Temporal Dynamics on Decoding Target Stimuli in Rapid Serial Visual Presentation using Magnetoencephalography

Chuncheng Zhang, Shuang Qiu, Shengpei Wang, Wei Wei, Huiguang He

***Abstract*— Rapid serial visual presentation (RSVP) is a high efficient paradigm in brain-computer interface (BCI). Target detection accuracy is the first consideration of RSVP-BCI. But the influence of different frequency bands and time ranges on decoding accuracy are still an open questions. Moreover, the underlying neural dynamic of the rapid target detecting process is still unclear. *Methods:* This work focused the temporal dynamic of the responses triggered by target stimuli in a static RSVP paradigm using paired structural Magnetic Resonance Imaging (MRI) and magnetoencephalography (MEG) signals with different frequency bands. Multivariate pattern analysis (MVPA) was applied on the MEG signal with different frequency bands and time points after stimuli onset. Cortical neuronal activation estimation technology was also applied to present the temporal-spatial dynamic on cortex surface. *Results:* The MVPA results showed that the low frequency signals (0.1 – 7 Hz) yield highest decoding accuracy, and the decoding power reached its peak at 0.4 second after target stimuli onset. The cortical neuronal activation method identified the target stimuli triggered regions, like bilateral parahippocampal cortex, precentral gyrus and insula cortex, and the averaged time series were presented.**

I. INTRODUCTION

Rapid serial visual presentation (RSVP) has been widely used in brain-computer interface (BCI) as a high efficient paradigm [1]. RSVP-BCI has been applied in many areas such as data categorization [2], face recognition [3] ,speller [4] and website evaluation [5]. The static RSVP refers a process to present static pictures at a high rate (e.g. 5-20 Hz) on a fixed position. The picture sequence consists of different types of visual stimuli, which can be classified as target and non-target stimuli [6]. In a RSVP paradigm, target stimuli can be designed to occur at a low probability, so that they will trigger event-related responses which can be detected by brain signals (e.g. electroencephalography (EEG) and magnetoencephalography (MEG)) [7, 8]. The detectable of target stimuli makes interfaceable between brain and computer using neural signal without actual motion.

The performance of RSVP-BCI is extremely rely on decoding accuracy of the target stimuli. Although the robust target event-related response can be obtained by averaging epoch responses [9], it is still challenge to directly detect target response using epoch response of single trail. As a result, pattern recognize algorithms have been used to discriminating target stimuli from non-target ones [10, 11, 25], known as multivariate pattern analysis (MVPA). Recent researches have focused on improving decoding accuracy by improving RSVP paradigm and machine leaning methods [10]. The dual [6] and motion-based [12] RSVP paradigm has been proposed to enhance the strength of target response. And xDAWN spatial filter [13] was used to enhance the signal-to-signal-plus-noise ratio of the signal. The convolutional neural network [14] and active learning [10] has also been used to improve the decoding performance.

However, little has been known about the temporal dynamics of the neural activity that triggered by target stimuli in RSVP. The influence of different frequency bands and time ranges on decoding accuracy are still an open questions. The EEG and MEG signal have high temporal resolution to support such rich frequency bands analysis. And MVPA applying on different time windows can be used to evaluate the decoding power in a temporal-resolution manner [7]. The investigation of temporal neural activity dynamics will benefit not only the decoding accuracy but also the understanding of neural processing of rapid object reorganization behind RSVP paradigm [15].

Besides the successful engineering applications of RSVP-BCI, the underlying neural activity is still unclear. Most RSVP-BCI are based on EEG [1], which has high temporal resolution and poor spatial resolution since it measures only the volume conduction currents [16]. The magnetoencephalography (MEG) signal has been used in RSVP research [6, 15], and shows advantage on uncovering high temporal resolution dynamic of brain activity [17]. Additionally, MEG signal can be paired with structural Magnetic Resonance Imaging (MRI) data to estimate cortical neuronal activation [18]. As a result, it provides a novel insight to assessing the temporal dynamic on the level of cortical neuronal activation [7].

In this paper, the temporal dynamic of target event-related responses in a static RSVP paradigm was investigated using paired structural MRI and MEG signal with different frequency bands. The MVPA was applied on MEG epoch responses to estimate the decoding power dynamic. The decoding power of target responses in several frequency bands were compared, and the comparisons of single time points were also presented. The origins of neuronal activation in the cortex were then estimated based on the averaged responses of target stimuli, to present the temporal-spatial dynamic of neuronal activation. And the brain regions activated by target stimuli were identified and their time series were presented.

