## Impact Statement

The reversal negativity (RN) and reversal positivity (RP) ERP components are known markers of perceptual reversals when viewing ambiguous figures. Are these components linked to perceptual processes or task/response demands? Our results demonstrate that the RP depends on response factors and thus cannot be linked purely to perceptual processes.

# Response Dependence of Reversal Related ERP Components in Perception of Ambiguous Figures

**Diane Abdallah[1] & Joseph L. Brooks[2]**

[1]School of Psychology, University of Kent

[2]School of Psychology, Keele University

Corresponding Author:

Dr Joseph L. Brooks

School of Psychology

Keele University

Keele

ST5 5BG

United Kingdom

j.l.brooks@keele.ac.uk

## Abstract

Perceptual multi-stability is characterised by alternating interpretations of an unchanging stimulus input. The reversal negativity (RN) and reversal positivity (RP) ERP components show differences in electrophysiological responses between trials on which participants experience a perceptual reversal of a multi-stable stimulus versus trials without a reversal (i.e., stable). However, it is unclear to what extent these two ERP components reflect reversal-related perceptual processing rather than task and response processes. To address this, we varied task and response requirements while measuring the RN and RP. In the standard reversal task, participants indicated whether they saw a perceptual reversal on each trial. In contrast, in the identity task participants reported perceived identity of the stimulus (e.g., face or vase) without any reference to reversals. In some blocks, reversal trials required a response whereas in other blocks stable trials required a response. We found that the RN appeared independently of task and response style. However, the early latency RP component was only present when participants responded manually. For non-response trials, a component was found during the same latency as the RP but with inverted polarity. Our results suggest that the early RP component is dependent on response-related processes rather than being a pure neural signature of perceptual processes related to endogenous perceptual reversals.

## 1. Introduction

The perception of multi-stable or bi-stable visual stimuli (a.k.a., ambiguous or reversible figures) spontaneously fluctuates, or "reverses", despite no corresponding change in the stimulus information (e.g., Leopold and Logothetis, 1999; Blake and Logothetis, 2002; Schwartz et al., 2012; Cao et al., 2014). Well-known multi-stable stimuli include the Necker Cube (or Necker Lattice version, Figure 1A; Necker, 1832) which reverses between facing either left or right, and Rubin's Faces-Vase (Figure 1D; Rubin, 1958/1915) which can be seen as either two profile faces or a vase. Perceptual reversals provide a valuable opportunity to dissociate changes in conscious experience from changes in sensory input. Thus, there has been significant interest in studying neural processes related to perceptual reversals (e.g., Kornmeier and Bach, 2004; Kornmeier & Bach, 2012; Hesselmann, et al., 2008; Brascamp, et al., 2018; Lumer & Rees, 1999; Pitts & Britz, 2011).

Because of its relatively high temporal resolution, the event-related potential (ERP) method has been widely used to study the neural correlates of perceptual reversals. For instance, in studies using the "manual response paradigm", participants continuously view an ambiguous figure and use a button press to indicate when they experience a perceptual reversal. After time-locking ERPs to the manual response, these studies have shown a P300-like parietal positivity that occurs before the response (e.g., Schiller, 1933; Basar-Eroglu, et al., 1993; Strüber and Herrmann, 2002; Strüber, et al., 2001) as well as changes in alpha and gamma band activity associated with perceptual reversals (e.g., Isoglu-Alkaç, et al., 2000; Isoglu-Alkaç and Strüber, 2006; Strüber and Herrmann, 2002; Mathes et al.,2006). However, because of the high variability in reaction times, the manual response paradigm does not reliably capture early post-onset reversal-related ERP components (e.g., Kornmeier and Bach, 2004; Isoglu-Alkaç et al., 1998).

Other studies have employed an "onset paradigm" which involves presenting a sequence of identical ambiguous stimuli separated by brief gaps (e.g., O'Donnell, et al., 1988; see Kornmeier & Bach, 2012 for a review). Participants then manually report whether their perception of the image reversed from one stimulus to the next (i.e., reversal trial) or remained the same (i.e., stable trial). When ERPs are time-locked to the onset of the second stimulus, two posterior reversal-related ERP components appear as differences between the reversal and stable waveforms.

First, the reversal positivity (RP; Kornmeier & Bach, 2005; 2006; Kornmeier, et al., 2007; Britz, et al., 2009) appears approximately 130 ms after stimulus onset as a more positive amplitude on reversal trials than on stable trials at, primarily, occipital electrode positions. The RP has been found for a range of multi-stable stimuli including the Necker Lattice (e.g., Kornmeier, et al., 2001; Kornmeier & Bach, 2005, 2006), the Necker cube (e.g., Kornmeier, Pfäffle, & Bach, 2011), Boring's Old/Young Woman (Kornmeier & Bach, 2014), and during binocular rivalry (e.g., Britz and Pitts, 2011). Importantly, the RP is present only for endogenous reversals (i.e., a change in perception between two identical ambiguous stimuli) and not for exogenous reversals (i.e., two different unambiguous stimuli in sequence). This suggests that the RP may be specifically linked to processes of internally-generated perceptual reversals rather than to changes in sensory input. Furthermore, the RP appears to be insensitive to low-level stimulus differences such as size (e.g., Kornmeier et al., 2011). However, perhaps because of its small amplitude, the RP has not been observed universally across studies (e.g., Pitts, et al., 2007; Intaitė, et al., 2010). Kornmeier and Bach (2012) have suggested that the RP is a marker of perceptual ambiguity detection or perceptual decision conflict that occurs only on reversal trials when perception is changing from one interpretation to another (Kornmeier et al., 2011).

A second reversal-related component, the reversal negativity (RN), appears approximately 260 ms post-stimulus and has been observed with a range of ambiguous stimuli (e.g., Kornmeier & Bach, 2004; 2005; Pitts, et al., 2007, 2008; Britz, et al., 2009; Intaite, et al., 2010). The RN is opposite in polarity to the RP with reversal trials having more negative amplitudes than stable trials in mostly occipital/parietal locations. However, the latency, duration, and spatial distribution of the RN can be sensitive to the length of the response interval (see Kornmeier & Bach, 2012 for review) suggesting that it can be modulated by strategic, non-perceptual factors. The functional role of the RN is still unclear. Although some work has dissociated it from attention-related ERP components such as the N2pc (e.g., Intaitė, et al., 2010), there is debate about whether it reflects higher-level, top-down influences over perceptual reversals (e.g., Pitts et al., 2008). Furthermore, it cannot be linked uniquely to endogenous perceptual reversals because it also appears for exogenous (i.e., stimulus-induced) reversals (e.g., Kornmeier & Bach, 2006).

The task typically used to evoke the RN and RP components is what we call the "Reversal Task". This task typically involves, intermittent presentation of identical ambiguous stimuli while participants report for each stimulus whether there was a reversal of perception, or not, between subsequent stimuli (i.e., a one-back task). The task of detecting reversals has similarities to a change detection task (e.g., Rensink, 2002; Cohen, et al., 2005) in which reversal trials contain a change to detect (relative to the last stimulus) and stable trials do not. Furthermore, in many RP and RN studies, behavioural results show that reversal trials are substantially less prevalent than stable stimuli (~ 30 vs 70% respectively; e.g., Kornmeier & Bach, 2004; Pitts et al., 2007). Thus, in terms of the task, reversal trials could be seen as relatively rare, task-relevant targets in a stream of more prevalent, non-target stable trials. ERP studies have shown that rare visual oddball events can elicit more negative amplitudes for rare targets compared to standard stimuli within a timeframe and scalp

distribution similar to that of the RN (N2b component; e.g., Potts, 2004; Courchesne, Hillyard, & Galambos, 1975). Furthermore, in a change detection task, visual awareness of a change has been associated with a posterior negative amplitude enhancement, similar to the RN, on trials with detected changes (e.g., Pazo-Álvarez, et al., 2017). Thus, it is possible that RN effects, when assessed using the reversal task in the form described above, could be related to visual oddball target detection processes rather than perceptual reversal processes.

Some authors (e.g., Pitts, Nerger, & Davis, 2007; Kornmeier & Bach, 2012; Kornmeier and Bach, 2004) have also highlighted similarities and potential equivalence between the RN and the Selection Negativity (SN) ERP component (e.g., Anllo-Vento & Hillyard, 1996). The SN appears as greater negative amplitudes on trials with attended features (e.g., reversals) compared to those with unattended features (e.g., stable trials). If reversal trials attract attention because of their task relevance whereas stable trials do not, then they could evoke an SN. In this account, the RN would be linked to general feature selection mechanisms instead of specifically to processes involved in perceptual reversal.

To investigate whether the RN and RP depend on task demands and target status differences between reversal and stable trials as described above, we directly compared the reversal task paradigm that has been used in many RP and RN studies (e.g., Kornmeir & Bach, 2004;Pitts, Nerger, & Davis, 2007) to an "Identity Task" (e.g., Pitts, et al., 2009; Pitts, Martínez, & Hillyard. 2010; Britz & Pitts, 2011). In the identity task, participants observed a sequence of 800 ms duration intermittently presented ambiguous stimuli (Necker Lattice in Experiment 1 or Rubin's Faces-Vase in Experiment 2; Figures 1 & 2). For each stimulus, participants were asked to identify which of two possible identities best matched their perception of the stimulus (i.e., Experiment 1: left-facing or right-facing for Necker Lattices; Experiment 2: face or vase). Although each trial can be later reclassified by the experimenter as a reversal or stable trial based on the sequence of responses across subsequent trials (e.g.,

face percept preceded by vase percept = reversal trial), the distinction between reversal and stable trials was not directly task relevant in the identity task. As task-related differences between the reversal and stable trials were minimised, this should allow the identity task to more purely isolate perceptual processing differences between reversal and stable trials. To determine whether this change in task demands affected RN and RP amplitude, we had participants do the identity task in half of the blocks of each experiment and the reversal task in the other half.

