Single-trial classification of neural responses evoked in rapid serial visual presentation: Effects of stimulus onset asynchrony and stimulus repetition

H. Cecotti[1,2], M.P Eckstein[2], B. Giesbrecht[2]

Abstract— **Rapid serial visual presentation (RSVP) tasks, in which participants are presented with a continuous sequence of images in one location, have been used in combination with electroencephalography (EEG) in a variety of Brain-Machine Interface (BMI) applications. The RSVP task is advantageous because it can be performed at a high temporal rate. The rate of the RSVP sequence is controlled by the stimulus onset asynchrony (SOA) between subsequent stimuli. When used within the context of a BMI, an RSVP task with short SOA could increase the information throughput of the system while also allowing for stimulus repetitions. However, reducing the SOA also increases the perceptual degradation caused by presenting two stimuli in close succession, and it decreases the target-to-target interval (TTI), which can increase the cognitive demands of the task. These negative consequences of decreasing the SOA could affect on the EEG signal measured in the task and degrade the performance of the BMI. Here we systematically investigate the effects of SOA and stimulus repetition (r) on single-trial target detection in an RSVP task. Ten healthy volunteers participated in an RSVP task in four conditions that varied in SOA and repetitions (SOA=500 ms, r=1; SOA=250 ms, r=2; SOA=166 ms, r=3; and SOA=100 ms, r=5) while processing time across conditions was controlled. There were two key results: First, when controlling for the number of repetitions, single-trial performance increases when the SOA decreases. Second, when the repetitions were combined, the best performance (AUC=0.967) was obtained with the shortest SOA (100 ms). These results suggest that shortening the SOA in an RSVP task has the benefit of increasing the performance relative to longer SOAs, and it also allows a higher number of repetitions of the stimuli in a limited amount of time.**

I. INTRODUCTION

Brain-machine interfaces (BMIs) based on the detection of event-related potentials (ERPs) in the electroencephalogram (EEG) signal require a task that evokes a consistent response that can be decoded reliably and quickly. Many tasks and paradigms have served as the basis for BMIs, but one that has been used a lot in the recent literature is rapid serial visual presentation (RSVP) [1]–[7]. In an RSVP task, a sequence of stimuli is presented at a single location (e.g., fixation point). Each individual stimulus in the sequence is presented briefly, and then is replaced by the next item in the sequence. Typical BMI variants of the RSVP require participants to attend to the sequence, and monitor for rare target stimuli that require a response, or are simply to be counted amongst more frequently occurring non-target stimuli.

Despite the advantages of the RSVP tasks for BMI, they have low information transfer rate (ITR) [8], due to the poor accuracy that is achieved for single-trial detection. Increasing the ITR remains one of the many challenges that must be addressed before BMIs could become an effective tool for high behavioral performance of healthy people. The ITR depends on both the time that is used for evoking a brain response and also the performance of the classifier for detecting the brain evoked response. Hence, it is important to manage the right tradeoff between the time that is required to present a visual stimulus, which should evoke a brain response depending on the type of stimulus, and the classifier accuracy.

There are three main parameters of the RSVP task that could influence ITR, and are independent of the classifier performance for single-trial detection: the presentation rate of the RSVP sequence, target probability, and target-to-target interval (TTI). A natural solution for increasing the ITR would be to increase the presentation rate of the RSVP sequence. The presentation rate is determined by the stimulus onset asynchrony (SOA), and increasing the presentation rate would require reducing the SOA between the stimuli. The consequence of reducing the SOA would be that more information would be presented to the observer per unit time. Previous studies indicated that visual processing needed during a go/no-go categorization task can be achieved under 150 ms [9]. Reducing SOA has several important implications. First, reducing the SOA increases the perceptual difficulty of the task and increases the likelihood for errors [10], [11]. Second, decreasing the SOA implies decreasing the TTI. The P3 amplitude increases when target-probability decreases, or when there is an increment in the number of non-targets preceding the target. Reductions in the TTI reduce the amplitude of the P3, and reductions in P3 amplitude would likely result in a poor discriminant feature in the classifier. In [12], it was suggested that ISI and probability do not independently affect P3 amplitude and that TTI offers a strong explanation of the reported relations between P3 amplitude and both ISI and probability.

