003; S *et al.*, 2014),  
30 independent component analysis (ICA) (Bigdely-Shamlo *et al.*, 2008; Blankertz *et al.*, 2011; Kumar  
31 and Sahin, 2013), or the xDAWN algorithm which maximizes the SNR between target and non-target  
32 stimuli classes (Rivet *et al.*, 2009; Rivet and Souloumiac, 2013; Cecotti, Eckstein and Giesbrecht,  
33 2014). In the case of image triage where the intention is to classify single-trial ERPs, spatial filters  
34 are used to enhance SNR and exploit spatial redundancy (e.g. Parra *et al.*, 2005). Yu *et al.* 2011 went  
35 a step further by utilizing a methodology that considers spatial and temporal features to ensure  
36 augmented single-trial detection accuracy (Yu *et al.*, 2011). Bilinear common spatial pattern (BCSP)  
37 was suggested to outperform Common Spatial Patterns (CSP) filters (composite and common spatial  
38 pattern filters) (Yu *et al.*, 2011). It should be noted however that CSP spatial filters were not designed  
39 to classify ERP but to classify oscillatory EEG activity. CSP are indeed ignoring the EEG time course  
40 – i.e., the ERP – and are thus suboptimal for RSVP-BCI. We would recommend using spatial filters  
41 dedicated to ERP classification, such as xDAWN, which were used successfully in many RSVP-BCI.  
42 Spatial filters are normally only performed on high-density EEG data which might be impractical in  
43 certain real-life applications (Parra *et al.*, 2005). High-density EEG data has been reported to increase

1 accuracy (Ušćumlić, Chavarriaga and Millán, 2013). Table 4 shows the most common method used  
2 for different application types.

3  
4 Face recognition applications differ from other applications as face images evoke different ERPs, in  
5 addition to the P300. Faces typically evoke a N170 component that changes between targets and non-  
6 targets (Maurer, Rossion and McCandliss, 2008; Luo *et al.*, 2010). The vertex positive potential (VPP)  
7 is also associated with face recognition (Zhang *et al.*, 2012). The midfrontal FN400 and later parietal  
8 FP600 components have been associated with familiarity and recollection, respectively, (MacKenzie  
9 and Donaldson, 2007). Specifically, the amplitude of FP600 (a positive deflection >500 ms post-  
10 stimulus) was found to significantly correlate with the extent of face familiarity (Touryan *et al.*, 2011).  
11 The use of spatial filters that utilize spatial and temporal features may act as an advantage over  
12 conventional spatial filters that only exploit spatial redundancy e.g. (Yu *et al.*, 2011). However, spatial  
13 filters can only be performed on high-density EEG data which might be impractical in certain real-  
14 life applications (Parra *et al.*, 2005).

### 15 16 17 5.7.3. Classification

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

36  
37 For the surveillance application, SVM achieved the highest percentage accuracies (Huang *et al.*, 2011;  
38 Weiden, Khosla and Keegan, 2012). For the RSVP-speller application, the most common method of  
39 classification used was Regularized Discriminant Analysis (RDA). RDA achieved an AUC  
40 performance of 0.948-0.973(Orhan *et al.*, 2011). Step Wise Linear Discriminant Analysis (SWLDA)  
41 was also used in RSVP-speller applications with high AUC performance and accuracies (0.82, 0.84,  
42 86%, 89%) (Hope *et al.*, 2013). In face recognition applications, the best AUC performance was  
43 produced using an SVM classifier (Cai *et al.*, 2013). Within this review, only one medical application  
44 was identified (Hope *et al.*, 2013) and researchers achieved high accuracy using a Fisher Discriminant  
45 Analysis. BLDA classifiers were also used, achieving high levels of accuracy (79%). The Spatially

1 Weighted Fisher Discriminant (SWFP) algorithm outperformed the Hierarchical Discriminant  
2 Component Analysis (HDCA) algorithm by 10% in categorization applications. Touryan *et al.* 2011  
3 demonstrated that EEG classification methods applied to categorization procedures can be adapted to  
4 rapid face recognition procedures (Touryan *et al.*, 2011). Window sizes post stimulus onset of 128,  
5 256 and 512 ms were fed into the classifiers. AUC values (average AUC = 0.945) are reported for the  
6 customized PCA models utilized to describe the changes in ERPs seen between familiar (famous and  
7 personal) and novel faces displayed for 500ms at a time. It is the customized version of these models  
8 i.e. the models developed for each participant using only that participant's data, which were shown to  
9 improve classification performance through the acknowledgment of discrete variability in the  
10 windowed ERP components.

11  
12  
13 Many of the BCI algorithms presented in tables 1 and 2 are linear, enabling simple/fast training with  
14 resilience to overfitting often caused by noise, implying suitability to single-trial EEG data  
15 classification. Nonetheless, linear methods can limit feature extraction and classification, and non-  
16 linear methods e.g. neural networks, are more versatile in modelling data of greater variability, also  
17 implying suitability to single-trial EEG data classification (Erdogmus, Mathan and Pavel, 2006;  
18 Yonghong Huang *et al.*, 2006; Lotte *et al.*, 2007). The use of neural networks, in particular deep  
19 neural network for RSVP-based BCI framework represents an attractive venture, and have shown  
20 promise over standard linear methods (Manor, Mishali and Geva, 2016; Huang *et al.*, 2017). A  
21 convolution neural network was shown to outperform a two-step linear classifier using the same  
22 dataset (Sadja *et al.*, 2014; Manor and Geva, 2015).

23  
24 The majority of studies reviewed investigate the effectivity of classifiers in identifying single-trial  
25 EEG correlates for target stimuli presented through an RSVP type paradigm. However, the spatial  
26 filtering technique as well as the type of classifier used has an impact on proficiency in detecting EEG  
27 of single trials (Bigdely-Shamlo *et al.*, 2008; Cecotti, Eckstein and Giesbrecht, 2014). For example,  
28 Independent Component Analysis reportedly identifies and divides multiple classes of non-brain  
29 response artefacts associated with eye and head movements, which would be useful for EEG de-  
30 noising during real-world applications when operators are mobile (Bigdely-Shamlo, Vankov *et al.*,  
31 2008).