To determine if the RN and RP are also affected by whether reversal trials are targets for response, we varied the response style across blocks. In half of the reversal task blocks, participants responded only when they perceived a reversal and withheld response on stable trials and in the other blocks, they adopted the opposite response style (e.g., Kornmeier & Bach, 2004). For the identity task, participants were instructed to respond only to one identity (e.g., faces) in half of the blocks and respond only to the other identity (e.g., vase) in the other half of the blocks. This means that we were able to observe trials with all combinations of trial type (reversal vs. stable) and response action (responded vs. no response) in both reversal and identity tasks. With this design we were able to compute the RN and RP when both types of trials were response targets (i.e., reversal, responded and stable, responded) and when both were not response targets (i.e., reversal, no response and stable, no response). This ensures that the reversal vs. stable trial comparisons which are used to compute the RN are not confounded by differences in response demands. Furthermore, we were able to determine whether status as a response target affects RN and RP amplitude.

## 2 Experiment 1: Necker Lattice

### 2.1 Methods

#### *2.1.1 Participants*

Twenty-five (17 female, 8 male) undergraduate psychology students were recruited from Keele University's Psychology Research Participation Time scheme and received partial course credit for participation. Participants had a mean age of 21.8 years (range: 18 to 48 years). All participants had normal or corrected-to-normal visual acuity (average 0.001 logMAR; Precision Vision Logarithmic ETDRS 2000 chart). Eight (of 25) participants were excluded from the analysis. Six of these were based on a priori exclusion criteria (see data analysis methods below). One exclusion was due to a participant becoming unwell and another was due to experimenter error resulting in data loss. The final sample included 17 participants (13 female; 4 male) with an average age of 22.8 years. The University Ethics Review Panel at Keele University and the Psychology Ethics Committee at the University of Kent approved this study. All participants gave informed consent ahead of participation.

### *2.1.2 Stimuli & Apparatus*

We used one ambiguous (Figure 1A) and two unambiguous versions of the Necker Lattice (Figure 1B-C; Necker, 1832; Kornmeier & Bach, 2004) as the visual stimuli in this study. All stimuli were presented centrally on a black background using a 24.5 inch BenQ Zowie XL2540 computer monitor at 120 Hz refresh rate and 1920 x 1080 pixels resolution. All stimuli were 5.52° x 5.52° and the maximum luminance of white portions was 373.43 cd/m$^2$ (CIE1931: x = 0.326, y = 0.325; ColorCal MKII Colorimeter, Cambridge Research Systems; Rochester, UK). Viewing distance was maintained at 57 cm using a chin rest. The experiment was controlled by PsychoPy2 V1.82.01 (Peirce, 2019). A 0.503° white cross at the centre of the screen served as fixation target ahead of and during stimulus presentation. Participant responses were recorded via a computer keyboard using the up arrow key.

**INSERT FIGURE 1 ABOUT HERE**

### *2.1.3 EEG Recording*

EEG scalp voltages were recorded at 1024 Hz using a 24 bit DC-coupled Biosemi ActiveTwo System (Biosemi; Amsterdam, Netherlands) with 64 Ag-AgCl active scalp electrodes, three active EOG electrodes (details below), and two active mastoid electrodes. Low pass filtering was performed in the Analogue-to-Digital Converter's decimation filter which had a 5th order sinc response with a -3 dB point at 1/5th of the sample rate. Participants wore a cloth cap with 64 10-10 system electrode positions (Seeck, et al., 2017; Nuwer, et al., 1998). Three EOG electrodes were used to measure right and left horizontal electro-oculogram (HEOG; just lateral to canthi) and vertical electro-oculogram (VEOG; 2 cm below the left eye). SignaGel (Parker Labs; https://www.parkerlabs.com/) was used as an electrolyte. Face electrode locations were prepared with an isopropyl alcohol wipe. All electrodes were adjusted to have offsets within a range of -10 to 10 mV of the common mode voltage. The Biosemi system does not require electrode impedance checks. Data were recorded relative to the CMS/DRL circuit and re-referenced offline (see data analysis methods below).

### *2.1.4 Procedure*

During the instructions, the experimenter ensured that each participant was able to distinguish the two interpretations of the Necker Lattice and explained that although the same picture would be presented repeatedly, they may perceive it differently on different trials. The experimenter also explained that some trials would have unambiguous stimuli. While the experimenters prepared the EEG cap, participants completed a practice test comprising eight blocks with 28 trials per block (8 unambiguous). The practice session included both the reversal and identity tasks (described below). After the practice trials, the experimenter calibrated the eyetracker and then the main experiment began.

**INSERT FIGURE 2 ABOUT HERE**

Each trial comprised a single stimulus presented binocularly for 800 ms followed by an ITI of 400 ms. This ITI was extended to 1000 ms if participants responded at any point during the 1200 ms (stimulus plus ITI) trial period (Figure 2). Within each block, approximately 90% of the trials contained the ambiguous Necker Lattice (Figure 1A) and the remaining ~10% contained unambiguous stimuli (Figure 1B-C; half right-facing). It is impossible to assess attention/accuracy in the ambiguous trials due to the subjective nature of ambiguous figure perception. Thus, unambiguous stimuli, for which there are objectively correct answers in both the identity and reversal tasks, were presented as an attention check throughout the trial sequence. Poor performance on these trials was used as an *a priori* exclusion criterion (see Data Analysis Methods section below). Trials with unambiguous stimuli were presented in either "stable" pairs (i.e., right-facing trial then right-facing trial or left-facing then left-facing) or "reversal" pairs (i.e., right-facing then left-facing or left-facing then right-facing). Stable and reversal pairs occurred equally often within the trial sequence. The unambiguous trial pairs appeared randomly throughout the sequence of ambiguous stimuli.

There were eight blocks of 150 trials each (4-5 minutes). Participants completed one of the two different tasks in each block. For each stimulus in the reversal task blocks, responses were based on whether the participant's perception on one trial matched (stable trial) or did not match (reversal trial) that of the immediately preceding stimulus. In contrast, in the identity task blocks, participant's responses were based on their perception of the Necker lattice orientation (i.e., front facing left or front facing right) without reference to the preceding stimulus. The reversal task was required in half (four) of the blocks and the identity task in the other half. The experimenter explained these tasks during the instructions period and the computer clearly indicated which task to complete before each block began.

Blocks also differed in the response style. For two of the four reversal task blocks, participants adopted a respond-to-reversals response style. This means that they only pressed a button on trials in which they experienced a reversal. They did not respond at all if their perception of the stimulus was the same as that of the previous stimulus (i.e., stable). For the other two reversal task blocks, participants adopted a respond-to-stable response style. This means that they only responded if their perception of the stimulus was the same as that of the preceding stimulus. They did not respond if there was a reversal in their perception. For identity task blocks, participants adopted a respond-to-left-facing response style for two of the four blocks and a respond-to-right-facing response style for the other two blocks. Thus, across the eight blocks of the experiment, there were four types of blocks: reversal task, respond-to-reversal; reversal task, respond-to-stability; identity task, respond-to-left-facing; and identity task, respond-to-right-facing. Each participant had a different random order of these block types. Participants were given clear on-screen instructions about the response style at the beginning of each block and performance on unambiguous trials was monitored to ensure compliance (see next subsection) and verbally instructed to report their first impression after stimulus onset. The full scripts including stimuli are available at https://osf.io/neum8/.

### 2.1.5 Data Analysis Methods

EEG recordings were manually checked for artefacts from eye movements and amplitude excursions exceeding ±100 mV and high-pass filtered offline at 0.1 Hz. On average, 2.86% (range: 0-4.95%) of trials were excluded due to these criteria. Mastoid electrodes were noisy in a substantial number of participants and thus were not available for referencing without substantial data loss. Data were referenced to the Cz electrode offline because: (a) its location is remote from many artefact sources (e.g., muscle, eyes), (b) has a central, non-lateralised position, (c) is remote from the posterior locations at which the RN

and RP are typically observed and the estimated source in inferior occipital-temporal cortex (Pitts, et al., 2009). Each trial was coded as either a reversal trial or a stable trial by considering whether there was a response on that trial and which block type it was in. For instance, in a reversal task block with respond-to-reversals response style, a trial would be marked as a reversal trial if there was a response during the trial and would be marked as a stable trial if there was no response. In identity task blocks, a trial would be marked as a reversal if there was a response on the trial but not on the trial before it or vice versa. If a trial had a response and the one before it did as well or both trials had no response, then this was marked as a stable trial. Once all trials were labelled, the data were segmented into 1200 ms epochs (-100 to 1100 ms) and sorted into eight conditions. These eight conditions were the combinations of three factors: task (reversal task or identity task), trial type (reversal trial or stable trial), and response action (responded or not) on the trial. All epochs were baseline corrected using the average amplitude from -100 to 0 ms and then epochs were averaged within each condition to form ERPs. ERPs were digitally filtered with a 25 Hz low-pass filter and averaged across participants to create grand average waveforms. Epochs for unambiguous trials were discarded because there were too few for analysis and they are not relevant to the questions of this paper.

Based on our a priori behavioural exclusion criteria (less than 70% correct on the unambiguous trials), no participants were excluded. This indicates that they correctly followed the response style and task instructions for each block. In addition, in order for a participant's data to be included in further analyses, at least 25 non-discarded trials per condition were required. Six participants were excluded because of the low number of non-discarded trials per trial type. The low number of trials for these participants was due to a combination of our EEG artefact rejection criteria and the participant's response pattern (i.e., too few reversal trials).

