Improving the Performance of SSVEP BCI with Short Response Time by Temporal Alignments Enhanced CCA

Aung Aung Phyo Wai, Min-Ho Lee, Seong Whan Lee, Cuntai Guan

Abstract—Steady State Visual Evoked Potentials (SSVEP) based Brain Computer Interface (BCI) provides high throughput in communication. In SSVEP-BCI, typically, higher accuracy can be achieved with a relatively longer response time. It is therefore a research topic to reduce the response time while keeping high accuracy. We propose a new method, temporal alignment enhanced Canonical Correlation Analysis (TACCA), followed by a decision fusion to improve classification accuracy with short response time. TACCA exploits linear correlation with non-linear similarity between steady-state responses and stimulus frequencies. We compare TACCA and three state-of-the-art methods using data from 54-subjects with response time ranging from 0.5 to 4 seconds. The evaluation results show that TACCA yields mean significant accuracy increase of 10-30% in all segment lengths, especially for the shorter time segment. One-way ANOVA tests show high significant differences between single and multiple phases in TACCA performance.

I. INTRODUCTION

The Brain Computer Interface (BCI) enables users to directly interact with either physical or virtual interfaces through brain signals without conventional pathways [1]. Among different brain sensing modalities, EEG-based BCI are widely adopted due to its portability, high temporal resolution, low-cost and high usability. Particularly, SSVEP BCI provides robust SNR (signal to noise ratio) with less dependency on time and phase synchronization resulting high accuracy and resilient to signal interferences [2]. But SSVEP also suffers from intra and inter subject variability as well as non-linearity. For reliable Steady-State Responses (SSR), SSVEP generally requires stimulus duration varied from a few seconds of single or multiple trials. This limits SSVEP in real-time usage as long exposure repetitive flickering stimuli cause user discomfort that could lead to accuracy deterioration [2].

Canonical Correlation Analysis (CCA) is commonly used baseline method for recognizing target frequency in SSVEP research [3]. Although CCA method does not requires training, some of its enhanced variants requires training in either subject-dependent or subject-independent manner [3, 4]. With non-stationary and noise prone nature of EEG, such personalized approaches might fail in real-time classification, although online adaptation can be applied for performance improvement [5]. Various pre- or post-processing methods are proposed to integrate with CCA improving accuracy [3]. Still, there is poor performance in short response time compared with long response time in both training and non-training approaches [3]. Training-based subject dependent and independent methods only marginally improve accuracy in short response time or short segment length. The huge gap in accuracy among different response times might be due to the poor linear relationship in short segment length. On the other hand, this might be due to non-linear origin of SSVEP responses resulting SSR with highly non-linear relationship [6]. Temporal alignment between two signals using Dynamic Time Warping (DTW) is successfully applied in waveform similarity measures to find non-linear mapping between measured signals and templates in speech and human activation recognition tasks [7].

Besides visual speller, our application interest for SSVEP is to objectively assess the visual field coverage and its integrity over time [8, 9]. The presence or absence of SSR with respect to multiple SSVEP stimuli can be regarded as objective measures in assessing one’s visual function. Such assessments require short test duration with no tedious training, preferably training-free, and detect SSR responses reliably in every subject. Short test brings better user experience that result better detection performance because long testing time for visual field assessment with existing methods is major obstacle in current clinical practice [2, 9]. But the current state of the art methods, even personalized training approaches, still could not provide high accuracy especially in short response time of 0.5 to 1 s data length [5].

To resolve this issue, we propose non-training method, TACCA that fuses intermediate decisions computed from linear correlation and non-linear mapping between EEG signals and stimulus reference signals. Our initial evaluation with 54-subjects SSVEP dataset shows superior accuracy improvements in segment with short response time. This improvement surpasses performance of those baseline and improved non-training methods [3, 10]. These promising results prove that non-training method can still achieve similar or better performance compare with those training methods; moving towards practical BCI for real-time visual speller and visual field assessment [1, 9].

II. PROPOSED METHODOLOGY

The proposed method consists of four steps that classifies the target class by maximizing the correlation and similarity between raw EEG and ideal reference signals [11]. Firstly, standard CCA is used to find maximum linear relationship. The maximum correlation coefficients, \( \rho_{\text{max},i} \) of respective stimulus frequencies, \( f_i \) are firstly determined. Secondly, the canonical variate matrix, \( \mathbf{W} \) is used to transform raw EEG signals into canonically correlated EEG signals as input to non-linear temporal alignment mapping. Standard DTW is implemented to find the minimum Euclidean distance by...
minimizing the warping path length. Multiple phases $\phi_k$ are added to CCA reference signals $\tilde{Y} = \sin(2 \pi f_t \phi_k + \phi_k)$ to include $k$ different phases as template signals, $\tilde{Y}$, to compute temporal similarity between canonical correlated EEG and phases-modified templates. Among distances $d_k$ computed from non-linear mapping, the stimulus frequency $f_i$ with the minimum distances $d_{m_{i,k}}$ are determined. The decision fusion complements independent decisions from CCA, $C_1$, and DTW, $C_2$ to compute final classification output, $C_{fuse(i,j)}$.