A second solution to increase the ITR would be to repeat the presentation of a visual stimulus to combine the classification scores and increase the robustness of the decision. This strategy is often used in BMI for increasing the detection of ERPs, like in the P300-speller [13], where

[1] School of Computing and Intelligent Systems, University of Ulster, Derry, Northern Ireland, UK h.cecotti@ulster.ac.uk

[2] Department of Psychological & Brain Sciences, Institute for Collaborative Biotechnologies, University of California Santa Barbara, Santa Barbara, CA, USA {firstname.lastname}@psych.ucsb.edu

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

several stimuli are presented in order to reliably select an item.

A third solution is to combine reductions in SOA with combining classifier scores from multiple presentations of the same stimulus. A key hurdle to overcome in this solution is to identify the best tradeoff between SOA and classification performance. A decrease of the SOA can allow the presentation of more images in a limited amount of time. Thus, decreasing the SOA can be used in two ways: First, for increasing the ITR of the BMI by decreasing the processing time, second, for repeating the visual stimuli without sacrificing time, by improving the detection by combining the classifier outputs corresponding to the presentation of two identical visual stimuli.

The purpose of the present study was to investigate the effect of the SOA on single-trial detection during an RSVP task and to determine the best tradeoff between the SOA and the number of repetitions (**r**) for improving the performance in a fixed amount of time. We addressed the different issues by testing ten healthy participants during an RSVP task with four conditions that differed in terms of the RSVP SOA, rate, and target repetitions.

II. Method

A. Participants

Ten healthy subjects (5 females, mean=19.5, sd=1) were recruited through the University of California, Santa Barbara (UCSB) online recruitment system and received either \$20 an hour or course credit for participation. All participants had normal or corrected-to-normal vision and provided informed consent prior to the experiment. The UCSB Human Subjects Committee approved all procedures.

B. Visual stimuli

Visual stimuli consisted of 900 color images (683×384 pixel). These images were taken from “Insurgency: Modern Infantry Combat” (Insurgency Team), a total conversion modification of the video game “Half-Life 2” (Valve corporation). The realistic images were separated into target scenes that contained a person (300 images) and non-target scenes that did not contain a person (600 images). Figure 1 depicts several examples of the images that were presented during the experiments. The images were presented on a 19 inch ViewSonic E90F CRT monitor with a resolution of 1024×768 pixels and a refresh rate of 60Hz. The images were centered on the screen (visual angle $\approx 26^o$). Participants were seated comfortably 60cm from the monitor in a darkened electromagnetically shielded chamber.

Fig. 1. Examples of visual stimuli (targets (left) vs. non-target (right)).

C. Procedure and design

The RSVP task was separated into different trials. At the beginning of each trial, a fixation cross was presented, when the observer was ready he or she initiated the sequence by pressing the enter button, after which the stimuli in the RSVP sequence were presented one after another in the same location. The task was to monitor the stream and to count the number of targets. The visual task was designed to be a target search for a rare item (a person in the image); the target probability was set to 10% and constant across trials. Each trial contained ten different images, one of them being a target. When there was the repetitions of the visual stimuli, images were presented in the same order for each repetition (0 ms interstimulus interval).

Each subject participated in an RSVP task performed under four conditions that correspond to four different SOA and four different number of repetition of the images. The number of repetitions was set in relation to the frequency rate to keep a constant processing time across conditions, and to avoid effects related to fatigue and the different duration of sessions. The set of parameters of the four conditions were: SOA=500 ms, f=2 Hz, r=1; SOA=250 ms, f=4 Hz, r=2; SOA=166 ms, f=6 Hz, r=3; and SOA=100 ms, f=10 Hz, r=5. In the next sections, we denote by $C_{f,r}$ the different conditions where f is the frequency and r is the number of repetitions. The parameters of the conditions were chosen such that the presentation added up time of a single visual stimulus is 500 ms across blocks for each of the four conditions. Participants performed two blocks for each condition. A block contained 100 trials, resulting in 1000, 2000, 3000, and 5000 images for the four conditions. The number of images that were contained in each trial was 10, 20, 30 and 50 for conditions $C_{2,1}$, $C_{4,2}$, $C_{6,3}$ and $C_{10,5}$, respectively. The order of the conditions was randomized and counterbalanced across participants.