In line with the analysis steps of previous experiments (Kornmeier & Bach, 2004; Pitts, Nerger & Davis, 2007) and to avoid inflating Type I error rate (e.g., Brooks, Zoumpoulaki, & Bowman, 2017; Kilner, 2013), we quantified the RN ERP component with an a priori region-of-interest (ROI) of 200-400ms in channels O1, O2, Oz, PO7, and PO8. The a priori ROI for the RP component was 100-200 ms in channels O1, O2, and Oz. Average amplitude was calculated separately for each channel within these temporal ROIs for each participant and condition. Results were submitted to two separate (one for RP and one for RN) repeated measures ANOVAs (5x2x2x2 for RN; 3x2x2x2 for RP) with Channel, Task (Reversal vs. Identity), Trial Type (Reversal vs. Stability) and Response Action (Response vs Non-Response) as factors.

## 2.2 Results

### *2.2.1 Response Time*

Response times (relative to stimulus onset) on ambiguous trials were evaluated with respect to the Trial Type (reversal vs. stable) and task (reversal vs. identity) factors. There were not sufficient data to analyse RTs for unambiguous trials. Response action was not a factor in the RT analysis because there were no responses in the non-response condition. For each participant, the mean reaction time was computed within each of the four conditions and then these were averaged across participants to create the mean reaction time in each condition (Figure 2C). Participants responded, on average, 676 ms (range = 570-760 ms; SE = 13 ms) after stimulus onset. A 2x2 Repeated Measures ANOVA found no significant effect of Task or Trial Type for response times (RTs), $p >$. 208. However, there was a significant interaction between Task and Trial Type, $F(1,16) = 15.40$, $p = .001$, $\eta_p^2 = .490$. This crossover interaction was examined by testing the simple effects of trial type in each task. In the reversal task, RTs for stable trials, $M = 670$ ms ($SE = 13$), were significantly faster than

those for reversal trials, $M$ = 700 ms ($SE$ = 15), $t(16)$= 2.89, $p$ = .011. In contrast, for the identity task, RTs for reversal trials, $M$ = 642 ms ($SE$ = 23), were faster than those for stable trials, $M$ = 692 ms ($SE$ = 13.5), $t(16)$= -2.49, $p$ = .024. We also assessed the interaction by testing the simple effects of task in each trial type. There was a significant difference in mean RT between the Reversal and Identity Tasks on reversal trials, $t(16) = 2.92$, $p = 0.01$ whereby participants indicated a faster reversal response in the Identity Task than in the Reversal Task (same means as above). There was however no difference in mean RTs between the Reversal and Identity Tasks on stable trials, $t(16) = -1.53$, $p = .146$. Corresponding analyses were not conducted for the unambiguous stimuli because there were too few trials. Mean RTs did not differ between left-facing (M = 682 ms, SE = 20) and right-facing (M = 670 ms, SE = 15) response trials in the identity task blocks, $t(16) = .727$, $p = .478$.

### *2.2.2 Unambiguous Lattices*

Across both the identity and reversal tasks, participants correctly responded to 85.95% ($SE$ = 1.74%) of the unambiguous lattice pairs. This is significantly greater than chance performance (50%), $t(16) = 20.64$, $p < .001$. In the identity task, participants had above chance orientation discrimination accuracy, 90.72% ($SE$= 1.70%), $t(16) = 23.99$, $p < .001$. In this task, there was a correct answer for every trial. This means that the responses to both stimuli that appeared sequentially were included in our calculations. In the Reversal Task, participants had above chance reversal detection accuracy, 76.72% ($SE$ = 3.32%), $t(16) = 8.05$, $p < .001$. As participants compared the second image to the preceding one in this task, there was only one correct answer per unambiguous pair.

### *2.2.3 Ambiguous Lattices*

In the Reversal Task, 33.66% (SE = 3.35%; Range = 15.30-66.16%) of trials, on average, were classified as reversal trials (66.34% stable). In the Identity Task, on average,

32.03% (SE = 3.24%; Range = 16.14- 71.61%) of trials were reversals (67.97% stable). There was not a significant difference between the percentage of reversals in the Reversal and Identity Tasks, $t(16) = .679$, $p = .507$. The reversal rate also did not differ between trials with a response (M = 32.74%; SE = 2.76%; Range = 17.55-60.96%) and those with no response (M = 33.91%; SE = 4.12%; Range = 12.93-77.23%), $t(16) = -.355$, $p < .742$. In the Identity Task only, 57.03% (*SE* = 1.99%; Range = 27.05-59.33%) of trials were reported as right-facing lattices (42.97% left-facing). A one sample t-test revealed that the percentage of right-facing trials reported is significantly greater than 50%, $t(16) = 3.526$, $p = .003$. Our results are consistent with previous studies which have found that participants have a bias for right-facing perception (Sundareswara & Schrater, 2008; Kornmeier, et al., 2009; Troje & McAdam, 2010).

### *2.2.4 Electrophysiological Results*

Figures 3A and 3B show the grand average ERP waveforms (ambiguous trials only; average of RN ROI electrodes) for reversal trials (solid lines) and stability trials (dashed lines) separately for the identity task and reversal task conditions. Figures 3C and 3D show the grand average reversal and stability trial ERP waveforms (ambiguous trials only; average of RN ROI electrodes) separately for the response and non-response trials. Results for the same four conditions using the RP ROI electrodes are shown in Figures 3E-3H. For all analyses below, the Greenhouse-Geisser correction was applied when Mauchly's test of sphericity was significant at the $p = .05$ level. Difference waves, scalp distributions, and ERPs at all electrodes are presented in the Supplementary Materials.

**INSERT FIGURE 3 ABOUT HERE**

### *2.2.5 Reversal Negativity (RN)*

Mean amplitude across the RN ROI was calculated in each condition for each participant (see Data Analysis Methods section). We conducted a 5x2x2x2 repeated measures ANOVA with the factors Channels, Task (Identity Task vs. Reversal Task), Trial Type (Reversal vs Stability), and Response Action (Response vs. Non-Response). There were significant main effects of Channels, $F(2.33,16) = 4.209$, $p = .018$, $\eta_p^2 = .208$, Trial Type, $F(1,16) = 18.272$, $p = .001$, $\eta_p^2 = .533$ (i.e., a significant RN), and Response Action, $F(1,16) = 16.496$, $p = .001$, $\eta_p^2 = .508$. To illustrate the significant RN indicated by the Trial Type effect, Figure 4A shows the mean (across participants) ERP amplitude for reversal (black bars) and stability (grey bars) trials separately for the identity and reversal tasks. There was a significant interaction between Channel and Trial Type, $F(4,16) = 4.346$, $p = .015$, $\eta_p^2 = .214$. To explore this interaction, we tested whether there was an effect of trial type at each electrode and found that there was (all $p < .003$). None of the other factors or interactions were significant, all $p > .150$ (see Supplementary Materials). We verified that there was a significant Reversal Negativity (i.e., difference between reversal and stability trials) in both the reversal task, $F(1,16) = 12.527$, $p = .003$, $\eta_p^2 = .439$ and in the identity task, $F(1,16) = 5.336$, $p = .035$, $\eta_p^2 = .250$ (Figure 4A).

**INSERT FIGURE 4 ABOUT HERE**

### *2.2.6 Reversal Positivity (RP)*

Mean amplitude across the RP ROI was calculated in each condition for each participant (see Data Analysis Methods section). A 3x2x2x2 repeated measures ANOVA with the factors Channels, Task, Trial Type (Reversal vs. Stability), and Response Action revealed a significant interaction between Trial Type and Response Action, $F(1,16) = 4.895$, $p = .042$, $\eta_p^2 = .234$. To illustrate this interaction, Figure 4B shows the average (across participants) mean RP ROI amplitude for reversal (black bars) and stability (grey bars) trials

separately for response and non-response trials. This analysis also revealed a marginally significant main effect of Task, $F(1,16) = 3.753$, $p = .071$, $\eta_p^2 = .190$. This main effect has no bearing on hypotheses related to the RP and thus will not be discussed further. None of the other factors or interactions were significant, all $p > .118$.

To explore the Trial Type x Response Action interaction (Figure 4B), we examined the effect of Trial Type separately for Response and Non-Response trials. This revealed a significant Trial Type effect (i.e., RP) in the Response trials, $F(1,16) = 4.803$, $p = .044$, $\eta_p^2 = .231$. There was no significant effect of Trial Type in the Non-Response trials, $F(1,16) = 2.892$, $p = .108$, $\eta_p^2 = .153$ (i.e., no RP).

### 2.3 Interim Discussion

We aimed to determine whether the reversal-related ERP components, RN and RP, are sensitive to task and response factors. We found that the RN was present in all conditions and was not modulated by task or response action factors. This means that the RN, as has been consistently observed in the reversal task across many studies, cannot be accounted for by task and response related differences between reversal and stable trials. In contrast, although the RP was not affected by task, it was only present on trials in which participants made a manual response. When the response style required participants to withhold response, there was no evidence of a significant RP component. Although the RP is uniquely present for endogenous reversals, our results suggest that it is not a pure measure of perceptual processing related to reversals. Instead, our results support the idea that the RP reflects response-related processes (e.g., Kornmeier & Bach, 2012).

Although the pattern of electrophysiological responses was the same between the identity and reversal tasks, the response time results differed between the two tasks. In the reversal task, participants responded more quickly to stability trials than reversal trials. This

result is similar to that observed by Kornmeier, Pfäffle, & Bach (2011) using the reversal task paradigm. Kornmeier and Bach (2011) have suggested that this RT difference could arise because reversal trials, compared to stability trials, involve a perceptual decision conflict in which the emerging perceptual interpretation of the ambiguous stimulus is incongruent with immediately preceding perceptual experiences and thus leads to longer processing times.