*Research supported by This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 61906188 and Grant 81701785, and in part by the Strategic Priority Research Program of CAS under Grant XDB32040200, and in part by the China Postdoctoral Science Foundation 2019M650893.

C.Z., S.Q., S.W., W.W., and H.H. are with Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Science, Beijing, China. (e-mail: chuncheng.zhang@ia.ac.cn)

II. EXPERIMENT

A. Subjects and RSVP experiment design

We recruited 10 college students to participant in the RSVP experiment (7 males and 3 females, aged 23.79 ± 3.6). All the subjects are right-handed, and have normal or corrected-to-normal vision, and had not prior experience in RSVP experiment. The study was approved by the ethical committee of Institute of Automation, Chinese Academy of Science. All subjects gave a written informed consent and received payment for their participation. The study was approved by the local ethics committee (Institute of Automation Chinese Academy of Sciences).

The RSVP experiment contained 11 consecutive sessions, and a session started when subject push a button when he or she was ready. Each session contained 14 blocks. During a block, 100 pictures were shown to the subject in random ordered sequences at a rate of 10 Hz, with no gaps between two consecutive stimuli. In which, 4 pictures were target pictures and the other 96 pictures were non-target pictures. As a result, the ratio between target and non-target pictures was set to 4%. Subjects were required to push the button as soon as they saw a target picture. The picture sequences were not only randomized but also followed the principle that the lags between each two target pictures were larger than 2 seconds.

All the pictures were selected from a dataset consisted of 1400 colored street scenes pictures. Picture size was 500×500 pixels^2. Target pictures (56 pictures) were the pictures containing pedestrians, and non-target pictures (1344 pictures) were the pictures containing no pedestrians (see Fig 1 (a) and (b)).

All the RSVP experiment were performed during MEG scanning, .and after that subjects were scanned with MRI scanner. Thus, all anatomical landmarks digitized in the MEG study were rendered identifiable in the MR images.

B. MEG acquisition

During MEG-scanning experiment, subjects performed RSVP experiment in a MEG scanner. MEG recordings were conducted in a magnetically shielded room with a whole-head CTF MEG system with 272 channels (MISL-CTF DSQ-3500, Vancouver, BC, Canada) at the MEG Center of Institute of Biophysics, Chinese Academy of Sciences. (see Fig 1 (c)) Prior to data acquisition, three coils were attached to the left and right pre-auricular points and nasion of each participant, and a head localization procedure was performed before and after each acquisition to locate the participant's head relative to the coordinate system fixed to the MEG system. Participants were asked to lie in a supine position, a projection screen was used to present visual stimuli during recording. The distance between subject's eyes and the screen was 680 mm, and the picture size shown on the screen was 150×150 mm (thus subtending a visual angle of 12.6×12.6 degrees^2).

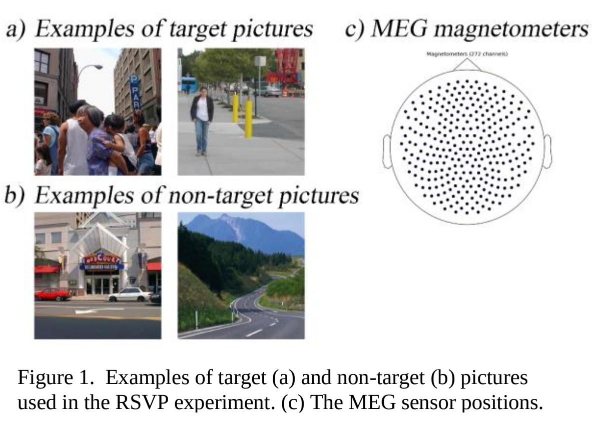

Figure 1. Examples of target (a) and non-target (b) pictures used in the RSVP experiment. (c) The MEG sensor positions.

C. Structural MRI acquisition

After MEG-scanning experiment, subjects were scanned with a 3.0 T MRI scanner (Siemens, Germany) at the MRI Center of Institute of Biophysics, Chinese Academy of Sciences, to acquire whole brain structural MRI scan of the subject. The field of view was 192×192 mm^2, and matrix size was 192×192, and slice thickness was 0.8 mm. Three fiduciary marks were placed in locations identical to the positions of the three coils used during MEG recording to allow for an accurate co-registration of the two data sets.