D. Signal acquisition

The EEG signal was measured for each subject from 32 Ag/AgCl sintered electrodes mounted in an elastic Biosemi headcap with active electrodes. The 32 electrodes were subsampled from the 10-10 system [14]. Additional electrodes were placed at the right and left mastoids, as well as 1 cm lateral to the left and right external canthi (horizontal), and above and below each eye (vertical) for the electrooculogram (EOG). A Biosemi ActiveTwo EEG amplifier was used for recording the signal. The EEG signal was sampled at 256 Hz and referenced offline to the average mastoid signal.

E. Signal processing

To classify the brain-evoked responses corresponding to the presence of a target or a non-target image, a set of features were extracted from the EEG signal to determine the presence of an ERP related to the target. To isolate the brain-evoked response on the target, the signal was first bandpassed filtered (Butterworth filter of order 4) with cutoff frequencies at 1 and 10.66 Hz. After the signal was downsampled to an equivalent of 32Hz, we considered a time segment of 625 ms of post-stimulus data (20 sampling points), which includes the main ERP components (N2, P3).

The next step consisted of enhancing the relevant signal using the xDAWN spatial filtering approach [15], [16]. In this method, spatial filters are obtained through the Rayleigh quotient by maximizing the signal-to-signal plus noise ratio. The result of this process provides N_f spatial filters, that are ranked in terms of their SSNR. For the classification, the first four spatial filters were used. The Bayesian linear discriminant analysis (BLDA) [17], [18] classifier was considered for the detection, with the first four spatial filters as input. For the classification, we considered two training conditions, because the classifier performance depends on the number of samples that were used during training. In the first training condition, the same number of trials were used for each condition, this number corresponds to the number of trials present in the first condition with only one repetition. In the second condition, all the available trials were considered. Indeed, for comparing the different conditions based on the ERP features, the number of trials must be the same for training the classifiers. However, for evaluating globally the best target detection system based on the same calibration time, we consider all the trials that are available in each condition.

F. Performance evaluation

Classifier performance was evaluated by using the area under the receiver-operator characteristic (ROC) curve (AUC). In addition, we also consider the evaluation of how the images are ranked because RSVP task can be used for the triage of images. In a block of N images that contain an image representing a target, it is important to know how the target image is ranked. Therefore, we determine the ranking error as err$_{rank}$, which represents the rank of the target image after sorting the scores of the classifier in descending order. If the target image is well detected, it will have the best score err$_{rank} = 1$, in the worst case err$_{rank} = N$; during the experiment, $N = 10$ corresponds to the number of different images per trial. Furthermore, as post-processing can be used for the final decision, we determine the top_i score ($1 \leq i \leq n$), which indicates the accuracy by considering the first i best images. For the statistical analysis based on repeated measures analysis of variance (ANOVA), we report the mean (M), standard error of the mean (SEM), repeated measures t-statistic (t), p-value (p).

III. RESULTS

A. Single-trial detection

For comparing single-trial trial performance across conditions, we first report the performance by considering the training condition with the same number of trials for each condition. SOA had a significant effect on classifier performance ($F(3, 9) = 13.73$, $p < 10e - 5$), with M=0.707 (SEM=0.028) for $C_{2,1}$, M=0.813 (SEM=0.019) for $C_{4,2}$, M=0.812, (SEM=0.017) for $C_{6,3}$, and M=0.835 (SEM=0.018) for $C_{10,5}$. Post-hoc t-tests after a Bonferoni correction revealed that single-trial detection was better with $C_{4,2}$ than $C_{2,1}$ ($t_9 = 4.643$, $p = 0.012$), the same way $C_{6,3}$ was better than $C_{2,1}$ (t_9=3.553, p=0.006), $C_{10,5}$ was better than $C_{2,1}$ ($t_9 = 4.494$, $p = 0.0015$).

We now consider all the available trials for training in each condition, *i.e.*, , $C_{10,5}$ has five more trials than $C_{2,1}$. Classifier performance for each individual and the group mean are shown in Figure 2. There was no difference across conditions ($F(3, 9) = 2.89$, $p = 0.05$), with M=0.793 (SEM=0.021) for $C_{2,1}$, M=0.830 (SEM=0.019) for $C_{4,2}$, M=0.823 (SEM=0.018) for $C_{6,3}$, and M=0.853 (SEM=0.018) for $C_{10,5}$. These results show that it is possible to decrease the SOA to 100 ms during an RSVP task without any significant decrease a performance.