$$CCA(X, Y) = \max_{W_x, W_y} \rho(U, V) \frac{W^T_x S_x W^T_x}{\sqrt{W^T_x S_x W^T_x}} \frac{W^T_y S_y W^T_y}{\sqrt{W^T_y S_y W^T_y}}$$ (1)

On the other hand, DTW method finds the non-linear mapping between input data $X$ and templates $\tilde{Y}$ through minimum distance measure [7]. DTW computes $d_k$ between canonical weighted EEG signals and reference with multiple templates consisting of phases $\phi_k$ as shown in Fig.1. For each $k$ phase, distance for each class $i$, $d_{m_{i,k}}$, is computed to detect the target class by minimizing path lengths $P_x, P_y$ using

$$DTW(X, Y) = \min_{P_x, P_y} \sum_{k=1}^{K} \left\| X_{k,m} - Y_{k,n} \right\|$$ (2)

The intermediate decisions, with incompleteness and uncertainty, from CCA and DTW will be fused to achieve reliable outcome [12]. So, the decisions of $\rho_{max}$ and $d_{min}$ for stimulus frequencies $f_i$ of all classes $F_i$ are fused to determine the target class. We implement complementary fusion to fuse outputs from CCA, $P_i$, and DTW, $d_{k,i}$ where correct class detection comparing with ground-truth from either method and detection agreement in two or more classes is treated as correct class recognition. So this fusion rule can possibly be varied from simple fusion logic to weighted voting fusion between $d_{min}$ and $\rho_{max}$ depending on the performance criteria and the complexity of intermediate decisions.

$$C_i = fuse(\arg \max_{F_i} \rho_i, \arg \min_{F_i,k} d_{k,i})$$ (3)

### III. Experiment and Evaluation Analysis

In existing SSVEP study, the datasets used in evaluation analysis usually include a small number of subjects with limited sessions, runs and trials [3, 13]. So we design SSVEP experiment with multiple subjects with high numbers of trials per class although number of targets is only four [14].

#### A. Experiment Design and Dataset Description

Our dataset includes 62-channels EEG data sampled at 1000 Hz recorded with 4-class SSVEP interface. This study was reviewed and approved by the Institutional Review Board at Korea University [1040548-KUIRB-16-159-A-2], and written informed consent was obtained from all participants before the experiments. A total of 54 subjects went through 4 sessions on different days. Each session has 100 trials where subject performs tasks with stimuli display and experiment instruction. Each trial consists of data segment where user gazes at intended flickering target that is one out of 4 frequencies (5.25, 6.67, 8.57, 12 Hz). The stimulus duration is set as 4s followed by 6s rest in each trial. Similar to other SSVEP analysis [3], we only use 10 channels from occipital and parietal areas (P-7/3/z/4/8, PO-1/z/2). The duration of response time ranges from 0.5 to 4 s from stimulus onset with an interval increment of 0.5s. Altogether, our SSVEP dataset include 5, 400 trials per class for individual segment length of 0.5 to 4 s to evaluate the average classification performance.

#### B. Analysis and Evaluation Scenario

The below Fig. 2 shows steps involved in our evaluation analysis. Firstly, multi-channels EEG signals, $X$ are down-sampled to 250Hz after 60Hz notch and 0.5 to 80Hz band pass filtering, and normalized to zero mean and unit variance. The reference, ground-truth stimulus flickering frequencies, signals are generated according to 4 stimulus frequencies and number of harmonics specified [3, 10].

![Figure 2. Evaluation Scenario](image)

We compare four non-training methods such as baseline CCA, DTW, FBCCA and proposed method, TA-CCA. Each method classifies the target class $\tilde{y}$ using min/max classifier, except TA-CCA and, decision fusion discussed in session II. Then, classification outputs compared with ground-truth are evaluated using loss function $L(y, \tilde{y})$ with two performance measures: accuracy and Information Transfer Rate (ITR).
The loss function is the sum of classification result, \( \hat{y} \) compared with ground-truth, \( y \) of total trials per class \( T \) for all classes \( K \) where the decision with minimum losses achieve the best performance.

\[
\min L(y, \hat{y}) = \sum_{k=1}^{K} \sum_{i=1}^{T} |y_i^k - \hat{y}_i^k| \quad (4)
\]

To investigate parameters affecting the performance of TACCA, we vary phases between single \((0)\) and multiple \(0, \frac{\pi}{2}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\) and number of harmonics \(2, 3, 4\) of stimulus frequencies. In temporal alignment analysis, the maximum samples for computing warping path are set as (1) same as segment length, (2) minimum segment length, and (3) 50% of segment length and Euclidean distance metric is used for all tests. Parameters for FBCCA follows 10 sub-bands with equal bandwidth covering multiple harmonics \([10]\). With classification outcomes of all methods with those varying parameters, we perform one-way ANOVA test with 5% confidence level to evaluate the effects of warping path length, harmonics and phases on classification accuracy.

