

# Collaborative Brain-Computer Interfaces to Enhance Group Decisions in an Outpost Surveillance Task

Saugat Bhattacharyya<sup>1</sup>, Davide Valeriani<sup>2</sup>, Caterina Cinel<sup>1</sup>, Luca Citi<sup>1</sup> and Riccardo Poli<sup>1</sup>

**Abstract**—We present a two-layered collaborative Brain-Computer Interface (cBCI) to aid groups making decisions under time constraints in a realistic video surveillance setting — the very first cBCI application of this type. The cBCI first uses response times (RTs) to estimate the decision confidence the user would report after each decision. Such an estimate is then used with neural features extracted from EEG to refine the decision confidence so that it better correlates with the correctness of the decision. The refined confidence is then used to weigh individual responses and obtain group decisions. Results obtained with 10 participants indicate that cBCI-assisted groups are significantly more accurate than groups using standard majority or weighing decisions using reported confidence values. This two-layer architecture allows the cBCI to not only further enhance group performance but also speed up the decision process, as the cBCI does not have to wait for all users to report their confidence after each decision.

## I. INTRODUCTION

Decisions are an important and ubiquitous aspect of our everyday lives. In some environments, decisions are critical in nature, and mistakes may result in extremely adverse outcomes, including loss of lives or resources. Yet, often an individual is asked to make decisions with either insufficient information or too much information and insufficient time to process it. In such cases, decisions made by groups are usually more effective than those made by individuals (wisdom of crowds) because groups exhibit better sensing and cognition capabilities [?], [?].

Group decision making can be further enhanced by weighing individual opinions with own estimates of their reliability, i.e., the corresponding reported confidence [?]. Such confidence estimates are determined by both conscious and unconscious factors, the unconscious mind often being quicker and more accurate in assessing own performance than the conscious one. In ideal conditions, the reported confidence correlates with the correctness of the decision [?], that is decision makers are more likely to be correct when they report a high degree of confidence than when they report a low confidence. However, there are many circumstances where reported confidence does not correlate with accuracy [?], [?], in which cases weighing decisions using reported confidence is not useful.

\*This work was funded by the Defence Science and Technology Laboratory (Dstl) on behalf of the UK Ministry of Defence (MOD).

<sup>1</sup>Saugat Bhattacharyya, Caterina Cinel, Luca Citi and Riccardo Poli are with the Brain Computer Interfaces and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, UK. Email: {saugat.bhattacharyya,ccinel,lciti,rpoli}@essex.ac.uk.

<sup>2</sup>Davide Valeriani is with the Department of Otolaryngology, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA. Email: davide\_valeriani@meei.harvard.edu

In previous work [?], we showed how collaborative Brain-Computer Interfaces (cBCIs) can be used to further improve group decision making. Our cBCIs used brain signals and response times (RT)<sup>1</sup> to estimate users' decision confidence while undertaking a visual matching task. Confidence estimates were then used to weigh individual decisions in making group decisions. The latter were found to be significantly better than either individual decisions and decisions made by equally-sized groups using standard majority. Similar results were later obtained with more realistic tasks based on static images [?], [?], [?]. In [?] we also showed that cBCI confidence estimates based on brain signals and RTs were more correlated with accuracy than the confidence reported by participants after each decision.<sup>2</sup>

This study extends our previous work on cBCIs for augmented decision-making in two directions. Firstly, we investigate the performance of our collaborative BCIs in making group decisions when users are exposed to video feeds from a realistic, dynamic environment (surveillance at a military outpost), and make decisions under time pressure. This is the very first application of a BCI to this type of problem. Secondly, we introduce a two-layer cBCI architecture, where the first layer estimates the confidence reported by the participants on a trial-by-trial basis, and the second layer uses this estimate together with neural features and RTs to weigh individual decisions and obtain group decisions.

