Proceedings of the
10th Graz Brain-Computer Interface Conference 2026

10.3217/978-3-99161-093-9-004

Confidence-weighted cumulative rCCA with post hoc re-analysis: unsupervised adaptive learning for calibration-free c-VEP BCI

J. Thielen[1]

[1]Machine Learning & Neural Computing department, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands

E-mail: jordy.thielen@donders.ru.nl

ABSTRACT: A brain–computer interface (BCI) typically requires a calibration period to train a decoding model, which is a tedious and time-consuming process. Previous studies have shown high performance of calibration-free decoding approaches, like adaptive reconvolution canonical correlation analysis (rCCA). rCCA can operate in an instantaneous mode, classifying trials independently without prior information, but can be improved by cumulatively learning from previously classified trials. Such cumulative strategy, however, relies on potentially unreliable pseudo-labels. In this study, two extensions to unsupervised rCCA are introduced. First, confidence-weighting ensures that model updates are driven primarily by high-confidence trials. Second, a post hoc re-analysis allows misclassifications to be corrected using a later, more accurate model. Four rCCA variants were evaluated on an existing code-modulated visual evoked potential (c-VEP) dataset, showing no significant effect of confidence-weighting, but a significant improvement of post hoc re-analysis. These findings pave the way for robust and fully calibration-free c-VEP BCI systems.

INTRODUCTION

Brain–computer interfaces (BCIs) based on code-modulated visual evoked potentials (c-VEPs) are among the fastest non-invasive systems using electroencephalography (EEG) [1]. In contrast to BCIs using the frequency-tagging stimulus protocol, which evoke narrow-band steady-state visual evoked potentials (SSVEPs), c-VEP BCIs use a noise-tagging protocol that evokes broadband neural responses [2]. This broadband nature enables high information transfer rates and robust decoding.

Beyond advances in stimulus design aimed at improving efficiency, usability and comfort [3, 4], substantial effort has been devoted to improving decoding for c-VEP BCIs. A limitation of many systems is their reliance on session-specific calibration data acquired during time-consuming passive training sessions [1]. Such calibration requirements hinder immediate use and reduce the practicality of BCIs in real-world applications.

To address this challenge, several strategies have been proposed that can work with small datasets [5]. For instance, the signal-to-noise ratio can be increased through optimized stimulation paradigms, like the noise-tagging stimulus protocol [1]. Or, prior knowledge can be incorporated into the decoding via careful regularization, like the reconvolution approach [6]. Furthermore, unsupervised methods may eliminate explicit calibration altogether, commonly referred to as zero-training or calibration-free approaches, such as unsupervised adaptive reconvolution [7].

Within the c-VEP domain, multiple calibration-free decoding methods have been explored. Relying on a circularly shifted m-sequence for stimulation, Turi et al. proposed estimating relative temporal lags between unclassified trials combined with a language model [8]. Thielen et al. introduced an adaptive reconvolution embedded in a canonical correlation analysis (CCA), which decodes by directly optimizing model fit [7]. Zheng et al. proposed narrow-band random stimulus sequences combined with filterbank CCA [9]. Finally, Santamaria-Vazquez et al. as well as Behboodi et al. investigated cross-participant transfer learning and foundation models for calibration-free decoding, optionally with fine-tuning [10, 11].

The present study builds on the calibration-free reconvolution CCA (rCCA) framework as proposed by Thielen et al. [7]. Recent work demonstrated that rCCA can operate in two unsupervised modes: instantaneous and cumulative [12, 13]. In the instantaneous mode, each trial is decoded independently using only that trial's data, making it computationally lightweight and robust to non-stationary data. While this mode is attractive for its simplicity and adaptability, its performance can be improved by allowing the model to learn from previously classified trials, yielding the cumulative variant [12, 13].

In earlier benchmarks, rCCA was evaluated not only on c-VEP data but also on a P300 event-related potential (ERP) dataset to assess its generalization capabilities beyond c-VEP [12, 13]. Additionally, rCCA was compared to unsupervised mean-difference maximization (UMM), a state-of-the-art method originating from the ERP domain [14]. Surprisingly, rCCA remained competitive with UMM even outside its original domain [12, 13].