III. METHODS

A. MEG and MRI Preprocessing

MEG data was recorded at a sampling rate of 1200 Hz, filtered between 0 and 600 Hz. We preprocessed the data using MNE software [19]. The first step of preprocessing was suppressing artificial noise using ICA method. Since ICA is sensitive to low-frequency drifts, a 1 Hz high-pass filter was used to suppress low frequency signal prior to ICA fitting. Then, the sources with large skewness, kurtosis and variance scores were marked and zeroed out from raw data. The data was then down sampled to 100 Hz. The down sampled data was then filtered by different band-pass filters to fetch data of several frequency bands. The bands used in this research were Delta band ($1 - 4Hz$), Theta band ($4 - 7\ Hz$), Alpha band ($8 - 12\ Hz$), and two custom bands: U07 band ($0.1 - 7\ Hz$) and U30 band ($0.1 - 30Hz$).

The filtered data were subsequently used extract short-time signal of each visual stimuli, represented as epoch responses. We extracted peri-stimulus data from $-200\ ms$ to $1200\ ms$ with respect to stimulus onset. In this experiment, the onset means the time of picture being presented. Every epoch response was baseline-corrected to the mean ($-200\ to\ 0\ ms$) from each channel. To be clear, there are 56 target and 1344 non-target epoch responses in one session. And an epoch response could be formatted into a matrix, $epoch \in R^{272 \times 140}$, 272 rows represented 272 channels and 140 columns represented as 140 time points. Additionally, all the epoch responses respected to target stimuli were averaged to produce evoked responses for visualization and cortical neuronal activation estimation.

B. Multivariate pattern analysis

MVPA was performed to evaluate whether target and non-target pictures can be discriminated from MEG epoch responses for individual subject. Although MEG data of 11 sessions were acquired for each subject, we only used the second to eleventh session in MVPA, since data from the first session may not stable since it was the first session subject performed experiment in MEG scanner. As a result, MEG epoch responses from 10 sessions was used in MVPA.

The MVPA was applied in a 10-folder cross-validation protocol. For each folder, epoch responses from one session were used as testing data, and epoch responses from other sessions were used as training data. In each folder, training and testing dataset were absolutely separated. Since non-target samples were hundred times more than target samples, we discarded non-target samples that occurred within one second range of target samples in training session. MVPA in one folder was separated into 3 steps, they were feature extraction, classifier training and validation.

Feature extraction was then applied to training data, using xDAWN algorithm [6]. The xDAWN spatial filters were fitted using epoch responses in training data to maximize the signal-to-signal-plus-noise ratio. The number of components was set to 6. The xDAWN [13] algorithm was using MNE software. The filters were then applied to training and testing data. For the epoch responses, the spatial components corresponding to target epoch responses were remained and other components was zeroed out. The spatial filtered epoch responses were considered as feature responses. An SVM classifier with RBF kernel was then trained using feature responses in training data. We used sklearn software [20] to apply SVM algorithm. The prior parameter $gamma$ was set as $'scale'$, the reciprocal of number of training samples timing variance of training data. Since non-target samples were hundred times more than target samples, we set $class\_weight$ option as $'balanced'$ to obtain a meaningful classifier. Finally, the SVM classifier was tested using feature responses in testing data. The predicted labels were compared to the true labels to obtain f1 and recall score of target sample, and accuracy score.

The MVPA on single time points was also performed to investigate the temporal dynamic of decoding power. It followed the same procedure as above except the classifier training and testing process. The spatial filtered epoch responses were segmented into 140 time points, and thus 140 SVM classifiers were trained and tested separately. As a result, the temporal dynamic of the scores were yielded.