## II. METHODOLOGY

### A. Participants

Ten healthy participants (four females, one left-handed, mean age =  $34.3 \pm 11.7$  years), with normal or corrected-to-normal vision and no reported history of epilepsy, took part in a decision-making experiment. All participants signed an informed consent form prior to taking part in the experiment. After the experiment, participants received a monetary compensation of GBP 12 plus an additional compensation of up to GBP 6 depending on performance (as explained in Sec. II-B). The total duration of the experiment was between 50 and 70 minutes, depending on the average RT of the participants and break durations. This research received ethical approval from the Ministry of Defence Research Ethics Committee (MoDREC) and the University of Essex in July 2017.

<sup>1</sup>RTs are known to reflect the decision confidence [?].

<sup>2</sup>Please, see [?] for a review on cBCIs and other forms of human cognitive augmentation.

## B. Stimuli and Procedure

Participants were shown a realistic video sequence, produced with the Unity game-engine, simulating the viewpoint of a soldier stationed at a military outpost and surveilling a clearing limited by an abandoned farmhouse at the back and vegetation on either side through a (simulated) night vision camera (Fig. 1). In each video (or trial), a character appeared from a distance, either from the house or from the adjoining vegetation, and brisk-walked towards the outpost. The video sequence was presented at a frame rate of 10 Hz. Participants were asked to decide, as quickly as possible, whether the character was wearing a helmet or a cap by clicking the left or the right mouse buttons, respectively. After each response, participants were asked to indicate (within 2 seconds) their decision confidence (termed *reported confidence* hereafter) on a scale from 0 (not confident) to 100 (very confident) by using the mouse wheel to vary the confidence in steps of 10.

In real-life situations, in the chosen decision task, an erroneous decision would lead to very negative consequences. In order to simulate this aspect at least to some degree, a reward/penalty system was used in the experiment. The system awards or takes away points based on the correctness of the decision and the corresponding RT. The points,  $p$ , obtained ( $p > 0$ ) or lost ( $p < 0$ ) in a trial were calculated as follows:

$$p = \begin{cases} \frac{1}{RT}, & \text{if the decision was correct,} \\ -|1 - \min(1, \frac{1}{RT})|, & \text{otherwise.} \end{cases}$$

For the typical values of RT recorded in the outpost experiment, this formula has the following characteristics: (a) when participants make correct decisions, they are rewarded more if their decision was quicker than if it was slower, because of the risks associated with allowing an unidentified person to get to close to the outpost, and (b) when participant give incorrect responses, more points are taken away if they are slower than if they are quicker, for similar reasons. Moreover, to further increase the time pressure on participants, the character disappeared after a random timeout, resulting in the trial being labelled as incorrect and a maximum point penalty being applied. The timeout with a mean of .9 s and standard deviation of 0.49 s is the same for all participants. At the end of the experiment, the number of points accumulated was converted into cash to determine the extra compensation for the participant (between GBP 0 and GBP 6).

Participants were comfortably seated at about 80 cm from an LCD screen. They were presented with 6 blocks of 60 trials (or video sequences). In each block, 30 trials showed a person wearing a helmet, and 30 trials showed a person wearing a cap (Fig. 1(b)) in a randomised manner. Prior to the start of the experimental session, participants underwent a brief training session of 15 trials (approximately 2 minutes) to familiarise with the task.

## C. Data Acquisition and Processing

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

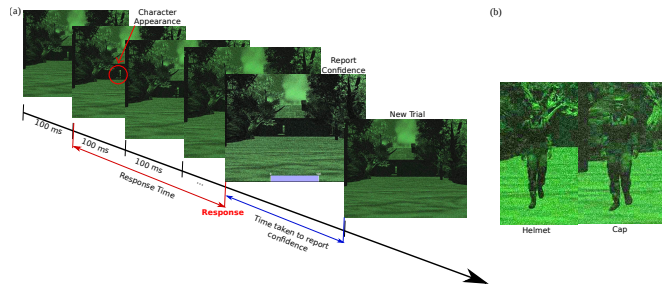


Fig. 1. (a) An example of video sequences presented to a participant in a single trial. (b) The characters appearing on-screen wearing a helmet (left) and a cap (right) in this experiment. The images have been enhanced for better clarity to readers.