32  
33 Additionally, (Cecotti, Eckstein and Giesbrecht, 2014) evaluated three classifiers using three different  
34 spatial filtering methods, so all in all twelve techniques were compared for three different RSVP  
35 paradigms. Marathe *et al.*, 2015 utilized an Active Learning technique in a bid to reduce the training  
36 samples required to calibrate the classifier. Active Learning is a partially supervised iterative learning  
37 technique reducing the amount labeled data during required for training. Recalibration depends on  
38 parameters such as, human attentiveness, physical surroundings or task-specific factors. Looking at  
39 the real world applicability of RSVP based BCI systems, (Marathe *et al.*, 2015) build upon work  
40 addressing the issue of thorough recalibration required for real-time BCI system optimization.

41 There is growing interest in the use of transfer learning (TL) for calibration reduction or suppression  
42 to encourage the real-world applicability of BCIs (Wang *et al.*, 2015). With TL, the EEG data or  
43 classifiers from a given domain is transformed in order to be applied to another domain, hence  
44 transferring data/classifier from one domain to another, possibly increasing the amount of data for the

target domain (Wang et al., 2015). For RSVP-BCI, this typically consists in combining EEG data or classifiers from different participants, in order to classify EEG data from another participant, for which very little or even no calibration EEG data is available. An unsupervised transfer method, namely Spectral Transfer with Information Geometry (STIG), ranked and collated unlabelled predictions from a group of information geometry classifiers which was established through training on individual participants (Waytowich et al., 2016). Waytowich et al, 2016 showed that STIG can be used for single-trial detection in ERP-based BCIs, eliminating the requirement for taxing data collection for training. With access to limited data, STIG outperformed alternative zero-calibration and calibration reduction algorithms (Waytowich et al., 2016). Within the BCI community conventional TL approaches still necessitate training for each condition, however methodologies have been applied to eradicate the need for subject-specific data calibration, where large-scale data is leveraged from other participants (Wei et al., 2016). This demarcates the potential for single-trial classification via unsupervised TL and user-independent BCI technology deployment.

#### 5.8. Suggested parameters

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

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

Parameter	Surveillance	RSVP-speller	Face Rec	Categorization/ Medical
<b>Stimuli No.</b>	>5000	>5000	2000	>4000
<b>% Targets</b>	~5-10	≤5	~10	10-25
<b>Stimulus presentation duration (ms)</b>	100-200	500	500	100-200

<b>Target examples</b>	Helipads, planes, vehicles, people etc.	Letters	Faces	Animals, mammograms etc.
<b>ERP component</b>	P300	P300	N170	P300
<b>Feature Extraction</b>	XDAWN	-	XDAWN	BCSP/XDAWN
<b>Classifier</b>	BLDA, SVM, LP, SP	RDA/ SWLDA	SVM	BLDA

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

## 6. Discussion and Conclusion

With the increasing intensity in RSVP-based BCI research there is a need for further standardization of experimental protocols, to compare and contrast development of the different applications described in this review. This will aid the realization of a platform which researchers can use to develop RSVP paradigms and compare their results and determine the optimal RSVP based BCI paradigm for their application type. This paper presents a review of the available research, the defining elements of the research and a categorization approach that will facilitate coordination efforts among researchers in the field. Research has revealed that using a combination of RSVP with BCI technology, allows the detection of targets at an expedited rate without detriment to accuracy.

Understanding the neural correlates of visual information processing can create symbiotic interaction between human and machine through BCIs. Further development of RSVP-based BCIs will depend on both basic and applied research. Within the last five years, there have been advancements in how studies are reported and a sufficient body of evidence exists in support of the development and application of RSVP BCIs. However, there is a need for the research to be developed further and standardized protocols applied, so that comparative studies can be done for progressive research. Many ERP reviews have been carried out, however, this paper focuses on RSVP visual search tasks with high variability in targets and the parameters used. This paper gives guidelines on which parameters impact performance but also on which parameters should be reported so that studies can be compared. It is important that design aspects shown in Tables 1 and 2 are reported and described

1 within each research study. It has been shown that RSVP based BCIs can be used in processing target  
2 images in multiple application types with a low-target probability, but consistency of reporting  
3 method renders it difficult to truly compare one paradigm to another or one parameter-setup to  
4 another.

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

8  
9 To maximize relatability to pre-existing literature in terms of keeping one feature that contributes to  
10 cognitive load constant, it is recommended that studies utilizing greater than one category type as  
11 targets to conduct the same study with just one target category in the first instance.

12  
13 For all applications, it is of course necessary to choose an epoch for single trial ERP classification  
14 corresponding to the temporal evolution of the most robust ERP components that are, on the whole,  
15 pre-established in the literature as associated with the specified task at hand i.e., target stimuli  
16 identification due to their infrequency, recognisability, relevancy or contents. However, whether the  
17 duration of stimuli presentation must extend beyond the latency between ERP component appearances  
18 relative to stimuli presentation is questionable.

19  
20 This review found a single medical application. More research in applying the RSVP-based BCI  
21 paradigm to high throughput screening within medicine is highly encouraged upon the basis that  
22 similarly complex imagery has been categorized relatively successfully in other applications e.g., side  
23 scan sonar imagery of mines or aircraft amongst birds eye view of maps in surveillance (Bigdely-  
24 Shamlo *et al.*, 2008; Barngrover *et al.*, 2016). The medical application of RSVP-based BCIs has  
25 immense potential in diagnostics and prognostics through recognition and tracking of established  
26 disease biomarkers, and accelerating high throughput health image screening.