This finding was notable because UMM incorporates several sophisticated mechanisms, like confidence-weighting. That is, in the cumulative mode, both rCCA and UMM rely on pseudo-labels, meaning that previously classified trials are used for learning, assuming that their predicted labels are correct. UMM explicitly accounts for

10.3217/978-3-99161-093-9-004

the uncertainty of pseudo-labels by applying confidence weighting, such that trials classified with higher confidence have a stronger influence on model adaptation, substantially improving its performance in a calibration-free scenario [14]. In contrast, rCCA treats all pseudo-labeled trials equally, but still performed on par with UMM on both c-VEP [12] and ERP [13] data.

Finally, when using a cumulative adaptive classifier, it is implicitly assumed that the model improves as more data are incorporated. Under this assumption, early unreliable pseudo-labels may be corrected by reanalyzing previously classified trials using a later, more accurate model. Similar post hoc re-analysis strategies have been shown to be beneficial in other unsupervised classification settings, including approaches based on expectation maximization [15] and learning from label proportions [16].

Motivated by these observations, the present study provides a first attempt to extend vanilla rCCA with a confidence-weighting and a post hoc re-analysis mechanism. The performance of these two extensions is systematically compared against the vanilla instantaneous and cumulative variants, with the aim to further improve calibration-free rCCA. All methods are evaluated on one existing open-access c-VEP dataset.

MATERIALS AND METHODS

Dataset: An open-access c-VEP dataset [17] was used as recorded by Thielen et al. [7]. From this dataset, only the offline part was used, in which 30 participants performed a copy-spelling task. EEG was recorded from eight electrodes positioned according to the 10–10 system (Fz, T7, T8, POz, O1, Oz, O2, Iz), using a Biosemi ActiveTwo amplifier at a sampling rate of 512 Hz.

Participants interacted with a 4×5 matrix speller displayed on a 12.9-inch iPad Pro with 60 Hz refresh rate and 1920×1080 resolution. Each key was modulated by a unique pseudo-random binary sequence derived from optimized modulated Gold codes [6], presented at full contrast (white for 1, black for 0). These sequences contained two flash durations (16.67 ms and 33.33 ms) and had a cycle length of 126 bits (2.1 s per cycle at 60 Hz).

Each participant completed five runs, each containing one trial per key in randomized order. Trials began with a 1 s green cue highlighting the target key, followed by 31.5 s of simultaneous flashing of all keys (i.e., 15 cycles), during which participants fixated on the target key. No feedback was provided. In total, each participant contributed 100 trials, containing 5 repetitions per key.

The EEG data were preprocessed as follows. First, a band-pass filter between 6–30 Hz was applied, as found optimal in previous work [12]. Second, the data were segmented into trials from -0.5 s to 32.0 s relative to stimulus onset. Third, the epochs were downsampled to 120 Hz. Finally, the initial and final 0.5 s of each trial were removed to mitigate filtering artefacts while maintaining accurate marker alignment, resulting in 31.5 s trials aligned with stimulation onset.

Algorithm 1 Instantaneous rCCA

Require: stimulus structure matrices $\{\mathbf{M}_i\}_{i=1}^{N}$
1: **for** trial $j = 1$ to K **do**
2: **for** stimulus $i = 1$ to N **do**
3: $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
4: $\rho_i \leftarrow \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
5: **end for**
6: $\hat{y}_j \leftarrow \arg\max_i \rho_i$
7: **end for**

rCCA classification: The current j-th trial is represented by $\mathbf{X}_j \in \mathbb{R}^{C \times T}$, where $C = 8$ denotes the number of EEG channels and T the number of EEG temporal samples. The objective of reconvolution canonical correlation analysis (rCCA) is to identify the attended target key $\hat{y}_j$ for trial j by maximizing the Pearson's correlation:

$$\hat{y}_j = \arg\max_i \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j, \mathbf{r}_i^\top \mathbf{M}_i) \tag{1}$$

