

PAPER • OPEN ACCESS

Improving P300 morphology through single-trial latency realignment: a comparative study of template-matching approaches

To cite this article: Ilaria Quattrocioni *et al* 2026 *J. Neural Eng.* **23** 036033

View the [article online](#) for updates and enhancements.

You may also like

- [Adapted wavelet transform improves time-frequency representations: a study of auditory elicited P300-like event-related potentials in rats](#)
Nelly Richard, Bettina Laursen, Morten Grupe *et al.*
- [Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy](#)
David E Thompson, Seth Warschausky and Jane E Huggins
- [Latency correction of event-related potentials between different experimental protocols](#)
I Iturrate, R Chavarriaga, L Montesano *et al.*



PAPER

OPEN ACCESS

RECEIVED
30 January 2026

REVISED
14 May 2026

ACCEPTED FOR PUBLICATION
3 June 2026

PUBLISHED
17 June 2026

Original content from
this work may be used
under the terms of the
Creative Commons
Attribution 4.0 licence.

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Improving P300 morphology through single-trial latency realignment: a comparative study of template-matching approaches

Ilaria Quattrocio^{1,2,3}, Valentina Caracci^{1,2,3}, Emanuela Rotondo^{1,2}, Emma Colamarino^{1,2},
Floriana Pichiorri², Angela Riccio², Febo Cincotti^{1,2}, Jlenia Toppi^{1,2,4,*} and Laura Astolfi^{1,4}

¹ Department of Computer, Control and Management Engineering, Sapienza University of Rome, Rome, Italy

² IRCCS Fondazione Santa Lucia, Rome, Italy

³ authors equally contributed.

⁴ authors equally supervised.

* Author to whom any correspondence should be addressed.

E-mail: jlenia.toppi@uniroma1.it

Keywords: EEG, event-related potentials, latency jitter, p300, template matching algorithm

Supplementary material for this article is available [online](#)

Abstract

Objective. Trial-to-trial latency variability—well known as latency jitter—is a major source of distortion in event-related potential (ERP) analysis, particularly for late cognitive components such as the P300. Several template-matching algorithms have been proposed to estimate single-trial latency and improve ERP reconstruction, but direct comparisons across different methodological approaches remain limited. This study provides a structured evaluation of three representative algorithms: the woody filter (WF), operating in the time domain; the adaptive wavelet filter (CWT-AWF), extending template matching to the time–frequency domain; and ReSync, a decomposition-based method that combines signal decomposition with time-restricted realignment. **Approach.** The algorithms were evaluated using surrogate EEG-like data with controlled amplitude ratios (reported as SNR) and known latency jitter, and real EEG recordings from healthy participants performing an auditory oddball task. Performance was assessed in terms of latency-estimation accuracy, latency variability, ERP morphology, and waveform quality. **Main Results.** Across simulated conditions, ReSync achieved significantly lower latency-estimation errors and reduced variability compared to WF and CWT-AWF, demonstrating robustness even at low SNR levels. Importantly, this advantage persisted when all methods were constrained within the same temporal window, indicating that performance gains are not solely attributable to time restriction. In real EEG data, all algorithms enhanced P300 morphology relative to non-aligned averages, but ReSync yielded the most consistent improvements, including the lowest latency jitter and stable latency distributions within a range consistent with previous findings. Complementary SNR analysis further indicated improved waveform quality when interpreted jointly with latency-based metrics. ReSync also remained stable across both single-channel and multi-channel realignment strategies. **Significance.** These findings highlight the advantage of combining decomposition and targeted realignment for mitigating ERP latency jitter. ReSync provides a reliable and morphology-preserving framework for single-trial ERP analysis, with potential applications in cognitive neuroscience, brain–computer interfaces, and clinical contexts.

1. Introduction

A substantial body of research in neurocognitive science has focused on investigating the brain's dynamic responses to external stimuli, with the aim

of deepening our understanding of brain function and cognition. Typically, a sensory input is presented and the subsequent neural response is recorded. This response, first described in 1964 [1, 2], consists of complex, millisecond-scale patterns known

as event-related potentials (ERPs). ERPs are a powerful tool for exploring how the brain modulates its activity in response to sensory stimulation. They reflect cognitive processes involved in stimulus evaluation and can reveal different stages of information processing—from early sensory encoding to higher-order cognitive operations [3]. Moreover, ERPs provide insight into the hierarchical organization of dynamic neural systems [4–7] and have proven valuable for both diagnostic applications in psychiatry and neurology and for brain–computer interface (BCI) systems used in communication and control [8–11].

ERPs appear as transient voltage deflections embedded within ongoing electroencephalographic (EEG) activity, time-locked to the onset of an external or peripheral stimulus. They can be observed as somatosensory, visual, or auditory responses, or as slow cortical potentials preceding voluntary movements [12]. These responses are measured directly from scalp EEG, which provides the temporal resolution required for this purpose.

Among ERPs, the P300 component is one of the most extensively studied. It manifests as a positive deflection reflecting attention and memory mechanisms involved in stimulus evaluation, such as attentional orientation, contextual updating, and response modulation. The P300 is typically elicited using an oddball paradigm, in which infrequent target stimuli are interspersed among frequent standard stimuli. In healthy young adults, the P300 shows maximal amplitude over centro-parietal sites and peaks around 300 ms after the onset of a rare stimulus [13, 14]. Because ERPs are much smaller in amplitude than the background EEG, signal processing methods are required to enhance their signal-to-noise ratio (SNR) [15]. The most widely used method is time-domain averaging, which computes a sample-by-sample, stimulus-locked average of multiple EEG epochs from repeated presentations of the same event [16]. This approach assumes that ERPs are stationary across trials—i.e. invariant in amplitude and latency—while unrelated EEG activity averages out as noise [17]. While effective for early sensory components such as the N100, this assumption breaks down for later components like the P300, whose timing reflects higher-order cognitive dynamics. In such cases, each stimulus repetition can produce responses with substantial trial-to-trial latency variability—a phenomenon known as latency jitter [18]. This variability may arise from intrinsic cognitive fluctuations [19] or from neural noise [20]. Latency jitter can distort average ERPs, resulting in attenuated amplitudes, shifted peaks, or even complete loss of the response [21]. It may also impair the performance of P300-based BCIs in both healthy users [22] and clinical populations such as patients with amyotrophic lateral sclerosis [23–25].

The detrimental effects of latency jitter have motivated the development of methods for estimating ERP latency at the single-trial level. Critically, these latency estimates are not merely descriptive: they are used to realign each trial to the estimated ERP peak rather than to the stimulus onset. This peak-based realignment is essential not only for reconstructing an ERP waveform that preserves its true morphology, but also for improving the detectability of the component itself, which may otherwise be compromised by latency jitter. In fact, jitter-induced smearing can attenuate or obscure ERP components to the point of hindering their reliable identification, an issue that is particularly relevant in clinical applications where the amplitude, timing, or absence of specific components carries diagnostic significance.

Early latency-correction approaches relied on time-domain filtering and peak detection but proved highly sensitive to low-frequency noise. Template-matching strategies, first introduced by Woody (1967 [26]), represented a major step forward by estimating single-trial latencies through cross-correlation with a predefined ERP template. However, these methods remained vulnerable to noise patterns resembling ERP morphology [27]. Later approaches such as the adaptive wavelet filter (AWF [28]) extended template matching into the time–frequency domain, leveraging wavelet decomposition to isolate event-related oscillatory activity [29, 30] and reduce interference from background noise.

A separate class of approaches—including principal component analysis, independent component analysis (ICA), and the residue iteration decomposition (RIDE)—attempts to address latency variability by decomposing ERPs into subcomponents [31, 32]. Although powerful, these methods depend on structural assumptions that can limit generalizability and may amplify noise when violated [33]. Recently, Ouyang proposed ReSync, a hybrid technique combining decomposition and template matching approaches by estimating latencies and reconstructing the ERP using trial-specific timing information [33]. A key feature of the ReSync approach is that latency estimation is performed within a predefined temporal window corresponding to the expected occurrence of the ERP component. This time-restricted strategy is designed to limit the influence of non-stimulus-locked activity and to reduce the risk of aligning trials to peaks occurring at implausible latencies, as well as to enable the separate realignment of ERP components occurring at different latencies. In contrast, classical template-matching approaches such as WF and AWF typically operate without explicit temporal constraints, potentially increasing their susceptibility to spurious peak alignment.

In multichannel EEG recordings, leveraging spatiotemporal information across electrodes can improve jitter correction by reducing cross-electrode

noise. Group-based or weighted channel correction strategies have also been shown to outperform single-electrode approaches in estimating ERP latencies [27].

Despite the wide range of methods proposed for ERP latency estimation, direct comparisons across different classes of template-matching approaches remain limited. In particular, time-domain (WF), time–frequency-domain (AWF), and decomposition-based (ReSync) methods have typically been evaluated in isolation, making it difficult to understand their relative strengths under controlled conditions. Similarly, the effect of applying latency correction at the single-channel level versus using spatially aggregated corrections across multiple electrodes has not been explicitly examined.

While several methods have been proposed to address latency jitter, a systematic comparison of their performance under controlled conditions remains limited. In particular, the impact of key methodological aspects—such as the presence of explicit temporal constraints on latency estimation and the use of spatially aggregated versus channel-specific realignment—has not been thoroughly investigated. The present study aims to provide a systematic evaluation of ERP realignment strategies for mitigating latency jitter and improving both P300 reconstruction and detectability. Specifically, the work addresses the following questions:

- i. How do time-domain, time–frequency, and decomposition-based template-matching algorithms compare in estimating single-trial ERP latency under varying noise conditions?
- ii. To what extent do these methods improve ERP morphology and component detectability in real EEG data, while preserving the temporal and structural characteristics of the signal?
- iii. How does the presence or absence of an explicit temporal constraint on latency estimation influence the robustness of realignment, particularly in the presence of spurious or non–stimulus-locked activity?
- iv. How does the application of latency correction at the single-channel level versus using spatially aggregated corrections across multiple electrodes influence realignment outcomes?

In this context, the present study aims not only to benchmark representative template-matching approaches operating in different domains (time, time–frequency, and decomposition-based), but also to identify the conditions under which these strategies succeed or fail in reconstructing and detecting the P300 component. By explicitly isolating and comparing these factors, the work provides a practical and conceptual framework for interpreting the behavior of latency realignment algorithms.

2. Methods

2.1. Template matching algorithms

The most common strategy for mitigating the impact of latency jitter on ERP reconstruction involves a three-step procedure: (i) estimating the ERP latency at the single-trial level, (ii) resynchronizing each trial based on the estimated latency rather than the stimulus onset, and (iii) applying time-domain averaging to the realigned trials to obtain an improved ERP waveform. Among these steps, reliable single-trial latency estimation is the most challenging, largely due to the low SNR of event-related responses in individual repetitions.

Template-matching algorithms provide an effective solution by estimating latency shifts through similarity measures computed between each trial and a predefined template, in either the time or time–frequency domain. The procedure typically begins by computing an initial template as the stimulus-locked average across trials. Next, a similarity function—commonly cross-correlation or cross-covariance—is calculated between the template and each single-trial waveform. The peak of this function identifies the temporal lag at which the single trial best matches the template. Each trial is subsequently shifted forward or backward by this lag, thereby aligning the ERP components across repetitions.