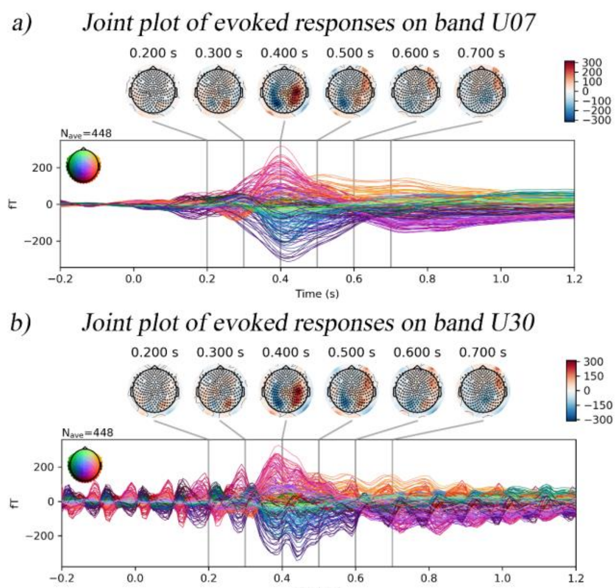

Figure 2. The joint plot of target evoked responses from one subject. It shows the time series of evoked responses of target stimuli and topographies from 0.2 to 0.7 seconds after target stimuli onset, under the frequency bands of 0.1 ~7 Hz (a) and 0.1 ~ 30 Hz (b).

C. Cortical neuronal activation estimation

While MVPA uncovered discriminative information encoded in MEG signals, one can be more interested in the underlying directly characterized origins of neuronal activation in the cortex.

Since we had acquired paired structural MRI and MEG data for every subject. The subject-specific cortical surfaces were build based on the MRI data. The surfaces building was using automatic segmentation of Freesurfer software (https://surfer.nmr.mgh.harvard.edu/). Then, a forward model was calculated to estimate the origins of the epoch responses by projecting the MEG data into cortical surfaces using the 'oct6' space. The coordinates of MEG and MRI data were incorporated into the new space. To obtain group neural activation, the subject-specific cortical surfaces were then transformed into average cortex in MNI space. The forward model computation and origins estimation was using MNE software. Finally, the evoked responses of MEG signals were mapped on the cortex using a dynamic statistical parametric mapping approach and time series of cortical neuronal activation were derived.

IV. RESULTS AND DISCUSSION

A. MEG signal visualization

The target stimuli evoked responses from one subject are shown in Fig 2. In U07 frequency band plot (Fig 2 (a)), it shows that no single peak window can be obtained from evoked responses. Despite temporal dynamic complexity, one can obtain the responses reached their peaks on the times from 0.3 to 0.5 seconds. In U30 band plot (Fig 2 (b)), a weaker waveform at 10 Hz frequency can be obtained, it is because the pictures were presented at the rate of 10 Hz.

B. MVPA scores

The MVPA scores from all five frequency bands are shown in Fig 3. The scores contain f1-score and recall score of target stimuli, and accuracy scores discriminating target and non-target stimuli. The different scores between each two bands were tested using paired t-test method. The star bars indicate the significant differences across bands. All the scores follow the same order across different bands, which is U07 > U30 > Delta > Theta > Alpha, and the scores of U07 bands are significantly higher than them of Delta, Theta and Alpha bands. The scores show no significant different between U07 and U30 band. The results infer that lower frequency signals contain stronger discriminating power.

The MVPA scores on time points are shown in Fig 4. The t-test method was applied to identify the time points that the scores significantly higher than chance level. The scores follow the order as the same as lateral. Epoch responses in U07 band yielded highest scores for almost all the single time points, and signals in Alpha band yielded the lowest. The peak scores are occurred at around 0.4 seconds after stimuli onset. The difference across bands happen after around 0.4 and 0.7 seconds. The signals from lower frequency bands (Delta, Theta and U07 bands) remain higher scores, when the scores of Theta band largely drop down.

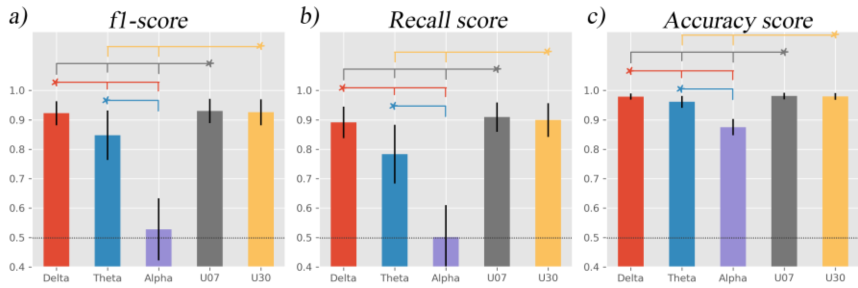

Figure 3. MVPA scores averaged across all the subjects and all the frequency bands. The scores are f1-score of target class (a), recall score of target class (b) and accuracy score (c). The dash lines indicate the chance level. The error bars indicate the standard variances. The horizon line respected to a bar (star symbol) indicates that it is significantly higher ($* \, p < 0.05$) than others pointed by the arrows.

