Proceedings of the
10th Graz Brain-Computer Interface Conference 2026
10.3217/978-3-99161-093-9-022

Optimizing Bit-wise Linear Decoding for Imperceptible Code-Modulated Visual Evoked Potentials (I-c-VEP)

M. A. Fodor[1,2], M. Tangermann[3], J. Thielen[3], I. Volosyak[1,2]

[1]Faculty of Technology and Bionics, Rhine-Waal University of Applied Sciences, Kleve, Germany
[2]Department of Informatics and Data Science, Graduate School for Applied Research in North Rhine-Westphalia, Bochum, Germany
[3]Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

E-mail: Ivan.Volosyak@hochschule-rhein-waal.de

ABSTRACT: Visual evoked potential (VEP)-based brain-computer interfaces (BCIs) are a well-established paradigm in the field, yet they remain limited by a fundamental usability constraint: they rely on perceptible low-frequency flicker, which induces visual fatigue and discomfort. To address this limitation, we previously proposed modulating an imperceptible high-frequency (60 Hz) carrier based on a pseudo-random binary code sequence. The resulting imperceptible code-modulated VEP (I-c-VEP) paradigm substantially reduces perceived flicker compared to conventional c-VEP approaches, while preserving key advantages over steady-state VEP (SSVEP) systems, namely efficient target scalability and robustness to narrowband interference.

In this study, we explore whether simple, efficient linear decoding pipelines suffice for bit-wise I-c-VEP classification across 16 participants. By validating our pipelines on strictly independent, temporally held-out data, we demonstrated that approaches based on linear ridge regression and linear discriminant analysis can achieve over 97% mean trial accuracy. These findings provide a solid foundation for efficient and high-performance real-time implementations of the I-c-VEP paradigm.

INTRODUCTION

Brain-computer interfaces (BCIs) can establish a direct communication pathway between the brain and external devices, enabling vital assistance for individuals with severe motor impairments by translating brain activity into control commands [1].

BCIs using the visual evoked potential (VEP) are favored for intuitive gaze control, offering robust evoked responses that can be reliably captured with non-invasive electroencephalography (EEG). The two prevalent VEP-based encoding strategies are steady-state VEP (SSVEP), which encodes targets by triggering VEPs through flashes at constant frequencies [2], and code-modulated VEP (c-VEP), which encodes targets by triggering VEPs in a pseudo-random sequence [3]. While SSVEP remains popular, c-VEP is often favored as it encodes information in the temporal structure of pseudo-random stimulation sequences rather than narrow spectral peaks. The use of unique sequences or circular shifts of a base sequence facilitates efficient target scalability, while also offering inherent robustness to narrowband interference [3].

Despite their performance, the broader adoption of VEP-based BCIs is constrained by a fundamental usability limitation: their reliance on high-contrast flashing as stimulation. Commonly, a black and white low-frequency (5–25 Hz) flicker is used to elicit robust neural responses. However, prolonged use induces significant visual fatigue, cognitive load, and discomfort [4], and the highly visible flicker can act as a distraction that can influence task performance [5].

To mitigate these adverse effects, one of the established avenues has been the adoption of high-frequency SSVEP paradigms (> 30 Hz) to reduce flicker perceptibility [4, 6]. Visual comfort generally increases with frequency until the critical flicker fusion (CFF) threshold is reached, at which point a temporally modulated light source is perceived as steady [6]. Stimulation above 50–60 Hz, termed rapid invisible frequency tagging (RIFT) [5], typically is no longer consciously perceived. Importantly, neurophysiological evidence demonstrates that the visual cortex continues to entrain to these frequencies even without conscious awareness [5, 6]. However, because the amplitude of the evoked response decreases with increasing stimulation frequency, the resulting signal-to-noise ratio (SNR) deteriorates at higher frequencies [5, 6]. Consequently, although RIFT offers a compelling route toward an imperceptible BCI stimulus, its practical adoption remains constrained by this fundamental SNR trade-off.

