2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Reduction of the ERP Measurement Time by a Weighted Averaging Using Deep Learning

Aoi Yoshida[1], Hikaru Sato[1], Siu Kang[1], Bunnoshin Ishikawa[2], and Tadanori Fukami[1]

Abstract— **In clinical examination, event-related potentials (ERPs) are estimated by averaging across multiple responses, which suppresses background EEG. However, acquiring the number of responses needed for this process is time consuming. We therefore propose a method for shortening the measurement time using weighted-average processing based on the output of deep learning. Using P300 as a representative component, here we focused on the shape of the ERP and evaluated whether our method emphasizes the P300 peak amplitude more than conventional averaging, while still maintaining the waveform shape and the P300 peak latency. Thus, using either CNN or EEGNet, the correlation coefficient reflecting the waveform shape, the peak P300 amplitude, and the peak latency were evaluated and compared with the same factors obtained from conventional waveform averaging. Additionally, the degree of background EEG suppression provided by our method was evaluated using the root mean square of the pre-stimulation waveform, and the number of fewer responses required for averaging (i.e., the reduction in measurement time) was calculated.**

The results showed that compared with P300 values obtained through conventional averaging, our method allowed for the same shape and response latency, but with a higher amplitude, while requiring a smaller number of responses. Our method showed that by using EEGNet, measurement time could be reduced by 13.7%. This corresponds to approximately a 40-second reduction for every 5 minutes of measurement time.

I. INTRODUCTION

ERPs are used in both clinical and research settings, especially in the fields of bioengineering, psychology and neuroscience. Furthermore, in recent years, it has been widely used in the field of brain-computer interface (BCI). ERPs are electrical potentials that occur transiently with respect to events that occur during cognitive processing. P300 is a positive potential that occurs approximately 300 ms after the presentation of sensory stimuli, and it is considered to be involved in human cognition and judgment.

Auditory oddball tasks are widely used in clinical examinations. In this type of task, two different tones are randomly presented such that one appears frequently and the other infrequently. Participants are instructed to press a button upon hearing the infrequent (oddball) tone. In addition to superimposing spontaneous EEGs on the ERP, ERPs are observed as waveforms mixed with artifacts such signals that originate from eye movements, not brain activity. Generally, ERPs are estimated by suppressing spontaneous EEG and artifacts, which is accomplished by averaging a large number of responses for target/non-target stimuli. (A stimulus required to perform a task is called a target stimulus, and the other stimulus is called non-target stimulus.) However, as the number of responses increases, the amount of time required to obtain the responses becomes problematical. Therefore, numerous methods for estimating or detecting ERPs from a single response and methods for preprocessing before averaging have been proposed. The exact ERP within a single-response waveform is unknown, and the estimation is difficult because it changes from moment to moment depending on the physical and mental state of the person, such as motivation and fatigue. The purpose of this study was not to estimate ERPs from single stimuli, but to obtain a waveform similar to the waveform obtained by conventional averaging in a shorter amount of time. If the measurement time can be shortened for clinical diagnoses, the physical and mental burden on the patient might be reduced. Because the appearance of ERPs fluctuates across responses, a calculation method is required in which responses containing clearly identifiable ERPs make larger contributions to the averaging. This type of weighting method will essentially exclude responses from the average if ERPs do not appear.

Regarding ERP averaging methods, Leonowicz et al. demonstrated robust averaging using a trimmed estimator that suppressed the effects of outliers[1]. This method assigns a small weight to extreme values located at both ends of the ordered data. Kotowsky et al. proposed robust weighted averaging based on criterion-function minimization for extracting ERPs[2]. This kind of correlation-based weighting serves to reduce the weight of corrupted responses.

In weighted averaging, obtaining the optimum weight is difficult. Therefore, we used deep learning to determine how reliable each response should be, and we added this information to the averaging process. In this study, we used EEGNet as one of the networks for deep learning. EEGNet is a network widely used for classification and detection in BCI, and is often used for comparison with other methods in performance evaluation[3], [4]. Recent reports have shown good results in the detection and classification of mortor imagery[5], [6]. Because the average waveform weighted by this reliability has a higher signal to noise ratio (the ratio of the ERP amplitude portion of the signal to components other than the ERP) than that obtained by conventional averaging, the measurement time can be reduced.

