Hybrid Collaborative Brain-Computer Interfaces to Augment Group Decision Making

Davide Valeriani*, Caterina Cinel* and Riccardo Poli*

*Brain Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, UK, {dvaler,ccinel,rpoli}essex.ac.uk

Abstract

Collaborative Brain-Computer Interfaces (cBCIs) have recently been used for neuroergonomics applications, such as improving group low-level decision making. This chapter describes a hybrid cBCI to augment group performance in two realistic target-detection tasks: visual search, where participants had to spot a polar bear in an arctic image including many penguins, and speech perception, where volunteers listened to audio recordings affected by noise and had to decide whether or not a target word was uttered. The cBCI aggregates individual behavioural responses according to confidence estimates obtained from neural signals and response times. Results indicate that the cBCI augments group performance in both tasks over traditional groups making decision with standard majority. Also, cBCI groups were superior to non-BCI groups using confidence values reported by the participants to weigh decisions in visual search, although the opposite was true in speech perception.

Keywords— brain-computer interfaces, decision making, EEG, neuroergonomics, human augmentation, speech perception, visual search

1 Introduction

Traditionally, brain-computer interfaces (BCIs) convert neural activity into commands, allowing people with severe disabilities to control external devices or to communicate using brain activity alone (Wolpaw & McFarland, 2004). However, in recent years, researchers have started to explore the use of neuroergonomic BCIs to augment human perception (Parasuraman & Rizzo, 2007; Matran-Fernandez & Poli, 2017). For example, in decision making, BCIs have been used to detect and correct user's errors (Parra, Spence, Gerson, & Sajda, 2003). Nevertheless, despite advances in neuroscience, the average individual assisted by a BCI still makes worse decisions than the average non-BCI user.

Years of research in decision making have shown how groups are generally better than individuals in making decisions (wisdom of crowds) (Bahrami et al., 2010; Kerr & Tindale, 2004), thanks to groups' augmented perception and cognition achieved by integrating different views and percepts through the interaction of their members. However, there are circumstances in which groups can perform worse than individuals, such as in the presence of strong leadership (Branson, Steele, & Sung, 2010; Locke & Anderson, 2015) or time constraints (Bahrami et al., 2010).

Previous research has investigated the possibility of using collaborative BCIs (cBCIs) (Wang & Jung, 2011) to improve single-user BCI performance through groups and also improve group decision making through BCIs. This idea has been tested, for example, in a simple discrimination task where groups of 8 cBCI-assisted observers were more accurate and faster than single non-BCI users (Eckstein, Das, Pham, Peterson, & Abbey, 2012). However, the performance of the cBCIs presented in that study was still worse than that achieved by non-BCI groups. Other studies have shown similar limitations (Cecotti & Rivet, 2014).

In recent research, we have shown that hybrid cBCIs could be used to improve group performance in decision making (Poli, Valeriani, & Cinel, 2014). We adopted a hybrid approach, where neural signals and response times (RTs) were used to predict the confidence level of each observer in a decision. These confidence estimates were then used to weigh individual behavioural responses (actual mouse-button presses) and build group decisions, which were significantly better than both individual ones and, for the first time, also decisions obtained by equally-sized non-BCI groups. This hybrid approach has been successfully tested with *low-level* decision tasks involving *visual perception* only (Poli et al., 2014; Valeriani, Poli, & Cinel, 2017), where very simple shapes were used.

This chapter investigates the possibility of using a hybrid cBCI to improve group performance in two tasks involving more realistic stimuli: *visual search* (Experiment 1), where users have to identify a polar bear in a scene including many distractors (penguins), and *complex speech perception* (Experiment 2), where participants have to decide whether certain target words are uttered in spoken sentences affected by noise. Group decisions are obtained by aggregating individual responses using either a simple majority or a weighted majority based on confidence estimated by (a) the participants after each decision or (b) a cBCI using neural signals and RTs.

2 Methods

2.1 Participants

Ten healthy volunteers (average age = 27.4 ± 5.5 years, 5 females) took part in Experiment 1. Ten native-English speakers healthy volunteers (average age = 24.9 ± 4.9 years, 2 females) with normal hearing did Experiment 2. All participants had normal or corrected-to-normal vision and gave written informed consent. This research received UK's MoD and University of Essex ethical approval in July 2014.