In a preceding study [7], we introduced the *imperceptible code-modulated VEP* (I-c-VEP) paradigm, in which a high-frequency (60 Hz) carrier used in RIFT is amplitude-modulated ("on" and "off") according to a predefined pseudo-random code sequence. Specifically, during the "off" state, the stimulus is displayed at a constant mean luminance, while during the "on" state, it flickers at 60 Hz between minimum and maximum luminance, reducing perceived flashing sensation as the target blends

10.3217/978-3-99161-093-9-022

into a more steady, quasi-uniform gray.

This approach bridges concepts from c-VEP and SSVEP research to address a fundamental usability limitation of VEP-based BCIs: visual discomfort caused by perceptible flicker. Although bit transitions themselves still introduce perceptible artifacts, this initial implementation may already represent one of the c-VEP BCI paradigms with the lowest amount of perceptible flicker to date, while preserving the key advantages of c-VEP approaches, efficient target scalability and robustness to narrowband interference.

When using code-modulated stimuli, one could classify the responses to unique events in the stimuli (i.e., bit-level decoding) instead of directly discriminating among multiple target classes (i.e., trial-level decoding). While decoding at the event level through models such as ridge regression [8] or reconvolution [9] can offer significant flexibility in stimulus design and high data efficiency, its adoption has been limited compared to prevalent and highly effective trial-level decoding strategies [3].

In this study, we explore whether simple, efficient linear bit-wise decoding pipelines suffice for the I-c-VEP paradigm. Across 16 participants, we evaluate ridge regression [8], shrinkage-LDA [10], block-Toeplitz LDA [11], and Riemannian tangent space LDA [12, 13]. We hypothesize that within this selection, we can identify a pipeline that achieves high accuracy while remaining data-efficient, providing a robust candidate for future implementations.

MATERIALS AND METHODS

Participants: The study was conducted at the Brain-Computer Interface laboratory of the Rhine-Waal University of Applied Sciences, in accordance with the Declaration of Helsinki, following approval from the Medical Faculty Ethics Committee of the Duisburg–Essen University (24-11957-BO). The full experiment involved 20 participants and included phases related to a separate study. Sixteen participants (9 female, 7 male, mean age of 24.81 ± 3.17 years) completed the full protocol within the scheduled 1.5-hour and were therefore included in the present analysis.

Hardware and Software: We used an Acer Nitro XV252Q F LCD monitor (1920×1080 resolution) at 360 Hz refresh rate, positioned 80 cm from the participant. To ensure timing and luminance stability, vertical synchronization (V-Sync) was enabled, while adaptive sync technologies and dynamic contrast settings were disabled. The panel operated at a maximum brightness with a gamma of 2.2, and with the "Overdrive" setting set to "Normal" to avoid overshoot artifacts. Stimulus presentation was done by a Dell workstation (AMD Ryzen 9 7950X3D CPU, NVIDIA RTX 4070 Ti Super GPU).

Continuous EEG data were acquired at 1200 Hz using a g.USBamp biosignal amplifier (Guger Technologies, Schiedlberg, Austria). Passive Ag/AgCl sensors (EasyCap GmbH, Wörthsee, Germany) were arranged in a 16-channel montage focused on the parieto-occipital cortex, adhering to the international 10–5 system: P7, P3, Pz, P4, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, O9, Iz, and O10. The reference and ground electrodes were placed at Cz and AFz, respectively, and impedances were maintained below 5 kΩ throughout the session.

Stimulus: The I-c-VEP paradigm displayed four gray square targets (7.08×7.08 cm; 250 px) at 360 Hz arranged horizontally on a black background with a 5.66 cm (200 px) spacing. Each target had a 1.23 cm (40 px) numeric label in the center. Trials presented two full cycles of a 31-bit m-sequence (2×2.583 s). Stimulation was implemented by gating a 60 Hz carrier signal on and off according to the m-sequence $s = 1111101100111000011010100100010$ at a presentation rate of 12 Hz (30 frames per bit; ≈ 83.33 ms). A logical bit '1' encoded the active 60 Hz flicker, realized by a square wave alternating 3 black and 3 white frames per cycle at the native 360 Hz refresh rate. Conversely, bit '0' encoded the static (off) state, displaying a mean luminance of 0.5 in linear color space (sRGB $(188, 188, 188)$) to ensure isoluminance (gray appearance) with the time-averaged intensity of the flicker (on) state.