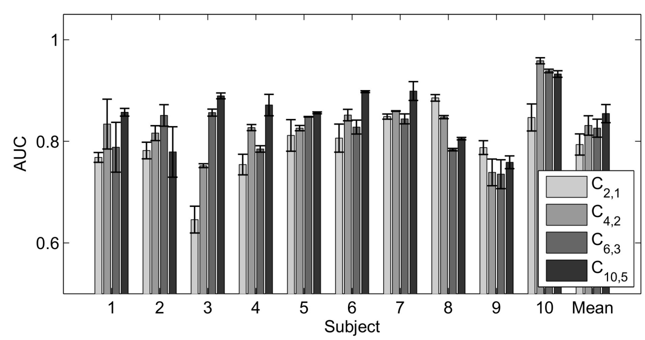

Fig. 2. AUC corresponding to single-trial detection, for each subject and each condition. The error bars indicate the standard error of the AUC across subjects.

B. Combined detection

The performance by considering the mean of the different decision scores across repetitions is presented in Fig. 3. As there is only one repetition for $C_{2,1}$, the results are identical with and without combination. There was a significant increase of the AUC by combining several trials: ($F(3, 9) = 30.64$, $p < 10 - 9$), with M=0.793 (SEM=0.066) for $C_{2,1}$, M=0.897 (SEM=0.049) for $C_{4,2}$, M=0.925 (SEM=0.049) for $C_{6,3}$, and M=0.967 (SEM=0.035) for $C_{10,5}$. Post-hoc t-tests after a Bonferoni correction indicated that increasing the number of repetitions improves target detection. Particularly, $C_{2,1}$ provides lower performance than $C_{4,2}$ ($t_9 = 6.26$, $p < 10e - 3$), $C_{6,3}$ ($t_9 = 5.18$, $p < 10e - 3$) and $C_{10,5}$ ($t_9 = 7.11$, $p < 10e - 3$). Similarly, $C_{10,5}$ provides better performance than $C_{4,2}$ ($t_9 = 5.15$, $p < 10e - 4$). The results show that it is better to have a short SOA, and to repeat the images, than to have a long SOA.

The error based on the rank for sorting blocks of 10 images, err$_{rank}$, was 2.86 for $C_{2,1}$, 1.93 for $C_{4,2}$, 1.70 for $C_{6,3}$, and 1.33 for $C_{10,5}$, showing that it possible to find the target among the two highest ranked responses. The accuracy by considering the n first best response is presented in Figure 4. A block of columns represents the accuracy if the target belongs to one of the n images that are supposed to contain the target.

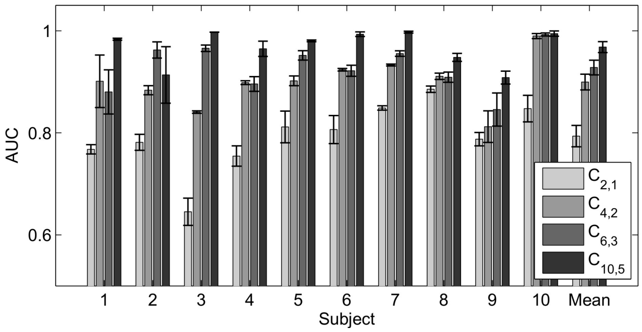

Fig. 3. AUC for each subject and each condition, after combination. The error bars indicate the standard error of the AUC across subjects.

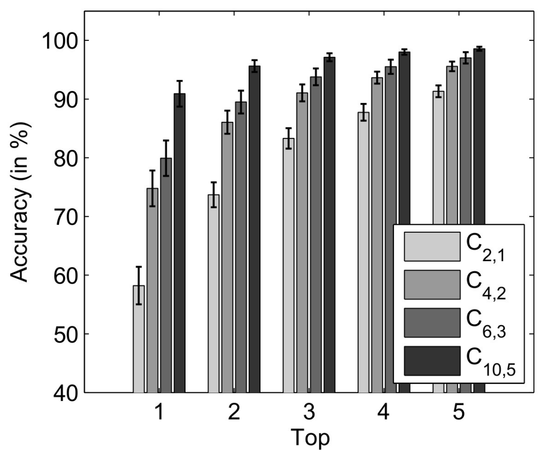

Fig. 4. Top results across subjects and for each condition, after combination of the decisions. The error bars indicate the standard deviation of the top results across subjects.