27  
28 Studies utilized varying image sizes, visual angles and participant distance from screen. Researchers  
29 are encouraged to report visual angle as it accounts for both images size and distance of participant  
30 from screen. A potential way to facilitate uniformity of these variables is to utilize a head mounted  
31 display (HMD) or Virtual Reality (VR) headset such as an Oculus Rift (Foerster *et al.*, 2016). The  
32 rapid visual information processing capacity is heavily dependent on visual parameters and use of an  
33 HMD headset would enable standardization of viewing distance, room lighting and visual angle  
34 (Foerster *et al.*, 2016). Use of a VR headset could distort electrode positions, nonetheless this affect  
35 could be easily mitigated. BCIs employing motion-onset visual evoked potentials (mVEP) have been  
36 utilized with VR headsets in neurogaming, and shown to be feasible (Beveridge, Wilson and Coyle,  
37 2016). The mVEP responses were evaluated in relation to mobile, complex and varying graphics  
38 within game distractors (Beveridge, Wilson and Coyle, 2016). (Foerster *et al.*, 2016) used the virtual  
39 reality device Oculus Rift for neuropsychological assessment of visual processing capabilities. This  
40 VR device is head-mounted and covers the entire visual field, thereby shielding and standardizing the  
41 visual stimulation and therefore may improve test-retest reliabilities. Compared to a CRT screen  
42 performances, visual processing speed, threshold of conscious perception and capacity of visual  
43 working memory did not differ significantly using the VR headset. VR headsets may therefore be  
44 applicable for standardized and reliable assessment and diagnosis of elementary cognitive functions  
45 in laboratory and clinical settings and maximise the opportunity to compare visual processing

1 components between individuals and institutions and to establish statistical norm distributions.  
2 Recently, a new VR-EEG combined headset with electrodes embedded in occipital areas for ERP  
3 detection has been reported for neurogaming ([www.neurable.com](http://www.neurable.com)). RVSP-based BCI paradigms may  
4 therefore benefit from the head mounted visual displays however a vision obscuring headset may not  
5 be appropriate in some contexts as it could limit the ability of the users, e.g. a person with disabilities,  
6 to communicate with their peers and environment. Such a headset may prevent the expressive or  
7 receptive use of non-verbal communication skills such as eye movement and facial expressions that  
8 are vital for users with non-verbal communication skills.

9  
10 Advancements towards RSVP of targets during moving sequences have shown promising results,  
11 although it is more difficult to study movie clips since the stimulus start event is not as clear. A  
12 remaining challenge in this area is for researchers to design signal processing tools that can deal with  
13 imprecise stimulus beginning/end (Cecotti, 2015). However, an advantage of moving mode is that the  
14 target stimulus remains on the screen for longer than with static mode, allowing participants the  
15 opportunity to confirm a target stimulus. Moving stimuli studies to date have been limited to  
16 surveillance applications so there is a need for further investigation in this area. Just over half the  
17 papers used button press mode in conjunction with one of the other modes, as not all of the studies  
18 are concerned with comparing EEG responses to motor responses. It is important to develop a scale  
19 in order to rank the difficulty of tasks. This will enable the comparison of paradigms that are at the  
20 same level. The key outcomes of this study are shown in Table 4, provided as suggested guidelines.  
21 These are suggested parameters that may be useful to researchers when designing RSVP-based BCI  
22 paradigms within the different application types. From this review, we can conclude that using these  
23 parameters will enable more consistent performance for the different application types and will enable  
24 improved comparison with new studies.

25  
26 In acknowledgment of the need for standardization of parameters for RSVP-based BCI protocols,  
27 Cecotti, Satp-Reinhold *et al.*, 2011 raise an interesting proposal stating that other parameters could  
28 be automatically prescribed in accordance with the chosen target likelihood; such as the optimal ISI  
29 length, classifiers and spatial filters (Cecotti, Sato-Reinhold, *et al.*, 2011). Such an infrastructure for  
30 parameter choices does not currently exist with studies focusing on the impact of different parameters.  
31 Future studies would benefit from engaging with iterative changes in design parameters. This would  
32 allow for a comparative study of the different design parameters and enable the identification of  
33 parameters that most affect the experimental paradigm. A study involving increasing the rate of  
34 presentation until classification starts to deteriorate significantly for various types of stimulus  
35 categories may indicate the maximum possible speed of RSVP-BCI. Additionally, a future  
36 development for RSVP-based BCIs might be to use real life imagery with numerous distractor stimuli  
37 amongst the target stimuli. This is a more difficult task but it would enhance paradigm reliability to  
38 real-life applications. Hybridizing RSVP BCIs with other BCI paradigms has also started to receive  
39 more attention (Kumar and Sahin, 2013). Users of this system navigate using motor imagery  
40 movements (left, right, up and down). Search queries are spelt using the Hex-O-Speller and results  
41 retrieved from a web search engine may be fed back to the user using RSVP. This study shows the  
42 potential benefits of the RSVP paradigm and how it may be used in order to aid physically impaired  
43 users. Eye-tracking can be used as an outcome measure to assess and enhance RSVP stimuli and



1 presentation modes. Specifically, using eye tracking researchers can establish where the participant's  
2 gaze is focused during erroneous trials and explore correlations between gaze variability and  
3 performance. With the RSVP-based BCI paradigm there is much scope to evaluate different data  
4 types/imagery. This is a fast growing field with a promising future. There are multiple opportunities  
5 and a large array of potential RSVP-BCI paradigm setups. Researchers in the field are therefore  
6 recommended to consider the literature to date and the comparative framework proposed in this paper.  
7