Experimental Protocol: The full experiment spanned approximately ≈90 min, including 20–30 min for the EEG setup and post-experiment procedures. For this study, two runs were conducted at the beginning and the end of the session to maximize their temporal distance. The first run was used for selecting and training the decoding models, and the second served as a hold-out test set. The interim period (50–60 min) was used for tasks of a separate study. This included a first sequential target selection task (≈6 min), another training (≈5 min), a second sequential target selection task (≈6 min), and approximately 10 min to collect subjective rating of various c-VEP stimuli. Each run contained 40 trials, during which participants sequentially fixated one of the four targets. A single trial comprised 5.16 s of continuous stimulation. Following each trial, a 2 s gaze-shift interval was provided to ensure participants had sufficient time to comfortably fixate the next target.

Preprocessing: For all models, the raw continuous EEG data ($f_{s\_raw} = 1200$ Hz) were high-pass filtered at 1 Hz (second-order Butterworth) to remove baseline drift, and notch-filtered at 50 and 100 Hz to mitigate power line interference. To prevent aliasing prior to downsampling, a fourth-order Butterworth low-pass filter was applied at 120 Hz (80% of the target 150 Hz Nyquist limit). The data were then decimated by a factor of 4 to a working sampling rate of $f_s = 300$ Hz. Finally, 300 ms epochs time-locked to the bit-transition onsets were extracted to capture the corresponding VEP morphology.

While focusing on decoding approaches, we evaluate multiple frequency bands to explore which ranges yield better performance and to account for potential synergies between specific pipelines and spectral filters. Three categories of frequency bands were considered: (i) broadband c-VEP filters (1–40 Hz, 1–65 Hz, and 1–90 Hz), (ii)

10.3217/978-3-99161-093-9-022

low, medium and high frequency bands (1–15 Hz, 15–40 Hz, and 40–90 Hz), and (iii) I-c-VEP-specific bands (55–65 Hz, and 58–62 Hz and 10–65 Hz, 20–65 Hz). The latter bands were intended to include the 60 Hz carrier as well as the broadband c-VEP response. The exclusion of the alpha band was motivated by the hypothesis that the perceptually subtle I-c-VEP stimulus may provide a lower visual drive than traditional flicker; consequently, endogenous alpha activity might be less suppressed. All filters were applied causally (forward-pass only) via cascaded second-order sections (SOS).

For spatial filtering, xDAWN [14] (2, 4, or 6 components) was applied across all LDA models to maximize the signal-to-signal-plus-noise ratio of the evoked responses. The ridge pipeline used canonical correlation analysis (CCA; 2 or 4 components), a common spatial filter for c-VEP decoding [9], to maximize correlation with class-specific templates, as in EEG2Code [8]. To isolate the 60 Hz carrier, common spatial patterns (CSP) [15] was evaluated alongside xDAWN to extract filters that maximize variance differences between classes. From the resulting squared amplitudes, either the full time-series or the mean power within a 210–270 ms post-transition window (centered on the observed peak of class discriminability) was extracted. Both raw and log-transformed inputs were evaluated.

Decoding Models: Five bit-wise linear decoding models were evaluated. For each bit, a single EEG epoch time-locked to the bit transition was extracted and linearly projected to a scalar output, which was subsequently used for bit-level decoding.

The first model, *linear ridge regression (Ridge)*, is a proven decoding approach for c-VEP [8]. It applies L_2 regularization to mitigate overfitting, with the penalty parameter (α) optimized via cross-validation on the training set. While inspired by the EEG2Code framework, our implementation extracts the epochs with a one-bit stride to yield exactly one prediction per bit, ensuring direct comparability with the other evaluated linear models.

The second, *shrinkage LDA (sLDA)* with Ledoit-Wolf shrinkage regularization of the covariance matrix, systematically shrinks the empirical covariance matrix toward a structured target, mitigating the ill-conditioning inherent to high-dimensional EEG feature spaces and limited training data [10].