Interestingly, we observed an RT difference in the opposite direction (i.e., reversal faster than stability) in the identity task blocks. This suggests that the hypothesized decision conflict associated with reversals in the reversal task may be task dependent. But what gives rise to the opposite effect in the identity task? This may be related to the fact that, in the identity task, reversal response trials were always preceded by a response-withheld trial. This is because, for an identity task trial to be coded as a reversal, the response on the preceding trial needed to be different (e.g., a non-response before a response). In contrast, stability response trials were always preceded by a trial with the same response. Thus, the differences could be due to strategic effects of preceding trials. Previous studies using the identity task have not reported RTs for reversal and stable trials separately (e.g, Pitts, et al., 2009). Thus, further work will be necessary to determine whether this RT response pattern is reliable and to develop a detailed understanding of it.

Before making firm conclusions, we set out to replicate the Experiment 1 results in an additional experiment using a different visual stimulus. Because it has been widely studied as an example of perceptual ambiguity (e.g., Pitts, Nerger, & Davis, 2007; Hesselmann, et al., 2008; Rassi, et al., 2019), we chose Rubin's faces-vase (Figure 1D) as the ambiguous stimulus for Experiment 2. Otherwise, the design of Experiment 2 was the same as that in Experiment 1. However, to make face and vase interpretations of the stimulus equally probable, the exact distance between the two edges (i.e., inter-edge distance, IED; see Figure

5) in the stimulus varied across participants. We used a pre-test procedure to select the IED for each participant that best approximated 50/50 faces/vase reports.

## 3 Experiment 2: Rubin's Faces-Vase

### 3.1 Methods

#### *3.1.1 Participants*

Twenty-five (14 females, 11 males; different from Experiment 1) undergraduate psychology students were recruited from Keele University's Psychology Research Participation Time scheme and received partial course credit for participation. Participants had a mean age of 21.76 years (range: 18-54 years). All participants had normal or corrected-to-normal visual acuity, average = 0.01 logMAR. Using the same exclusion criteria as Experiment 1, data from six participants were excluded. In addition to that, we excluded two participants from the analyses due to technical issues relating to the recording of these participants' EEG data. The final sample included 17 participants (9 females; 8 males) with an average age of 20.4 years. Ethics and consent arrangements were identical to Experiment 1.

#### *3.1.2 Stimuli & Apparatus*

The apparatus and software were the same as in Experiment 1. The stimulus on each trial was either a white outline version of Rubin's ambiguous Faces-Vase (Figure 1D) or one of two unambiguous versions of the Faces-Vase (Figure 1E-F). To create the unambiguous stimuli, we adjusted the ambiguous image to include T-junction partial occlusion cues. Each stimulus was presented centrally on a black background. All stimuli were 4.62° vertically. The horizontal frame width (FW; Figure 5) and inter-edge distance (IED; distance between the nose tips, Figure 5) for each participant's main experiment stimulus were determined by a pre-test to maximise ambiguity of the stimulus (see details in procedure below). The average

FW in the main experiment was 9.86° (range: 7.96-13.23°). The average IED in the main experiment was 2.96° (range: 2.55-3.61°). A 0.503° grey cross at the centre of the screen served as fixation target before and during stimulus presentation. Participants made key responses on a computer keyboard. EEG and eye-tracking methods for acquisition and analysis were the same as in Experiment 1.

**INSERT FIGURE 5 ABOUT HERE**

### *3.1.3 Procedure*

While the experimenters setup the EEG cap, each participant undertook a 15 minute pre-test to determine the maximally ambiguous configuration of the face-vase stimulus (i.e., closest to 50% faces) using the configural cue of small area (e.g., Rubin, 1958/1915; Castro, Lazareva, Vecera, & Wasserman, 2010; Harrower, 1936). All combinations of three frame widths (7.96°, 10.60°, and 13.23°) and 8 IEDs (2.13°, 2.34°, 2.55°, 2.76°, 2.97°, 3.19°, 3.40° and 3.61°) were presented to participants for 800 ms each. There were four blocks of 96 trials each with 16 repetitions of each IED and FW combination. In the pre-test, participants had to report via button press whether they perceived faces (left arrow key) or a vase (right arrow key) on each trial (i.e., identity task). The values of IED and FW which resulted in percentage of vase reports closest to 50% were used in the main experiment. If the participant's percentage of vase percepts was below 30% or above 70%, then the pre-test was repeated. Three of the 25 participants repeated the pre-test and suitable values obtained on the second run.

The main experiment procedure was similar to that of Experiment 1 except that the Necker lattice stimuli were replaced with faces-vase stimuli. Response options in the identity task for Experiment 2 were faces or vase. Two of the identity task blocks were respond-to-face blocks whereas the other two were respond-to-vase blocks. The ambiguous faces-vase

stimulus appeared on 90% of the trials. The remaining trials were either face-biased (5%) or vase-biased (5%) stimuli. Unambiguous stimuli were presented in either stable pairs (face-biased then face-biased; vase-biased then vase-biased) or reversal pairs (face-biased then vase-biased; vase-biased then face-biased). Stable and reversal pairs occurred equally often. Unambiguous pairs were distributed randomly throughout the sequence of ambiguous stimuli. There were four types of blocks with each presented twice: reversal task, respond-to-reversal; reversal task, respond-to-stability; identity task, respond-to-faces; and identity task, respond-to-vase. Each participant has a different random order of the blocks.

### *3.1.4 Data Analysis Methods*

The steps taken to pre-process, exclude and analyse the data were the same as in Experiment 1. On average, 9.15% (range: 4.42-17.28%) of the trials were excluded due to our exclusion criteria (see Experiment 1). Six out of 25 participants were excluded from Experiment 2 due to insufficient trials in at least one condition (e.g. low number of stability trials in one of the conditions). ROIs and the factorial design were the same as in Experiment 1.

## 3.2 Results

### *3.2.1 Response Time*

For processing and analysis of the response times, we used the same approach as in Experiment 1. On average, participants responded 700ms (range = 620-757ms, SE = 11.0) after stimulus onset. A 2x2 Repeated Measures ANOVA revealed that there was no significant main effect of Task or Trial Type on response times, $p > .107$. However, there was a significant interaction between Task and Trial Type, $F(1,16) = 22.822$, $p < .001$, $\eta_p^2 = .588$ (Figure 5B). This crossover interaction was examined by testing the simple effects of trial type in each task. RT for reversal trials ($M = 663$ ms, $SE = 16$) was significantly faster than

for stability trials ($M$ = 718, $SE$ = 9) in the identity task, $t(16)$= -3.406, $p$ = .004. However, there was not a significant difference between reversal ($M$ = 697, $SE$ = 21) and stability ($M$ = 695, $SE$ = 8) trial RTs in the reversal task, $t(16)$= .08, $p$ = .937. These results replicate the Experiment 1 RT results in the Identity task and suggest that this novel RT effect warrants further investigation in future studies. Interestingly, the Experiment 1 reversal task RT effect did not replicate in Experiment 2. The reason for this is unclear. Overall, many RP and RN studies do not conduct detailed analysis of RTs. Future work should look at RTs to see whether there are replicable behavioural patterns that can shed light on theoretical questions. Corresponding analyses were not conducted for the unambiguous stimuli because there were too few trials. In identity task blocks, there was no significant difference between face response trials (M = 702 ms, SE = 36) and vase response trials (M = 705 ms, SE = 29), $t(16)$ = .08, $p$ = .937.

### *3.2.2 Unambiguous Faces-Vase Stimuli*

Overall, participants correctly responded to 83.17% (SE = 1.79%) of the unambiguous Face-Vase pairs. In the identity task, participants correctly discriminated the stimulus on 87.99% (SE = 3.13%) of trials. This was significantly greater than 50%, $t(16)$ = 17.01, $p < .001$. Participants scored 73.7% (SE = 4.28%) on reversals vs stability discrimination in the reversal task and this was significantly greater than 50%, $t(16)$ = 8.99, $p$ <.001.

### *3.2.3 Ambiguous Faces-Vase Stimuli*

In the Reversal Task, 32.21% (SE = 3.05%; Range = 10.68-50.34%) of trials were classified as reversal trials (67.78% stable). In the Identity Task, on average, 41.86% (SE = 2.78%; Range = 22.52-63.27%) of trials were reversals (58.13% stable). There was a significant difference between the percentage of reversals in the reversal and identity tasks,

$t(16) = -3.224$, $p < .005$. The reversal rate did not differ between trials with a response (M = 35.49%; SE = 2.74%; Range = 20.50-58.05%) and those with no response (M = 37.62%; SE = 2.60%; Range = 18.81-54.49%), $t(16) = -1.272$, $p < .222$. In the identity task, 50.52% (*SE* = 2.47%; Range = 29.10-66.42%) of trials, on average, were reported as faces (49.48% vase) and this was not significantly different from 50%, $t(16) = .209$, $p = .837$. These results demonstrate that the pre-test was effective at making the stimulus maximally ambiguous.

### *3.2.4 Electrophysiological Results*

Figures 6A and 6B show the grand average ERP waveforms (ambiguous trials only; average of RN ROI electrodes) for reversal trials (solid lines) and stability trials (dashed line) separately for the identity task and reversal task conditions. Figures 6C and 6D show the grand average reversal and stability trial ERP waveforms (ambiguous trials only; average of RN ROI electrodes) separately for the response and non-response trials. Results for the same four conditions using the RP ROI electrodes are shown in Figures 6E-6H. The RP and RN ROI time windows are indicated with shading. For all of the analyses below, the Greenhouse-Geisser correction was applied when Mauchly's test of sphericity was significant at the $p = .05$ level. Difference waves, scalp distributions, and ERPs at all electrodes are presented in the Supplementary Materials.