## 8 **References**

- 9 Acqualagna, L. and Blankertz, B. (2011) 'A gaze independent spelling based on rapid serial visual  
10 presentation', in *2011 Annual International Conference of the IEEE Engineering in Medicine and  
11 Biology Society*. IEEE, pp. 4560–4563. doi: 10.1109/IEMBS.2011.6091129.
- 12 Acqualagnav, L. *et al.* (2010) 'A novel brain-computer interface based on the rapid serial visual  
13 presentation paradigm', in *2010 Annual International Conference of the IEEE Engineering in  
14 Medicine and Biology*. IEEE, pp. 2686–2689. doi: 10.1109/IEMBS.2010.5626548.
- 15 Barngrover, C. *et al.* (2016) 'A Brain-Computer Interface (BCI) for the Detection of Mine-Like  
16 Objects in Sidescan Sonar Imagery', *IEEE Journal of Oceanic Engineering*, 41(1), pp. 124–139. doi:  
17 10.1109/JOE.2015.2408471.
- 18 Beveridge, R., Wilson, S. and Coyle, D. (2016) '3D graphics, virtual reality, and motion-onset visual  
19 evoked potentials in neurogaming', in Damien Coyle (ed.) *Brain-Computer Interfaces: Lab  
20 Experiments to Real-World Applications*. Vol 228. Elsevier, pp. 329–353. doi:  
21 10.1016/bs.pbr.2016.06.006.
- 22 Bigdely-Shamlo, N. *et al.* (2008) 'Brain Activity-Based Image Classification From Rapid Serial  
23 Visual Presentation', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5),  
24 pp. 432–441. doi: 10.1109/TNSRE.2008.2003381.
- 25 Blankertz, B. *et al.* (2011) 'Single-trial analysis and classification of ERP components — A tutorial',  
26 *NeuroImage*, 56(2), pp. 814–825. doi: 10.1016/j.neuroimage.2010.06.048.
- 27 Bohannon, A. W. *et al.* (2017) 'Collaborative image triage with humans and computer vision', *2016  
28 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference  
29 Proceedings*, pp. 4046–4051. doi: 10.1109/SMC.2016.7844866.
- 30 Boksem, M. A. S., Meijman, T. F. and Lorist, M. M. (2005) 'Effects of mental fatigue on attention:  
31 an ERP study.', *Brain research. Cognitive brain research*, 25(1), pp. 107–16. doi:  
32 10.1016/j.cogbrainres.2005.04.011.
- 33 Cai, B. *et al.* (2013) 'A rapid face recognition BCI system using single-trial ERP', in *In Neural  
34 Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, p. (pp. 89–92).
- 35 Cai, B. *et al.* (2013) 'A rapid face recognition BCI system using single-trial ERP', in *2013 6th  
36 International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, pp. 89–92. doi:  
37 10.1109/NER.2013.6695878.
- 38 Cecotti, H., Sato-Reinhold, J., *et al.* (2011) 'Impact of target probability on single-trial EEG target  
39 detection in a difficult rapid serial visual presentation task.', *Conference proceedings : ... Annual  
40 International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE  
41 Engineering in Medicine and Biology Society. Conference, 2011*, pp. 6381–6384.
- 42 Cecotti, H., Kasper, R. W., *et al.* (2011) 'Multimodal target detection using single trial evoked EEG  
43 responses in single and dual-tasks.', *Conference proceedings : ... Annual International Conference of  
44 the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology*

1 *Society. Conference*, 2011, pp. 6311–4. doi: 10.1109/IEMBS.2011.6091557.

2 Cecotti, H. *et al.* (2012) ‘Multiclass classification of single-trial evoked EEG responses’, in *2012*  
3 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE,  
4 pp. 1719–1722. doi: 10.1109/EMBC.2012.6346280.

5 Cecotti, H. (2015) ‘Toward shift invariant detection of event-related potentials in non-invasive brain-  
6 computer interface’, *Pattern Recognition Letters*. Elsevier Ltd., 66, pp. 127–134. doi:  
7 10.1016/j.patrec.2015.01.015.

8 Cecotti, H. (2016) ‘Single-Trial Detection With Magnetoencephalography During a Dual-Rapid  
9 Serial Visual Presentation Task’, *IEEE Transactions on Biomedical Engineering*, 63(1), pp. 220–227.  
10 doi: 10.1109/TBME.2015.2478695.

11 Cecotti, H., Eckstein, M. P. and Giesbrecht, B. (2012) ‘Effects of performing two visual tasks on  
12 single-trial detection of event-related potentials.’, *Conference proceedings : ... Annual International*  
13 *Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine*  
14 *and Biology Society. Conference*, 2012, pp. 1723–1726.

15 Cecotti, H., Eckstein, M. P. and Giesbrecht, B. (2014) ‘Single-trial classification of event-related  
16 potentials in rapid serial visual presentation tasks using supervised spatial filtering.’, *IEEE*  
17 *transactions on neural networks and learning systems*. Institute of Electrical and Electronics  
18 Engineers Inc., 25(11), pp. 2030–42. doi: 10.1109/TNNLS.2014.2302898.

19 Chahine, G. and Krekelberg, B. (2009) ‘Cortical contributions to saccadic suppression’, *PLoS ONE*,  
20 4(9).

21 Chennu, S. *et al.* (2013) ‘The cost of space independence in P300-BCI spellers.’, *Journal of*  
22 *neuroengineering and rehabilitation*. BioMed Central Ltd., 10(1), p. 82. doi: 10.1186/1743-0003-10-  
23 82.

24 Cohen, M. (2014) *Analyzing Neural Time Series Data: Theory and practice*. MIT Press.

25 Cohen, M. X. (2014) *Analyzing Neural Time Series Data: Theory and Practice (Issues in Clinical*  
26 *and Cognitive Neuropsychology)*. The MIT Press.

27 Diamond, M. R., Ross, J. and Morrone, M. C. (2000) ‘Extraretinal control of saccadic suppression.’,  
28 *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(9), pp. 3449–  
29 55.

30 Dias, J. C. and Parra, L. C. (2011) ‘No EEG evidence for subconscious detection during Rapid Serial  
31 Visual Presentation’, *2011 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*.  
32 Ieee, pp. 1–4. doi: 10.1109/SPMB.2011.6120108.

33 Erdogmus, D., Mathan, S. and Pavel, M. (2006) ‘Comparison of Linear and Nonlinear Approaches  
34 on Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks’, *The 2006 IEEE*  
35 *International Joint Conference on Neural Network Proceedings*, pp. 1136–1142. doi:  
36 10.1109/IJCNN.2006.1716229.

37 Fawcett, T. (2006) ‘An introduction to ROC analysis’, *Pattern Recognition Letters*, 27(8), pp. 861–  
38 874.

39 Files, B. T. and Marathe, A. R. (2016) ‘A regression method for estimating performance in a rapid  
40 serial visual presentation target detection task.’, *Journal of neuroscience methods*, 258, pp. 114–23.  
41 doi: 10.1016/j.jneumeth.2015.11.003.

42 Fine, E. M. and Peli, E. (1995) ‘Scrolled and rapid serial visual presentation texts are read at similar  
43 rates by the visually impaired.’, *Journal of the Optical Society of America. A, Optics, image science,*  
44 *and vision*, 12(10), pp. 2286–92.