The third, *block-Toeplitz LDA (btLDA)* assumes that the covariance between two time points depends only on their temporal distance rather than absolute position, and that covariance decreases with temporal distance. This assumption allows to regularize the covariance matrix into a block-Toeplitz structure with drastically reduced free parameters. Applying this temporal regularization alongside standard shrinkage significantly improves performance, particularly in limited-data scenarios [11].

Fourth, *Riemannian tangent space LDA (tsLDA)* was implemented [12]. Standard spatial covariance estimation discards the time-locked temporal morphology critical for decoding VEPs. To map this temporal morphology onto the Riemannian manifold, "super-epochs" were constructed by concatenating the raw single-trial EEG epochs with target-specific template responses. These templates were generated by calculating the grand average of the time-locked training epochs for each respective class (i.e., the mean evoked responses for bit '0' and bit '1'). Following this, spatial covariance estimation was performed directly on the super-epochs, an approach proven successful for the c-VEP domain [13] to effectively embed the temporal phase information directly into the spatial covariance matrix. Resulting symmetric positive definite (SPD) matrices were projected into the tangent space using the Fréchet mean of the training data as a fixed reference, enabling robust classification via sLDA.

Finally, *feature-level (early) and score-level (late) fusion* strategies were assessed to integrate complementary temporal, spatial, and spectral information. Early-fusion concatenated features extracted across a filter bank into a unified vector, which was standardized using parameters estimated from the training set and classified by a single sLDA model. Late-fusion aggregated the outputs of independently trained base models. Raw decision scores (hyperplane distances) from each model were z-scored using the mean and standard deviation derived from the training set, averaged, and transformed via a logistic sigmoid function to yield a final posterior probability. This approach effectively fuses predictions in a straightforward way without requiring a separate meta-classifier.

Evaluation: Final pipeline configurations were selected in a staged, domain-informed way to avoid exhaustive permutation testing. To ensure zero data leakage, pipeline configuration and final evaluation were strictly decoupled. Pipelines were ranked and selected based on 5-fold cross-validation performance on the training run, using the normalized area under the learning curve (nAUC; computed with a 4-trial step size) as the primary metric to capture both accuracy and data efficiency.

The selected models from each approach were then evaluated once on the temporally held-out test data (second run, recorded ≈1 h later). Learning curves were computed using a fine-grained 1-trial step size to assess performance under both rapid (≈55 s; 8 trials) and full (≈4.75 min; 40 trials) training scenarios. Decoding curves were generated per-bit (83.33 ms steps) to track classification accuracy over the trial duration. We report the final peak accuracy (±SD) and a "plateau point"—defined as the earliest evaluation time not significantly lower than the peak (one-sided Wilcoxon, $p > 0.05$). Differences between models were assessed via two-sided Wilcoxon signed-rank tests with step-down Holm-Bonferroni correction. Finally, simulated information transfer rate (ITR; see `https://bci-lab.hochschule-rhein-waal.de/en/itr.html`) was calculated to quantify system throughput.