2048 Hz, referenced to the mean of the electrodes placed on the earlobes, and band-pass filtered between 0.15 to 40 Hz to reduce electrical noise. Artefacts caused by eye-blinks and other ocular movements were reduced using a standard subtraction algorithm based on correlations to the averages of the differences between channels Fp1-F1 and Fp2-F2. Then, for each trial, the data were segmented into response-locked epochs starting 1 s before the response and lasting for 1.5 s. The epochs were then baseline corrected using the average voltage recorded in a time window of 200 ms before the stimulus and detrended. Epochs were then low-pass filtered with a cut off frequency of 14 Hz with an optimal Finite Impulse Response (FIR) filter designed with the Remez exchange algorithm. Finally, the data were down-sampled to 32 Hz. As we did in previous work [?], [?], [?], [?], each epoch was then labelled as *confident* or *not confident*, depending on whether the participant had made a correct or incorrect decision in that trial, respectively.

Common Spatial Pattern (CSP) [?] was used to extract neural features from each EEG epoch. The main idea behind CSP is to transform the multi-channel EEG data into a lower-dimensional spatial subspace using a projection matrix that maximises the variance of two-class signal matrices on the training set. The variances for the two classes (i.e., correct and incorrect) are largest in the first and the last rows of the projection matrix. So, the logarithm of the variances of the first and the last spatial subspaces were used as neural features in the cBCI.

RTs were measured (in seconds) by time-stamping the clicks of an ordinary USB mouse.

## D. Estimation of Reported Confidence

Decision-support systems that rely on reported confidence present the drawback that decisions can only be made after the process of assessing and reporting individual confidence values is complete, which delays the decision by an additional few seconds. This is also a drawback in any cBCI system that uses the reported confidence as an additional feature. To overcome this issue, in this work we have used a logistic regression model using L2 normalization and a regularization rate  $\lambda=1$  to estimate the reported confidence from the RTs on a trial by trial basis.

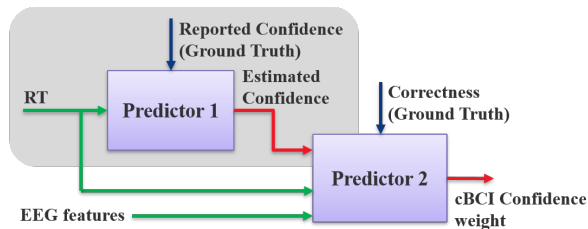


Fig. 2. The two-layered cBCI system. Predictor 1 is trained to estimate the reported confidence using RT for each decision. After training its output is fed into Predictor 2 to estimate the decision confidence.

This model is labelled a “Predictor 1” in Fig. 2. It presents the advantage of speeding up decisions by not requiring participants to report their confidence after training.

### E. cBCI Confidence Prediction

Once an confidence-estimation model is trained, it outputs an estimated reported confidence that was used as a feature in place of the actual reported confidence in a second logistic regression model labelled as “Predictor 2” in Fig. 2, which used L2 normalization and a regularization rate  $\lambda=0.001$ .

This model also used RTs as well as the two neural features extracted via CSP as inputs. In this case, for each participant, the model was used to predict the cBCI confidence (i.e., the correctness) of each decision.

In this study, an eight-fold cross validation procedure was used, to ensure that the results were not affected by over-fitting. More specifically, in each fold, cross validation split the data into training, validation and test sets. The training set was used to compute the CSP projection matrix and train the first layer of the cBCI system (“Predictor 1”). The second layer (“Predictor 2”) was then trained using the validation sets, and finally the cBCI confidence was computed using the test sets.