In Section II of the current paper, we describe how to

[1]Hikaru Sato, Aoi Yoshida, Siu Kang, and Tadanori Fukami are with the Graduate School of Science and Engineering, Yonezawa, Yamagata 992-8510, Japan {t212170m, t212014m}@st.yamagata-u.ac.jp, {siu, fukami}@yz.yamagata-u.ac.jp

[2]Bunnoshin Ishikawa is with the Department of Neuropsychiatry, Hotokukai Utsunomiya Hospital, Utsunomiya, Tochigi 320-8521, Japan houtokukai_uh@yahoo.co.jp

obtain the weight for weighted averaging using deep learning and how to estimate the measurement time. In Section III, we compare two types of weighted averaging methods with conventional averaging. In Section IV, we discuss how the proposed methods shorten the measurement time, and finally state the conclusion of this study.

II. Method

A. EEG measurement

EEGs were recorded using a Nihon Koden polygraph (EEG-1100) with a 0.3-s time constant, a 60-Hz high-cut filter, and a 97.5-nV quantization level. We obtained EEG signals at a sampling frequency of 200 Hz from 19 electrodes placed according to the international 10/20 system. Monopolar derivation was obtained from bilateral reference electrodes attached to the corresponding earlobes. EEG measurement took place at Hotokukai Utsunomiya Hospital after obtaining informed consent from the participants and approval from the hospital's ethics committee. Fifty healthy individuals participated in the study.

We used the oddball paradigm as an experimental design to elicit the ERP. Two types of auditory stimuli were presented: a target stimulus and a non-target stimulus. Participants were instructed to push a button when they heard the target stimulus. Responses for 30 target stimuli (2-kHz tone bursts) and 120 non-target stimuli (1-kHz tone bursts) were obtained for each participant. The average interval between tones was 2 s and the total measurement time was approximately 5 min. This trial was repeated twice per person. Therefore, the number of responses obtained in this study was approximately 3,000 for target stimuli and 12,000 for non-target stimuli.

B. ERP estimation by weighted averaging

Data recorded from 100 ms before stimulus onset to 1,000 ms after the onset were analyzed. The ground zero μV was defined as the average during the pre-stimulus epoch (100 ms before stimulus onset). We used two kinds of neural network models: convolutional neural network (CNN) and EEGNet[7], as shown in Fig.1. The CNN network shown in Fig.1(a) has a structure in which convolution and pooling are repeated twice for the input data, and then a fully connected network is connected. Input data to the network were a potential matrix of 19 electrodes $\times$ 125 sample points, which was constructed after down-sampling the 1-s EEG data measured at a sampling frequency of 1,000 Hz to 125 Hz. The EEGNet network shown in Fig.1(b) has a structure suitable for learning an EEG signal, with filtering in both the spatial axis direction and then the time axis direction. The input data and output elements are the same as for CNN. Data from 32 of the 50 participants (64%) were used as learning data, data from 8 (16%) were used as validation data, and data from the remaining 10 (20%) were used as test data. We confirmed that the participants correctly performed the task for all the EEG responses, included in the training data. Here, learning was performed by labeling responses to the target stimulus as 1 and those to the nontarget stimulus as 0. The training data and test data were divided such that responses from the same participant were not included in both training and test datasets. Because EEGs have very small electrical potentials, learning does not proceed well if these values are input without preprocessing. Therefore, before learning, we standardized the data such that the mean potential was 0 and the variance was 1. To compare the learning results of the CNN and EEGNet networks, we used the same parameters as much as possible. The input size was 19 $\times$ 125, the batch size was 64, and the number of epochs was set to a maximum of 500, while preventing overfitting by early stopping.

We compared the conventionally averaged waveform with the waveform obtained through weighted averaging up to n-th target response using the output from deep learning as a weight, $y_n(t)$ using the equation,

$$y_n(t) = \frac{\sum_{i=1}^{n} \omega_i s_i(t)}{\sum_{i=1}^{n} \omega_i} \quad (1)$$

, where the i-th response and weight are $s_i(t)$ and ω_i, respectively.