RESULTS

The section first describes the staged selection of the

10.3217/978-3-99161-093-9-022

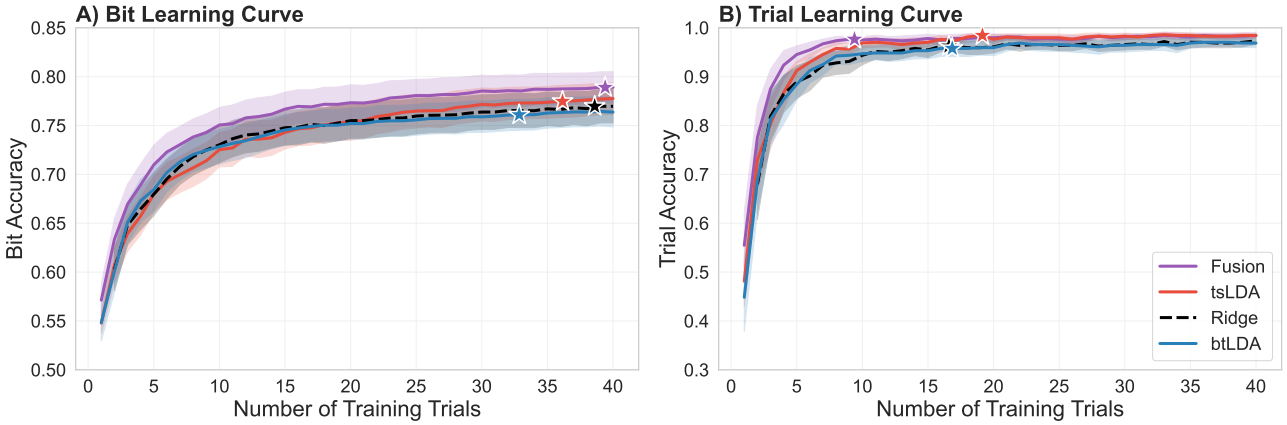

Figure 1: **Grand average offline learning curves across $N = 16$ participants.** **(A)** Bit-wise (binary) classification accuracy as a function of training trials. **(B)** Trial-level (4-class) decoding accuracy evaluated using the full 5.16 s trial duration. Solid lines represent the mean performance across subjects, while the shaded regions indicate ± 1 standard error of the mean (SEM). The star markers denote the plateau point for each curve, defined statistically via a one-sided Wilcoxon signed-rank test as the earliest time point not significantly worse ($p > 0.05$) than the peak accuracy.

pipelines and then reports and compares the performance of the selected, best implementation for each pipeline. To maintain clarity, the learning and decoding curves present only the best configurations for each approach: the *Ridge*, *btLDA*, *tsLDA*, and the *Late-fusion* LDA model, all evaluated on the held-out test set. Figure 1 illustrates the mean bit- and trial-level accuracies as a function of training data volume, while Figure 2 depicts trial decoding accuracies over evaluation time. Of the 62 bits comprising each trial, the final 4 were excluded to prevent classification of data from the post-stimulus interval. Consequently, a single 40-trial run yielded $40 \times 58 = 2{,}320$ epochs, corresponding to one epoch per evaluated bit.

Optimal Models:

Time-Domain Pipeline (btLDA): First, *sLDA* was evaluated with the broadband c-VEP and I-c-VEP-specific frequency bands using xDAWN with 4 spatial components. The broadband 20–65 Hz filter yielded the highest performance (CV-nAUC: 76.0%), significantly outperforming the second-best 10–65 Hz band (CV-nAUC: 74.8%; $W = 7.0$, $p = .001$). Consequently, using the 20–65 Hz band and xDAWN (4 components), we compared *sLDA* against *btLDA*. The *btLDA* model demonstrated significantly better data efficiency, achieving a higher CV-nAUC (76.5% vs. 76.0%; $W = 22.0$, $p = .016$). Going forward with *btLDA* and xDAWN, we evaluated the number of spatial components (2, 4, and 6). Extracting 2 components yielded the highest CV-nAUC (76.7%), though the difference compared to 4 components was not statistically significant ($p = .159$). Consequently, the final optimal *btLDA* model used a 20–65 Hz bandpass filter, xDAWN extracting 2 components.

Riemannian Pipeline (tsLDA): For the tsLDA pipeline, we utilized xDAWN with two components and evaluated three bandpass configurations: the previously established 20–65 Hz band, a broad 1–90 Hz band, and an early-fusion approach. The latter concatenated features from low (1–15 Hz), mid (15–40 Hz), and high (40–90 Hz) bands prior to classification. This *Early-fusion* configuration significantly outperformed the 20–65 Hz band, yielding a CV-nAUC of 76.9% compared to 74.3% ($W = 1.0$, $p < .001$). Consequently, the early-fusion xDAWN implementation with two components was selected as the optimal tsLDA pipeline.

60 Hz Carrier Pipeline: To isolate the carrier (55–65 Hz), we compared spatial filters (xDAWN vs. CSP), component counts (2 vs. 4), and features (raw vs. log-power; binned vs. full time-series). xDAWN significantly outperformed CSP ($W = 0.0, p < .001$), and raw power proved superior to log-transformation. Temporal binning was used as it minimized dimensionality with negligible ($W = 66.0, p = .940$) CV-nAUC loss compared using the full time-series. Using binned raw power, 2-component xDAWN yielded a significantly higher CV-nAUC than 4-component (64.4% vs. 64.2%; $W = 18.0, p = .008$). This configuration achieved $81.4\% \pm 20.4\%$ trial accuracy with 40 training trials.