Here, $\mathbf{w}_i \in \mathbb{R}^C$ is a spatial filter and $\mathbf{r}_i \in \mathbb{R}^L$ is the evoked response of length L samples, both associated with the i-th stimulus. The matrix $\mathbf{M}_i \in \mathbb{R}^{L \times T}$ is a Toeplitz representation of the stimulus sequence, encoding the onset, duration, and possible overlap of evoked responses [6]. Similar to previous work [7, 12], a 300 ms response to a short flash, a 300 ms response to a long flash, and a 300 ms response to the onset of stimulation at the start of a trial were modeled at 120 Hz, such that $L = 3 \cdot 0.3 \cdot 120 = 108$. All unsupervised adaptive rCCA classifiers considered in this work use Equation 1 for decoding. They differ only in how the spatial filter $\mathbf{w}_i$ and temporal response $\mathbf{r}_i$ are learned and updated, potentially using previously classified trials.

Instantaneous rCCA: The instantaneous rCCA (Algorithm 1) does not use information from previous trials and thus treats each trial independently. For a given trial j, it estimates sequence-specific spatial filters $\mathbf{w}_i$ and evoked responses $\mathbf{r}_i$ for each of the $N = 20$ candidate stimulus sequences $i \in 1, \ldots, N$, using only the current trial $\mathbf{X}_j$:

$$\mathbf{w}_i, \mathbf{r}_i = \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j, \mathbf{r}_i^\top \mathbf{M}_i) \tag{2}$$

Consequently, both the parameter estimation in Equation 2 and the prediction in Equation 1 are performed on the same data $\mathbf{X}_j$, once for each candidate stimulus. Effectively, classification is done by selecting the stimulus that yields the highest model fit to the observed EEG.

Cumulative rCCA: The cumulative rCCA (Algorithm 2) incorporates information from previous trials and can therefore improve classification performance over time as more data become available. That us, for a given trial j, it estimates sequence-specific spatial filters $\mathbf{w}_i$ and evoked responses $\mathbf{r}_i$ for each of the $N = 20$ candidate stimulus sequences $i \in 1, \ldots, N$, using all data observed up to and including the current trial $\mathbf{X}_j$.

This setting introduces a key challenge: rCCA is inherently supervised and requires labeled training data for

10.3217/978-3-99161-093-9-004

Algorithm 2 Cumulative rCCA

Require: stimulus structure matrices $\{\mathbf{M}_i\}_{i=1}^{N}$
for trial $j = 1$ to K **do**
  for stimulus $i = 1$ to N **do**
    $\mathbf{S}_j \leftarrow [\mathbf{X}_0, \ldots, \mathbf{X}_j]$
    $\mathbf{D}_i \leftarrow [\mathbf{M}_{\hat{y}_0}, \ldots, \mathbf{M}_{\hat{y}_{j-1}}, \mathbf{M}_i]$
    $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{S}_j,\ \mathbf{r}_i^\top \mathbf{D}_i)$
    $\rho_i \leftarrow \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
  end for
  $\hat{y}_j \leftarrow \arg\max_i \rho_i$
end for

calibration. In the absence of ground-truth labels in a calibration-free setting, a pseudo-labeling approach is used, where the predicted labels from previous trials are treated as true labels. The cumulative learning objective is then given by:

$$\mathbf{w}_i, \mathbf{r}_i = \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{S}_j, \mathbf{r}_i^\top \mathbf{D}_i) \tag{3}$$

Here, $\mathbf{S}_j = [\mathbf{X}_0, \ldots, \mathbf{X}_j]$ denotes the concatenation of all EEG trials up to the current trial j, and $\mathbf{D}_i = [\mathbf{M}_{\hat{y}_0}, \ldots, \mathbf{M}_{\hat{y}_{j-1}}, \mathbf{M}_i]$ is the corresponding concatenation of stimulus structure matrices, using pseudo-labels $\hat{y}_k$ for $k < j$ and the current candidate i.

By definition, the first trial is still processed using instantaneous rCCA (Equation 2) as no previous trials are available. From the second trial onward, the model progressively accumulates training data and updates its parameters accordingly (Equation 3). Importantly, prediction for each trial is always performed using Equation 1, which considers only the current trial data, while learning leverages the full history.