The following subsections describe the woody filter (WF), the AWF, and the ReSync method, and detail how each algorithm performs latency estimation and realignment on both synthetic and real EEG datasets. To improve clarity, scalar, vector, and matrix quantities are denoted using distinct notation. In particular, matrix variables are indicated in boldface, while scalar and vector quantities are represented using standard notation. This convention is adopted throughout the manuscript to avoid ambiguity.

2.1.1. WF

The WF is the earliest template-matching approach proposed for reducing latency variability in ERP analysis [26]. It is an adaptive, iterative algorithm operating in the time domain, which uses the stimulus-locked average across trials as an initial template and relies on cross-correlation to estimate the latency of each single-trial ERP.

Let the dataset:

$$x[n] = [x_1(n), x_2(n), \dots, x_K(n)], n = 1, \dots, N \quad (1)$$

consist of K EEG responses time-locked to identical stimulus events recorded from a single electrode, where $x[n]$ denotes the set of single-trial EEG responses, $x_i(n)$ denotes the i th single-trial response in the discrete-time domain, and N is the number of samples per trial.

WF performs alignment of the single-trial responses with respect to a template $x_T(n)$ that is

computed as the stimulus-locked average across trials:

$$x_T(n) = \frac{1}{K} \sum_{i=1}^K x_i(n), n = 1, \dots, N. \quad (2)$$

For each trial, the following steps are executed:

- 1) *Cross-correlation computation.* The cross-correlation $\Phi_i(m)$ between the template $x_T(n)$ and the i th trial is computed as:

$$\Phi_i(m) = \sum_{n=1}^N x_T^*(n) x_i(n+m) \quad (3)$$

with $m = -N, \dots, N$, where m denotes the correlation lag.

- 2) *Latency estimation.* The estimated latency shift $\hat{s}_i$ corresponds to the lag at which the cross-correlation reaches its maximum:

$$\hat{s}_i = \arg \max_m \Phi_i(m). \quad (4)$$

- 3) *Trial realignment.* Each trial is shifted by $\hat{s}_i$ to obtain the aligned waveform $x_i^A(n)$.

2.1.2. AWF

The AWF extends the classical WF approach to the time–frequency domain by operating on the wavelet decomposition of the signal [28]. Instead of aligning ERPs directly in the time domain, AWF applies the WF procedure independently to each multi-resolution scale obtained via wavelet transform. Each scale corresponds to a specific frequency band and can be interpreted as a bandpass-filtered representation of the signal. Because activity at a given scale is assumed to contain a single dominant feature, this multiscale approach enables the preservation of multiple ERP-related components across different frequencies.

The goal of AWF is to compensate for trial-to-trial variations in latency and phase within the scales that carry ERP features of interest, reducing the negative effect of higher-frequency activity on ERP identification at single-trial level. This allows the algorithm to perform realignment with reduced distortion, particularly when both low-frequency ERP components and high-frequency artifacts (e.g. myogenic activity) are present [18].

The AWF procedure consists of two main steps:

- 1) *Wavelet decomposition.* Wavelet transform is computed for each trial $x^j[n]$ to obtain its multi-resolution time–frequency representation.
- 2) *Scale-wise WF alignment.* The WF algorithm is applied independently to the wavelet coefficients at each selected scale, estimating and correcting latency shifts at the resolution corresponding to that frequency band.

2.1.3. ReSync

ReSync is a template-matching algorithm that, like the WF, operates in the time domain, but introduces two key innovations: (i) it employs a signal decomposition framework to isolate the single-trial, stimulus-locked ERP component using a timing function, and (ii) it restricts realignment to a specific time window, ensuring that only the portion of the signal containing the ERP of interest is modified [33].

For each electrode, the continuous EEG signal is modeled as:

$$x(\tilde{n}) = \sum_{n=1}^N c(n) d(\tilde{n} - n) + \varepsilon, \tilde{n} = 1, \dots, N \cdot K \quad (5)$$

where $x(\tilde{n})$ is the concatenated recording containing K trials, each of length N , $c(n)$ is the stimulus-locked ERP component, $d(\tilde{n})$ is the timing function (1 at stimulus onset, 0 otherwise), ε denotes background noise.

In matrix form, the model becomes:

$$\mathbf{X} = \mathbf{D} \cdot \mathbf{C} + \varepsilon. \quad (6)$$

From this formulation, the ERP component $\mathbf{C}$ can be reconstructed by least-squares estimation:

$$\mathbf{C} = (\mathbf{D}^T \mathbf{D})^{-1} \cdot \mathbf{D}^T \cdot \mathbf{X} \quad (7)$$

where $\mathbf{D}^T$ denotes the transpose of the timing matrix.

ReSync applies latency estimation and realignment only within the time window where the ERP is expected, following these steps:

- 1) *Template computation.* The ERP template is defined as the reconstructed component $\mathbf{C}$.
- 2) *Windowing.* A time interval of interest is selected, and the template is windowed accordingly, yielding $\mathbf{C}_T$.
- 3) *Dominant frequency estimation.* The dominant frequency of $\mathbf{C}_T$ is determined by identifying the peak of its Fourier Transform, reflecting the characteristic spectral content of the ERP.
- 4) *Cross-covariance computation.* For each trial, the cross-covariance function $\Phi_i^{\text{COV}}(m)$ is computed between $\mathbf{C}_T$ and the corresponding windowed trial:

$$\Phi_i^{\text{COV}}(m) = \sum_{n=1}^N \left(\mathbf{C}_T^*(n) - \frac{1}{N} \sum_{n=1}^N \mathbf{C}_T^*(n) \right) \times \left(x_i(n+m) - \frac{1}{N} \sum_{n=1}^N x_i(n) \right) \quad (8)$$

with $m = -N/2, \dots, N/2$, where the reduced lag range avoids estimates dominated by high-variance regions.

- 5) *Frequency-based filtering.* $\Phi_i^{\text{COV}}(m)$ is low-pass filtered using a cutoff equal to the dominant frequency plus 1 Hz, preserving only the spectral content relevant to the ERP.
- 6) *Latency estimation.* The single-trial latency is estimated as the lag that maximizes the filtered cross-covariance. These estimates are used to construct a resynchronized timing function $d^R(\tilde{n})$.
- 7) *Signal reconstruction.* The realigned EEG trace X^R is reconstructed by replacing D with D^R in equation (6):

$$X^R = D^R \cdot C + \varepsilon \quad (9)$$

where D^R is the resynchronized timing matrix obtained from the estimated single-trial latencies.

2.2. Comparing performances of realignment algorithms on surrogate EEG-like data—

The performance of the three algorithms was first assessed using surrogate EEG-like data simulating a typical P300 ERP experiment, in which the P300 component was affected by latency jitter across trials. The simulation study consisted of the following steps:

- 1) *Generation of surrogate data.* Multi-trial EEG-like signals were generated, each containing P300 waveforms with an imposed well-known latency shift (for details see *par.* 2.2.1). Three SNR conditions were simulated (factor SNR: 0.1, 0.5, 1). Two scenarios were considered: SINGLE including a single P300 component; SINGLE + SPURIOUS including a P300 component accompanied by an additional, non-stimulus-locked spurious peak.
- 2) *Algorithm application.* The three realignment algorithms (factor ALGORITHM: WF, AWF, ReSync) were applied offline to the surrogate datasets generated in step 1, under both the SINGLE and SINGLE + SPURIOUS scenarios and across all SNR conditions.
- 3) *Performance evaluation.* Algorithm performance was quantified by evaluating each method's ability to estimate the imposed latency jitter at the single-trial level.

To ensure statistical robustness, the entire simulation procedure was repeated $W = 100$ times for every combination of factors.

2.2.1. Generation of EEG-like signals

EEG-like simulated signals containing P300 waveforms were generated using a linear additive model:

$$x_i^j(t) = s_i^j(t) + \text{SNR} \cdot C_i^j(t), \quad i = 1, \dots, K \text{ and } j = 1, \dots, W \quad (10)$$

where $x_i^j(t)$ denotes the EEG response of the i th trial in the j th iteration, $s_i^j(t)$ represents the EEG background activity, and $C_i^j(t)$ is the ERP component centered at the latency $t_{\text{ERP},i}^j$ (defined below), scaled by the SNR factor.

In the simulation framework, the parameter referred to as SNR is defined as the ratio between the peak amplitude of the ERP component and that of the background activity at the single-trial level. This parameter is used to control the relative prominence of the ERP signal during data generation and does not correspond to a conventional SNR estimate based on signal energy or variance. To facilitate interpretation, the corresponding SNR values were also estimated *a posteriori* on the averaged waveforms using a standard ERP-based definition (ratio between P300 peak amplitude and baseline standard deviation) [34]. The imposed amplitude ratios (0.1, 0.5, 1) correspond approximately to SNR values of ~ 2.8 , ~ 10.4 , and ~ 16.2 in the SINGLE scenario, and ~ 3.0 , ~ 7.0 , and ~ 12.4 in the SINGLE + SPURIOUS scenario, respectively.

The background activity $s_i^j(t)$ was generated using an autoregressive (AR) model driven by white noise (optimal order equal to 20). The AR coefficients were estimated from real EEG data recorded at electrode Cz during resting-state conditions in a healthy subject.

For each iteration, the ERP component $C_i^j(t)$ was modeled as a unit-amplitude Gaussian waveform centered at:

$$t_{\text{ERP},i}^j = t^* + t_i^j, \quad i = 1, \dots, K \text{ and } j = 1, \dots, W \quad (11)$$

where t^* is the expected latency of the ERP (e.g. 300 ms for the P300), and t_i^j is a trial-specific shift. The shifts t_i^j were drawn from a gamma distribution with mean and standard deviation equal to the imposed jitter (50 ms). For each trial, this shift was generated independently, so that the latency of each ERP component varied autonomously across repetitions. The values of t_i^j were randomly drawn from a gamma distribution, ensuring that all shifts were strictly positive [18]. This choice reflects the assumption that ERP latency variability occurs relative to a minimum physiologically plausible latency of the P300 component and therefore manifests as trial-to-trial delays rather than anticipatory responses. The parameters of the gamma distribution were selected such that the resulting distribution of ERP latencies across trials exhibited both a mean and a standard deviation equal to the imposed jitter value (50 ms). This procedure guaranteed that the overall latency jitter—defined as the standard deviation of the single-trial peak latencies—matched the desired jitter level by construction, while maintaining a positively skewed latency distribution consistent with empirical observations of cognitive ERP components.