2.2 Experiments

Participants underwent a sequence of 8 blocks of 40 trials each, for a total of 320 trial for each experiment. Each trial started with the presentation of a fixation cross for 1 s, to allow participants to get ready for the stimulus and the EEG signals to go back to the baseline. Then, in Experiment 1 a display containing an image of an arctic environment was presented, followed by a black and white 24×14 checkerboard mask shown for 250 ms. Participants had to decide whether or not a polar bear was present by pressing the left or the right mouse button, respectively. In Experiment 2, an audio recording was played and participants had to decide whether or not one of the target words ("route", "check", "grid", "lookout", "side", "trucks", "village") was uttered. After the response, the participants of both experiments were asked to indicate their confidence in that decision (0–100% in steps of 10%) using the mouse wheel. This had to be done within 4 s of the response. The mouse was controlled using the preferred hand. RTs from the stimulus onset were recorded.

For Experiment 1, the set of stimuli was similar to the one used in (Valeriani, Poli, & Cinel, 2015) and consisted of manually-created realistic images representing an arctic environment containing a variable number of penguins (distractors) and, possibly, a polar bear (target). The resulting dataset contained 68 stimuli with the target and 10 without it. In Experiment 2, we used audio recordings consisting of 41 sentences containing one target word and 42 sentences without any target word. Between 4 and 20 words (average length 9.3 ± 2.8 words) were uttered in each audio recording, which were recorded from a member of the army (male, native-English speaker). The duration of the audio recordings was between 2.19 and 8.75 s (average duration 4.3 ± 1.4 s). Two sets of 415 stimuli were created from these audio recordings: "standard" and "high-noise". For each audio recording, we created five versions by superimposing multiple types of noise on the original audio files, including white noise, environmental noise, volume changes, speed change, change of sampling rate, and audio drop-outs, all of which are typical of real-world military communications. The difference between the standard and high-noise sets is that the stimuli in the latter were generally more affected by noise than the former ones. Noise was added using the Pydub library (`www.pydub.com`).

The same sequence of sentences was used in the experiment for all participants in order to simulate, offline, concurrent group decisions. In Experiment 2, however, the difficulty of the audio tracks was dynamically varied by picking them from either the "standard" or the "high-noise" set, so as to keep the accuracy of all participants not too far from 80%. Stimuli containing the target were presented in 25% and 50% of the trials of Experiment 1 and 2, respectively. Volunteers were comfortably seated at about 80 cm from a LCD screen and participants of Experiment 2 were wearing in-ear earphones. In Experiment 2, volunteers undertook a memorisation task before starting the actual experiment to memorise the set of target words. Participants were then familiarised with the task by undertaking two sessions of 10 trials each in both experiments. Preparation and task familiarisation took approximately 40 minutes, while each experiment took about 35 minutes.

2.3 Data Recording and Group Decisions

A Biosemi ActiveTwo EEG system was used to record the neural signals from 64 electrode sites following the 10-20 international system. The EEG data were sampled at 2,048 Hz, referenced to the mean of the electrodes placed on the earlobes and band-pass filtered between 0.15 and 40 Hz. Artefacts caused by eye movements were corrected by applying a method based on correlations to the average difference between channels Fp1–F1 and Fp2–F2.

For each trial, response-locked epochs starting 1 s before the user's response and lasting 1.5 s were extracted from the EEG data recorded at each channel for Experiment 1, and at locations FC5, C5, CP5, TP7, T7 for Experiment 2, as in the auditory experiment we expected that key information could be found in the neural signals recorded in the left temporal lobe (Zatorre, 2012). Stimulus-locked epochs starting on the onset of the stimulus and lasting 1.5 s were also extracted for Experiment 1, while this representation of the EEG data was not useful in Experiment 2 as the audio recordings had different length and the target word could be uttered at any time within them. The data of each epoch were passed through a filter with a pass band of 0–6 Hz, a stop band of 8–1024 Hz and finally down-sampled to 16 Hz. Similarly to (Valeriani et al., 2017), the epochs were split into training and test sets using 10-fold cross-validation. Each epoch of the training set was assigned a label representing the correctness of the decision of the participant in that trial. We then used Local Temporal Correlation Common Spatial Patterns (LTCCSP) (Zhang et al., 2013) to extract two neural features from each epoch representing the decision confidence of the user (Valeriani et al., 2017). Hence, in Experiment 1 the feature vector was composed of four LTCCSP features (two for each type of epochs) and the RT, while for Experiment 2 we only had two LTCCSP features and the RT (to which the length of the audio recording was subtracted to remove its dependency on the stimulus at hand). The feature vector associated with each decision was then fed into a Least Angle Regressor (LARS) (Efron & Hastie, 2004) to obtain the confidence estimate.