Confidence-weighted cumulative rCCA: The cumulative rCCA uniformly aggregates data from all previous trials to update its parameters, implicitly assuming that all pseudo-labels are correct. In practice, this assumption may be violated. Pseudo-labels may be unreliable due to noise in the EEG signals, even with a well-specified model. Or, pseudo-labels may be unreliable due to model inaccuracies even when the EEG data is of good quality, for instance when too little data are available.

To reduce the impact of unreliable pseudo-labels, confidence-weighted cumulative rCCA (Algorithm 3) is introduced using a confidence-weighting mechanism like used in UMM [14]. The key idea is to assign greater influence to trials that are classified with high confidence and to downweight trials with low confidence during parameter updates.

The confidence c_j of trial j is defined as the normalized correlation margin between the highest winning correlation ρ_w and the second-highest runner-up correlation ρ_r, similar to the margin dynamic stopping method used by Thielen et al. [6]:

$$c_j = \frac{\rho_w - \rho_r}{\sigma}, \tag{4}$$

Algorithm 3 Confidence-weighted cumulative rCCA

Require: stimulus structure matrices $\{\mathbf{M}_i\}_{i=1}^{N}$
for trial $j = 1$ to K **do**
  for stimulus $i = 1$ to N **do**
    $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
    $\rho_i \leftarrow \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
  end for
  $c_j \leftarrow (\rho_w - \rho_r)/\sigma$
  for stimulus $i = 1$ to N **do**
    $\mathbf{S}_j \leftarrow [c_0\mathbf{X}_0, \ldots, c_j\mathbf{X}_j]$
    $\mathbf{D}_i \leftarrow [c_0\mathbf{M}_{\hat{y}_0}, \ldots, c_{j-1}\mathbf{M}_{\hat{y}_{j-1}}, c_j\mathbf{M}_i]$
    $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \mathrm{corr}(\mathbf{w}_i^\top \mathbf{S}_j,\ \mathbf{r}_i^\top \mathbf{D}_i)$
    $\rho_i \leftarrow \mathrm{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
  end for
  $\hat{y}_j \leftarrow \arg\max_i \rho_i$
end for

Here, σ is the standard deviation of all N correlations except the highest winning correlation ρ_w. Thus, c_j increases when the winning candidate is well separated from the runner-up, and when the overall non-maximum correlation distribution is narrow, both indicating higher classification confidence.

These confidence values are used as linear weights in the cumulative learning process. Specifically, the accumulated data matrices in Equation 3 are redefined as $\mathbf{S}_j = [c_0\mathbf{X}_0, \ldots, c_j\mathbf{X}_j]$ and $\mathbf{D}_i = [c_0\mathbf{M}_{\hat{y}_0}, \ldots, c_{j-1}\mathbf{M}_{\hat{y}_{j-1}}, c_j\mathbf{M}_i]$, such that high-confidence trials contribute more strongly to the parameter estimation.

Since the confidence c_j of the current trial must be available prior to model updating, confidence values for all trials are obtained using an instantaneous rCCA. The resulting correlation scores are then used to compute c_j, after which the model is updated accordingly with confidence weighting (Equation 3).

Confidence-weighted cumulative rCCA with post hoc re-analysis: Confidence-weighted cumulative rCCA reduces the influence of unreliable pseudo-labels by downweighting low-confidence trials. Still, even with reduced weights, such trials still contribute to learning and may introduce errors if their pseudo-labels are incorrect.

By construction, cumulative rCCA learns incrementally across trials, such that the model is expected to improve over time. This opens the possibility of correcting earlier mistakes by reclassifying past trials using a later, presumably more accurate model, a process referred to as post hoc re-analysis [15, 16]. In principle, this strategy can rectify early misclassifications and improve the reliability of subsequent updates, since later model adaptations are then based on more accurate pseudo-labels.

In confidence-weighted cumulative rCCA with post hoc re-analysis (Algorithm 4), after classifying the current j-th trial and updating the model, all previous trials are reclassified by applying the updated model via Equation 1, and their pseudo-labels are replaced accordingly, affecting model updates for subsequent trials.