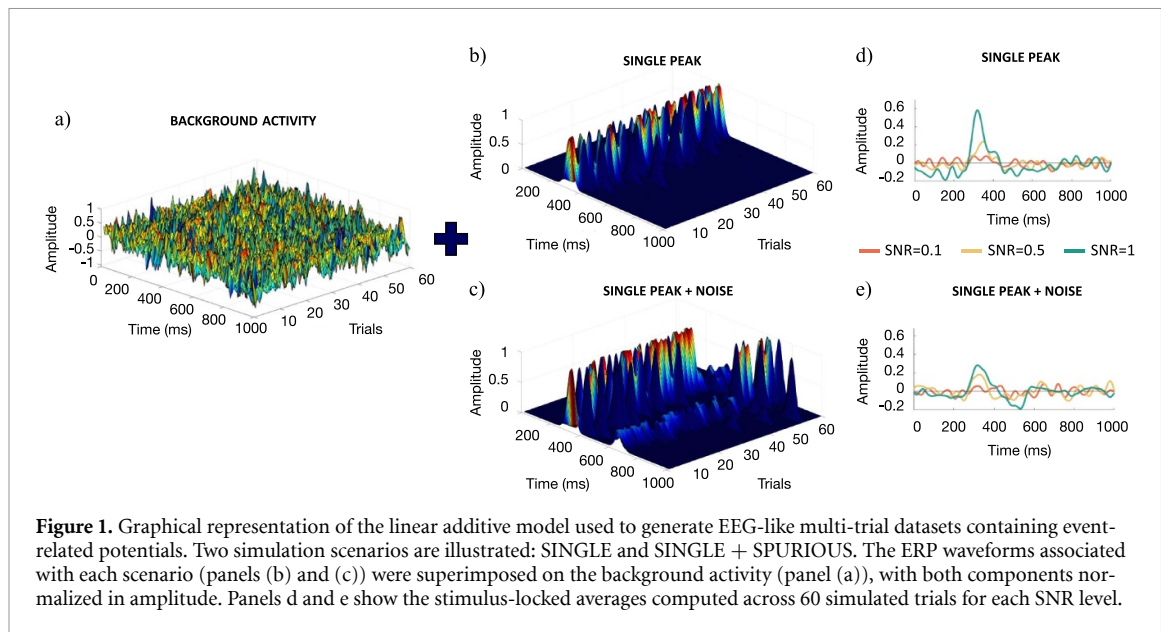


Figure 1. Graphical representation of the linear additive model used to generate EEG-like multi-trial datasets containing event-related potentials. Two simulation scenarios are illustrated: SINGLE and SINGLE + SPURIOUS. The ERP waveforms associated with each scenario (panels (b) and (c)) were superimposed on the background activity (panel (a)), with both components normalized in amplitude. Panels d and e show the stimulus-locked averages computed across 60 simulated trials for each SNR level.

ERP components were then added to the background activity for both simulated conditions. In particular, in the SINGLE scenario only a peak was imposed with t^* equal to 300 ms simulating the presence of a P300 component for all the observations, while in the SINGLE + SPURIOUS scenario two Gaussian peaks were superimposed on the background activity, with t_1^* equal to 300 ms for the P300 peak, and t_2^* equal to 600 ms for the spurious peak. In 80% of the trials, the P300 amplitude was four times larger than that of the spurious peak; in the remaining 20%, the spurious peak amplitude was four times larger. This asymmetric ratio was chosen to avoid generating two peaks that would both appear equally plausible as stimulus-locked responses due to a high repetition probability, while still introducing a realistic subset of trials dominated by non-task-related activity. The SINGLE + SPURIOUS scenario was specifically designed to evaluate the effect of latency estimation outside the expected temporal window of the P300. In particular, the additional peak was positioned outside the plausible latency range of the component of interest in order to model an interfering deflection occurring at implausible latencies. This configuration allows us to explicitly assess the risk of unconstrained realignment algorithms aligning trials to non-stimulus-locked peaks, potentially leading to artificial amplitude enhancement. A graphical representation of the two scenarios is provided in figure 1. For both conditions, each dataset consisted of $K = 60$ trials, generated according to equations (10) and (11). The choice of 60 trials reflects the typical number of target stimuli used in auditory oddball paradigms [35–37].

Finally, to match the pre-processing applied to real EEG data (see section 2.3.4), simulated trials were band-pass filtered using a fourth-order Butterworth filter in the range 1–20 Hz.

The code used to generate the synthetic data and perform the analyses is available at this link.

2.2.2. Application of jitter correction algorithms

EEG-like surrogate data were used to assess the ability of the three realignment algorithms to estimate single-trial ERP latency shifts. The estimated latencies were directly compared with the ground-truth shifts imposed during data generation. Accurate recovery of these shifts is a prerequisite for effective latency jitter correction in template-matching frameworks.

The three algorithms (WF, AWF, ReSync) were applied to the surrogate datasets. For each scenario, the full procedure was repeated 100 times. For every SNR level, algorithm, and iteration, we collected the single-trial latency estimates required to align each trial with the template.

The AWF procedure for this study was therefore implemented as follows:

- 1) Each trial was transformed into the wavelet domain across the entire bandwidth of the filtered signal (1–20 Hz), using a continuous wavelet transform (CWT) with Morlet function as mother wavelet, selected for its effectiveness in capturing low-frequency ERP components such as the P300 [38].
- 2) The analysis was restricted to a narrow frequency range centered around the realignment frequency, defined as the frequency corresponding to the PSD peak, extended ± 0.5 Hz to create a 1 Hz bandwidth. If this estimated frequency did not fall within the characteristic P300 range (2–4 Hz), it was replaced with 3 Hz, which is widely used in P300 analyses [33, 39]. This choice reflects a trade-off between isolating ERP-related oscillations and preserving the broadband nature of the component. The selection of this frequency range

was further supported by previous work from our group, in which different bandwidths were systematically evaluated and the chosen range was found to optimize latency estimation performance [40]. This approach therefore aims to balance spectral specificity and robustness to noise at the single-trial level.

- 3) A template was computed using only the wavelet coefficients within the selected realignment frequency range, allowing the realignment to focus on the dominant frequency component while reducing the influence of unrelated activity.
- 4) WF was applied at the corresponding single scale, yielding a latency shift estimate for each trial.

In the original formulation, AWF relies on a discrete wavelet decomposition to isolate the frequency components associated with the ERP of interest. In the present study, we implemented an adapted version of AWF based on the CWT, in order to obtain a more flexible and fine-grained time–frequency representation. This choice allows for a more precise selection of the frequency range associated with the P300 component, which is known to be predominantly represented in the low-frequency bands. It should be noted that, due to the use of CWT instead of the original discrete wavelet framework, the present implementation should be considered an adaptation of the AWF method rather than a strict replication.

For clarity, this implementation will hereafter be referred to as CWT-AWF throughout the manuscript. The CWT used in the CWT-AWF implementation was computed using the MATLAB Wavelet Toolbox (MathWorks, Natick, MA, USA).

For the ReSync algorithm, the time window used for latency estimation was selected according to the expected P300 latency, which typically falls within 300–400 ms [13, 14]. To accommodate the imposed jitter, a broader window of [200, 500] ms was used during realignment.

To explicitly assess the role of temporal constraints, an additional analysis was performed in which all algorithms were applied within the same temporal window ([200–500] ms), corresponding to the expected latency range of the P300 component. This control condition allowed us to isolate the effect of time-restriction from the intrinsic properties of the different realignment approaches.

2.2.3. Computation of performance parameters

To evaluate the performance of the three alignment algorithms, the mean of the imposed single-trial latencies was first computed across all K observations for each iteration j (denoted as $\bar{t}_j$). This value represents the ground-truth latency of the template for that iteration. The estimated latency for each trial was then compared to this reference value. For each iteration j , an error metric was defined as:

$$Err_j = \frac{1}{K} \sum_{i=1}^K |\tilde{t}_j^i - \bar{t}_j| \quad (12)$$

with $i = 1, \dots, K$ and $j = 1, \dots, W$. $\tilde{t}_j^i$ denotes the latency estimated for the i th trial during iteration j , and K is the total number of simulated trials in each dataset ($K = 60$) and W is the number of repetitions ($W = 100$).

2.2.4. Statistical analysis

To compare the performance of the three algorithms across the two simulated conditions (SINGLE and SINGLE + SPURIOUS), a statistical analysis was performed considering algorithm, SNR, and scenario as factors. Since the distributions of the performance metrics deviated from normality (Shapiro–Wilk test, $\alpha = 0.05$), non-parametric tests were employed.

To assess the effect of SNR, the Kruskal–Wallis test was applied separately for each algorithm and scenario; significant results were followed by Mann–Whitney U post hoc tests to identify pairwise differences among SNR levels. To evaluate the effect of the algorithm, the Friedman test was performed for each SNR level and scenario, with Wilcoxon matched-pairs tests used for post hoc comparisons among algorithms. To assess the influence of the scenario, Mann–Whitney U tests were applied for each algorithm and SNR combination. For all statistical tests, the significance threshold was set to $\alpha = 0.05$ and adjusted using the false discovery rate (FDR) correction to control for multiple comparisons.

All statistical analyses were performed for both implementations of the alignment methods: the original formulation and the time-constrained condition in which all algorithms were restricted to the same temporal window ([200–500] ms). This dual analysis enabled a direct comparison between unconstrained and windowed realignment, allowing the effect of temporal constraints to be evaluated independently of the specific algorithm.

An overview of all statistical analyses performed in the study is reported in table 1.

2.3. Comparing performances of realignment algorithms on real EEG data from healthy subjects

Participants—EEG data were acquired from 10 healthy volunteers (6 males; mean age = 26.5 ± 6.2 years) at Fondazione Santa Lucia (Rome, Italy) to evaluate the performance of the alignment algorithms. All participants reported no history of neurological or psychiatric disorders, no alcohol or drug abuse, and had normal or corrected-to-normal vision.

All were right-handed. The study protocol was approved by the local Ethics Committee (CE PROG. 603/2017) and conducted in accordance with the Declaration of Helsinki. All participants provided a written informed consent before taking part to the study.

Table 1. Overview of statistical analyses performed in the study.

Analysis context	Dependent variable	Factors	Statistical test	Post-hoc test	Purpose
Simulated data (per scenario and SNR)	Error	ALGORITHM (3 levels)	Friedman test	Wilcoxon matched-pairs	Compare performance across algorithms
Simulated data (per scenario and algorithm)	Error	SNR (3 levels)	Kruskal–Wallis test	Mann–Whitney U	Assess effect of noise level
Simulated data (per SNR and algorithm)	Error	SCENARIO (2 levels)	Mann–Whitney U	—	Compare SINGLE vs SINGLE + SPURIOUS
Real data (single shift)	P300 amplitude	ALGORITHM (4 levels)	Friedman test	Wilcoxon matched-pairs	Evaluate amplitude enhancement
Real data (single shift)	Latency shift	ALGORITHM (3 levels)	Friedman test	Wilcoxon matched-pairs	Assess mean latency correction
Real data (single shift)	Latency jitter	ALGORITHM (3 levels)	Friedman test	Wilcoxon matched-pairs	Assess variability reduction
Real data (single shift)	SNR	ALGORITHM (4 levels)	Friedman test	Wilcoxon matched-pairs	Compare SNR levels across algorithms
Real data (median shift, 3/9 channels)	SIR across channels	ALGORITHM (3 levels)	Friedman test	Wilcoxon matched-pairs	Compare dispersion across methods
Real data (median shift, per algorithm)	SIR across channels	NUMBER OF CHANNELS (2 levels)	Mann–Whitney U	—	Evaluate effect of spatial aggregation

2.3.1. Experimental setup

A standard auditory P300 oddball paradigm was implemented using Cubase Essential (Steinberg Media Technologies GmbH, 2008). The stimulus sequence consisted of 420 complex high-frequency tones (440 + 880 + 1760 Hz; standard stimuli) and 60 complex low-frequency tones (247 + 494 + 988 Hz; rare stimuli), pseudo-randomly interspersed.