### F. Making Group Decisions

The decision  $D_{g,i}$  of group  $g$  in trial  $i$  is made as follows:

$$D_{g,i} = \text{sign}\left(\sum_{p=1}^m c_{p,i} \cdot d_{p,i}\right),$$

where  $m$  is the size of the group,  $d_{p,i}$  is the decision of participant  $p$  in trial  $i$  ( $d_{p,i} = -1$  for “confident”,  $d_{p,i} = +1$  if “not confident”), and  $c_{p,i}$  is the corresponding confidence computed by the cBCI.

Groups of size  $m=2, \dots, 10$  were formed offline by considering the  $\binom{10}{m}$  combinations of the 10 participants. For comparison purposes, the average group performance was computed when using: (a) cBCIs based on the two CSP features only, (b) cBCIs using the two CSP features and the RT, (c) cBCIs using CSPs, RT and reported confidence, (d) cBCIs using CSPs, RTs and *estimated* reported confidence (the system proposed here), (e) groups weighing decision using the confidence estimated by a logistic regression using only RTs as features, (f) traditional groups using standard majority, and (g) groups weighing individual decisions using the reported confidence (confidence majority).

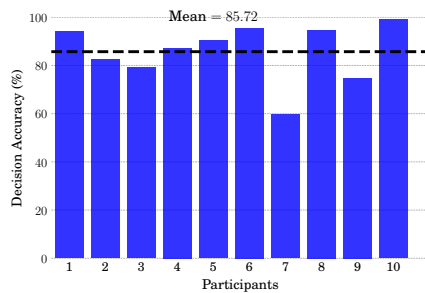


Fig. 3. Decision accuracy for each participant of the experiment.

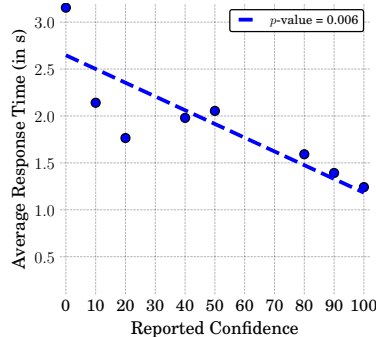


Fig. 4. Average RT across participants as a function of reported confidence, and corresponding regression line. The  $p$ -value is shown in the key.

## III. RESULTS

### A. Individual Performance

Fig. 3 shows the participants’ accuracies. The average accuracy (dashed line in the figure) was 85.72%. The standard deviation was 11.42%, indicating that individual performance levels varied widely. Indeed, the decision accuracy ranged from 59.72% to 99.17%, with some participants (2, 3, 7 and 9) finding the tasks particularly challenging.

### B. RTs as Predictors of Reported Confidence

Fig. 4 shows a plot (and the corresponding regression line) of the average RT (across all participants) for different values of reported confidence (0, 10, 20, etc.). Note that no participant reported a confidence of 30, 60 and 70 during the experiment. A *negative correlation* ( $r = -0.86385$ ) was observed between response time and the reported confidence. The regression line (slope =  $-1.4644$ , intercept =  $2.6472$ ) does clearly indicate that RTs are inversely proportional to the reported confidence, i.e., participants take longer to respond when they are not confident than when they are.

### C. Group Performance

Fig. 5 reports the performance of groups of sizes 1 to 10 obtained using confidence-based methods using different combinations of CSP and behavioural features (RT, reported confidence and estimated reported confidence), standard majority and confidence majority. Note that the lines labelled as RT, cBCI(CSPs) and cBCI(CSPs + RT) are almost exactly on top of each other due to the similarity in performance