Because averaging is generally performed on approximately 30 target stimuli in clinical examinations, we used 30 responses to generate averaged waveforms. In the equations, $x_{30}(t)$ is the averaged waveform from 30 responses by conventional averaging. The following three indices were calculated for each participant and used to compare our weighted-averaging methods with conventional averaging.

(a) Correlation coefficient

This index evaluates similarity of shape between two waveforms.

$$CC_n = \frac{\sum_{\forall t \in T} x_{30}(t) y_n(t)}{\sqrt{\sum_{\forall t \in T} x_{30}(t)^2} \sqrt{\sum_{\forall t \in T} y_n(t)^2}} \quad (2)$$

Here, T represents the time during which all sample points in the analysis were obtained.

(b) P300 peak amplitude

P300 peak amplitude is an important index for clinical examinations. We therefore defined the peak-amplitude ratio between two waveforms as,

$$A_n = \frac{A_{y_n}}{A_{x_{30}}} \times 100 \quad (\%) \quad (3)$$

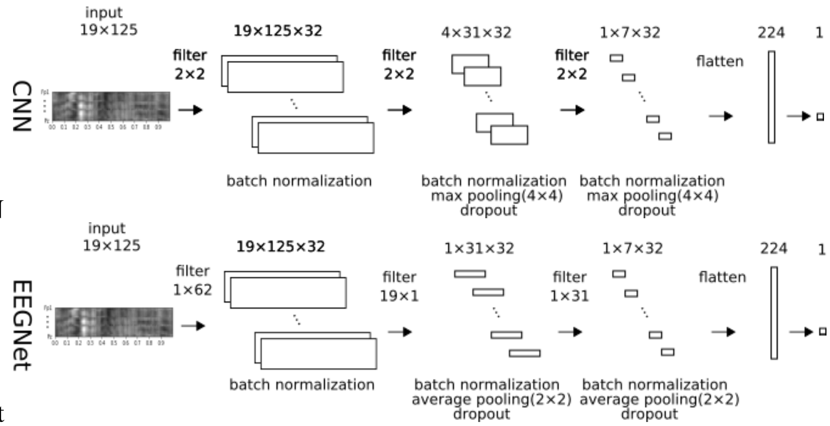
, where $A_{x_{30}}$ and A_{y_n} are the maximum amplitudes for averaged waveforms based on conventional and weighted averaging, respectively.

(c) P300 peak latency

P300 peak latency is also important, and we defined the peak-latency ratio between two waveforms as,

$$L_n = \frac{L_{y_n}}{L_{x_{30}}} \times 100 \quad (\%) \quad (4)$$

Here, $L_{x_{30}}$ and L_{y_n} are the maximum latencies for averaged waveforms based on conventional and weighted averaging, respectively.



(a) CNN

(b) EEGNet

Fig. 1. Structures for the CNN and EEGNet networks.

C. Estimating the shortening of measurement time

Assuming that the electrical potential distribution at an arbitrary time follows a Gaussian distribution with a mean of zero, it becomes 0μV by averaging an infinite number of responses. Because ERPs are evoked by a sensory stimulus, it cannot exist during the pre-stimulation period. Here, when the set of all sample points in the 100-ms pre-stimulus period is T' and the number of sample points included in the period is N_s, the root mean square (RMS) of $x(t)$ can be calculated by the following equation.

$$RMS = \frac{1}{N_s} \sum_{\forall t \in T'} x(t)^2 \tag{5}$$

The closer this value is to 0, the more background EEG is suppressed. We estimated the number of trials required by the weighted-averaging method to obtain an RMS equivalent to that generated by conventional averaging. The difference between this value and the conventional value (30), gives us the number of responses that can be reduced by weighted averaging. With this result, we can estimate how much time can be saved.

III. Results

Figure 2 shows the relationship between the number of responses that each averaging method uses and the correlation coefficient between waveforms. When the number of responses was 30, the correlation coefficient was 0.977 for CNN and 0.990 for EEGNet. The result for the P300 peak-amplitude ratio is shown in Fig.3. It ratio was 111.4% for CNN and 111.1% for EEGNet. Similarly, the P300 peak-latency ratio was 102.2% for CNN and 98.4% for EEGNet (Fig.4). Here, the result for conventional averaging can be obtained by replacing y_n with x_n in equations (2) to (4).