1 Foerster, R. M. *et al.* (2016) ‘Using the virtual reality device Oculus Rift for neuropsychological  
2 assessment of visual processing capabilities’, *Scientific Reports*. Nature Publishing Group, 6(1), p.  
3 37016. doi: 10.1038/srep37016.

4 Forster, K. I. (1970) ‘Visual perception of rapidly presented word sequences of varying complexity’,  
5 *Perception & Psychophysics*, 8(4), pp. 215–221. doi: 10.3758/BF03210208.

6 Gerson, A. D., Parra, L. C. and Sajda, P. (2005) ‘Cortical origins of response time variability during  
7 rapid discrimination of visual objects’, *NeuroImage*, 28(2), pp. 342–353.

8 Gerson, A. D., Parra, L. C. and Sajda, P. (2006) ‘Cortically coupled computer vision for rapid image  
9 search.’, *IEEE transactions on neural systems and rehabilitation engineering : a publication of the*  
10 *IEEE Engineering in Medicine and Biology Society*, 14(2), pp. 174–9. doi:  
11 10.1109/TNSRE.2006.875550.

12 Gonsalvez, C. L. and Polich, J. (2002) ‘P300 amplitude is determined by target-to-target interval.’,  
13 *Psychophysiology*, 39(3), pp. 388–96.

14 Guo, C., Voss, J. L. and Paller, K. A. (2005) ‘Electrophysiological correlates of forming memories  
15 for faces, names, and face–name associations’, *Cognitive Brain Research*, 22(2), pp. 153–164. doi:  
16 10.1016/j.cogbrainres.2004.08.009.

17 Hart, B. M. *et al.* (2013) ‘Attention in natural scenes: contrast affects rapid visual processing and  
18 fixations alike.’, *Philosophical transactions of the Royal Society of London. Series B, Biological*  
19 *sciences*. The Royal Society, 368(1628), p. 20130067. doi: 10.1098/rstb.2013.0067.

20 Healy, G. and Smeaton, A. F. (2011) ‘Optimising the number of channels in EEG-augmented image  
21 search’, in *Proceedings of HCI 2011 - 25th BCS Conference on Human Computer Interaction*. British  
22 Computer Society, pp. 157–162.

23 Hild, K. E. *et al.* (2011) ‘An ERP-based brain-computer interface for text entry using rapid serial  
24 visual presentation and language modeling’, in *ACL HLT 2011 - 49th Annual Meeting of the*  
25 *Association for Computational Linguistics: Human Language Technologies, Proceedings of Student*  
26 *Session*, pp. 38–43.

27 Hope, C. *et al.* (2013) ‘High throughput screening for mammography using a human-computer  
28 interface with rapid serial visual presentation (RSVP)’, in Abbey, C. K. and Mello-Thoms, C. R. (eds)  
29 *Proceedings of SPIE - The International Society for Optical Engineering*, p. 867303. doi:  
30 10.1117/12.2007557.

31 Huang, L. *et al.* (2017) ‘BHCR: RSVP target retrieval BCI framework coupling with CNN by a  
32 Bayesian method’, *Neurocomputing*, 238, pp. 255–268. doi: 10.1016/j.neucom.2017.01.061.

33 Huang, Y. *et al.* (2007) ‘A fusion approach for image triage using single trial erp detection.’, in  
34 *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, p. pages 473  
35 –476.

36 Huang, Y. *et al.* (2008) ‘Large-scale image database triage via EEG evoked responses’, *IEEE*  
37 *International Conference on Acoustics, Speech and Signal Processing*, pp. 429–432.

38 Huang, Y. *et al.* (2011) ‘A framework for rapid visual image search using single-trial brain evoked  
39 responses’, *Neurocomputing*, 74(12-13), pp. 2041–2051. doi: 10.1016/j.neucom.2010.12.025.

40 Johnson, R. (1986) ‘A triarchic model of P300 amplitude.’, *Psychophysiology*, pp. 367–384. doi:  
41 10.1111/j.1469-8986.1986.tb00649.x.

42 Kindermans, P.-J. *et al.* (2014) ‘Integrating dynamic stopping, transfer learning and language models  
43 in an adaptive zero-training ERP speller’, *Journal of Neural Engineering*, 11(3), p. 035005. doi:  
44 10.1088/1741-2560/11/3/035005.

45 Kranczioch, C., Debener, S. and Engel, A. K. (2003) ‘Event-related potential correlates of the

1 attentional blink phenomenon', *Cognitive Brain Research*, 17(1), pp. 177–187. doi: 10.1016/S0926-  
2 6410(03)00092-2.

3 Kumar, S. and Sahin, F. (2013) 'Brain computer interface for interactive and intelligent image search  
4 and retrieval', in *2013 High Capacity Optical Networks and Emerging/Enabling Technologies*. IEEE,  
5 pp. 136–140. doi: 10.1109/HONET.2013.6729772.

6 Lamme, V. A. F. and Roelfsema, P. R. (2000) 'The distinct modes of vision offered by feedforward  
7 and recurrent processing', *Trends in Neurosciences*, 23(11), pp. 571–579. doi: 10.1016/S0166-  
8 2236(00)01657-X.

9 Leutgeb, V., Schäfer, A. and Schienle, A. (2009) 'An event-related potential study on exposure  
10 therapy for patients suffering from spider phobia.', *Biological psychology*, 82(3), pp. 293–300. doi:  
11 10.1016/j.biopsycho.2009.09.003.

12 Lin Zhimin, Ying Zeng, Hui Gao, Li Tong, Chi Zhang, Xiaojuan Wang, Qunjian Wu, and B. Y.  
13 (2017) 'Multi-Rapid Serial Visual Presentation Framework for EEG-based Target Detection',  
14 *BioMed Research International*.

15 Liu, R. and Kwon, M. (2016) 'Integrating oculomotor and perceptual training to induce a  
16 pseudofovea: A model system for studying central vision loss.', *Journal of vision*. Association for  
17 Research in Vision and Ophthalmology, 16(6), p. 10. doi: 10.1167/16.6.10.