10.3217/978-3-99161-093-9-004

Algorithm 4 Confidence-weighted cumulative rCCA with post hoc re-analysis

Require: stimulus structure matrices $\{\mathbf{M}_i\}_{i=1}^{N}$
for trial $j = 1$ to K **do**
  for stimulus $i = 1$ to N **do**
    $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \text{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
    $\rho_i \leftarrow \text{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
  end for
  $c_j \leftarrow (\rho_w - \rho_r)/\sigma$
  for stimulus $i = 1$ to N **do**
    $\mathbf{S}_j \leftarrow [c_0\mathbf{X}_0, \ldots, c_j\mathbf{X}_j]$
    $\mathbf{D}_i \leftarrow [c_0\mathbf{M}_{\hat{y}_0}, \ldots, c_{j-1}\mathbf{M}_{\hat{y}_{j-1}}, c_j\mathbf{M}_i]$
    $\mathbf{w}_i, \mathbf{r}_i \leftarrow \arg\max_{\mathbf{w}_i, \mathbf{r}_i} \text{corr}(\mathbf{w}_i^\top \mathbf{S}_j,\ \mathbf{r}_i^\top \mathbf{D}_i)$
    $\rho_i \leftarrow \text{corr}(\mathbf{w}_i^\top \mathbf{X}_j,\ \mathbf{r}_i^\top \mathbf{M}_i)$
  end for
  $\hat{y}_j \leftarrow \arg\max_i \rho_i$
  $\mathbf{w}, \mathbf{r} \leftarrow \mathbf{w}_{\hat{y}_j}, \mathbf{r}_{\hat{y}_j}$
  for trial $k = 1$ to $j - 1$ **do**
    $\hat{y}_k \leftarrow \arg\max_i \text{corr}(\mathbf{w}^\top \mathbf{X}_k,\ \mathbf{r}^\top \mathbf{M}_i)$
  end for
end for

Please note that the confidence values c_j are not affected by this post hoc re-analysis, because they are generally estimated via an instantaneous rCCA.

Analysis: This study systematically compared the performance of four calibration-free decoding methods based on an unsupervised adaptive rCCA: instantaneous rCCA (rcca_i), cumulative rCCA (rcca_c), confidence-weighted cumulative rCCA (rcca_cw), and confidence-weighted cumulative rCCA with post hoc re-analysis (rcca_cwp). All methods were evaluated on an existing c-VEP dataset, with trials processed in their original chronological order to simulate an online decoding scenario and retaining any non-stationarity present in the data.

Performance was quantified using decoding curves, defined as classification accuracy as a function of trial duration. Classification accuracy was evaluated at sixteen trial durations: from 1.05 s to 8.4 s in steps of 1.05 s, from 10.5 s to 16.8 s in steps of 2.1 s, and from 18.9 s to 31.5 s in steps of 4.2 s.

Statistical significance of paired differences between methods at a particular trial duration was assessed using a one-sided Wilcoxon signed-rank test.

RESULTS

For all four methods, the classification accuracy as a function of decoding time, referred to as the decoding curve, is shown in Figure 1. In addition, Table 1 reports the earliest decoding time at which a given grand-average accuracy level is reached, and Table 2 summarizes the grand-average accuracy at selected decoding times.

Overall, all methods show a clear and monotonic improvement in performance with increasing decoding time, indicating that access to longer EEG segments consistently enhances classification accuracy.

Table 1: **Grand average decoding time.** Listed are the trial durations in seconds to reach a certain grand average classification accuracy.

	0.70	**0.80**	**0.85**	**0.90**	**0.95**
rcca_i	6.30	8.40	10.50	14.70	27.30
rcca_c	3.15	3.15	3.15	5.25	7.35
rcca_cw	3.15	3.15	4.20	4.20	6.30
rcca_cwp	3.15	3.15	4.20	4.20	6.30

Table 2: **Grand average classification accuracy.** Listed are the grand average classification accuracy reached at distinct trial durations in seconds.