Linear Ridge Regression (Ridge): For the ridge regression model, evaluating the number of CCA components revealed that extracting 2 components significantly outperformed 4 components (76.9% vs. 76.5%; $W = 21.0$, $p = .013$).

Early and Late Fusion: The aim was to combine the best complementary frequency bands and model architectures. In the *Early-fusion* approach, concatenating the 20–65 Hz band with narrowband 60 Hz features improved upon the isolated pipelines, achieving a CV-nAUC of 77.4%. The *Late-fusion* approach, which aggregated the probabilistic outputs of the best-performing *btLDA*, *tsLDA*, and 60 Hz models, yielded the highest overall performance (CV-nAUC: 78.9%). This configuration significantly outperformed the next best model (*tsLDA*) ($W = 0.0$, $p < .001$), establishing it as the best performing model.

Table 1 provides the mean bit-wise classification accuracies, nAUC scores, and trial-level accuracies (T = 40) for

10.3217/978-3-99161-093-9-022

Table 1: **Offline classification performance using the full 40-trial training set.** Accuracies are reported as % (±SD). Bit nAUC is the normalized area under the bit-wise learning curve.

Model	Bit Acc.	Bit nAUC	Trial Acc.
Ridge	77.0 ± 6.8	74.2	97.3 ± 3.0
btLDA	76.4 ± 6.1	73.6	96.9 ± 3.7
tsLDA	77.8 ± 6.5	73.9	98.4 ± 1.5
Early-fusion	77.8 ± 6.9	74.5	97.5 ± 3.7
Late-fusion	78.9 ± 6.4	75.9	98.4 ± 2.2

the models evaluated on the hold-out test set.

Learning: Notably, the *Late-fusion* model significantly outperformed all other models in both bit-wise nAUC ($p < .001$) and final bit accuracy ($p < .002$, Holm-Bonferroni corrected). Figure 1 demonstrates that all models achieved high trial-level accuracies given 40 training trials. While the *Late-fusion* model achieved the highest final trial accuracy ($98.4\% \pm 2.2\%$), it did not differ significantly from the *tsLDA* ($p = .353$) or *btLDA* ($p = .096$) baselines at this terminal point after multiple comparison correction. However, the *Late-fusion* model demonstrated superior data efficiency, achieving a significantly higher trial-level nAUC than all baselines ($p < .01$, Holm-Bonferroni corrected).

We defined a performance plateau as the earliest point not significantly below peak performance (one-sided Wilcoxon signed-rank test, $p > 0.05$). The *Late-fusion* model converged the fastest, reaching 97.7% at just $T = 9$ trials. In contrast, other models required more data: *Early-fusion* plateaued at $T = 16$ (96.1%), *Ridge* and *btLDA* at $T = 17$ (96.2% and 95.8%), and *tsLDA* eventually plateauing at its peak accuracy of 98.4% at $T = 19$.

Decoding: When constrained to only 8 trials of training data (Figure 2A), *Late-fusion* achieved the highest performance, plateauing at $T = 3.42$ s (96.6%) with a final accuracy of $97.3\% \pm 3.5\%$. In this limited-data regime, *Late-fusion* significantly outperformed all other models across both decoding nAUC ($p < .01$) and final accuracy ($p < .05$, Holm-Bonferroni corrected). Under this same training constraint, the final accuracies of the other models were lower: $95.8\% \pm 4.6\%$ for *tsLDA*, $94.2\% \pm 6.4\%$ for *btLDA*, and $93.8\% \pm 9.9\%$ for *Early Fusion*.