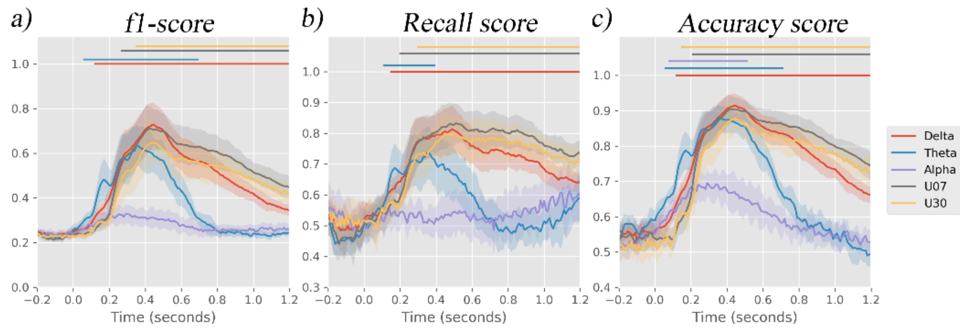

Figure 4. Temporal resolution MVPA scores averaged all the subjects and all the frequency bands. The scores are f1-score of target class (a), recall score of target class (b) and accuracy score (c). The shadows indicate the standard variances. The horizon bars indicate the times that the score is significantly larger than change level ($* \, p < 0.05$).

The results infer a temporal dynamic of decoding power on MEG epoch responses, not only on the manner of amplitude of single-channel but also on the discriminating power of multi-channel combining of MEG sensor signals.

C. *Cortical neuronal activation*

The averaged neural activation on the cortical surfaces from several time points are shown in Fig 5. The activation was estimated based on the evoked responses under U07 band. The values were tested using t-test method, the significant large values ($* \, p < 0.05$) were plotted on the surface.

Specifically to visual cortex, the activation in the ventral pathway were activated but the dorsal pathway, it may inferred that pedestrians detection RSVP task more involving the ventral visual stream rather than the dorsal one. The averaged time series of estimated cortical neuronal activation of strongly activated brain regions are shown in Fig 6. The brain regions were located based on cortical parcellation of Freesurfer software. The time series of regions were the average of all the surface vertex in the region. We identified the strong activated regions based on the rule that the mean amplitude of the time series were larger than two times of its standard variance value.

It shows that bilateral parahippocampal cortex were strongly activated. It is consistent with previous researches that parahippocampal cortex involve in processing which was inevitably triggered by identify pedestrians in target picture [21]. The precentral gyrus were mildly activated since the subjects were required to push a button as soon as they saw a target picture [22]. And the mild activation in precentral gyrus, which is also known as a part of supplementary motor area, may infer the motor preparing process to quickly push the button. Specific to left hemisphere, activation on cingulate cortex, caudal anterior cingulate gyrus and isthmus cingulate gyrus, were detected. It may also because of the target detection task in RSVP experiment, since the cingulate cortex has been extensively implicated in selective attention to target stimuli that requires responding [23].

The bilateral insula cortex were strongly activated. However, the insular cortex is a richly connected structure that functions as a cortical hub involved in multimodal sensory processing and autonomic control [24]. Their role in RSVP task was not clear, and worth further investigation.

V. Conclusion

In this paper, the temporal dynamic of target event-related responses in a static RSVP paradigm was investigated using MEG signal with different frequency bands. The MVPA results showed that the U07 band signals (0.1-7 Hz) yielded highest decoding accuracy, and further uncover the decoding power dynamic reached its peak at around 0.4 second after target stimuli onset. The cortical neuronal activation identified the target stimuli triggered regions, like bilateral parahippocampal cortex, precentral gyrus and insula cortex. And their roles in rapid target recognize processing of RSVP paradigm need further investigation.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61976209, Grant 61906188 and Grant 81701785, and in part by the Strategic Priority Research Program of CAS under Grant XDB32040200, and in part by the China Postdoctoral Science Foundation 2019M650893.