**INSERT FIGURE 6 ABOUT HERE**

### *3.2.5 Reversal Negativity (RN)*

Mean amplitude across the RN ROI was calculated as in Experiment 1. A 5x2x2x2 repeated measures ANOVA with the factors Channels, Task, Trial Type, and Response Action revealed significant main effects of Trial Type, $F(1,16) = 15.461$, $p = .001$, $\eta_p^2 = .491$, and Response Action, $F(1,16) = 16.495$, $p = .001$, $\eta_p^2 = .507$. There was also a significant interaction between Channels and Task, $F(1.369, 16) = 12.395$, $p < .001$, $\eta_p^2 =$

.437. This interaction and the main effect of Response Action were not explored further because they do not involve the trial type factor and thus have no bearing on the hypotheses regarding the RN. No other factors or interactions were significant, $p > .161$ (see Supplemental Materials). To illustrate the significant RN indicated by the Trial Type effect, Figure 7A shows the mean grand average ERP amplitudes for reversal and stability trials separately for the identity and reversal tasks.

Although the Trial Type factor did not significantly interact with Task, in order to address our a priori hypotheses regarding task differences, we tested the effect of trial type separately in the two tasks. The effect of trial type was only marginally significant in the Reversal Task, $F(1,16) = 4.230$, $p = 0.056$, $\eta_p^2 = .209$ but significant in the Identity Task, $F(1,16) = 14.499$, $p = .002$, $\eta_p^2 = .475$.

**INSERT FIGURE 7 ABOUT HERE**

### *3.2.6 Reversal Positivity (RP)*

Mean amplitude across the RP ROI was calculated as in Experiment 1. A 3x2x2x2 repeated-measures ANOVA with the factors Channels, Task, Trial Type and Response Action revealed a significant main effect of Channels, $F(1.529,16) = 7.477$, $p = .005$, $\eta_p^2 = .318$. All other main effects were not significant, all $p > .081$ (see Supplemental Materials). There was a significant interaction between Trial Type and Response Action, $F(1,16) = 9.485$, $p = .007$, $\eta_p^2 = .372$. To further analyse this interaction, we examined the effect of trial type separately for Response and Non-Response trials (Figure 7B). For Response Trials, there was a significant Trial Type effect in the typical direction of the Reversal Positivity (i.e., reversal trials > stable trials), $F(1,16) = 5.266$, $p = .036$, $\eta_p^2 = .248$. There was also a significant Trial Type effect in the Non Response trials, $F(1,16) = 12.350$, $p = .003$, $\eta_p^2 = .436$. However, this effect was opposite in polarity to the typical RP polarity (i.e., stable >

reversal; Figure 7B). There was also a marginally significant interaction between Channels and Response Action, $F(1.258,16) = 3.686$, p = .061, $\eta_p^2 = .187$. However, this interaction was not explored further because it did not involve the trial type factor and thus has no bearing on the hypotheses for the RP component. No other interactions were significant, $p > .143$.

***4 Comparison of Experiments 1 and 2 ERP Results & Average Reference Results***

To determine whether the RN and RP results were similar across Experiments 1 and 2, we conducted two 2x2x2x2 mixed-factors ANOVAs with Trial Type, Task, Response Action, and Experiment/Illusion (between-subjects; Necker Lattice vs. Faces-Vase) as factors. We collapsed the data over all other factors because none of these interacted with the Trial Type factor in Experiments 1 or 2. Because the RN and RP components are defined by the trial type effect (i.e., the difference between reversal and stability trials), only factors that interact with that factor have any bearing on our hypotheses about these ERP components. Thus, below we present only significant effects and interactions that involve the trial type factor. The full ANOVA results tables are available in the supplemental materials.

For the RN analysis, there was a significant main effect of Trial Type, $F(1, 32) = 31.579$, $p < .001$, $\eta_p^2 = .497$, reflecting the RN component. The Trial Type effect reflects a significant RN ERP component. The interaction between Trial Type and Experiment was not significant, $F(1,32) = .449$, $p = .507$, which is consistent with there being no difference in the RN between Experiments 1 and 2. For the RP analysis, there was a significant interaction between the Response Action and Trial Type factors, $F(1,32) = 14.025$, $p = .001$, $\eta_p^2 = .305$. These factors did not significantly interact with the Experiment factor, $F(1,32) = .397$, $p = .533$. This is consistent with there being the same general pattern of effects across the two

experiments. Full ANOVA results for the combined analysis are available in the Supplemental Materials.

In the analyses presented for Experiments 1 and 2 and the combined analysis above, we used a Cz reference for the reasons outlined in section 2.1.5 (Data Analysis Methods). However, Cz has not been commonly used as a reference in studies of the RN and RP. Previous studies of the RN and RP have used a variety of references including average mastoids (e.g., Pitts, et al., 2009), average/linked ears (e.g., Kornmeier and Bach, 2004; 2005; Britz, Landis, & Michel, 2009), average reference (e.g., Pitts, et al., 2007, 128 channels; Britz & Pitts, 2011, 64 channels), averaged T7/T8 (e.g., Kornmeier & Bach, 2014), and nose (e.g., Intaitė, et al., 2010). To assess whether our results were dependent on the choice of reference, we conducted the combined analysis reported above with an average reference of the 64 scalp electrodes (see Supplemental Materials for list of electrode locations). We were not able to conduct analyses with ear references because these locations were not recorded due to lack of ear clips to secure the electrodes.

Within the RN time window, we found a significant main effect of Trial Type, $F(1,32) = 16.707$, $p < .001$, $\eta_p^2 = .343$, such that ERP amplitudes were more negative on reversal trials than on stability trials. No other factors significantly interacted with the Trial Type factor (all $p > .148$; see Supplemental Materials Table S7). This is congruent with what we found in our earlier analyses with the Cz reference and demonstrates that the RN is not affected by task or response factors. For the RP time range, we found a significant Trial Type x Response interaction, $F(1,32) = 15.05$, $p < .001$, $\eta_p^2 = .320$ (see Supplementary Materials Figure S1). To understand this interaction, we examined the effect of trial type separately for reversal and non-reversal trials. For the response trials, ERP amplitude on reversal trials was significantly more positive than on stability trials (i.e., a normal polarity RP), $t(33) = 3.835$, $p < .001$. In contrast, for non-response trials, stability trials had significantly more positive ERP

amplitude than reversal trials (i.e., a reverse polarity RP), $t(33) = -2.949$, $p = .006$. There was not a significant main effect of Trial Type, $F(1,32) = .210$, $p = .650$. No other factors significantly interacted with Trial Type (all $p > .302$; Supplementary Materials Table S8). These results demonstrate that the same pattern of results using the average reference as those found using the Cz reference.

## 5 General Discussion

The results of Experiment 2 with Rubin's Face-Vase were similar to the results of Experiment 1. We found that both the RN and RP were unaffected by whether participants performed the reversal task or the identity task. We had hypothesised that the reversal task, which has been commonly used to evoke the RN and RP, might create critical task and response-related differences between reversal and stable trials. In particular, reversal trials may stand out as relatively rare targets amongst more common, non-target stable trials[1]. In turn, these differences between reversal and stable trials could be responsible for the differences between reversal and stable trial ERP amplitudes that comprise the RN and RP. Our results suggest that the RN and RP cannot be explained by our hypothesised task-induced differences between reversal and stability conditions. This is because both components were present with the same amplitude when participants performed the identity task which was designed to reduce these differences.

Although we intended the identity task to better equate task and response factors between reversal and stable trials, this was not complete. For instance, the percentage of reversals (~33%) was equal in the two tasks and thus there is potential for reversal trials to stand out as rare oddballs even in the identity task. This could account for the similarity of

[1] The RN has been observed for exogenous perceptual reversals (i.e., a switch from one unambiguous variant of the stimulus to another) with 50% probability of reversal (Kornmeier, et al., 2004). This indicates that the RN can be observed when reversals are not rare. However, it is unclear whether this applies to the endogenous reversals which are the topic of this study.

our RN results between the identity and reversal tasks. However, we would argue that the rarity of reversals was less salient in the identity task than in the reversal task because the task did not explicitly require monitoring for them. One way to address this issue would be to test whether the RN/RP differs depending on reversal rate. Unfortunately, we did not have sufficient data to compute the RN/RP for different reversal rates across our participants, but future research should explore this. Reversal and stable trials also differed in their response actions in the identity task. Specifically, reversal trials in the identity task always required a change in response action relative to the previous trial whereas stable trials had the same response action as the preceding trial[2]. Thus, a reversal trial in the identity task necessarily involved a task-relevant change in motor behaviour that was not required for stable trials. This was not the case for the reversal task. Reversal and stable trials could require a response action change or no change depending on the preceding trial type[3]. Therefore, the fact that the RN does not differ between the reversal and identity tasks suggests that it is not sensitive to the variations in response actions between the reversal and identity tasks.