Figure 5 shows the RMS during the 100-ms period before stimulus onset. The results indicate that the degree of background EEG-noise suppression using conventional averaging (30 responses) was equivalent to the amount of suppression obtained using CNN weighted averaging with 27.5 responses, and EEGNet weighted averaging using 25.9 responses. Here, these values were obtained by linear interpolation. Finally, Fig.6 shows waveforms obtained by conventional averaging using 30 responses and those obtained using weighted averaging and a reduced number of responses. A statistical significance test was performed between the weighted average method and the conventional method in terms of correlation coefficient, peak amplitude, and peak latency. Here, paired t-test was performed on three indexes obtained from the averaged waveform of 26 responses. As a result, a significant difference ($p < 0.05$) was observed in the peak amplitude in both EEGNet and CNN, but no significant difference was observed in the correlation coefficient and the peak latency. The results confirmed that even if the number of responses is reduced, the waveform shape is not significantly affected.

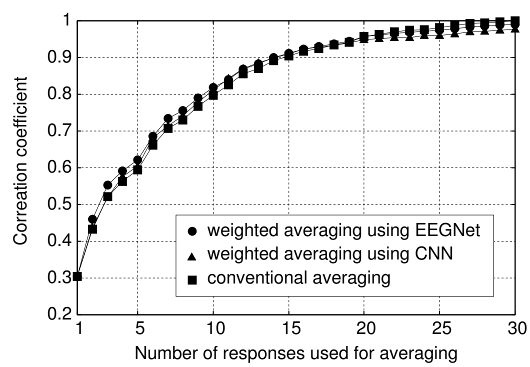


Fig. 2. Correlation coefficients for the averaged waveform obtained from 30 responses.

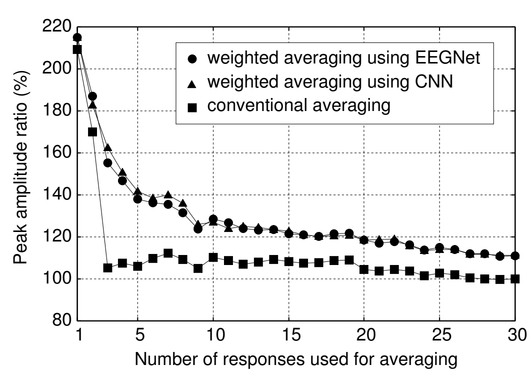


Fig. 3. Peak-amplitude ratio for the averaged waveform obtained from 30 responses.

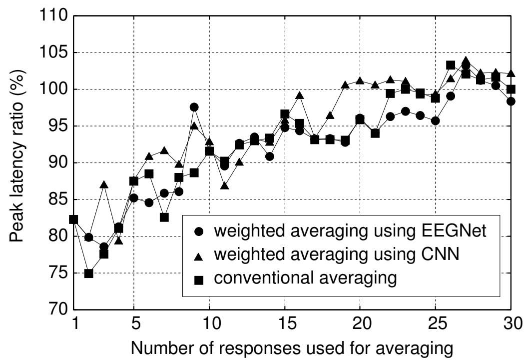


Fig. 4. Peak-latency ratio for the average waveform obtained from 30 responses.

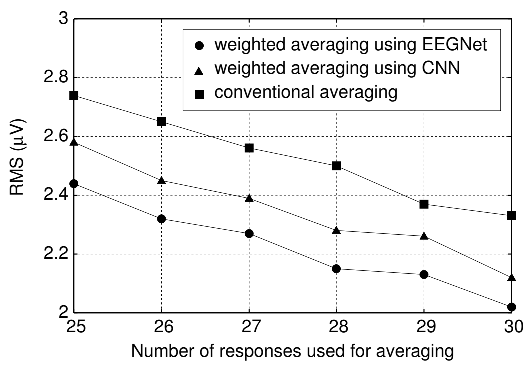


Fig. 5. RMS during the pre-stimulation period.