IV. RESULTS AND DISCUSSION

Overall, TACCA outperforms baseline CCA, DTW and FBCCA methods in all response times as shown in Fig. 3. TACCA achieves average accuracy increment of 26.34% and 12.4% compared with CCA and FBCCA methods. Especially high accuracy improvements in response time of 0.5 to 1 second though different percentage of higher accuracy in other response time. One-way ANOVA test shows significant difference in accuracy among methods \((F(3,572)=57.0247, p<0.001)\). The Fig. 3 shows DTW and TACCA accuracies in different phases as multiple-phases option is important for performance improvements especially in short response time. But, there is no difference in accuracy between single and six phases for CCA and FBCCA methods; showing accuracies at single phase only for both methods in Fig. 3.

Both DTW and TACCA achieve ITR higher than 150 bits/min in 0.5s segment as shown in Fig 4. Even, the ITR of TACCA with single phase reference input achieves 6.2 to 19.43 bits/min higher than FBCCA and CCA respectively. But the average ITR difference between single and multiple phases of TACCA at 0.5 and 1 s length is 139 and 46 bits/min respectively. This highlights the importance of multiple phases applied in reference signals for TACCA.

B. Discussion

The proposed classification method, TACCA relies on decision fusion with intermediates decision from DTW with multiple phases and CCA. From analysis with our datasets, TACCA outperforms tremendously in SSVEP segment with short response time in terms of accuracy and ITR compared with other methods as shown in Fig. 3 and Fig. 4. Besides non-training approach, TACCA also does not require additional parameters to fine-tune the performance. Besides conventional EEG pre-processing and stimulus frequencies, no further knowledge and adjustment of SSVEP parameters required. The accuracy outliers in TACCA as shown in Fig.
5 are mainly due to drop in accuracy using single phase (0°) in reference signals. The same reason is also applicable where overall mean DTW accuracy is poorer than that of CCA and FBCCA methods compared with Fig. 3. These results show less variability in accuracy among subjects with TACCA method especially reference signals of six phases.

![Figure 5. Performance Comparison of different methods under different phase, harmonic and warping length distance options.](image)

The optimal parameters of TACCA with two harmonics, six phases and 50% of data length for warping path achieve accuracy and ITR of 91.57% and 175 bits/min at 0.5 s. The possible limitation of TACCA is high computational cost of dynamic programming used in DTW that might affect online SSVEP BCI performance. But this can be alleviated by adopting efficient DTW algorithm [15]. In our dataset, number of targets is limited to four stimuli unlike other datasets with dense numbers of targets [2, 3]. With current promising results, we will need to further evaluate how TACCA performs with number of classes higher than four and dense frequency resolution in both offline and online classification tasks. In SSVEP based target classification, non-training approach is generally regarded as lower overall accuracy than training methods [5]. So we will compare and evaluate the performance of TACCA with existing state of the art training methods as well as adaptation methods as the next step [3, 5, 16]. TACCA can further be improved in different aspects such as adding phase angles more than six, using enhanced CCA methods as well as improved temporal alignment mapping [12, 15]. Furthermore, we will need to evaluate computation time of TACCA and real-time target detection compared with state-of-the-art (SOTA) methods [3, 16]. In this study, we purposely omitted SOTA method in [16] for fair comparison as that method requires training.

The only parameter to set with TACCA is the number of phases used in algorithm as both harmonics and warping path length do not affect the performance significantly. Although multiple phases achieve higher accuracy, it is important to select the optimal number of phases to trade-off among accuracy, computational costs and user comfort [2]. Still, we have to further validate TACCA with other datasets different stimulus frequencies and frequency resolution to understand full potentials of fusing decisions from linear and non-linear signals relationship.

V. CONCLUSION

For real-world SSVEP applications, it is important to detect and classify responses within a shortest time possible with no subject training and less variation in performance. Existing methods generally achieve high accuracy in long segment length with training required. So we propose TACCA that leverages linear correlation and non-linear mapping of SSR with decision fusion improving accuracy in short response time. Besides advantage of no training, no extra hyper-parameters adjustment requires with TACCA. Our offline analysis results show significant performance improvement in TACCA compared with three non-training methods. With these promising results, we will further improve and evaluate with multiple datasets to validate the performance improvements. We are hoping that the proposed TACCA based non-training method fulfills the needs of applying SSVEP BCI in real world applications.

REFERENCES