In line with the results of Experiment 1, we found that the RP in Experiment 2 was sensitive to response action. Specifically, the RP was present, as normal, on trials in which participants responded due to task instructions. In contrast, non-response trials showed a significant difference between reversal and stable trials in the RP time window but with an opposite polarity to the RP present on response trials (i.e., stable > reversal). This result in the non-response condition of Experiments 2 is different from that observed in Experiment 1

[2] For instance, in a respond-faces identity task block, a face percept followed by a vase percept (i.e., the second trial being a reversal trial) would require a response on the first trial and then withholding response on the subsequent trial (i.e., a change in response action). In contrast, a face percept followed by a face percept (i.e., the second trial being a stable trial) would require a response on both trials (i.e., no change in response action)

[3] For instance, in a respond-reversal reversal task block, a reversal trial preceded by another reversal trial would require the same response twice in a row (i.e., no change in response action) whereas a reversal trial preceded by a stable trial would require no response and then a response (i.e., a change in response action). Similarly, a stable trial preceded by another stable trial would require the same action whereas a stable trial preceded by a reversal trial would require a change in response action.

where there was no significant difference in the RP time window for non-response trials. We conducted an overall analysis combining both datasets from the two experiments to test whether this discrepancy would be reflected in a Trial Type x Response Action x Experiment interaction, but it was not. We cannot account for this minor discrepancy between the RP results in Experiments 1 and 2. It could be that this arises from slight processing or time course differences between the faces-vase figure and the Necker lattice. Nonetheless, it is clear that the RP is sensitive to response action instructions. In contrast to the RP, we found across two studies that the RN is insensitive to response action as has been suggested by previous studies (e.g., Kornmeier & Bach, 2004). However, through visual inspection of the Experiment 2 results, the RP difference (reversal > stability) in response trials appears to extend beyond its 100-200 ms window into the subsequent RN time window (e.g., Figure 6C & 6G). This was not the case in Experiment 1 (Figures 3C & 3G). Because the RP is opposite in polarity to the RN, any extension of it into the RN timeframe may partially explain why the RN was less reliable in Experiment 2 (i.e., only significant in the identity task) compared to Experiment 1. To our knowledge, previous studies of the RP have not used the faces-vase stimulus or compared RP amplitude across multiple stimuli in the same study. Thus, it is not clear whether this possible modulation of the RN by the RP is due to noise in our data or a systematic inter-stimulus difference in the RP duration. Further work will be needed to evaluate this possibility.

Our observed difference in the RP between response and non-response trials cannot be attributed to the extended ITI (from 400 to 1000 ms; see methods) after responses in our procedure. Responses occurred, on average, approximately 700 ms after stimulus onset and the extension of the ITI occurred only after the response occurred. Thus, the extended ITI occurred well after the RP on these trials was already over. This means that the extended ITI

and response-related brain activity may have affected the subsequent trial but not retrospectively affected the response trial itself.

We based our response style design (i.e., in some blocks, respond only when a reversal is perceived and, in other blocks, respond only when stability is perceived) on previous studies (e.g., Kornmeier & Bach, 2004; 2005). By counterbalancing (across blocks) whether reversal or stability was the target for response, these studies ensured that the reversal vs. stable trial comparison was not confounded with response target status as would have been the case if one had adopted a single response style across the experiment (e.g., respond only to reversal). Nonetheless, some of those studies did not statistically assess whether response style affected the results. Here, we showed that response does not affect the RN. In contrast though, averaging over response is not appropriate for studying the RP. Averaging over response and non-response trials eliminated the RP in our results (i.e., no main effects of trial type). This effect of response, in addition to the RP's relatively short duration, may account for why some researchers have found this component difficult to observe.

The measurement of the RP and RN in each of our conditions depends on participants accurately responding in line with their subjective perceptual experience on each trial and following response rules that varied from one block to another. Failure to respond accurately, due to inattention, misunderstanding or another reason, could lead to misclassification of reversal trials as stable and vice versa. In turn, this would affect the RN and RP components. For instance, in the reversal task, if participants reversed their response mapping on a large proportion of trials, this could eliminate or reverse the polarity of the RN and RP

components[4]. Could such response mistakes account for the modulation of the RP in our results? We believe that this is unlikely for two reasons. First, any systematic response mistakes that modulated the RP would have also affected the RN amplitude because both components are derived from the same waveform in each condition. However, we saw no corresponding modulations of the RN by response (Trial Type x Response Interaction: Exp. 1, $p = .798$; Exp. 2, $p = .437$; see Supplemental Materials). Second, if participants were mixing up reversal and stability trial labels we would expect that, on average, the 30/70 balance of reversal/stability trials should reverse to 70/30 (or at least move in that direction) in the no response trials compared to the response trials. We saw no suggestion of this effect in the behavioural data for ambiguous stimuli (see section 2.2.3 and 3.2.3).

It is worth noting that there are some differences between our study design and previously published studies of the RN and RP. First, many studies have used a paired-stimulus paradigm in which two ambiguous figures are presented in quick succession and a reversal/stable decision is made only after the second stimulus (e.g., Intaitė, et al., 2010). This differs from our paradigm in which every single stimulus was judged against the preceding one in a continuous sequence. This was done because it was more time efficient and thus allowed us to address the task and response factors using a within-subjects design and within a single testing session. We see no reason why our observed effect of response on the RP would not generalise across these two versions of the paradigm. Second, our within-subjects design involved task and response style switching between the blocks and may have highlighted task and response factors to participants. Nonetheless, aside from the RP modulation by response, we observed similar RN and RP effects to previous studies which maintained a single task throughout (e.g., Kornmeier & Bach, 2004; 2005, Pitts et al., 2009).

[4] Response mistakes in the identity task would never reverse the sign of the RP component. If you reverse all of the trial labels (e.g., wrong button assignment), then the reversals would actually all still be in the same position in the sequence.

Further work using a between-subjects design will be needed to directly address the effect of task switching between blocks on the RP and RN. A between-subjects design would also allow more trials per condition and allow greater signal-to-noise ratio in individual conditions. Finally, some previous studies have used peak amplitudes to quantify the RN and RP amplitudes (e.g., Kornmeier & Bach, 2004) whereas others, including the present work, have used average amplitudes across an ROI window (e.g., Pitts, Nerger, & Davis, 2007). Our choice was based on the ERP methods literature which cautions against use of peak amplitude and suggests that average amplitude over a window is more robust to noise than peak measures (e.g., Clayson, Baldwin, & Larson, 2013). Because of this, we did not do a peak amplitude analysis but this difference needs to be considered when comparing our data to other studies.

Our results challenge any interpretation of the RP as purely related to the perceptual processing responsible for reversals. Otherwise, it should occur regardless of response action given that the perceptual processes related to reversal must occur, in some form, regardless of the task and response action required of the participant. However, it is possible that these perceptual processes are different between response and non-response trials and that the observed differences in the RP reflect this. In this case, the RP would still reflect perceptual processes involved in generating reversals but there presumably would then be another ERP correlate of the perceptual processes generating reversals on non-response trials. It is also possible that the RP effect that we observed is related to attention which has been shown previously to modulate the amplitude of early perceptual ERP components (for reviews see: Hillyard, Vogel, & Luck, 1998; Luck, Woodman, & Vogel, 2000). On non-response reversal trials (i.e., respond stable), reversals arguably fall outside of the attentional focus because they are less task relevant. Although the perceptual reversal occurs, the RP reversal-related perceptual signal may be attenuated in line with its task relevance. Because of the relatively

small amplitude of the RP, this may have made the RP unobservable in our study given our sample size and trial numbers. Although possible, this account is not congruent with the polarity reversal of the RP component that we observed in Experiment 2. Further work will be necessary to confirm whether this polarity reversal is robust and whether the effect can be attributed to attention.

Alternatively, the RP may arise from response-dependent, non-perceptual processes that are not directly involved in generating the reversals themselves but instead only co-occur with reversal processes. Other researchers have suggested that the RP may be related to perceptual ambiguity detection or perceptual decision conflict by the visual system (e.g., Kornmeier & Bach, 2012; Kornmeier, Pfäffle, & Bach, 2011). As visual input of an ambiguous stimulus is received on each trial, this is combined with any internal noise/bias and compared against the perceptual expectation set up by the interpretation of the preceding stimulus. On stable trials, sensory input and noise is presumably consistent with the preceding interpretation, perceptual ambiguity is low, and no decision conflict arises. In contrast, on reversal trials, the visual system detects a conflict between the state of perceptual affairs for the new stimulus and the expectation established by the preceding stimulus interpretation. This signals a state of ambiguity and a potentially different perceptual decision from the preceding stimulus. If this is an accurate description of the source of the RP, then our results suggest that this ambiguity/conflict signal is modulated by whether the trial requires a response or not. It is not clear why this would be the case.

It is not possible to tease apart the perceptual and non-perceptual accounts above based on our data. Nonetheless, given that the response-dependence of the RP has not previously been demonstrated, our work provides a clear foundation for further investigation of this issue. Furthermore, our results demonstrate that response arrangements must be considered carefully when studying the RP. In particular, collapsing across response and non-

response can hide a significant source of variability in the data even in relatively early latency ERP components.

## References

Anllo-Vento, L. & Hillyard, S.A. (1996). Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics, 58(2)*, 191-206. https://doi.org/10.3758/BF03211875

Basar-Eroglu, C., Strüber, D., Stadler, M., and Kruse,E. (1993). Multistable visual perception induces a slow positive EEG wave. *International Journal of Neuroscience, 73*, 139–151. https://doi.org/10.3109/00207459308987220

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13. https://doi.org/10.1038/nrn701

Brascamp, J., Sterzer, P., Blake, R., & Knapen, T. (2018). Multistable perception and the role of the frontoparietal cortex in perceptual inference. *Annual review of psychology, 69*, 77-103. https://doi.org/10.1146/annurev-psych-010417-085944

Britz, J., Landis, T., & Michel, C. (2009). Right Parietal Brain Activity Precedes Perceptual Alternation of Bistable Stimuli. *Cerebral Cortex, 19(1)*, 55-65. https://doi.org/10.1093/cercor/bhn056

Britz, J., & Pitts, M. A. (2011). Perceptual reversals during binocular rivalry: ERP components and their concomitant source differences. *Psychophysiology, 48*(11), 1490-1499. https://doi.org/10.1111/j.1469-8986.2011.01222.x

Brooks, J. L., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating Type I error rate. *Psychophysiology*, 54(1), 100– 113. https://doi.org/10.1111/psyp.12682

Cao, R., Braun, J., & Mattia, M. (2014). Stochastic Accumulation by Cortical Columns May Explain the Scalar Property of Multistable Perception. *Physical Review Letters, 113(9)*. https://doi.org/10.1103/physrevlett.113.098103