Stimuli had a duration of 50 ms, with 5 ms rise/fall time, an intensity of 70 dB, and an inter-stimulus interval (ISI) of 850 ms. Participants passively listened to the auditory stream delivered binaurally through sound-isolating earphones (Shure SE-115 CL, sensitivity of 105 dB, impedance of 16 Ω and frequency range of 22 Hz–17.5 kHz) equipped with soft flex sleeves.

2.3.2. EEG recordings

EEG signals were recorded from 26 active electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FCz, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2) arranged according to an extended 10–20 System (GAMMASys; Guger Technology OG, Austria). The reference was placed at the left mastoid and ground at Fpz. Four additional electrodes were used to record eye movements. Signals were amplified (g.USBamp, Guger Technology), band-pass filtered (0.5–70 Hz), digitized at 512 Hz, and stored for offline analysis (Brain Vision Recorder v1.05.0005; Brain Products GmbH, Germany).

2.3.3. Pre-processing of EEG traces

Data were offline referenced to the common average over all electrodes. EEG traces were then band-pass filtered in the range [1–20] Hz (IIR phase shift-free

Butterworth filters with a slope of 48 dB/oct), ocular artifacts were removed by means of ICA, then traces were segmented into [−200; 1000] ms windows according to the stimulus onset separately for *rare* and *standard* stimuli. A baseline correction was applied by using the signal mean value within 200 ms before the stimulus onset, and artifacts were removed by means of an automatic threshold criterion (i.e. signal segments exceeding 80 μ V in absolute value were excluded from the analysis). All pre-processing was performed using Vision Analyzer 1.0 (Brain Products GmbH).

2.3.4. Evaluating the effect of the realignment on ERP characteristics

To evaluate how the three alignment algorithms, influence ERP morphology, we focused on target trials, which are expected to elicit a reliable P300 response in an auditory oddball paradigm [41]. Two electrode montages were considered: a midline set (Fz, Cz, Pz), commonly used in P300 studies [36], and a nine-channel set (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4), providing broader spatial sampling of P300-related activity. For each participant, and each channel in the two sets, single-trial latency shifts were estimated using WF, CWT-AWF, and ReSync. For CWT-AWF, a subject-specific realignment frequency was identified as the median of the PSD peak frequencies across Fz, Cz, and Pz, ensuring alignment within the frequency band most representative of each individual. Realignment was applied only to the wavelet coefficients within this target scale, and the full-band signal (1–20 Hz) was then reconstructed using the inverse wavelet transform.

For ReSync, realignment was restricted to the [200–500] ms interval—consistent with the expected P300 latency range—to avoid affecting portions of the waveform unrelated to the component of interest. This selective operation aims to realign only stimulus-locked activity while preserving the rest of the signal.

The EEG traces were realigned according to three spatial strategies: (i) specific shift, where each electrode received its own latency correction; (ii–iii) median shift, where the same latency correction—computed as the median across either three or nine electrodes—was applied to all channels. To summarize, 9 conditions were obtained, corresponding to the combination of three algorithms (WF, CWT-AWF, ReSync) and three spatial strategies (specific shift, median shift 3ch, median shift 9ch). Grand averages (GAs) were computed to examine the effect of realignment on group-level ERP morphology for the three algorithms and the different spatial strategies.

To quantify the effect of realignment on ERP morphology, four metrics were extracted:

- i) *P300 amplitude*, measured as the maximum value of the across-trial average within the expected P300 time window. This parameter reflects the degree of peak enhancement produced by realignment, capturing the recovery of amplitude otherwise lost due to jitter-induced smearing;
- ii) *P300 latency shift*, computed as the across-trial mean of the single-trial latency estimates for each channel;
- iii) *P300 latency jitter*, defined as the standard deviation of these latency estimates, representing the dispersion of realigned peak latencies across trials.
- iv) *P300 SNR*, defined as the ratio between the P300 peak amplitude and the standard deviation of the pre-stimulus baseline, providing an additional index of signal quality and alignment effectiveness [34].

The combined use of latency shift and jitter allows us to assess whether the increase in amplitude obtained after realignment is driven by the correct alignment of P300 peaks consistent with expected latency ranges, or by the unintended alignment of spurious deflections. Accordingly, optimal realignment performance is expected to be characterized by an increased P300 amplitude, latency shifts centered close to zero, and low latency jitter. Deviations from this pattern, such as large average shifts or excessively high jitter values, indicate that peak enhancement may be driven by the alignment of spurious or non-stimulus-locked activity rather than by improved temporal precision of the genuine ERP component. Importantly, amplitude is therefore not interpreted in isolation, but jointly with latency-based metrics to mitigate the risk of waveform distortion.

In addition, for the median-shift spatial strategy, we calculated the semi-interquartile range (SIR) across channels to evaluate the consistency of P300 latency shift across the scalp.

Finally, GA waveforms were qualitatively compared to assess how specific versus median-based latency corrections influenced waveform morphology and P300 detectability.

2.3.5. Statistical analysis

Before conducting statistical comparisons, we examined the distributions of all extracted parameters—P300 amplitude, P300 latency shift, P300 latency jitter, P300 SNR and across-channel SIR—and verified that none followed a normal distribution (Shapiro–Wilk test, $\alpha = 0.05$). Consequently, all analyses were performed using non-parametric statistical methods.

To assess the influence of the alignment algorithm on P300 amplitude and P300 SNR, we applied the Friedman test with four within-subject levels (non-realigned, WF, CWT-AWF, ReSync), allowing us to evaluate whether any realignment procedure produced a measurable enhancement relative to the unaligned data. A similar Friedman analysis was performed on the P300 latency shift and P300 latency jitter parameters; in this case, the factor ALGORITHM included only the three realignment methods (WF, CWT-AWF, ReSync), and tests were carried out separately for Fz, Cz, and Pz to capture potential topographical differences.

We also examined how consistently each method produced latency estimates across channels by analyzing the across-channel SIR values. Here, the factor ALGORITHM (WF, CWT-AWF, ReSync) was treated as a within-subject variable, while the factor NUMBER OF CHANNELS (3 vs 9 channels) served as a between-condition comparison. The Friedman test was used to evaluate algorithm-driven differences, whereas comparisons across channel sets were conducted with the Mann–Whitney U test. When significant effects were detected, Wilcoxon matched-pair tests were used for post hoc pairwise comparisons. To reduce the risk of false positives associated with multiple testing, we set a significance level $\alpha = 0.05$ corrected using the FDR procedure.

3. Results

3.1. Testing of alignment algorithms on EEG-like surrogate data

Effect of the algorithm and SNR on performance—tables 2 and 3 summarize the results of the Friedman and Kruskal–Wallis tests, which were conducted to evaluate the influence of the algorithm and SNR factors on latency-estimation error, separately for the SINGLE and SINGLE + SPURIOUS scenarios. Both factors showed a significant effect on the error metric

Table 2. Results of the Friedman test computed on the latency-estimation error in the SINGLE and SINGLE + SPURIOUS scenarios, considering ALGORITHM (WF, CWT-AWF, ReSync) as within-factor. The analysis was repeated for each SNR level (0.1, 0.5, 1) and for both the original (unconstrained) formulation and the time-constrained implementation in which all algorithms were restricted to the same temporal window ([200–500] ms). See figures 2(a) and (b) for the corresponding boxplot representations.

Factor: ALGORITHM (dof = 2)							
ALGORITHM VERSION	SCENARIO	SNR = 0.1		SNR = 0.5		SNR = 1	
		Chi Sqr.	<i>p</i>	Chi Sqr.	<i>p</i>	Chi Sqr.	<i>p</i>
<i>Original</i>	<i>SINGLE</i>	178.88	<0.0001	169.22	<0.0001	151.62	<0.0001
	<i>SINGLE + SPURIOUS</i>	178.89	<0.0001	185.28	<0.0001	157.22	<0.0001
<i>Time-constrained</i>	<i>SINGLE</i>	177.38	<0.0001	150.72	<0.0001	150.72	<0.0001
	<i>SINGLE + SPURIOUS</i>	182.00	<0.0001	151.28	<0.0001	150.50	<0.0001

Table 3. Results of the Kruskal–Wallis test computed on the latency-estimation error in the SINGLE and SINGLE + SPURIOUS scenarios, considering SNR (3 levels: 0.1, 0.5, 1) as between-factor. The analysis was performed separately for each algorithm (WF, CWT-AWF, ReSync). See figures 2(a) and (b) for the corresponding boxplot representation.

Factor: SNR (dof = 2)							
ALGORITHM VERSION	SCENARIO	WF		CWT-AWF		ReSync	
		H	<i>p</i>	H	<i>p</i>	H	<i>p</i>
<i>Original</i>	<i>SINGLE</i>	206.22	<0.0001	156.52	<0.0001	113.36	<0.0001
	<i>SINGLE + SPURIOUS</i>	182.75	<0.0001	135.78	<0.0001	90.01	<0.0001
<i>Time-constrained</i>	<i>SINGLE</i>	190.079	<0.0001	91.12	<0.0001	113.36	<0.0001
	<i>SINGLE + SPURIOUS</i>	89.27	<0.0001	5.79	0.0552	90.0052	<0.0001

across all conditions. Focusing first on the effect of the algorithm under the original formulation (table 2; figure 2(a)), clear differences emerged across methods. At low and medium SNR values (0.1 and 0.5), CWT-AWF outperformed WF, yielding significantly smaller errors.

At the highest SNR level (1.0), however, the performance of the two methods converged, and their error values did not differ significantly. Across all SNR levels, ReSync consistently achieved the lowest error, outperforming both WF and CWT-AWF regardless of noise conditions. This highlights the robustness of ReSync's decomposition-based framework, even when the input signal is heavily contaminated by noise.

When all algorithms were constrained within the same temporal window ([200–500] ms) (figure 2(b)), a general reduction in latency-estimation error was observed for both WF and CWT-AWF, particularly at low and medium SNR levels. This indicates that part of the performance differences observed in the unconstrained case is attributable to the absence of explicit temporal restrictions. Despite this improvement, ReSync maintained superior performance across all SNR conditions, confirming that its advantage is not solely due to the use of temporal constraints, but also to its ability to isolate stimulus-locked activity through decomposition.

The effect of SNR itself was also highly significant for all three algorithms (table 3). As expected, error values decreased as SNR increased, reflecting

improved detectability of the ERP peak under less noisy conditions, for all the three algorithms.

Effect of the scenario on the performance of realignment algorithms—The results of the statistical analysis conducted to investigate the influence of the scenario on the algorithms' performance for each method and SNR level are summarized in table 4 and in figure 3. Focusing first on the original formulation (figure 3(a)), the scenario had a clear and consistent impact on the WF algorithm. For all SNR values, WF performed significantly worse in the SINGLE + SPURIOUS condition compared to the SINGLE condition, reflecting its sensitivity to the presence of spurious peaks. In contrast, CWT-AWF was less affected by the additional peak. A significant scenario effect emerged only at the highest SNR level (SNR = 1), suggesting that CWT-AWF can distinguish the true ERP component reasonably well when the underlying signal is strong. ReSync showed the greatest robustness to scenario differences. At the lowest SNR level (0.1), the scenario had no significant effect on its performance. A significant difference emerged only at SNR = 0.5 and 1, indicating modest sensitivity to the additional peak at higher SNRs. Even in these cases, however, ReSync maintained considerably lower error values relative to the other algorithms.