With the full 40-trial training set (Figure 2B), *Late-fusion* plateaued at $T = 3.42$ s (98.4%) with a final accuracy of $98.4\% \pm 2.2\%$. It maintained a higher nAUC than *Ridge*, *btLDA*, and *Early-fusion* ($p \leq .001$), though the difference against *tsLDA* was not significant ($p = .052$). Similarly, *Late-fusion*'s final accuracy did not significantly differ from *tsLDA* ($p = .353$) or *btLDA* ($p = .096$) after correction (see the final column of Table 1).

DISCUSSION

In the section, we address the key findings of the staged pipeline selection process and highlight the *Late-fusion* approach, which achieved the strongest performance.

The superior performance with the 20–65 Hz band may suggest that the subtle I-c-VEP stimulus induces a weaker visual drive, leaving low-frequency endogenous alpha activity less suppressed. Consistent with Riemannian theory [12], *tsLDA* benefited from a broader spectrum due to its manifold metric, which is inherently robust to the high-variance noise that typically degrades Euclidean classifiers.

While the 81.4% trial accuracy of the narrowband 60 Hz classifier confirms the carrier's discriminative value, the superior performance of the broadband models demonstrates that lower frequencies contain informative evoked components, providing essential complementary features.

The success of the *Late-fusion* underscores the benefits of combining fundamentally different models. Because individual models can provide partially orthogonal views of the neural data, their errors are likely not fully correlated. Furthermore, regularization in the base models may prevent highly confident misclassifications. Consequently, when averaging them, the scores of misclassifications may sometimes be suppressed by the confident correct predictions of complementary models, especially when aggregating multiple robust pipelines.

Regarding data efficiency, the *Late-fusion* model peaked at 98.4% accuracy (40 training trials) but reached a statistical plateau at just nine trials (62.4 s training time; 97.7% accuracy), demonstrating that calibration time can be substantially reduced. Future work should extend this toward zero-training paradigms for maximum efficiency [9].

Using all 40-trials for training, the *Late-fusion* model yielded an 88.8% accuracy at the 1.17 s decoding mark. Assuming a 1.5–1.0 s gaze-shift interval, this translates to an offline simulated ITR of 34.0–41.9 bits/min. While competitive for a four-target interface, the 12 Hz m-sequence presentation rate bottlenecks the ITR, as the long bit duration increases the minimum time required for target discrimination. Future studies should explore higher presentation rates to reduce the required decoding time and maximize system throughput. Finally, generating and aggregating multiple predictions per bit—similar to the decoding strategy seen in the EEG2Code framework [8]—could further enhance classification robustness. Furthermore, as this study establishes that the underlying code patterns remain reliably decodable, future research should investigate various high-frequency carriers and phase-shifting techniques to further enhance signal decodability and system throughput.

CONCLUSION

In this study, bit-wise linear decoding strategies for the imperceptible code-modulated visual evoked potentials paradigm were selected through a staged, domain-informed calibration process and validated on an independent hold-out dataset recorded one hour post-training across 16 participants. Both ridge regression and regularized LDA models exceeded 97% mean trial-level accuracy, with a *Late-fusion* model, which aggregated the