18 Lotte, F. *et al.* (2007) 'A review of classification algorithms for EEG-based brain–computer  
19 interfaces', *Journal of Neural Engineering*. IOP Publishing, 4(2), pp. R1–R13. doi: 10.1088/1741-  
20 2560/4/2/R01.

21 Lucas, H. D., Chiao, J. Y. and Paller, K. A. (2011) 'Why Some Faces won't be Remembered: Brain  
22 Potentials Illuminate Successful Versus Unsuccessful Encoding for Same-Race and Other-Race  
23 Faces', *Frontiers in Human Neuroscience*, 5, p. 20. doi: 10.3389/fnhum.2011.00020.

24 Luck, S. J. (2005) *An introduction to the event-related potential technique*. MIT Press.

25 Luck, S., Woodman, G. and Vogel, E. (2000) 'Event-related potential studies of attention.', *Trends  
26 in cognitive sciences*, 4(11), pp. 432–440. doi: Doi 10.1016/S1364-6613(00)01545-X.

27 Luo, W. *et al.* (2010) 'Three stages of facial expression processing: ERP study with rapid serial visual  
28 presentation.', *NeuroImage*. NIH Public Access, 49(2), pp. 1857–67. doi:  
29 10.1016/j.neuroimage.2009.09.018.

30 MacKenzie, G. and Donaldson, D. I. (2007) 'Dissociating recollection from familiarity:  
31 Electrophysiological evidence that familiarity for faces is associated with a posterior old/new effect',  
32 *NeuroImage*, 36(2), pp. 454–463. doi: 10.1016/j.neuroimage.2006.12.005.

33 Maguire, J. F. and Howe, P. D. L. (2016) 'Failure to detect meaning in RSVP at 27 ms per picture',  
34 *Attention, Perception, & Psychophysics*. Attention, Perception, & Psychophysics, (April), pp. 1405–  
35 1413. doi: 10.3758/s13414-016-1096-5.

36 Manor, R. and Geva, A. B. (2015) 'Convolutional Neural Network for Multi-Category Rapid Serial  
37 Visual Presentation BCI', *Frontiers in Computational Neuroscience*, 9, p. 146. doi:  
38 10.3389/fncom.2015.00146.

39 Manor, R., Mishali, L. and Geva, A. B. (2016) 'Multimodal Neural Network for Rapid Serial Visual  
40 Presentation Brain Computer Interface.', *Frontiers in computational neuroscience*. Frontiers Media  
41 SA, 10, p. 130. doi: 10.3389/fncom.2016.00130.

42 Marathe, A. *et al.* (2015) 'Improved Neural Signal Classification in a Rapid Serial Visual Presentation  
43 Task using Active Learning', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*,  
44 4320(March 2016), pp. 1–1. doi: 10.1109/TNSRE.2015.2502323.

- 1 Marathe, A. R. *et al.* (2014) ‘Confidence metrics improve human-autonomy integration’, *Proceedings*  
2 *of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI’14*, pp. 240–241.  
3 doi: 10.1145/2559636.2563721.
- 4 Marathe, A. R. *et al.* (2015) ‘The effect of target and non-target similarity on neural classification  
5 performance: a boost from confidence’, *Frontiers in Neuroscience*, 9(August), pp. 1–11. doi:  
6 10.3389/fnins.2015.00270.
- 7 Marathe, A. R., Ries, A. J. and McDowell, K. (2014) ‘Sliding HDCA: Single-trial eeg classification  
8 to overcome and quantify temporal variability’, *IEEE Transactions on Neural Systems and*  
9 *Rehabilitation Engineering*, 22(March), pp. 201–211. doi: 10.1109/TNSRE.2014.2304884.
- 10 Mathan, S. *et al.* (2008) ‘Rapid image analysis using neural signals’, *Proceeding of the twenty-sixth*  
11 *annual CHI conference extended abstracts on Human factors in computing systems*. New York, New  
12 York, USA: ACM Press, p. 3309. doi: 10.1145/1358628.1358849.
- 13 Matran-Fernandez, A. and Poli, R. (2014) ‘Collaborative Brain-Computer Interfaces for Target  
14 Localisation in Rapid Serial Visual Presentation’, *6th Computer Science and Electronic Engineering*  
15 *Conference, CEEC 2014 - Conference Proceedings*, pp. 127–132.
- 16 Maurer, U., Rossion, B. and McCandliss, B. D. (2008) ‘Category specificity in early perception: face  
17 and word n170 responses differ in both lateralization and habituation properties.’, *Frontiers in human*  
18 *neuroscience*. Frontiers Media SA, 2, p. 18. doi: 10.3389/neuro.09.018.2008.
- 19 McCarthy, G. and Donchin, E. (1981) ‘A metric for thought: a comparison of P300 latency and  
20 reaction time.’, *Science (New York, N.Y.)*, 211(4477), pp. 77–80.
- 21 Ming, D. *et al.* (2010) ‘Time-locked and phase-locked features of P300 event-related potentials  
22 (ERPs) for brain–computer interface speller’, *Biomedical Signal Processing and Control*, 5(4), pp.  
23 243–251. doi: 10.1016/j.bspc.2010.08.001.
- 24 Mohedano, E. *et al.* (2014) ‘Object Segmentation in Images using EEG Signals’, in *Proceedings of*  
25 *the ACM International Conference on Multimedia - MM ’14*. New York, New York, USA: ACM  
26 Press, pp. 417–426. doi: 10.1145/2647868.2654896.
- 27 Mohedano, E. *et al.* (2015) ‘Exploring EEG for Object Detection and Retrieval’, pp. 1–4. doi:  
28 10.1145/2671188.2749368.
- 29 Oken, B. S. *et al.* (2014) ‘Brain–Computer Interface With Language Model–Electroencephalography  
30 Fusion for Locked-In Syndrome’, *Neurorehabilitation and Neural Repair*, 28(4), pp. 387–394. doi:  
31 10.1177/1545968313516867.
- 32 Oliva, A. (2005) ‘Gist of the Scene’, *Encyclopedia of Neurobiology of Attention*. Elsevier, San Diego,  
33 CA.
- 34 Orhan, U. *et al.* (2011) ‘Fusion with language models improves spelling accuracy for ERP-based  
35 brain computer interface spellers.’, *Conference proceedings : ... Annual International Conference of*  
36 *the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology*  
37 *Society. Conference, 2011*, pp. 5774–5777.
- 38 Orhan, U. *et al.* (2012) ‘Improved accuracy using recursive bayesian estimation based language  
39 model fusion in ERP-based BCI typing systems.’, *Conference proceedings : ... Annual International*  
40 *Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine*  
41 *and Biology Society. Conference, 2012*, pp. 2497–2500.
- 42 Orhan, U. *et al.* (2012) ‘RSVP keyboard: An EEG based typing interface’, *ICASSP, IEEE*  
43 *International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 645–648.  
44 doi: 10.1109/ICASSP.2012.6287966.
- 45 Orhan, U. *et al.* (2013) ‘Offline analysis of context contribution to ERP-based typing BCI