When all algorithms were constrained within the same temporal window ([200–500] ms) (figure 3(b)), the impact of the scenario was markedly reduced for both WF and CWT-AWF. In particular, the performance gap between the SINGLE and

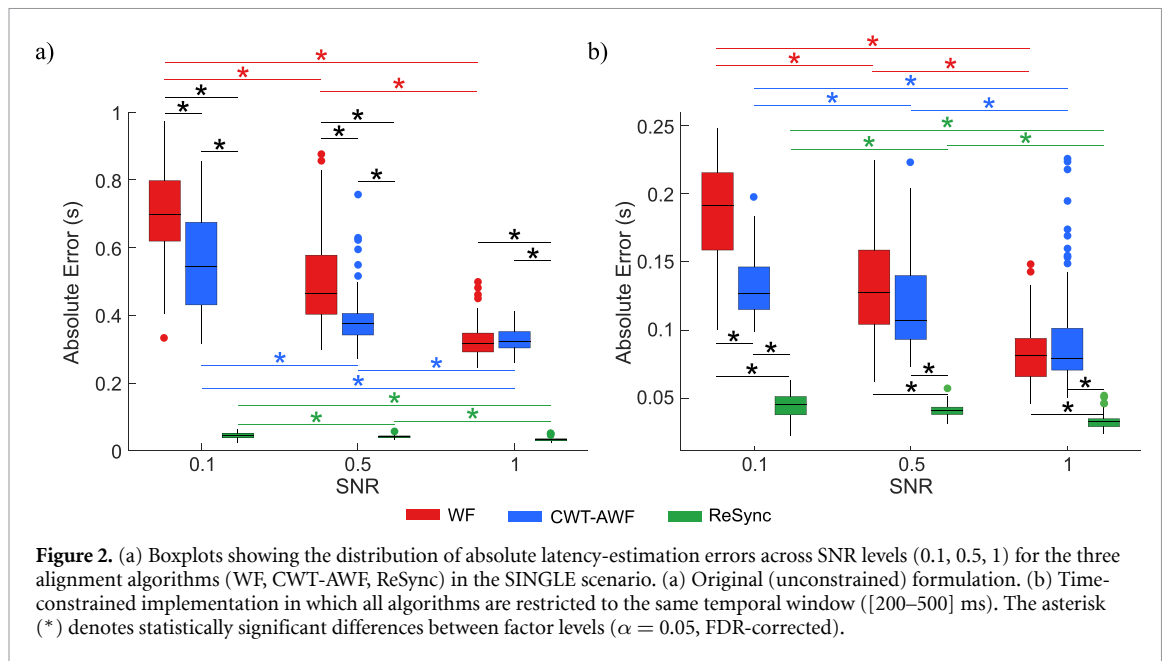


Table 4. Results of the Mann–Whitney U test computed on the latency-estimation error, considering SCENARIO (2 levels: SINGLE, SINGLE + SPURIOUS) as between-factor. The analysis was performed separately for each algorithm (WF, CWT-AWF, ReSync) and for each SNR level (0.1, 0.5, 1), under both the original (unconstrained) formulation and the time-constrained implementation ([200–500] ms). See figures 3(a)–(b) for the corresponding boxplot representations.

		Factor: scenario (dof = 1)					
ALGORITHM VERSION	SCENARIO	WF		CWT-AWF		ReSync	
		Z adj.	p	Z adj.	p	Z adj.	p
Original	SNR = 0.1	2.38	0.017	0.22	0.828	0.95	0.343
	SNR = 0.5	3.93	<0.0001	1.10	0.270	4.20	<0.0001
	SNR = 1	5.82	<0.0001	4.06	<0.0001	4.94	<0.0001
Time-constrained	SNR = 0.1	−0.086	0.932	−0.056	0.955	−0.95	0.343
	SNR = 0.5	−4.85	<0.0001	−4.49	<0.0001	−4.20	<0.0001
	SNR = 1	−9.42	<0.0001	−8.83	<0.0001	−4.94	<0.0001

SINGLE + SPURIOUS conditions decreased across all SNR levels, especially at lower SNR values, indicating that temporal constraints limit the influence of spurious peaks on latency estimation. Despite this improvement, residual differences between scenarios remained visible, particularly at higher SNR levels. ReSync continued to exhibit the highest robustness across conditions, with minimal performance degradation and consistently lower error values compared to the other methods.

3.2. Testing of alignment algorithms on real EEG data from healthy subjects

3.2.1. Effect of realignment on ERP characteristics

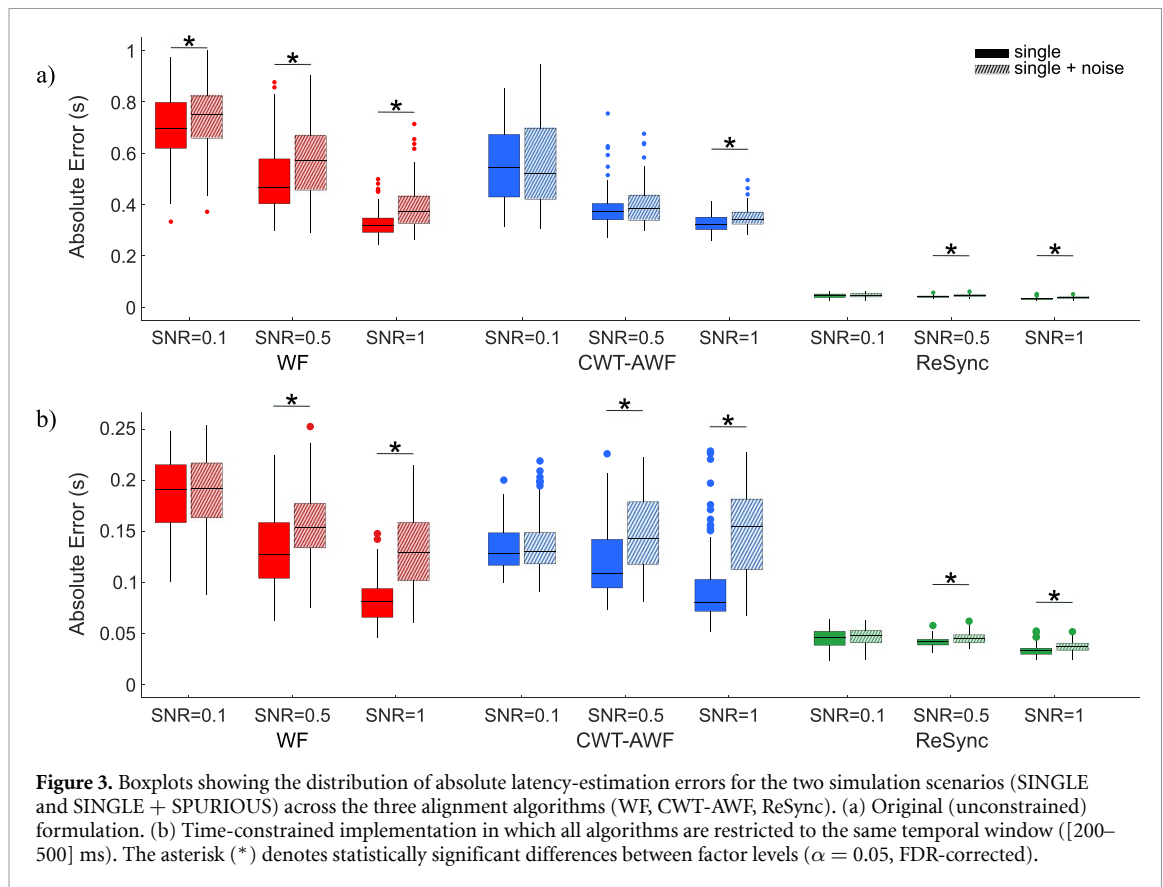
Figure 4(a) compares the grand-average (GA) waveforms obtained at Pz for the three alignment algorithms and for the raw (non-aligned) data. Because the dataset was collected from healthy participants, the P300 component is already clearly visible even without realignment, as the natural jitter in this population is relatively small. Nonetheless, applying the

alignment algorithms leads to a noticeable enhancement of the P300 morphology.

All three methods increase the peak amplitude relative to the non-aligned case, with WF producing the largest amplification—although often accompanied by broader oscillatory patterns—while CWT-AWF and ReSync induce more moderate yet cleaner enhancements.

The statistical evaluation of P300 amplitude, mean latency shift, and latency jitter across the three midline electrodes (Fz, Cz, Pz) is summarized in table 5 and illustrated in figures 4(b)–(d). The alignment method significantly influenced all three parameters at all scalp locations, with the exception of the latency-shift metric at Pz, where the effect was marginal ($p = 0.045$).

As shown in figure 4(b), the raw data consistently displayed the lowest P300 amplitudes (median values of approximately 4–6.5 μV). All three alignment algorithms significantly increased these values. WF produced the largest increases (≈ 10 –13 μV), followed by ReSync (≈ 7 –10 μV), whereas CWT-AWF yielded



the smallest amplification ($\approx 5.5\text{--}7.5 \mu\text{V}$). These patterns closely match the qualitative observations from figure 3(a): while all algorithms enhance the waveform, WF tends to strongly amplify the peak, whereas ReSync and CWT-AWF produce a lower P300 enhancement. Latency-shift distributions (figure 4(c)) further highlight the differences among methods. CWT-AWF systematically produced values centered around ~ 200 ms, while WF and ReSync yielded distributions centered closer to 0 ms. An exception emerged at Cz for WF, where the estimated shifts were centered around ~ 150 ms.

Dispersion differed markedly across algorithms: WF exhibited the widest spread, CWT-AWF showed intermediate variability and ReSync demonstrated the tightest and most stable estimates.

These trends were confirmed by statistical comparisons. CWT-AWF differed significantly from both WF and ReSync at Fz and Cz, while WF and ReSync did not differ significantly from each other. Latency jitter (figure 4(d)) revealed the most striking contrasts. WF exhibited extremely large jitter values (centered around 700 ms), followed by CWT-AWF (centered around 400 ms). ReSync, by contrast, produced smaller jitter values, centered near 50 ms, with very narrow distributions. ERP SNR results (figure 4(e)) show that all methods tend to improve signal quality compared to the non-aligned condition, although with different levels of variability. ReSync provides more stable SNR values across channels. However,

SNR should be interpreted alongside latency-based metrics, as increases in SNR alone may also reflect alignment of spurious activity.

A detailed quantitative summary (median and interquartile range) of the results in panels b–e of figure 4 is provided in the Supplementary Materials (table S1).

3.2.2. Effect of number of channels on realignment algorithms performance

To further investigate how spatial information influences the quality of latency correction, we examined the effect of the number of electrodes used to compute the median shift.