	1.05	**2.10**	**4.20**	**10.50**	**31.50**
rcca_i	0.07	0.27	0.60	0.87	0.96
rcca_c	0.25	0.54	0.87	0.97	0.97
rcca_cw	0.22	0.55	0.91	0.97	0.97
rcca_cwp	0.24	0.59	0.93	0.97	0.97

A first prominent observation is that instantaneous rCCA (rcca_i) yields lower performance than each of the cumulative variants. In particular, its decoding accuracy is significantly lower than that of cumulative rCCA (rcca_c) up to and including 23.1 s ($p < .05$). At longer decoding times (27.3 s and 31.5 s), the differences are no longer statistically significant ($p = .199$ and $p = .463$, respectively), meaning that instantaneous rCCA can reach equal ceiling performance with sufficiently long single-trials.

Interestingly, no statistically significant differences are observed between confidence-weighted cumulative rCCA (rcca_cw) and cumulative rCCA (rcca_c) at any decoding time. While rcca_cw shows a small absolute improvements over rcca_c at some time points (2.1 s and 4.2–6.3 s) and slightly worse performance at others (1.05 s and 3.15 s), these effects are minor and disappear for decoding times of 7.35 s and longer, after which both methods yield essentially identical performance.

In contrast, improvements are observed for confidence-weighted cumulative rCCA with post hoc re-analysis (rcca_cwp). Compared to rcca_c, rcca_cwp achieves significantly higher accuracy at 4.2 s (0.93 vs. 0.87, $p = .010$) and 6.30 s (0.96 vs. 0.94, $p = .043$), with no significant differences at other time points ($p > .05$). Importantly, rcca_c is never significantly better than rcca_cwp. Similarly, when compared to rcca_cw, the post hoc re-labeled variant performs significantly better from 2.1 s to 5.25 and 7.35 s to 8.4 s ($p < .05$), while no significant differences are observed at other decoding times. Again, at no time point does rcca_cw outperform rcca_cwp.

DISCUSSION

In this study, an extension of adaptive rCCA was proposed that incorporates confidence-weighted model updates and post hoc re-analysis. Using an existing c-VEP dataset, four calibration-free decoding strategies

10.3217/978-3-99161-093-9-004

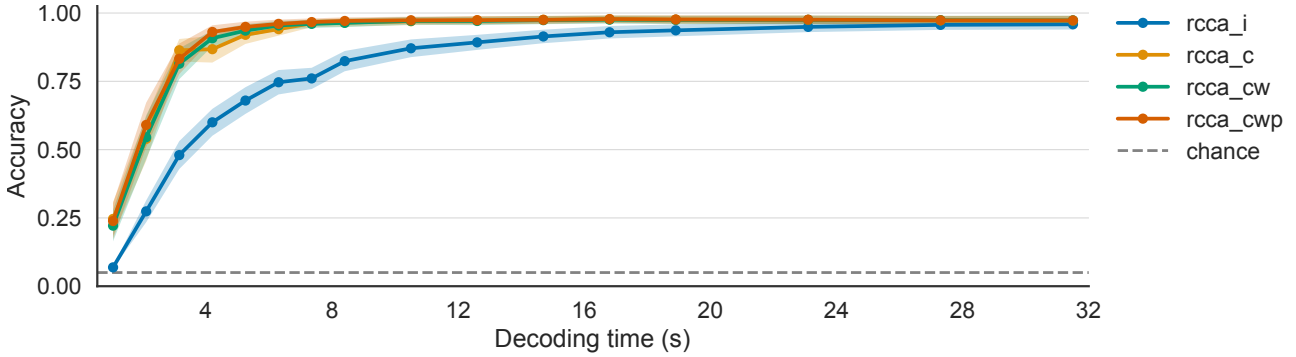

Figure 1: **Grand average decoding curve.** The mean calibration-free classification accuracy and its standard error are shown across varying single-trial durations from 1.05 s to 31.5 s. Curves indicate performances for instantaneous rCCA (rcca_i), cumulative rCCA (rcca_c), confidence-weighted cumulative rCCA (rcca_cw), and confidence-weighted cumulative rCCA with post hoc re-analysis (rcca_cwp). The dashed gray line denotes the theoretical chance level (0.05 %).

were evaluated: instantaneous rCCA, cumulative rCCA, confidence-weighted cumulative rCCA, and confidence-weighted cumulative rCCA with post hoc re-analysis. These methods differ in how, and to what extent, previously classified data are exploited for model adaptation.