IV. Discussion and conclusion

This study had two goals. The first was to assess the effect of the SOA on single-trial detection during an RSVP task for target detection. We observed that the SOA has affected single-trial detection, with an increase of the performance when the SOA decreases. The second goal was to assess tradeoff between SOA and number of repetitions to increase the ITR during RSVP tasks. By considering experiments conducted with a constant amount of time, we have shown that it is better to present images at a faster rate and to repeat the image presentations.

At the BMI application level, the results have highlighted the advantage of considering a short SOA during an RSVP task. An SOA of 100 ms has two advantages. First, it allows the same performance than an SOA of 500 ms for single-trial detection. Second, a low SOA allows the repetition of the stimuli for increasing the performance. In application settings where the time is limited or when the ITR shall be improved, an SOA of 100 ms with 5 repetitions represent efficient parameters because the condition $C_{10,5}$ could provide an accuracy over 90%. Such parameters could be applied in other RSVP tasks, *i.e.*, virtual keyboard, target detection systems. While encouraging, what is unclear from our work is whether the short SOA approach is stable over time and whether it is robust to changes in task difficulty. Moreover, different results may be expected during long sessions when people cannot keep their attention for a long time. Whereas it is possible to decrease the SOA, and to obtain good performance with few sessions, further investigations need to be carried out with sessions lasting several hours.

Acknowledgment

This research was supported by the Institute for Collaborative Biotechnologies through contract W911NF-09-D-0001 from the U.S. Army Research Office.

References

[1] N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig, "Brain activity-based image classification from rapid serial visual presentation," *IEEE Trans. on Neural Systems and Rehab. Eng.*, vol. 16, no. 5, pp. 432–41, 2008.

[2] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, "Effects of performing two visual tasks on single-trial detection of event-related potentials," *34nd International IEEE Conf. EMBC*, pp. 1–4, 2012.

[3] H. Cecotti, M. Eckstein, and B. Giesbrecht, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Trans. Neural Networks and Learning Systems*, pp. 1–13, 2014.

[4] L. C. Parra, C. Christoforou, A. D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. G. Philiastides, and P. Sajda, "Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 95–115, 2008.

[5] E. A. Pohlmeyer, J. Wang, D. C. Jangraw, B. Lou, S. Chang, and P. Sajda, "Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases," *J. Neural Eng.*, vol. 8, p. 036025, 2011.

[6] M. G. Philiastides and P. Sajda, "EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making," *The Journal of Neuroscience*, vol. 27, no. 48, pp. 13 082–91, 2007.

[7] J. Touryan, L. Gibson, H. J. Horne, and P. Weber, "Real-time measurement of face recognition in rapid serial visual representation," *Frontiers in Psychology*, vol. 2, no. 42, pp. 1–8, 2011.

[8] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, pp. 767–791, 2002.

[9] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–2, 1996.

[10] K. H. Kim, J. H. Kim, J. Yoon, and J. K. Y, "Influence of task difficulty on the features of event-related potential during visual oddball task," *Neuroscience Letters*, vol. 445, no. 2, pp. 179–83, 2008.

[11] W. Prinzmetal, A. Zvinyatskovskiy, P. Gutierrez, and L. Dilem, "Voluntary and involuntary attention have different consequences: The effect of perceptual difficulty," *The quaterly journal of experimental psychology*, vol. 62, no. 2, pp. 352–369, 2009.

[12] R. J. Croft, C. J. Gonsalvez, C. Gabriel, and R. J. Barry, "Target-to-target interval versus probability effects on P300 in one-and two-tone tasks," *Psychophysiology*, vol. 40, pp. 322–328, 2003.

[13] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, pp. 510–523, 1988.

[14] F. Sharbrough, G. Chatrian, and R. P. e. a. Lesser, "Guidelines for standard electrode position nomenclature," *Bloomfield, IL: American EEG Society*, 1990.

[15] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain-computer interface," *IEEE Trans Biomed Eng.*, vol. 56, no. 8, pp. 2035–43, 2009.

[16] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, and J. Mattout, "A robust sensor selection method for P300 brain-computer interfaces," *J. Neural Eng.*, vol. 8, p. 016001, 2011.

[17] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–447, 1992.

[18] U. Hoffmann, J. Vesin, K. Diserens, and T. Ebrahimi, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 115–125, 2008.