Castro, L., Lazareva, O.F., Vecera, S.P., Wasserman, E.A. (2010). Changes in area affect figure-ground assignment in pigeons. *Vision Research*, *50*, 497–508. https://doi.org/10.1016/j.visres.2009.12.016

Clayson, P. E., Baldwin, S. A., & Larson, M. J. (2013). How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. *Psychophysiology*, 50, 174-186. https://doi.org/10.1111/psyp.12001

Cohen, E.H., Barenholtz, E., Singh, M., Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision, 5(4)*, 3. https://doi.org/10.1167/5.4.3

Courchesne, E., Hillyard, S.A., Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology, 39(2)*, 131-143. https://doi.org/10.1016/0013-4694(75)90003-6

Harrower, M. R. (1936). Some Factors Determining Figure-Ground Articulation. *British Journal of Psychology. General Section, 26*, 407-424. https://doi.org/10.1111/j.2044-8295.1936.tb00806.x

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences, 105*(31), 10984-10989. https://doi.org/10.1073/pnas.0712043105

Hillyard, S.A., Vogel, E.K., & Luck, S.J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society: Biological Sciences*, 33, 1257-1270. https://doi.org/10.1098/rstb.1998.0281

Intaitė, M., Koivisto, M., Rukšėnas, O., & Revonsuo, A. (2010). Reversal negativity and bistable stimuli: Attention, awareness, or something else? *Brain and Cognition, 74*(1), 24-34. https://doi.org/10.1016/j.bandc.2010.06.002

Isoglu-Alkaç, Ü., Başar-Eroğlu, C., Ademoglu, A., Demiralp T., Miener M., Stadler M. (1998) Analysis of the electroencephalographic activity during the Necker cube reversals by means of the wavelet transform. *Biological Cybernetics, 79*, 437–42. https://doi.org/10.1007/s004220050492

İşoğlu-Alkaç, Ü., Başar-Eroğlu, C., Ademoğlu, A., Demiralp, T., Miener, M., & Stadler, M. (2000). Alpha activity decreases during the perception of Necker cube reversals: an application of wavelet transform. *Biological Cybernetics, 82*(4), 313-320. https://doi.org/10.1007/s004220050585

İşoğlu-Alkaç, Ü. & Strüber, D. (2006). Necker cube reversals during long-term EEG recordings: sub-bands of alpha activity. *International Journal of Psychophysiology, 59*, 179–189. https://doi.org/10.1016/j.ijpsycho.2005.05.002

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*, 124(10), 2062– 2063. https://doi.org/10.1016/j.clinph.2013.03.024

Kornmeier, J., & Bach, M. (2004). Early neural activity in Necker-cube reversal: Evidence for low-level processing of a gestalt phenomenon. *Psychophysiology, 41*(1), 1-8. https://doi.org/10.1046/j.1469-8986.2003.00126.x

Kornmeier, J., & Bach, M. (2005). The Necker cube—an ambiguous figure disambiguated in early visual processing. *Vision research, 45*(8), 955-960. https://doi.org/10.1016/j.visres.2004.10.006

Kornmeier, J., & Bach, M. (2006). Bistable perception—along the processing chain from ambiguous visual input to a stable percept. *International Journal of Psychophysiology, 62*(2), 345-349. https://doi.org/10.1016/j.ijpsycho.2006.04.007

Kornmeier, J., & Bach, M. (2012). Ambiguous figures–what happens in the brain when perception changes but not the stimulus. *Frontiers in Human Neuroscience, 6*, 51. https://doi.org/10.3389/fnhum.2012.00051

Kornmeier, J., & Bach, M. (2014). EEG correlates of perceptual reversals in Boring's ambiguous old/young woman stimulus. *Perception, 43(9),* 950-962. https://doi.org/10.1068/p7741

Kornmeier, J., Ehm, W., Bigalke, H., & Bach, M. (2007). Discontinuous presentation of ambiguous figures: How interstimulus-interval durations affect reversal dynamics and ERPs. *Psychophysiology, 44*(4), 552-560. https://doi.org/10.1111/j.1469-8986.2007.00525.x

Kornmeier, J., Hein, C., & Bach, M. (2009). Multistable perception: When bottom-up and top-down coincide. *Brain and Cognition, 69(1)*, 138-147. https://doi.org/10.1016/j.bandc.2008.06.005

Kornmeier, J., Heinrich, S.P., Atmanspacher, H., & Bach, M. (2001). The reversing "Necker Wall" – a new paradigm with reversal entrainment reveals an early EEG correlate. In: ARVO 2001 Annual Meeting. Investigative Ophthalmology & Visual Science (ARVO Supplement). p. 409.

Kornmeier, J., Pfäffle, M., & Bach, M. (2011). Necker cube: Stimulus-related (low-level) and percept-related (high-level) EEG signatures early in occipital cortex. *Journal of vision, 11*(9), 12-12. https://doi.org/10.1167/11.9.12

Leopold, D., & Logothetis, N. (1999). Multistable phenomena: changing views in perception. Trends in Cognitive Sciences, 3(7), 254-264. https://doi.org/10.1016/s1364-6613(99)01332-7

Luck, S.J., Woodman, G.F., & Vogel, E.K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4, 432-440. https://doi.org/10.1016/S1364-6613(00)01545-X

Lumer, E. D., & Rees, G. (1999). Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proceedings of the National Academy of Sciences, 96*(4), 1669-1673. https://doi.org/10.1073/pnas.96.4.1669

Mathes, B., Strüber, D., Stadler, M. A., & Basar-Eroglu, C. (2006). Voluntary control of Necker cube reversals modulates the EEG delta-and gamma-band response. *Neuroscience Letters, 402*(1-2), 145-149. https://doi.org/10.1016/j.neulet.2006.03.063

Nuwer, M.R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J.-M., Hinrichs, H., Ikeda, A., Fransisco, J.C., Rappelsburger, P. (1998). I.F.C.N. standards for digital recording of clinical EEG. *Electroencephalography & Clinical Neurophysiology*, 106, 259-261. https://doi.org/10.1016/S0013-4694(97)00106-5

O'Donnell, B. F., Hendler, T., Squires, N. K. (1988). Visual evoked potentials to illusory reversals of the Necker cube. *Psychophysiology, 25*, 137–143. https://doi.org/10.1111/j.1469-8986.1988.tb00976.x

Pazo-Álvarez, P., Roca-Fernández, A., Gutiérrez-Domínguez, F. J., & Amenedo, E. (2017). Attentional Modulation of Change Detection ERP Components by Peripheral Retro-Cueing. *Frontiers in human neuroscience*, *11*, 76. https://doi.org/10.3389/fnhum.2017.00076

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203. 10.3758/s13428-018-01193-y

Pitts, M. A., & Britz, J. (2011). Insights from intermittent binocular rivalry and EEG. *Frontiers in Human Neuroscience, 5*, 107. https://doi.org/10.3389/fnhum.2011.00107

Pitts, M. A., Gavin, W. J., & Nerger, J. L. (2008). Early top-down influences on bistable perception revealed by event-related potentials. *Brain and Cognition, 67*(1), 11-24. https://doi.org/10.1016/j.bandc.2007.10.004

Pitts, M.A., Martínez, A., Stalmaster, C., Nerger, J.L. and Hillyard, S.A. (2009), Neural generators of ERPs linked with Necker cube reversals. *Psychophysiology*, 46: 694-702. https://10.1111/j.1469-8986.2009.00822.x

Pitts, M. A., Martínez, A., & Hillyard, S. A. (2010). When and where is binocular rivalry resolved in the visual cortex? *Journal of Vision*, 10(14), 25. https://doi.org/10.1167/10.14.25

Pitts, M., Nerger, J., & Davis, T. (2007). Electrophysiological correlates of perceptual reversals for three different types of multistable images. *Journal Of Vision, 7(1)*, 6. https://doi.org/10.1167/7.1.6

Potts, G.F. (2004). An ERP index of task relevance evaluation of visual stimuli. Brain and *Cognition, 56(1)*, 5-13. https://doi.org/10.1016/j.bandc.2004.03.006.

Rassi, E., Wutz, A., Müller-Voggel, N., Weisz, N. (2019) Prestimulus feedback connectivity biases the content of visual experiences. *Proceedings of the National Academy of Sciences*, 116 (32), 16056-16061. https://10.1073/pnas.1817317116

Rensink, R.A. (2002). Change Detection. *Annual Review of Psychology*, 53, 245-277. https://doi.org/10.1146/annurev.psych.53.100901.135125

Rubin, E. (1958/1915). Figure and ground. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194–203). Princeton, NJ: Van Nostrand.

Schiller, P. V. (1933). Stroboskopische Alternativversuche. *Psychologische Forschung, 17*, 179–214.

Schwartz, J., Grimault, N., Hupe, J., Moore, B., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 367(1591), 896-905. https://doi.org/10.1098/rstb.2011.0254

Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., He, B., Beniczky, S. (2017). The standardized EEG electrode array of the IFCN. Clinical Neurophysiology, 128(10), 2070-2077. https://doi.org/10.1016/j.clinph.2017.06.254

Strüber, D., Basar-Eroglu, C., Miener, M., & Stadler, M. (2001). EEG gamma-band response during the perception of Necker cube reversals. *Visual Cognition, 8*(3-5), 609-621. https://doi.org/10.1080/13506280143000151

Strüber, D., & Herrmann, C. S. (2002). MEG alpha activity decrease reflects destabilization of multistable percepts. *Cognitive Brain Research, 14*(3), 370-382. https://doi.org/10.1016/S0926-6410(02)00139-8

Sundareswara, R.1. & Schrater, P.R. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of Vision, 8*, 12. https://doi.org/10.1167/8.5.12

Troje, N. F., & McAdam, M. (2010). The Viewing-from-Above Bias and the Silhouette Illusion. *I-Perception*, 143–148. https://doi.org/10.1068/i0408

## Author Note

Address of corresponding author. Joseph L. Brooks, School of Psychology, Dorothy Hodgkin Building, Keele University, Keele, ST5 5BG, United Kingdom. Email: J.L.Brooks@keele.ac.uk. We thank Mitchell Wakeman, Melissa McDermott, Tigan Schofield, Zahra Syed, and Josephine Greenhill for assistance in data collection, Frank Gasking and John Allen for PsychoPy support, and Kim Dundas for theoretical discussions during lab meeting. DA was partially supported by a Postgraduate Research Bursary from the University of Kent while conducting this project and JLB was supported by a Research Development Award from the Faculty of Natural Sciences at Keele University.