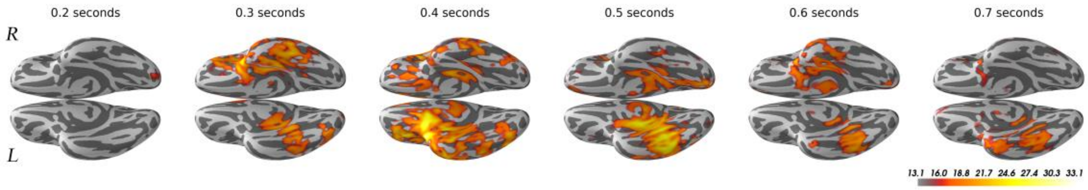

Figure 5. The estimated cortical neuronal activation of the evoked responses under U07 band from 0.2 to 0.7 seconds. The colored areas indicate the averaged activation across subjects, and only the significant large values (*p<0.05) are plotted on the surface.

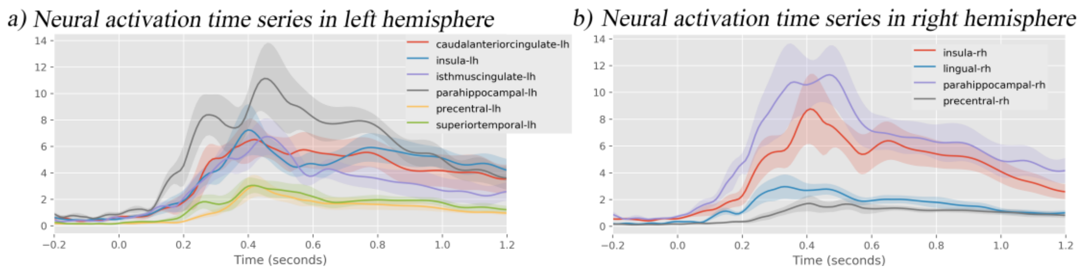

Figure 6. The time series of averaged neural activation of strong activated brain regions in left (a) and right (b) hemispheres. The strong activated regions were identified based on the rule that the mean amplitude of the time series were larger than two times of its standard variance value. The shadows indicate the standard variances.

REFERENCES

[1] S. Lees *et al.*, "A review of rapid serial visual presentation-based brain-computer interfaces," *J Neural Eng,* vol. 15, no. 2, p. 021001, Apr 2018, doi: 10.1088/1741-2552/aa9817.

[2] G. Guerrero and D. P. Calvillo, "Animacy increases second target reporting in a rapid serial visual presentation task," *Psychon Bull Rev,* vol. 23, no. 6, pp. 1832-1838, Dec 2016, doi: 10.3758/s13423-016-1040-7.

[3] H. Wang, P. Sun, C. Ip, X. Zhao, and S. Fu, "Configural and featural face processing are differently modulated by attentional resources at early stages: an event-related potential study with rapid serial visual presentation," *Brain Res,* vol. 1602, pp. 75-84, Mar 30 2015, doi: 10.1016/j.brainres.2015.01.017.

[4] H. W. Cao, K. Y. Yang, and H. M. Yan, "Character Decomposition and Transposition Processes of Chinese Compound Words in Rapid Serial Visual Presentation," *Front Psychol,* vol. 8, p. 483, 2017, doi: 10.3389/fpsyg.2017.00483.

[5] J. W. Owens, B. S. Chaparro, and E. M. Palmer, "Exploring website gist through rapid serial visual presentation," *Cogn Res Princ Implic,* vol. 4, no. 1, p. 44, Nov 20 2019, doi: 10.1186/s41235-019-0192-1.

[6] H. Cecotti, "Single-Trial Detection With Magnetoencephalography During a Dual-Rapid Serial Visual Presentation Task," *IEEE Trans Biomed Eng,* vol. 63, no. 1, pp. 220-7, Jan 2016, doi: 10.1109/TBME.2015.2478695.

[7] A. K. Robinson, T. Grootswagers, and T. A. Carlson, "The influence of image masking on object representations during rapid serial visual presentation," *Neuroimage,* vol. 197, pp. 224-231, Aug 15 2019, doi: 10.1016/j.neuroimage.2019.04.050.