An important observation is that, while in general yielding lower performance than its cumulative alternatives, instantaneous rCCA ultimately reaches a similar ceiling performance as cumulative rCCA, given sufficient data. Because instantaneous rCCA relies exclusively on the current trial, it does not benefit from cross-trial learning. This makes instantaneous rCCA the simplest and most robust variant with respect to non-stationarities, as it adapts solely to local signal statistics.

Consistent with earlier findings [12, 13], cumulative rCCA significantly outperforms instantaneous rCCA. This confirms that learning from previously classified trials substantially improves model performance. However, cumulative rCCA relies on pseudo-labels, assuming that previously predicted labels are correct, which introduces the risk that unreliable labels degrade model estimates.

To mitigate this issue, the confidence-weighted cumulative rCCA was introduced, which downweights low-confidence trials and emphasizes high-confidence ones during model updates. Contrary to expectations and to earlier results with UMM [14], this confidence-weighting did not yield significant performance gains over vanilla cumulative rCCA. Although rCCA and UMM are structurally similar, differences in the confidence metric and in how weighting is applied (e.g., on averages or also covariance estimates) may explain this discrepancy.

While confidence-weighting reduces the influence of unreliable trials, such trials may still carry incorrect pseudo-labels. Therefor, post hoc re-analysis was introduced, which reconsiders previously classified trials after each model update. It is assumed that a more informed model produces more accurate pseudo-labels. Indeed, confidence-weighted cumulative rCCA with post hoc re-analysis significantly outperformed both vanilla and confidence-weighted cumulative rCCA at shorter trial durations, indicating that iterative correction can meaningfully improve calibration-free decoding, in line with other studies showing such improvement [15, 16].

It is worth noting that the instantaneous and cumulative rCCA results reported here are slightly higher than those in earlier work [12], despite using the same dataset and methods. The only difference lies in the downsampling rate: 120 Hz in the present study versus 180 Hz previously. The lower sampling rate may have reduced model capacity and the risk of overfitting, leading to marginally improved performance.

Although the proposed extensions demonstrate the feasibility of confidence-weighting and post hoc re-analysis for calibration-free rCCA, several open questions remain. First, the confidence metric used here, the normalized correlation margin, was inspired by earlier dynamic stopping work [6] and by UMM [14]. While it clearly separated correct from incorrect predictions, its lack of impact on performance suggests that alternative confidence measures may be more effective. Furthermore, the confidence-weighting was applied at the trial-level decoding in this work. Instead, UMM applies the confidence-weighting at the epoch-level which may influence the impact it has on the decoding process [14].

Second, confidence values were estimated using instantaneous rCCA to ensure availability for the current trial. An alternative is to derive confidence from the most recent cumulative model, although preliminary experiments in this study did not show improvements.

Third, post hoc re-analysis was limited to updating pseudo-labels only. If confidence values were estimated using cumulative rCCA, these too could be re-estimated, potentially yielding additional benefits.

Fourth, the cumulative variants maintain a complete history of all previous trials. This is memory-intensive and may become impractical in long sessions. Moreover, retaining all past data may be suboptimal under non-stationary conditions. Future work could explore sliding windows like first-in–first-out buffers, or exponential forgetting mechanisms to balance memory constraints, robustness as well as adaptability.

Finally, both this and previous studies benchmark

10.3217/978-3-99161-093-9-004

calibration-free methods using offline datasets [12, 13]. While this enables controlled comparisons, generalization to real-world use must ultimately be validated in online experiments, as demonstrated in earlier work [7]. In such calibration-free online scenario, decoding curves cannot be estimated, and static stopping times cannot be pre-calibrated. Instead, dynamic stopping strategies are required, ideally also in a calibration-free manner. The normalized correlation margin used here could serve this purpose, although it requires a predefined threshold [6]. Alternative dynamic stopping approaches that avoid explicit calibration [7] warrant further investigation.