- 1 performance.’, *Journal of neural engineering*, 10(6), p. 066003. doi: 10.1088/1741-  
2 2560/10/6/066003.
- 3 Paller, K. A. and Wagner, A. D. (2002) ‘Observing the transformation of experience into memory.’,  
4 *Trends in cognitive sciences*, 6(2), pp. 93–102.
- 5 Parra, L. *et al.* (2008) ‘Spatiotemporal Linear Decoding of Brain State’, *IEEE Signal Processing*  
6 *Magazine*, 25(1), pp. 107–115. doi: 10.1109/MSP.2008.4408447.
- 7 Parra, L. C. *et al.* (2005) ‘Recipes for the linear analysis of EEG’, *NeuroImage*, 28(2), pp. 326–341.  
8 doi: 10.1016/j.neuroimage.2005.05.032.
- 9 Polich, J. and Donchin, E. (1988) ‘P300 and the word frequency effect’, *Electroencephalography and*  
10 *Clinical Neurophysiology*, 70(1), pp. 33–45. doi: 10.1016/0013-4694(88)90192-7.
- 11 Poolman, P., Frank, R. M., Luu, P., Pederson, S. M., and Tucker, D. M. (2008) ‘A single-trial analytic  
12 framework for eeg analysis and its application to target detection and classification.’, *NeuroImage*,  
13 42(2):787 – 798.
- 14 Potter, M. C. *et al.* (2002) ‘Recognition memory for briefly presented pictures: The time course of  
15 rapid forgetting.’, *Journal of Experimental Psychology: Human Perception and Performance*, 28, pp.  
16 1163–1175. doi: 10.1037//0096-1523.28.5.1163.
- 17 Potter, M. C. *et al.* (2014) ‘Detecting meaning in RSVP at 13 ms per picture.’, *Attention, perception*  
18 *& psychophysics*, 76(2), pp. 270–9. doi: 10.3758/s13414-013-0605-z.
- 19 Primativo, S. *et al.* (2016) ‘Perceptual and Cognitive Factors Imposing “Speed Limits” on Reading  
20 Rate: A Study with the Rapid Serial Visual Presentation’, *PLOS ONE*. Edited by K. Paterson. Public  
21 Library of Science, 11(4), p. e0153786. doi: 10.1371/journal.pone.0153786.
- 22 Raymond, J. E., Shapiro, K. L. and Arnell, K. M. (1992) ‘Temporary suppression of visual processing  
23 in an RSVP task: An attentional blink?’, *Journal of Experimental Psychology: Human Perception*  
24 *and Performance*, 18, pp. 849–860. doi: 10.1037/0096-1523.18.3.849.
- 25 Rensink, R. A. (2000) ‘When Good Observers Go Bad: Change Blindness, Inattentional Blindness,  
26 and Visual Experience’, *Change*, 6(9), pp. 458–468.
- 27 Rivet, B. *et al.* (2009) ‘xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–  
28 Computer Interface’, *IEEE Transactions on Biomedical Engineering*, 56(8), pp. 2035–2043. doi:  
29 10.1109/TBME.2009.2012869.
- 30 Rivet, B. and Souloumiac, A. (2013) ‘Optimal linear spatial filters for event-related potentials based  
31 on a spatio-temporal model: Asymptotical performance analysis’, *Signal Processing*. Elsevier, 93(2),  
32 pp. 387–398.
- 33 Rosenthal, D. *et al.* (2014) ‘Evoked neural responses to events in video’, *IEEE Journal on Selected*  
34 *Topics in Signal Processing*, 8(3), pp. 358–365. doi: 10.1109/JSTSP.2014.2313022.
- 35 Rousselet, G. A., Thorpe, S. J. and Fabre-Thorpe, M. (2004) ‘How parallel is visual processing in the  
36 ventral pathway?’, *Trends in Cognitive Sciences*, 8(8), pp. 363–370. doi: 10.1016/j.tics.2004.06.003.
- 37 Rugg, M. D. *et al.* (1998) ‘Dissociation of the neural correlates of implicit and explicit memory.’,  
38 *Nature*, 392(6676), pp. 595–598. doi: 10.1038/33396.
- 39 S, G. *et al.* (2014) ‘Spatiotemporal representations of rapid visual target detection: a single-trial EEG  
40 classification algorithm.’, *IEEE transactions on bio-medical engineering*. IEEE Computer Society,  
41 61(8), pp. 2290–303. doi: 10.1109/TBME.2013.2289898.
- 42 Saavedra, C. and Bougrain, L. (2012) ‘Processing Stages of Visual Stimuli and Event-Related  
43 Potentials’, *The NeuroComp/KEOpS’12 workshop*.
- 44 Sajda, P. *et al.* (2010) ‘In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled Computer

1 Vision', *Proceedings of the IEEE*, 98(3), pp. 462–478. doi: 10.1109/JPROC.2009.2038406.

2 Sajda, P. *et al.* (2010) 'In a blink of an eye and a switch of a transistor: cortically coupled computer  
3 vision', *Proceedings of the IEEE*, 98.

4 Sajda, P., Gerson, a. and Parra, L. (2003) 'High-throughput image search via single-trial event  
5 detection in a rapid serial visual presentation task', *First International IEEE EMBS Conference on  
6 Neural Engineering, 2003. Conference Proceedings.*, pp. 7–10. doi: 10.1109/CNE.2003.1196297.