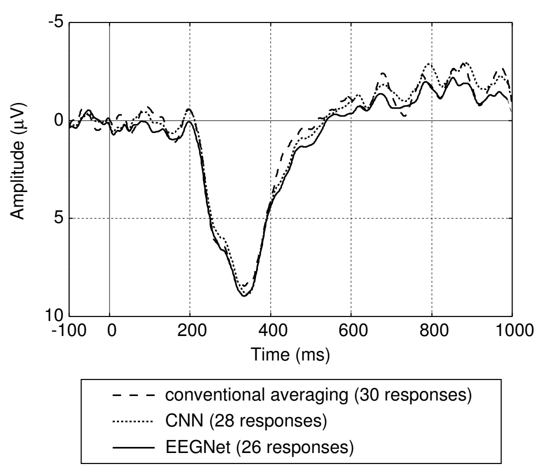


Fig. 6. ERP waveform estimated by conventional and weighted averaging.

IV. Discussions

Regarding the correlation coefficient, although EEGNet exhibited a slightly higher value than CNN, the values were similar and indicate that the waveform shape is close to that generated by conventional averaging. For the maximum P300-amplitude ratio, both CNN and EEGNet exceeded 100%, indicating a larger value than that obtained from 30 responses averaged by conventional averaging. Weighted averaging resulted in larger amplitudes because it assigned large weights to the responses in which the appearance of P300 was clear and easily distinguishable, and small weights to the responses in which it was not. As for the peak latency, both CNN and EEGNet produced values that were close to 100% of the 30 conventionally averaged waveform responses. However, the latencies provided by EEGNet remained more stable with respect to the number of responses than did those obtained via the CNN network.

Thus, weighted averaging using the output of deep learning increases P300 peak amplitude while requiring a smaller number of responses, but it does not significantly affect the shape or latency of the ERP, relative to that obtained via conventional averaging. From the RMS results, we can expect that measurement time will be shortened by 8.4% ((30 – 27.5) / 30) for CNN and 13.7% ((30 – 25.9) / 30) for EEGNet. This means that in general, EEGNet can save approximately 41.1 seconds for every 5 minutes of conventional measurement time.

V. Conclusions

In this study, we attempted to reduce the need to collect enough data for accurate auditory ERP measurement. Our new method uses weighted averaging based on the output of deep learning. We showed that the time can be reduced by 13.7% using our method with EEGNet. Here, we applied this method to data from healthy participants, but in the future, we would like to examine whether this method can be applied to patient data as well and also compare this method with other averaging methods. Moreover, we would like to build an online system that can terminate the measurement as soon as the ERP that can be used for diagnosis is obtained during the measurement.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 19K12077. We thank Adam Phillips, PhD, from Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript.

References

[1] Z. Leonowicz, J Karvanen and S. L. Shishkin, ”Trimmed estimators for robust averaging of event-related potentials,” J Neurosci Methods, vol.142, no.1, pp.17–26, 2005.

[2] K. Krzysztof, S. Katarzyna and L. Jacek, ”Improved robust weighted averaging for event-related potentials in EEG,” Biocybern Biomed Eng., vol. 39, pp.1036–1046, 2019.

[3] S. Panwar, P. Rad, T. P. Jung, Y. Huang, ”Modeling EEG data distribution with a Wasserstein generative adversarial network to predict RSVP events.” IEEE Trans. Neural Syst. Rehabil. Eng., vol.28, no.8, pp.1720-1730, 2020.

[4] X. Xiao, M. Xu, J. Han, E. Yin, S. Liu, X. Zhang, T. P. Jung, D. Ming, ”Enhancement for P300-speller classification using multi-window discriminative canonical pattern matching.” J. Neural Eng., vol.18, no.4, 046079, 2021.

[5] O. Avilov, S. Rimbert, A. Popov, L. Bougrain, ”Deep learning techniques to improve intraoperative awareness detection from electroencephalographic signals.” Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp.142-145, 2020.

[6] X. Wu, W. Zhang, Z. Fu, R. T. H. Cheung, R. H. M. Chan, ”An investigation of in-ear sensing for motor task classification.” J. Neural Eng., vol.17, no.6, 066010, 2020.

[7] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung and B. J. Lance, ”EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces,” J Neural Eng., vol.15, 056013, 2018.