As shown in figure 5(a), the dispersion of the latency shifts differed significantly depending on the alignment algorithm for both the 3-channel configuration ($\chi^2 = 12.6$, $p = 0.0018$) and the 9-channel configuration ($\chi^2 = 15.2$, $p = 0.0005$). Post hoc comparisons revealed that WF and CWT-AWF produced substantially larger cross-channel variability, whereas ReSync consistently yielded SIR values close to zero—indicating highly stable latency estimates across electrodes. Interestingly, the number of electrodes used to compute the median shift (3 vs 9 channels) did not significantly affect the dispersion for any of the three algorithms (WF: $p = 1.00$; CWT-AWF: $p = 0.36$; ReSync: $p = 0.76$).

This suggests that the benefits of median-based realignment are not strongly dependent on the size of

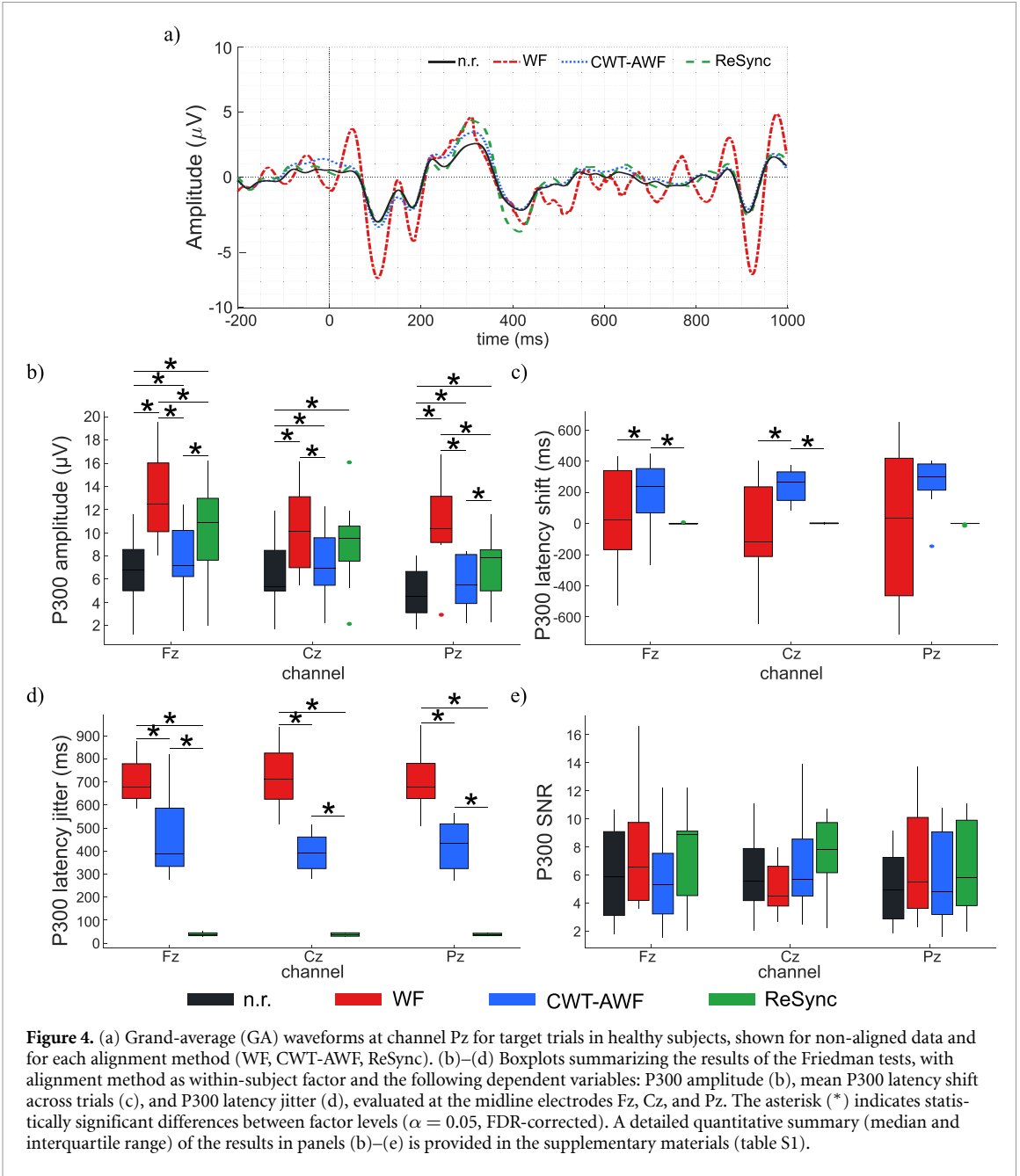
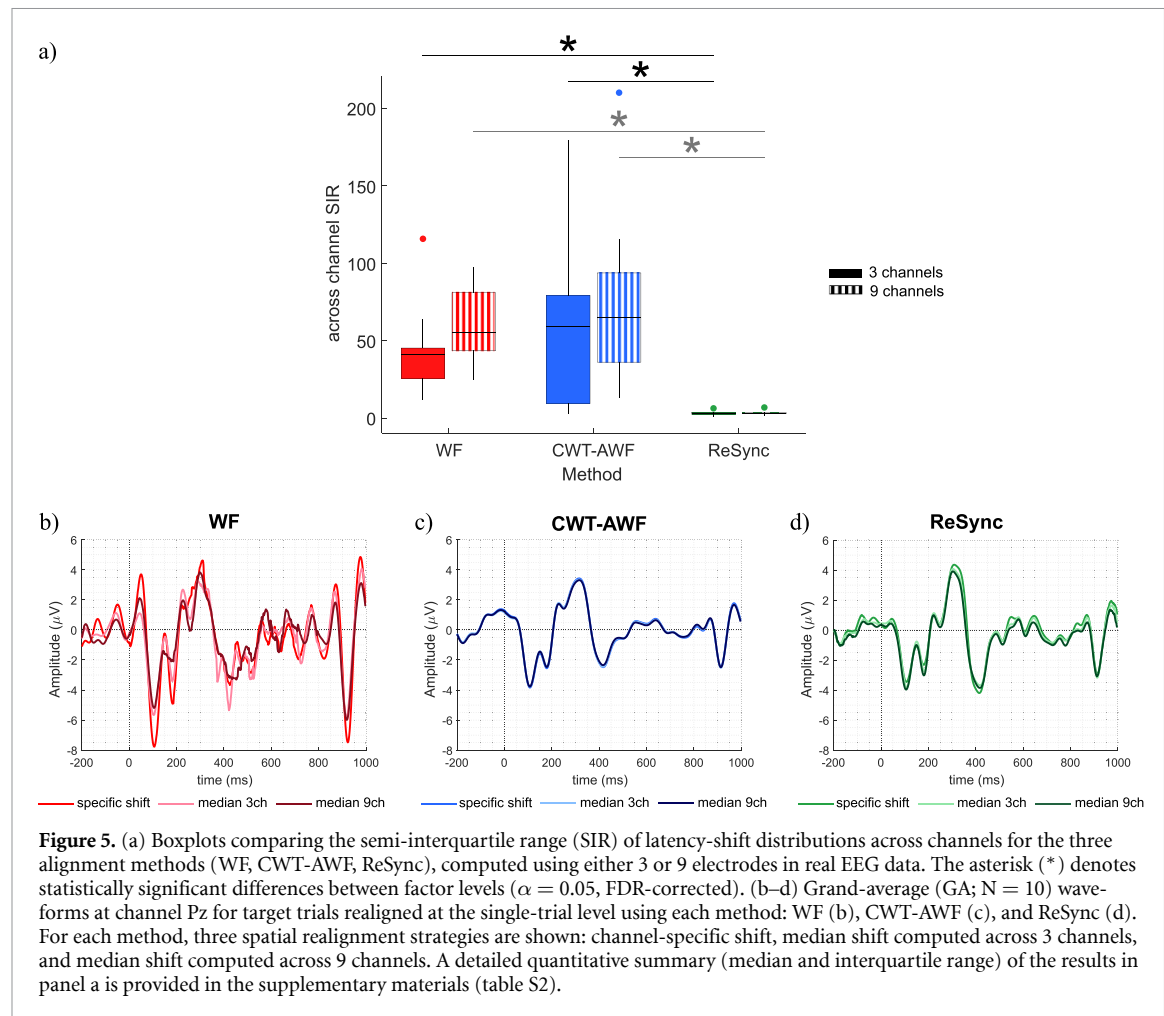


Table 5. Results of the Friedman tests computed on three ERP parameters: P300 amplitude, P300 latency shift, and P300 latency jitter, P300 SNR. For P300 amplitude and P300 SNR, the within-subject factor ALGORITHM included four levels (raw, WF, CWT-AWF, ReSync). For latency shift and latency jitter, the factor included three levels (WF, CWT-AWF, ReSync). Analyses were performed separately for each scalp location (Fz, Cz, Pz). See figures 4(b)–(d) for the corresponding boxplot representations.

	Fz		Cz		Pz	
	Chi Sqr.	p	Chi Sqr.	p	Chi Sqr.	p
P300 AMPLITUDE (<i>dof</i> = 3)	28.08	<0.0001	19.32	0.00023	25.56	0.000010
P300 LATENCY SHIFT (<i>dof</i> = 2)	7.40	0.025	9.60	0.0082	6.20	0.045
P300 LATENCY JITTER (<i>dof</i> = 2)	18.20	0.00011	20.00	0.00005	20.00	0.00005
P300 SNR (<i>dof</i> = 3)	8.76	0.0327	8.28	0.0406	2.76	0.430

the electrode set and that the intrinsic characteristics of each algorithm play a much larger role in determining cross-channel consistency.

Figures 5(b)–(d) illustrate how these statistical differences translate into P300 morphology at Pz. In all three alignment methods, the P300 component remains clearly identifiable under both the specific shift and median shift conditions. The WF realignments continue to show broader oscillatory behavior and larger peak amplitudes, consistent with previous observations. The CWT-AWF waveforms, by contrast, are nearly identical across the



specific-shift, 3-channel median, and 9-channel median conditions, indicating that CWT-AWF is largely insensitive to the choice of spatial aggregation.

For WF and ReSync, however, applying a specific shift to each channel results in a more pronounced P300 peak than the median-shift cases. This suggests that per-channel realignment preserves finer temporal details that may be attenuated when a common median shift is applied across electrodes. Nevertheless, even under the median-shift condition, ReSync continues to produce clean and temporally sharp P300 waveforms, reinforcing its robustness across spatial configurations.

A detailed quantitative summary (median and interquartile range) of the results in panels a of figure 5 is provided in the Supplementary Materials (table S2).

4. Discussion

This study compared three template-matching algorithms -WF, CWT-AWF, and ReSync—designed to mitigate ERP latency jitter and improve identification of the P300 component. While these methods share the use of an ERP template, they differ considerably

in their operating domains: WF in the time domain, CWT-AWF in the time–frequency domain, and ReSync through a decomposition-based reconstruction within a restricted time window. Beyond a pure performance comparison, the aim of this work was to systematically benchmark these approaches under controlled conditions in order to isolate key factors influencing realignment performance. In particular, the analysis focused on the role of temporal constraints on latency estimation and on the impact of channel-specific versus spatially aggregated correction strategies—two aspects that have been only marginally addressed in previous studies. This perspective allows the results to be interpreted not only in terms of relative performance, but also in terms of the mechanisms underlying successful or misleading realignment.