7 Sasane, S. and Schwabe, L. (2012) 'Decoding of EEG Activity from Object Views : Active Detection  
8 vs . Passive Visual Tasks'.

9 Serre, T., Oliva, A. and Poggio, T. (2007) 'A feedforward architecture accounts for rapid  
10 categorization.', *Proceedings of the National Academy of Sciences of the United States of America.*  
11 National Academy of Sciences, 104(15), pp. 6424–9. doi: 10.1073/pnas.0700622104.

12 Simons, D. J. and Levin, D. T. (1997) 'Change blindness', *Trends in Cognitive Sciences*, 1(7), pp.  
13 261–267. doi: 10.1016/S1364-6613(97)01080-2.

14 Smith, M. E. (1993) 'Neurophysiological Manifestations of Recollective Experience during  
15 Recognition Memory Judgments', *Journal of Cognitive Neuroscience*, 5(1), pp. 1–13. doi:  
16 10.1162/jocn.1993.5.1.1.

17 Stoica, A. *et al.* (2013) 'Multi-brain fusion and applications to intelligence analysis', in *Proceedings  
18 of SPIE - The International Society for Optical Engineering*, p. 87560N. doi: 10.1117/12.2016456.

19 Touryan, J. *et al.* (2011) 'Real-time measurement of face recognition in rapid serial visual  
20 presentation', *Frontiers in Psychology*, 2(March), pp. 1–8. doi: 10.3389/fpsyg.2011.00042.

21 Ušćumlić, M., Chavarriaga, R. and Millán, J. D. R. (2013) 'An iterative framework for EEG-based  
22 image search: robust retrieval with weak classifiers.', *PloS one*, 8(8), p. e72018. doi:  
23 10.1371/journal.pone.0072018.

24 Vidal, J. J. (1973) 'Toward direct brain-computer communication.', *Annual review of biophysics and  
25 bioengineering*, pp. 157–80. doi: 10.1146/annurev.bb.02.060173.001105.

26 Wang, J. *et al.* (2009) 'Brain state decoding for rapid image retrieval', *Proceedings of the seventeen  
27 ACM international conference on Multimedia - MM '09*. New York, New York, USA: ACM Press,  
28 p. 945. doi: 10.1145/1631272.1631463.

29 Ward, R., Duncan, J. and Shapiro, K. (1997) 'Effects of similarity, difficulty, and nontarget  
30 presentation on the time course of visual attention.', *Perception & psychophysics*, 59(4), pp. 593–  
31 600.

32 Waytowich, N. R. *et al.* (2016) 'Spectral Transfer Learning Using Information Geometry for a User-  
33 Independent Brain-Computer Interface', *Frontiers in Neuroscience*. Frontiers, 10, p. 430. doi:  
34 10.3389/fnins.2016.00430.

35 Weiden, M., Khosla, D. and Keegan, M. (2012) 'Electroencephalographic detection of visual saliency  
36 of motion towards a practical brain-computer interface for video analysis', in *Proceedings of the 14th  
37 ACM international conference on Multimodal interaction - ICMI '12*. New York, New York, USA:  
38 ACM Press, p. 601. doi: 10.1145/2388676.2388800.

39 Wolpaw, J. R. and Wolpaw, E. W. (2012) *Brain-computer interfaces : principles and practice*.  
40 Oxford University Press. doi: 10.1093/acprof:oso/9780195388855.001.0001.

41 Won, D.-O. *et al.* (2017) 'Shifting stimuli for brain computer interface based on rapid serial visual  
42 presentation', in *2017 5th International Winter Conference on Brain-Computer Interface (BCI)*.  
43 IEEE, pp. 40–41. doi: 10.1109/IWW-BCI.2017.7858152.

44 Yazdani, A. *et al.* (2010) 'The impact of expertise on brain computer interface based salient image  
45 retrieval', in *2010 Annual International Conference of the IEEE Engineering in Medicine and*

1 *Biology*. IEEE, pp. 1646–1649. doi: 10.1109/IEMBS.2010.5626655.  
2 Yonghong Huang *et al.* (2006) ‘Comparison of Linear and Nonlinear Approaches on Single Trial  
3 ERP Detection in Rapid Serial Visual Presentation Tasks’, in *The 2006 IEEE International Joint*  
4 *Conference on Neural Network Proceedings*. IEEE, pp. 1136–1142. doi:  
5 10.1109/IJCNN.2006.246818.  
6 Yovel, G. and Paller, K. A. (2004) ‘The neural basis of the butcher-on-the-bus phenomenon: when a  
7 face seems familiar but is not remembered’, *NeuroImage*, 21(2), pp. 789–800. doi:  
8 10.1016/j.neuroimage.2003.09.034.  
9 Yu, K. *et al.* (2011) ‘A Bilinear Feature Extraction Method for Rapid Serial Visual Presentation  
10 Triage’. doi: 10.1109/IWBE.2011.6079025.  
11 Yu, K. *et al.* (2014) ‘The analytic bilinear discrimination of single-trial EEG signals in rapid image  
12 triage.’, *PLoS one*. Public Library of Science, 9(6), p. e100097. doi: 10.1371/journal.pone.0100097.  
13 Zander, T. O. *et al.* (2010) ‘Enhancing Human-Computer Interaction with Input from Active and  
14 Passive Brain-Computer Interfaces’, in Tan, D. S. and Nijholt, A. (eds) *Brain -Computer Interfaces*.  
15 Springer London, pp. 181–199. doi: 10.1007/978-1-84996-272-8\_11.  
16 Zander, T. O. and Kothe, C. (2011) ‘Towards passive brain–computer interfaces: applying brain–  
17 computer interface technology to human–machine systems in general’, *Journal of Neural*  
18 *Engineering*, 8(2), p. 025005. doi: 10.1088/1741-2560/8/2/025005.  
19 Zhang, Y. *et al.* (2012) ‘A novel BCI based on ERP components sensitive to configural processing  
20 of human faces.’, *Journal of neural engineering*, 9(2), p. 026018. doi: 10.1088/1741-  
21 2560/9/2/026018.  
22