[8] S. Lees, P. McCullagh, L. Maguire, F. Lotte, and D. Coyle, "Speed of rapid serial visual presentation of pictures, numbers and words affects event-related potential-based detection accuracy," *IEEE Trans Neural Syst Rehabil Eng,* Nov 18 2019, doi: 10.1109/TNSRE.2019.2953975.

[9] S. Yi *et al.*, "Emotional noun processing: an ERP study with rapid serial visual presentation," *Plos One,* vol. 10, no. 3, p. e0118924, 2015, doi: 10.1371/journal.pone.0118924.

[10] A. R. Marathe, V. J. Lawhern, D. Wu, D. Slayback, and B. J. Lance, "Improved Neural Signal Classification in a Rapid Serial Visual Presentation Task Using Active Learning," *IEEE Trans Neural Syst Rehabil Eng,* vol. 24, no. 3, pp. 333-43, Mar 2016, doi: 10.1109/TNSRE.2015.2502323.

[11] M. Zijing, Y. Wan Xiang, and H. Yufe, "Design of deep convolutional networks for prediction of image rapid serial visual presentation events," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2017, pp. 2035-2038, Jul 2017, doi: 10.1109/EMBC.2017.8037252.

[12] D. O. Won, H. J. Hwang, D. M. Kim, K. R. Muller, and S. W. Lee, "Motion-Based Rapid Serial Visual Presentation for Gaze-Independent Brain-Computer Interfaces," *IEEE Trans Neural Syst Rehabil Eng,* vol. 26, no. 2, pp. 334-343, Feb 2018, doi: 10.1109/TNSRE.2017.2736600.

[13] 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, Aug 2009, doi: 10.1109/TBME.2009.2012869.

[14] R. Manor and A. B. Geva, "Convolutional Neural Network for Multi-Category Rapid Serial Visual Presentation BCI," *Front Comput Neurosci,* vol. 9, p. 146, 2015, doi: 10.3389/fncom.2015.00146.

[15] Y. Mohsenzadeh, S. Qin, R. M. Cichy, and D. Pantazis, "Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway," *Elife,* vol. 7, Jun 21 2018, doi: 10.7554/eLife.36329.

[16] F. Destoky *et al.*, "Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope," (in English), *Neuroimage,* vol. 184, pp. 201-213, Jan 1 2019, doi: 10.1016/j.neuroimage.2018.09.006.

[17] P. Tewarie *et al.*, "Tracking dynamic brain networks using high temporal resolution MEG measures of functional connectivity," (in English), *Neuroimage,* vol. 200, pp. 38-50, Oct 15 2019, doi: 10.1016/j.neuroimage.2019.06.006.

[18] F. Chella *et al.*, "The impact of improved MEG-MRI co-registration on MEG connectivity analysis," (in English), *Neuroimage,* vol. 197, pp. 354-367, Aug 15 2019, doi: 10.1016/j.neuroimage.2019.04.061.

[19] A. Gramfort *et al.*, "MEG and EEG data analysis with MNE-Python," *Front Neurosci-Switz,* vol. 7, no. 267, pp. 1-13, 2013.

[20] A. Abraham *et al.*, "Machine Learning for Neuroimaging with Scikit-Learn," *Front Neuroinform,* vol. 8, pp. 14-14, 2014.

[21] R. A. Epstein, A. Harris, D. A. Stanley, and N. Kanwisher, "The Parahippocampal Place Area: Recognition, Navigation, or Encoding?," *Neuron,* vol. 23, no. 1, pp. 115-125, 1999.

[22] T. A. Yousry *et al.*, "Localization of the motor hand area to a knob on the precentral gyrus. A new landmark," *Brain,* vol. 120, no. 1, pp. 141-157, 1997.

[23] G. Bush, P. Luu, and M. I. Posner, "Cognitive and emotional influences in anterior cingulate cortex," *Trends Cogn Sci,* vol. 4, no. 6, pp. 215-222, 2000.

[24] D. Borsook *et al.*, "The Insula A "Hub of Activity" in Migraine," *The Neuroscientist,* vol. 22, no. 6, pp. 632-652, 2016.

[25] Wang, Haofei, Bertram E. Shi, and Yiwen Wang. "Convolutional Neural Network for Target Face Detection using Single-trial EEG Signal." *International Conference of the IEEE Engineering in Medicine and Biology Society (2018)*: 2008-2011.