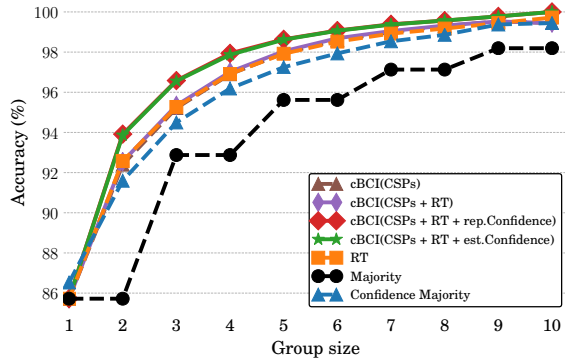


Fig. 5. Accuracies for groups of sizes 1 to 10 when using (a) the two CSP features only (cBCI(CSPs)) in brown, (b) the two CSP features and the RT (cBCI(CSPs + RT)) in purple, (c) the two CSP features, the RT and the reported confidence (cBCI(CSPs + RT + rep.Confidence)) in red, (d) the two CSP features, the RT and the estimated reported confidence (cBCI(CSPs + RT + est.Confidence)) in green, (e) only RTs as features (RT) in orange, (f) standard majority in black, and (g) confidence majority in blue.

of corresponding systems. The same happens for the lines labelled cBCI(CSP + RT + rep.Confidence) and cBCI(CSP + RT + est.Confidence). Generally, all confidence-based systems performed significantly better than both standard majority and confidence majority (Wilcoxon signed-rank test  $p < 0.05$ ) for groups of sizes 2 to 9. The biggest increase in accuracy was for groups of size 2 (pairs) where the *minimum* absolute improvement over standard majority is 6.68%. This is due to such systems breaking ties better than random. Also, note that cBCI(CSPs + RT)-based decisions are statistically better than corresponding RT-based decisions for groups of sizes 6, 7 and 8, and non-significant otherwise, showing that CSPs may provide a small improvement over RT alone.

The cBCI based on CSP and RT features (red line in Fig. 5) is significantly superior to the cBCI based only on neural features for groups of sizes up to 7 ( $p < 0.05$ ) and better than the RT-based decision system for groups of sizes up to 8 ( $p < 0.03$ ). The cBCI based solely on neural features are significantly better than the RT-based decision system for groups of sizes 6, 7 and 8 ( $p < 0.02$ ). When using the reported or estimated confidence as additional features for the cBCIs (green and orange lines in Fig. 5, respectively), groups become significantly better than the cBCIs based only on CSP and RT features ( $p < 0.035$ ) for all groups of sizes up to 9. The cBCIs based on reported and estimated confidences had almost identical performance.

#### IV. CONCLUSION

In this paper, for the first time, we have shown that collaborative BCIs can, very markedly, improve group decision making when users are asked to analyse video feeds of a difficult, realistic, dynamic environment, simulating night-vision surveillance at a military outpost, under time-pressure conditions. Such an environment forces the users to balance decision speed with accuracy, which is almost always required in real-world critical situations, such as threat detection. All cBCIs were statistically superior to

decisions based on majority, the biggest increase in accuracy occurring in pairs, which are also the groups where a cBCI approach is easiest to implement. However, all cBCIs were also statistically superior to weighing decisions based on reported confidence.

While reported confidence alone could not compete with the cBCIs, when used as an extra feature in input to the cBCI, further significant improvements in performance were obtained. However, systems relying on reported confidence produced group decisions with a significant delay, particularly in large groups, as the group decision could only be made after the slowest user had reported his/her confidence. In many environments where decisions are time sensitive, such as in the video surveillance of an outpost studied here, such an additional delay would be unacceptable.

We solved this problem by introducing a two-layer cBCI architecture, where the first layer estimates the confidence reported by the participants from their RTs in each decision, and the second uses this estimate together with neural features and RTs to weigh individual decisions and obtain group decisions. This has completely removed the issue mentioned above without influencing accuracy in any significant way.

In future research, we plan to test our cBCI methods in other decision-making tasks involving multi-sensory stimuli, multi-choice decision-making tasks and textual reasoning tasks. Furthermore, we aim to include other physiological features in the confidence estimation, such as galvanic skin response and pupil dilation.

#### REFERENCES