4.1. Comparison of performances on simulated data

Our findings based on simulated data confirm that latency jitter significantly impacts ERP morphology and detectability for late cognitive components such as the P300.

CWT-AWF showed better resilience due to its reliance on multiscale wavelet decomposition, which

is known to capture event-related oscillatory patterns more effectively than traditional time-domain analysis. In fact, wavelet-based approaches naturally align with the oscillatory nature of ERPs [29], particularly within the delta–theta frequency range characteristic of the P300 [30]. In the present study, CWT-AWF was further optimized by selecting a dedicated frequency range tailored to surrogate and real datasets [42]. This targeted approach enhanced the algorithm's ability to capture the ERP-specific frequency content while suppressing unrelated neural activity [18]. Despite these improvements, CWT-AWF still did not match the consistency and accuracy achieved by ReSync.

The results obtained in the SINGLE + SPURIOUS condition further emphasize the importance of constraining latency estimation within the expected temporal window of the ERP component. When this constraint is not enforced, alignment algorithms may preferentially lock onto peaks occurring at implausible latencies, leading to artificial amplitude enhancement that does not reflect genuine neural responses. In this sense, the simulation framework was specifically designed to make this effect explicit and to highlight the role of time-restricted realignment strategies.

The additional analysis performed with a shared temporal constraint further clarifies the role of windowing in the presence of spurious components. Constraining the search to the expected P300 latency range reduces the impact of the non-stimulus-locked peak, particularly for methods such as WF and CWT-AWF that do not inherently restrict the temporal domain of realignment. This results in improved performance and reduced sensitivity to the SINGLE + SPURIOUS condition, especially under low SNR. However, the persistence of residual differences between scenarios, even under identical temporal constraints, indicates that windowing alone is not sufficient to fully mitigate the influence of spurious activity. In this context, the superior robustness of ReSync highlights the importance of combining temporal constraints with decomposition-based strategies, which more effectively isolate stimulus-locked components from background and spurious signals.

While the simulation results provide a clear and controlled comparison of the different methods, it should be noted that the adopted model does not fully capture the complexity of real EEG signals, including variability in waveform morphology, non-stationarity, and the presence of overlapping components. As a result, the observed differences between methods may be accentuated compared to real data. Therefore, the simulation results should be interpreted as a benchmark under controlled conditions, rather than as a direct reflection of real-world performance. This consideration further motivates the validation of the methods on real EEG recordings, as discussed in the following section.

4.2. Comparison of performances on real EEG data

Analyses on real EEG recordings from healthy participants aligned with the simulation findings. All three algorithms enhanced the P300 amplitude relative to the non-aligned data, reflecting the benefit of jitter correction (figure 4(a)). However, the quality of realignment varied considerably. WF produced the largest amplitude increases but introduced greater variability and visible distortions, likely due to the over-alignment of spurious or non-time-locked features. CWT-AWF produced more moderate enhancements, benefiting from better noise suppression, but still exhibited broader latency distributions. In contrast, ReSync consistently offered a balanced improvement, enhancing P300 amplitude while maintaining temporal and morphological integrity. The method produced the smallest latency jitter values—around 50 ms—and maintained narrow distributions. These values are not only substantially reduced compared to those produced by WF and CWT-AWF but also fall within range typically reported for P300 responses in healthy subjects [22, 23, 43] highlighting ReSync's ability to preserve genuine neural variability rather than amplifying algorithmic artifacts.

At first glance, the observation that WF yields the largest amplitude enhancement despite exhibiting the highest latency variability may appear counterintuitive. This apparent discrepancy highlights that increased P300 amplitude does not necessarily reflect improved alignment. In particular, methods such as WF may produce sharper average waveforms by aligning trials to prominent but inconsistently timed peaks, including spurious or non-stimulus-locked deflections. This results in inflated amplitude despite large latency variability.

The distribution of latency shifts provides further insight into the different realignment behaviors of the three algorithms when interpreted in conjunction with the evaluation metrics defined in section 2.3.4. Importantly, increases in P300 amplitude should not be interpreted in isolation, as they may also arise from the alignment of spurious or non-stimulus-locked activity. In this study, amplitude enhancement was therefore evaluated jointly with latency shift and latency jitter. Discrepancies between these metrics—such as large amplitude increases accompanied by broad or systematically shifted latency distributions—were interpreted as indicative of potential waveform distortion, rather than genuine improvement in ERP alignment.

WF produced latency shifts spanning a wide range, approximately from -200 to $+300$ ms, indicating that trials were realigned both by advancing and delaying the waveform. This bidirectional and highly dispersed behavior reflects the absence of explicit constraints on peak selection: WF does not impose restrictions on frequency content, nor does it limit realignment to a specific temporal

window. Consequently, latency estimates show large dispersion and are often far from the expected P300 range, a pattern consistent with the high jitter values observed. Within the proposed metric framework, this combination—large amplitude increases accompanied by wide latency dispersion—suggests that peak enhancement may be driven, at least in part, by alignment to spurious or non-stimulus-locked deflections rather than by improved temporal precision.

CWT-AWF exhibited a more constrained behavior, producing exclusively positive latency shifts ranging approximately from 50 to 350 ms. The time-frequency decomposition improves the identification of physiologically relevant ERP-related activity, reducing the likelihood of aligning to irrelevant deflections. However, the phase-based realignment performed in the wavelet domain introduces a systematic delay in the reconstructed signal, resulting in a latency-shift distribution centered around ~ 200 ms. Although this approach reduces bidirectional dispersion compared to WF, the resulting shifts remain substantially displaced from zero and are associated with moderate-to-high jitter. According to the metric framework, this pattern reflects partial improvement in peak selection but limited control over the absolute timing of the realigned component.

ReSync displayed a qualitatively different and more favorable profile. By combining signal decomposition with a time-restricted realignment strategy, ReSync explicitly constrained latency estimation to the expected P300 temporal window and targeted stimulus-locked activity. This resulted in latency shifts tightly clustered around zero and markedly reduced jitter. When considered alongside the observed amplitude enhancement, this combination of metrics indicates genuine recovery of the P300 component, rather than noise-driven peak inflation [33]. Consistently, the ERP SNR analysis provides complementary evidence, indicating that ReSync achieves higher and more stable signal quality across channels when interpreted jointly with its reduced latency jitter and near-zero latency shifts. Importantly, the low jitter values obtained with ReSync fall within the range typically reported for P300 responses in healthy subjects, supporting both accurate reconstruction and reliable detection of the ERP while remaining consistent with expected physiological variability [22, 23].

4.3. Evaluation of the median shift across multiple channels

The study also explored the impact of applying latency correction at the channel level versus using median-based corrections across multiple channels [27]. To our knowledge, this is the first time that the influence of channel-level versus spatially aggregated latency correction has been systematically examined in the context of ERP realignment algorithms. This

analysis revealed that ReSync maintained stable performance regardless of whether specific or median corrections were applied, whereas WF showed clear benefits from median-shift strategies—likely due to cross-electrode noise averaging. CWT-AWF remained largely unaffected by this factor (figures 4(b)–(d)). Overall, these findings suggest that spatial aggregation may be particularly useful for algorithms that are more susceptible to noise and cross-channel variability, while robust methods such as ReSync preserve performance even when channel-specific variability is minimized. The comparison between channel-specific and spatially aggregated latency correction strategies has not been systematically investigated in the context of ERP realignment. For this reason, the present study focuses on simple and well-defined approaches to establish a clear baseline for comparison. Future work should extend this framework to more advanced strategies, such as weighted combinations of channels based on signal quality or spatial proximity, to further exploit the spatial structure of EEG signals.

4.4. Practical recommendations for method selection

Based on the present findings, some practical considerations can be drawn to guide the selection of ERP realignment methods under different experimental conditions. When the primary objective is to improve ERP detectability while preserving reliable timing estimates—particularly in clinical or applied contexts where false positives and distorted latency estimates are critical—methods that explicitly constrain realignment in both time and signal structure are preferable. In this regard, ReSync appears well suited for scenarios characterized by low SNR, substantial latency variability, or the presence of spurious deflections, as it consistently combines amplitude enhancement with near-zero latency shifts and jitter values consistent with previous findings.

Time-domain approaches such as WF may be adequate in high-SNR conditions with minimal interference, where latency variability is limited and ERP morphology is stable. However, caution is warranted when interpreting large amplitude increases obtained with WF, as these may be accompanied by wide latency dispersion and reflect noise-driven alignment rather than genuine improvement in temporal precision. The use of spatially aggregated (median-based) corrections can partially mitigate these effects and may be beneficial when WF is applied to multichannel recordings.

Time-frequency methods such as CWT-AWF offer a compromise between simplicity and robustness by leveraging spectral constraints to improve peak selection. They may be advantageous when the frequency characteristics of the ERP component are well defined and when moderate improvements in alignment are sufficient. Nevertheless, the systematic

latency delays observed with CWT-AWF suggest that this approach may be less suitable when absolute timing information is critical.

More generally, the present results highlight the importance of jointly evaluating amplitude enhancement together with latency shift and latency jitter when assessing realignment performance. Increases in ERP amplitude should be interpreted cautiously if accompanied by large shifts or high jitter, as these patterns may indicate alignment to non-stimulus-locked activity. Conversely, balanced improvements across these metrics provide stronger evidence of genuine recovery of the underlying neural response.

4.5. Limitations and impact

Despite the strengths of the present work, several limitations should be acknowledged. First, the comparison focused exclusively on offline template-matching approaches—WF, CWT-AWF, and ReSync—which, while representative of distinct methodological families, do not encompass the full spectrum of modern latency-correction techniques. Alternative strategies, including machine learning-based approaches or unsupervised decomposition methods such as ICA or RIDE, may offer additional insights and should be explored in future studies. Second, all three methods rely on the presence of a predefined ERP template, which assumes consistency across trials. In paradigms involving learning, fatigue, or otherwise non-stationary responses, this assumption may not hold. In such cases, template-based methods could misalign components or miss relevant dynamic features.

Another limitation concerns the use of an autoregressive model to generate background activity in the simulated data. Although this approach allows for controlled manipulation of signal properties and preserves realistic spectral characteristics, it does not fully capture the non-stationary and spatially complex nature of real EEG noise. Future work should investigate the use of real resting-state EEG as background activity to further improve the ecological validity of the simulations, while carefully balancing this with the need to retain ground-truth control over latency variability.

Moreover, the ReSync algorithm's reliance on a user-defined realignment window introduces an element of task- and subject-specific tuning. Although this improves specificity, it may require adaptation across different paradigms or participant groups, which could limit generalizability.

A further methodological aspect concerns the selection of the frequency band in the CWT-AWF approach. In the present study, a relatively narrow band centered around the dominant spectral component was used to enhance the identification of ERP-related activity at the single-trial level. This choice is supported by previous work from our group, in which different frequency ranges were systematically

evaluated and shown to affect latency estimation performance [40]. While this strategy improves robustness to noise, it may also limit the ability of the method to capture the full broadband nature of the P300 component. This reflects an inherent trade-off between spectral specificity and completeness of the reconstructed signal. Future studies should therefore systematically investigate the sensitivity of the method to bandwidth selection, as well as to other wavelet parameters, in order to further improve robustness and generalizability.