All possible groups of size 2–10 were formed offline by combining the ten participants. Each group decision was then obtained by considering the sign of the weighted sum of its members' decisions (*weighted majority*), where the confidence values estimated by the hybrid cBCI were used as weights. We also simulated group decisions based on the reported confidence provided by participants after each decision by using these values (i.e., values in the set $\{0.0, 0.1, \ldots, 1.0\}$) to weigh individual decisions. The performance of these two weighted-majority-based systems was compared with that of non-BCI groups using standard majority to make decisions.

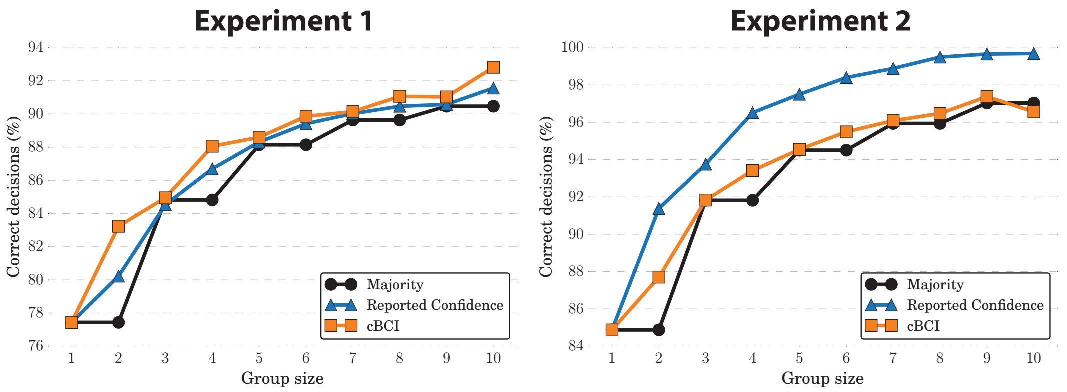

Figure 1: Percentage of correct decisions achieved by groups of different sizes using standard majority (black line) or a weighted majority using either the confidence reported by the participants (blue line) or the confidence estimated by the cBCI (orange line) for Experiment 1 (left) and 2 (right).

3 Results

3.1 Individual Performance

In Experiment 1, participants made correct decisions in $77.5 \pm 9.1\%$ of the trials, while in Experiment 2 they correctly identified target words in $84.9 \pm 3.2\%$ of the trials, showing the difficulty of both tasks for a single decision maker. We should note that, in Experiment 2, the average performance slightly deviated from the targeted one (80%) as some participants were still able to perform well in the task, despite the use of high-noise stimuli.

3.2 Group Performance

Figure 1 shows the percentage of correct decisions achieved by groups of increasing size using one of the three methods described in Section 2.3 for the two experiments.

We used the Wilcoxon signed-rank test to statistically compare the group performance of the three methods for each experiment. Naturally, statistical significance could not be reached for groups of size 10 as we only have one such group. Hence, we focus our analyses on groups of size 2–9.

In Experiment 1, groups assisted by our hybrid cBCI were significantly superior to equally-sized groups using standard majority for all group sizes ($p < 0.02$) and to groups using the weighted majority based on reported confidence values for even sizes ($p < 0.006$). The two weighted-majority methods performed on par for odd-sized groups. Moreover, the method based on the reported confidence was significantly better than majority for group sizes 2, 4, 5, 6, 7, 8 ($p < 0.002$), while the two methods were on par for size 3 and 9.

In Experiment 2, however, groups making decisions using the reported confidence were significantly superior to cBCI-assisted and majority-based groups for all group sizes 2–9 ($p < 0.003$). Groups assisted by the cBCI were better ($p < 0.006$) than majority-based groups for all sizes except 3, where the former was nearly statistically significantly better than the latter ($p = 0.076$).