10.3217/978-3-99161-093-9-022

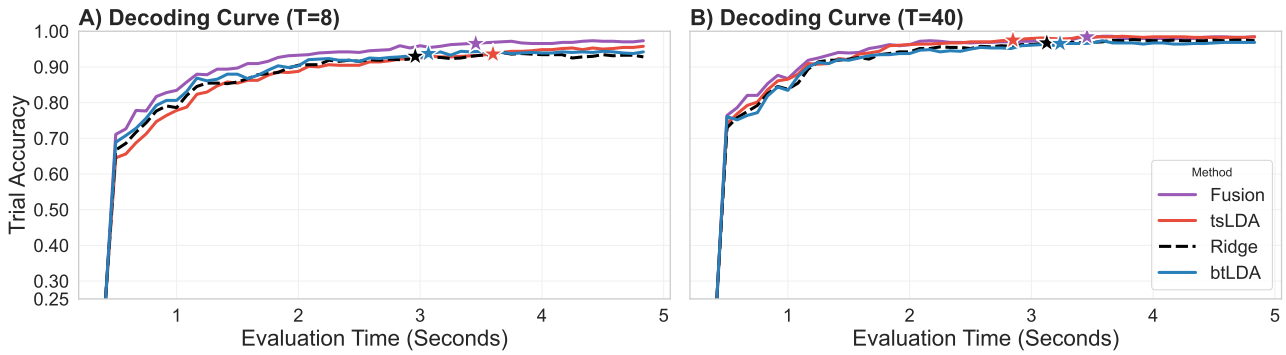

Figure 2: **Average offline decoding curves** as a function of evaluation time for $N = 16$ participants, 4 targets. **(A)** Model performance in a limited-data scenario, trained on only 8 trials (≈55 s of training). **(B)** Model performance trained on the full dataset (40 trials). The star markers denote the plateau point for each curve, defined statistically via a one-sided Wilcoxon signed-rank test as the earliest time point not significantly worse ($p > .05$) than the peak accuracy. The theoretical chance level is 25%.

outputs of independently trained base models, achieving a peak accuracy of 98.4%. These results establish that linear bit-wise decoding can provide a robust and computationally efficient foundation for high-performance, real-time I-c-VEP brain-computer interfaces.

ACKNOWLEDGEMENTS
This work was supported by the European Union's research and innovation program under the Marie Skłodowska-Curie grant agreement No 101118964, the "The Friends of the University Rhine-Waal - Campus Cleve" association, and the Dutch Research Council (NWO) (project number 024.005.022).

AUTHOR CONTRIBUTIONS
MAF: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft; MT, JT: Conceptualization, Methodology, Validation, Writing - review & editing, Supervision; IV: Funding acquisition, Writing - review & editing, Supervision.

REFERENCES
[1] Wolpaw JR, R. Millán J del, Ramsey NF. Chapter 2 - Brain-computer interfaces: Definitions and principles. In: Brain-Computer Interfaces. Elsevier, 2020, 15–23.
[2] Vialatte FB, Maurice M, Dauwels J, Cichocki A. Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. Progress in Neurobiology. 2010;90(4):418–438.
[3] 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(6):061002.
[4] Azadi Moghadam M, Maleki A. Fatigue factors and fatigue indices in SSVEP-based brain-computer interfaces: a systematic review and meta-analysis. Frontiers in Human Neuroscience. 2023;17.
[5] Brickwedde M, Bezsudnova Y, Kowalczyk A, Jensen O, Zhigalov A. Application of rapid invisible frequency tagging for brain computer interfaces. Journal of Neuroscience Methods. 2022;382:109726.
[6] Liu K et al. A high-frequency SSVEP-BCI system based on a 360 Hz refresh rate. Journal of Neural Engineering. 2023;20(4):046042.
[7] Fodor MA, Volosyak I. Towards Visual-Fatigue-Free BCI with Imperceptible Visual Evoked Potentials (I-VEP). In: 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Oct. 2025, 3813–3818.
[8] Nagel S, Spüler M. Asynchronous non-invasive high-speed BCI speller with robust non-control state detection. Scientific Reports. 2019;9(1).
[9] Thielen J, Marsman P, Farquhar J, Desain P. From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface. Journal of Neural Engineering. 2021.
[10] Blankertz B, Lemm S, Treder M, Haufe S, Müller, K-R. Single-trial analysis and classification of ERP components — A tutorial. NeuroImage. 2011;56(2):814–825.
[11] Sosulski J, Tangermann M. Introducing block-Toeplitz covariance matrices to remaster linear discriminant analysis for event-related potential brain–computer interfaces. Journal of Neural Engineering. 2022;19(6):066001.
[12] Barachant A, Bonnet S, Congedo M, Jutten C. Multiclass Brain–Computer Interface Classification by Riemannian Geometry. IEEE Transactions on Biomedical Engineering. 2012;59(4):920–928.
[13] Ying J, Wei Q, Zhou X. Riemannian geometry-based transfer learning for reducing training time in c-VEP BCIs. Scientific Reports. 2022;12(1).
[14] Rivet B, Souloumiac A, Attina V, Gibert G. xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain–Computer Interface. IEEE Transactions on Biomedical Engineering. 2009;56(8):2035–2043.
[15] Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering. 2000;8(4):441–446.