CONCLUSION

This study evaluated calibration-free adaptive decoding for c-VEP BCIs by extending reconvolution canonical correlation analysis (rCCA) with confidence-weighted updates and post hoc re-analysis. Four variants were compared: instantaneous rCCA, cumulative rCCA, confidence-weighted cumulative rCCA, and confidence-weighted cumulative rCCA with post hoc re-analysis.
The results confirm that cumulative learning substantially improves decoding performance over instantaneous classification. Furthermore, while confidence-weighting alone did not significantly outperform vanilla cumulative rCCA, the addition of post hoc re-analysis yielded consistent improvements, demonstrating that iterative correction of pseudo-labels is an effective strategy in calibration-free settings.
Overall, these findings show that adaptive rCCA can be made more robust by explicitly addressing pseudo-label uncertainty, thereby further reducing the need for calibration and advancing practical plug-and-play c-VEP BCIs.

ACKNOWLEDGEMENTS

This work was supported by the Dutch Research Council (grant 024.005.022). This project has received funding from the European Union's research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101118964.

REFERENCES

[1] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain–computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. Journal of Neural Engineering. 2021;18:061002.
[2] Gao S, Wang Y, Gao X, Hong B. Visual and auditory brain–computer interfaces. IEEE transactions on Biomedical Engineering. 2014;61(5):1436–1447.
[3] Thielen J. Addressing BCI inefficiency in c-VEP-based BCIs: A comprehensive study of neurophysiological predictors, binary stimulus sequences, and user comfort. Biomedical Physics & Engineering Express. 2025.
[4] Dehais F, Castillos KC, Ladouce S, Clisson P. Leveraging textured flickers: A leap toward practical, visually comfortable, and high-performance dry EEG code-VEP BCI. Journal of Neural Engineering. 2024;21(6):066023.
[5] Tangermann M et al. Learning from small datasets–review of workshop 6 of the 10th international BCI meeting 2023. Journal of Neural Engineering. 2025.
[6] Thielen J, Broek P van den, Farquhar J, Desain P. Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. PLOS One. 2015;10(7):e0133797.
[7] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials for brain–computer interface. Journal of Neural Engineering. 2021;18(5):056007.
[8] Turi F, Gayraud T, Clerc M. Auto-calibration of c-VEP BCI by word prediction. HAL Open Science. 2020.
[9] Zheng L, Dong Y, Tian S, Pei W, Gao X, Wang Y. A calibration-free c-VEP based BCI employing narrow-band random sequences. Journal of Neural Engineering. 2024;21:026023.
[10] Santamaría-Vázquez E, Martínez-Cagigal V, Ruiz-Gálvez R, Martín-Fernández A, Pascual-Roa B, Hornero R. Towards calibration-free user-friendly c-VEP-based BCIs: An exploratory study using deep-learning. In: Converging Clinical and Engineering Research on Neurorehabilitation V. Springer Nature Switzerland: Cham, 2025, 685–689.
[11] Behboodi M, Kinney-Lang E, Etemad A, Kirton A, Abou-Zeid H. Leveraging foundation models for calibration-free c-VEP BCIs. In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2025, 4564–4571.
[12] Thielen J, Sosulski J, Tangermann M. Exploring new territory: Calibration-free decoding for c-VEP BCI. In: 9th Graz Brain-Computer Interface Conference 2024. 2024, 325–330.
[13] Thielen J, Tangermann M. Exploring new territory II: Calibration-free decoding for ERP BCI. In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2025, 3788–3793.
[14] Sosulski J, Tangermann M. UMM: Unsupervised mean-difference maximization. arXiv preprint arXiv:2306.11830. 2023.
[15] Kindermans PJ, Schreuder M, Schrauwen B, Müller KR, Tangermann M. True zero-training brain-computer interfacing–an online study. PLOS One. 2014;9(7):e102504.
[16] Hübner D, Verhoeven T, Schmid K, Müller KR, Tangermann M, Kindermans PJ. Learning from label proportions in brain-computer interfaces: Online unsupervised learning with guarantees. PLOS One. 2017;12(4):e0175856.
[17] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface. Version 3. 2023. [Online]. Available: `https://doi.org/10.34973/9txv-z787`.