4 Conclusions

In this chapter, we have shown that it is possible to improve group performance in realistic decision-making tasks involving visual or auditory stimuli with a hybrid collaborative BCI. The cBCI estimates the decision confidence of each participant from the neural signals and the RTs, and uses these values to weigh individual responses and obtain group decisions. These results confirm that our previous findings obtained with low-level decision tasks extend to more complex tasks.

Moreover, we have seen that in the visual search task the confidence estimated by the cBCI provided a better prediction of correctness than the reported confidence, while in the speech perception task it was the reported confidence the best predictor of correctness. This suggests that the best confidence estimate varies across tasks. However, we should note that in the speech perception task the cBCI was only relying on EEG data from five electrode locations (to promote generalisation) and response-locked epochs (due to the variability of the target words positions in the sentences) to estimate the decision confidence. In future research we will focus on complementing the feature set used for confidence estimation with (a) neural features extracted from the time, frequency, and time-frequency domains (e.g., wavelets), and (b) features extracted from other physiological measures related to decision making (e.g., skin conductance and pupil dilation).

Acknowledgements

This work was supported by the Defence and Security National PhD programme through DSTL.

References

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally Interacting Minds. *Science*, *329*(5995), 1081–1085.

Branson, L., Steele, N. L., & Sung, C.-H. (2010). When two heads are worse than one: Impact of group style and information type on performance evaluation. *Journal of Business and Behavioral Sciences*, *22*(1), 75–84.

Cecotti, H., & Rivet, B. (2014). Subject Combination and Electrode Selection in Cooperative Brain-Computer Interface Based on Event Related Potentials. *Brain Sciences*, *4*(2), 335–55.

Eckstein, M. P., Das, K., Pham, B. T., Peterson, M. F., & Abbey, C. K. (2012). Neural decoding of collective wisdom with multi-brain computing. *NeuroImage*, *59*(1), 94–108.

Efron, B., & Hastie, T. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499.

Kerr, N. L., & Tindale, R. S. (2004). Group Performance and Decision Making. *Annual Review of Psychology*, *55*(1), 623–655.

Locke, C. C., & Anderson, C. (2015). The downside of looking like a leader: Power, nonverbal confidence, and participative decision-making. *Journal of Experimental Social Psychology*, *58*, 42–47.

Matran-Fernandez, A., & Poli, R. (2017, apr). Brain-Computer Interfaces for Detection and Localization of Targets in Aerial Images. *IEEE Transactions on Biomedical Engineering*, *64*(4), 959–969.

Parasuraman, R., & Rizzo, M. (2007). *Neuroergonomics.* New York, New York, USA: Oxford University Press.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2003). Response error correction - A demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 173–177.

Poli, R., Valeriani, D., & Cinel, C. (2014, jul). Collaborative Brain-Computer Interface for Aiding Decision-Making. *PLoS ONE*, *9*(7), e102693.

Valeriani, D., Poli, R., & Cinel, C. (2015, apr). A collaborative Brain-Computer Interface for improving group detection of visual targets in complex natural environments. In *2015 7th international ieee/embs conference on neural engineering (ner)* (pp. 25–28). Montpellier: IEEE.

Valeriani, D., Poli, R., & Cinel, C. (2017, jun). Enhancement of Group Perception via a Collaborative Brain-Computer Interface. *IEEE Transactions on Biomedical Engineering*, *64*(6), 1238–1248.

Wang, Y., & Jung, T.-P. (2011). A collaborative brain-computer interface for improving human performance. *PLOS ONE*, *6*(5), e20422.

Wolpaw, J. R., & McFarland, D. J. (2004, dec). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(51), 17849–54.

Zatorre, R. J. (2012). Neural Specializations for Tonal Processing. *Annals of the New York Academy of Sciences*, *930*(1), 193–210.

Zhang, R., Xu, P., Liu, T., Zhang, Y., Guo, L., Li, P., & Yao, D. (2013). Local Temporal Correlation Common Spatial Patterns for Single Trial EEG Classification during Motor Imagery. *Computational and Mathematical Methods in Medicine*, *2013*, 1–7.