An important limitation of the present study is that validation on real EEG data was performed on a relatively small and homogeneous group of healthy young participants. Although the within-subject design and the consistency of the observed effects support the robustness of the findings, the limited sample size may constrain their generalizability. In clinical populations, ERP components such as the P300 are often characterized by delayed latencies, increased jitter, and altered morphology. These factors may affect the performance of alignment algorithms and could require more flexible or adaptive temporal constraints. Future studies should therefore evaluate the proposed approaches in larger cohorts and publicly available benchmark datasets, as well as in clinical populations, where accurate single-trial detection may be even more critical.

These findings also have potential implications for BCI applications, where latency variability can significantly reduce the detectability of target ERPs at the single-trial level. In this context, realignment methods such as ReSync could be integrated as a preprocessing step to enhance temporal consistency and improve SNR prior to classification. In practical implementations, latency estimation could be derived from calibration data and applied adaptively to incoming trials. Notably, the use of time-restricted realignment may help prevent alignment to non-target or spurious responses, which is particularly relevant in asynchronous or noisy BCI settings. However, further work is needed to evaluate these approaches in fully online and unlabeled scenarios.

5. Conclusion

Overall, the results demonstrated that ReSync—by integrating decomposition-based processing with time-restricted realignment—consistently outperforms traditional time-domain and time-frequency approaches in terms of latency estimation accuracy, robustness to noise, and preservation of ERP morphology. While WF and CWT-AWF provide improvements compared to conventional averaging, they remain more vulnerable to misalignment and the influence of spurious peaks, particularly under low SNR conditions. In contrast, ReSync's stable performance across all simulated and real-data

scenarios underscores the advantages of combining spatial and temporal precision within a targeted alignment framework.

Importantly, this study provides the first systematic comparison between decomposition-based and classical template-matching approaches for ERP jitter correction, as well as the first evaluation of channel-level versus spatially aggregated latency correction strategies. These novel analyses highlight that advanced jitter correction methods such as ReSync not only enhance P300 detectability but also yield plausible single-trial latency estimates in healthy subjects—something not achieved by traditional methods. Importantly, the additional analyses demonstrate that temporal constraints alone are not sufficient to ensure accurate realignment. While windowing improves the performance of classical methods, ReSync consistently outperforms them by combining time-restricted alignment with decomposition-based signal reconstruction, resulting in greater robustness to noise and spurious activity.

Taken together, these findings support the adoption of next-generation jitter correction techniques in ERP research and applied domains—including BCIs and clinical diagnostics—where precise single-trial timing is essential. Future work should investigate how these methods can be integrated into real-time pipelines and extended to more diverse experimental settings and clinical populations.

Acknowledgment

This work was partially supported by Sapienza University of Rome (RM123188F229EC72; AR1241906ECF784C; AR1241906ECF8D8E) and the Italian Ministry of Health (PNRR-POC-2023-12377627).

Data availability statement

The data cannot be made publicly available upon publication because they are not available in a format that is sufficiently accessible or reusable by other researchers. The data that support the findings of this study are available upon reasonable request from the authors.

Supplementary tables S1–S2 available at <https://doi.org/10.1088/1741-2552/ae7766/data1>.

ORCID iDs

Emma Colamarino  0000-0002-5375-8339
 Floriana Pichiorri  0000-0003-1225-7025
 Angela Riccio  0000-0002-4098-9461
 Jlenia Toppi  0000-0002-8279-1699

References

- [1] Sutton S, Braren M, Zubin J and John E 1965 Evoked-potential correlates of stimulus uncertainty *Science* **150** 1187–8
- [2] Walter W G, Cooper R, Aldridge V J, McCallum W C and Winter A L 1964 Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain *Nature* **203** 380–4
- [3] Comanducci A et al 2020 Clinical and advanced neurophysiology in the prognostic and diagnostic evaluation of disorders of consciousness: review of an IFCN-endorsed expert group *J. Clin. Neurophysiol.* **131** 2736–65
- [4] Deco G, Jirsa V K, Robinson P A, Breakspear M and Friston K 2008 The dynamic brain: from spiking neurons to neural masses and cortical fields *PLoS Comput. Biol.* **4** e1000092
- [5] Beim Graben P, Gerth S and Vasishth S 2008 Towards dynamical system models of language-related brain potentials *Cogn. Neurodyn.* **2** 229–55
- [6] Kiebel S J, David O and Friston K J 2006 Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization *NeuroImage* **30** 1273–84
- [7] Kiebel S J, Garrido M I, Moran R J and Friston K J 2008 Dynamic causal modelling for EEG and MEG *Cogn. Neurodyn.* **2** 121–36
- [8] Rezeika A, Benda M, Stawicki P, Gemblar F, Saboor A and Volosyak I 2018 Brain–computer interface spellers: a review *Brain Sci.* **8** 57
- [9] Riccio A, Pichiorri F, Schettini F, Toppi J, Riseti M, Formisano R, Molinari M, Astolfi L, Cincotti F and Mattia D 2016 Interfacing brain with computer to improve communication and rehabilitation after brain damage *Prog. Brain Res.* **228** 357–87
- [10] Wang Y, Nakanishi M and Zhang D 2019 EEG-based brain-computer interfaces *Adv. Exp. Med. Biol.* **1101** 41–65
- [11] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *J. Clin. Neurophysiol.* **113** 767–91
- [12] Sanei S and Chambers J A 2007 *Event-Related Potentials EEG Signal Processing* (Wiley) pp 127–59
- [13] Polich J and Kok A 1995 Cognitive and biological determinants of P300: an integrative review *Biol. Psychol.* **41** 103–46
- [14] Polich J 2007 Updating P300: an integrative theory of P3a and P3b *Clin. Neurophysiol.* **118** 2128–48
- [15] Vincent A 1992 Methods for improving the signal-to-noise ratio of endogenous-evoked potentials *Integr. Physiol. Behav. Sci.* **27** 54–65
- [16] Dawson G D 1954 A summation technique for the detection of small evoked potentials *Electroencephalogr. Clin. Neurophysiol.* **6** 65–84
- [17] Sanei S and Chambers J A 2013 *EEG Signal Processing* (Wiley) (<https://doi.org/10.1002/9780470511923>)
- [18] Mouraux A and Iannetti G D 2008 Across-trial averaging of event-related EEG responses and beyond *Magn. Reson. Imaging.* **26** 1041–54
- [19] Dhawale A K, Smith M A and Ölvéczky B P 2017 The role of variability in motor learning *Annu. Rev. Neurosci.* **40** 479–98
- [20] Faisal A A, Selen L P J and Wolpert D M 2008 Noise in the nervous system *Nat. Rev. Neurosci.* **9** 292–303
- [21] Ouyang G and Zhou C 2020 Characterizing the brain's dynamical response from scalp-level neural electrical signals: a review of methodology development *Cogn. Neurodyn.* **14** 731–42
- [22] Aricò P, Aloise F, Schettini F, Salinari S, Mattia D and Cincotti F 2014 Influence of P300 latency jitter on event related potential-based brain-computer interface performance *J. Neural Eng.* **11** 035008
- [23] Zisk A H, Borgheai S B, McLinden J, Hosni S M, Deligani R J and Shahriari Y 2021 P300 latency jitter and its correlates in people with amyotrophic lateral sclerosis *Clin. Neurophysiol.* **132** 632–42

- [24] Vucic S 2021 P300 jitter latency, brain-computer interface and amyotrophic lateral sclerosis *Clin. Neurophysiol.* **132** 614–5
- [25] Riccio A, Simione L, Schettini F, Pizzimenti A, Inghilleri M, Belardinelli M O, Mattia D and Cincotti F 2013 Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis *Front. Hum. Neurosci.* **7** 732
- [26] Woody C D 1967 Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals *Med. Biol. Eng.* **5** 539–54
- [27] Ouyang G, Hildebrandt A, Sommer W and Zhou C 2017 Exploiting the intra-subject latency variability from single-trial event-related potentials in the P3 time range: a review and comparative evaluation of methods *Neurosci. Biobehav. Rev.* **75** 1–21
- [28] Browne M and Cutmore T R H 2000 Adaptive wavelet filtering for analysis of event-related potentials from the electroencephalogram *Med. Biol. Eng. Comput.* **38** 645–52
- [29] Rol Başar E, Schürmann M A, Demiralp T A, Başar-Eroglu C A and Ademoglu A H 2001 Event-related oscillations are ‘real brain responses’—wavelet analysis and new strategies *Int. J. Psychophysiol.* **39** 91–127
- [30] Başar-Eroglu C, Başar E, Demiralp T and Schürmann M 1992 P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review *Int. J. Psychophysiol.* **13** 161–79
- [31] Foti D, Hajcak G and Dien J 2009 Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA *Psychophysiology* **46** 521–30
- [32] Ouyang G, Sommer W and Zhou C 2015 Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE) *Psychophysiology* **52** 839–56
- [33] Ouyang G 2020 ReSync: correcting the trial-to-trial asynchrony of event-related brain potentials to improve neural response representation *J. Neurosci. Methods* **339** 108722
- [34] Maidhof C, Rieger M, Prinz W and Koelsch S 2009 Nobody is perfect: ERP effects prior to performance errors in musicians indicate fast monitoring processes *PLoS One* **4** e5032
- [35] D’avanzo C, Schiff S, Amodio P and Sparacino G 2011 A Bayesian method to estimate single-trial event-related potentials with application to the study of the P300 variability *J. Neurosci. Methods* **198** 114–24
- [36] Duncan C C, Barry R J, Connolly J F, Fischer C, Michie P T, Näätänen R, Polich J, Reinvang I and Van Petten C 2009 Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400 *Clin. Neurophysiol.* **120** 1883–908
- [37] Iyer D and Zouridakis G 2007 Single-trial evoked potential estimation: comparison between independent component analysis and wavelet denoising *Clin. Neurophysiol.* **118** 495–504
- [38] Schneider R V *et al* 2015 Choice of mother wavelets in CWT spectral decomposition SEG Technical Program Expanded Abstracts 2015 (GeoScienceWorld) pp 2957–61
- [39] McCarthy G and Donchin E 1981 A Metric for Thought: a Comparison of P300 Latency and Reaction Time *Science* **211** 77–80
- [40] Quattrocchi I, Caracci V, Riccio A, Galiotta V, D’ippolito M, Cincotti F, Toppi J and Astolfi L Decomposition frequency optimization in wavelet-based template matching algorithms to manage P300 latency jitter. 2024 *2024 46th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* pp 1–4
- [41] Picton T W 1992 The P300 wave of the human event-related potential *J. Clin. Neurophysiol.* **9** 456–79
- [42] Jaskowski P and Verleger R 1999 Amplitudes and latencies of single-trial ERP’s estimated by a maximum-likelihood method *IEEE Trans. Biomed. Eng.* **46** 987–93
- [43] Thompson D E, Warschausky S and Huggins J E 2012 Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy *J. Neural Eng.* **10** 016006