Figure Captions

*Figure 1.* Examples of the Necker lattice (A) based on Kornmeier & Bach (2004) and Rubin's Faces-Vase (D) ambiguous figures used in Experiments 1 and 2, respectively. (B) A biased version of the Necker lattice which is typically perceived with its front face towards the upper left. (C) A biased Necker lattice with its front face towards the lower right. (E) A modified faces-vase image biased towards the face interpretation. (F) A faces-vase image biased towards the vase interpretation.

Figure 2. (A) A sequence of three ambiguous stimulus trials in Experiment 1. Intermittent Necker Lattices were presented in a continuous sequence for 800 ms each and separated by an ITI fixation cross of either 400 ms, if no response was made to the stimulus within 1200 ms of stimulus onset, or 1000 ms if a response was made with 1200 ms of stimulus onset. (B) Each row (black background) of this table shows a sequence of correct responses (respond or not) for three stimuli given the task and response style condition (different rows; indicated in column 2) and the participant's subjective perception of the ambiguous stimulus (different columns; indicated in row 2). In reversal task trials, the participant's response depended on their perception of both a given stimulus and the one that precedes it. (C) Mean reaction times as a function of task (identity vs. reversal) and trial type (reversal vs. stability) in Experiment 1.

Figure 3. Each panel plots the grand mean ERP waveforms for reversal (solid line) and stability trials (dashed line) in Experiment 1 (Necker Lattice) under different response and task conditions. Panels A-D are averaged over the RN ROI (O1, Oz, O2, PO7, PO8) and panels E-H are averaged over the RP ROI (O1, Oz, O2). (A & E) ERP waveforms in the reversal task (collapsed over response); (B & F) ERP waveforms in the identity task

(collapsed over response); (C & G) ERP waveforms for the response trials (collapsed over task); (D & H) ERP waveforms for the non-response trials (collapsed over task). The boxes indicate the ROI time-windows used to quantify the RN (dashed outline, light grey fill) and RP (solid outline, dark grey fill) ERP component mean amplitudes (see Data Analysis Methods section). Figure 4 shows ROI mean amplitudes. For each plot in this figure, corresponding difference wave plots, scalp maps, and ERPs for all channels are available in the Supplemental Materials.

Figure 4. ERP results for the RN and RP for Experiment 1 (Necker Lattice). A significant RN or RP in each condition (i.e., difference between black and grey bars) is indicated with ** ($p<.01$) or * ($p<.05$) above/below the bars. Error bars represent the standard error of the mean. (A) Mean amplitudes (with SEMs) in the RN ROI (200-400 ms; O1, O2, Oz, PO7, PO8 average) for Reversal Trials (in black) and Stability Trials (in grey) separately for the identity task and reversal task. (B) Mean reversal and stability ERP amplitudes within the RP ROI (100-200 ms; O1, O2, Oz average) is shown separately for Response Trials and Non-Response trials to illustrate the significant Response Action x Trial Type interaction.

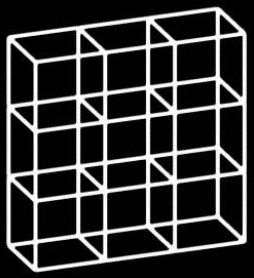
Figure 5. (A) The ambiguous faces-vase stimulus used in Experiment 2. The frame width (FW) and inter-edge distance (IED) were adjusted for each participant in a pre-test participant to determine the maximally ambiguous stimulus for use in the main experiment. (B) Mean reaction times as a function of task (identity vs. reversal) and trial type (reversal vs. stability) in Experiment 2.

Figure 6. Each panel plots the grand mean ERP waveforms for reversal (solid line) and stability trials (dashed line) in Experiment 2 (Faces-Vase) under different response and task

conditions. Panels A-D are averaged over the RN ROI (O1, Oz, O2, PO7, PO8) and panels E-H are averaged over the RP ROI (O1, Oz, O2). (A & E) ERP waveforms in the reversal task (collapsed over response); (B & F) ERP waveforms in the identity task (collapsed over response); (C & G) ERP waveforms for the response trials (collapsed over task); (D & H) ERP waveforms for the non-response trials (collapsed over task). The boxes indicate the ROI time-windows used to quantify the RN (dashed outline, light grey fill) and RP (solid outline, dark grey fill) ERP component mean amplitudes (see Data Analysis Methods section). Figure 7 shows ROI mean amplitudes. For each plot in this figure, corresponding difference wave plots, scalp maps, and ERPs for all channels are available in the Supplemental Materials.

Figure 7. ERP results for RN and RP for Experiment 2 (Faces-Vase). A significant RN or RP in each condition (i.e., difference between black and grey bars) is indicated with ** ($p<.01$) or * ($p<.05$) above the bars. Error bars represent the standard error of the mean. (A) Mean amplitudes (with SEMs) in the RN ROI (200-400 ms; O1, O2, Oz, PO7, PO8 average) for Reversal Trials (in black) and Stability Trials (in grey) separately for the identity task and reversal task. (B) Mean reversal and stability ERP amplitudes within the RP ROI (100-200 ms; O1, O2, Oz average) is shown separately for Response Trials and Non-Response trials to illustrate the significant Response Action x Trial Type interaction.

(A)



Ambiguous Necker Lattice

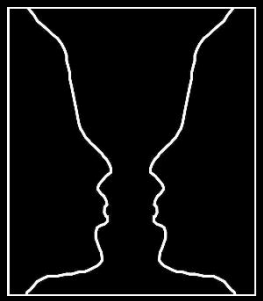
(B)

Biased Upper Left Facing

(C)

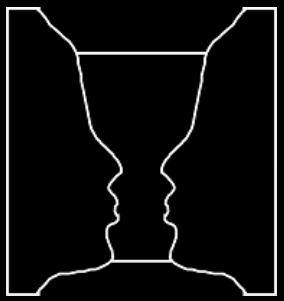
Biased Lower Right Facing

(D)



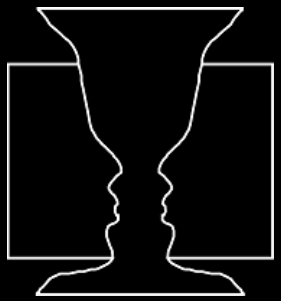
Ambiguous Faces-Vase

(E)



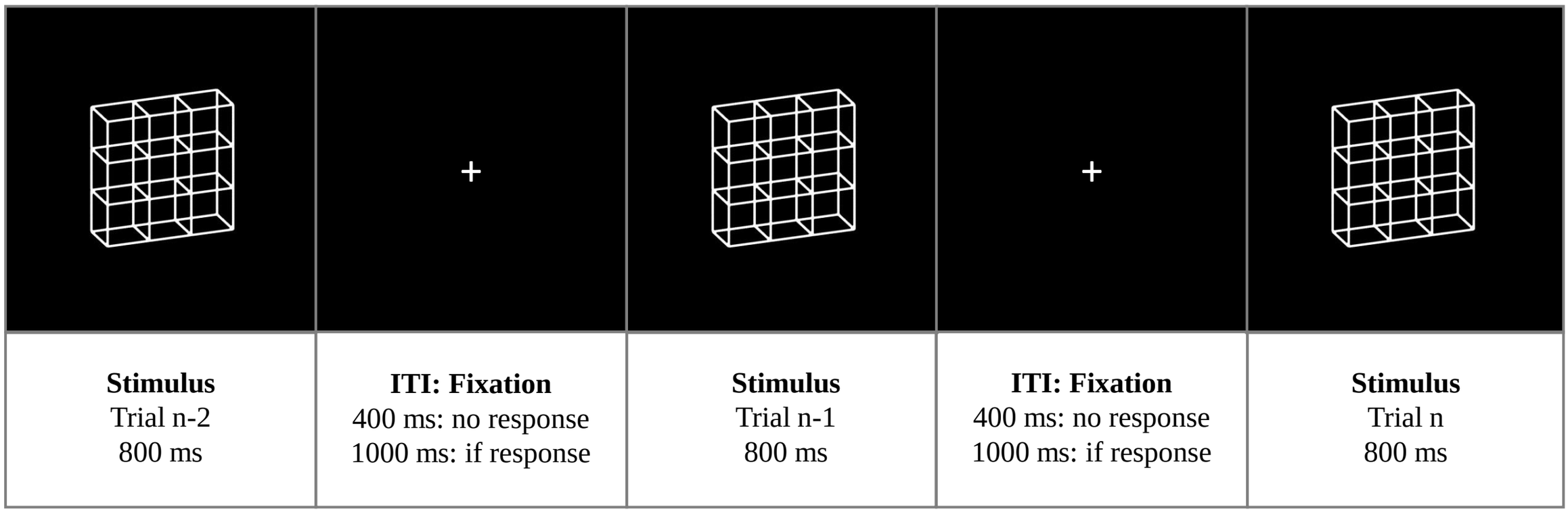
Face Biased

(F)



Vase Biased

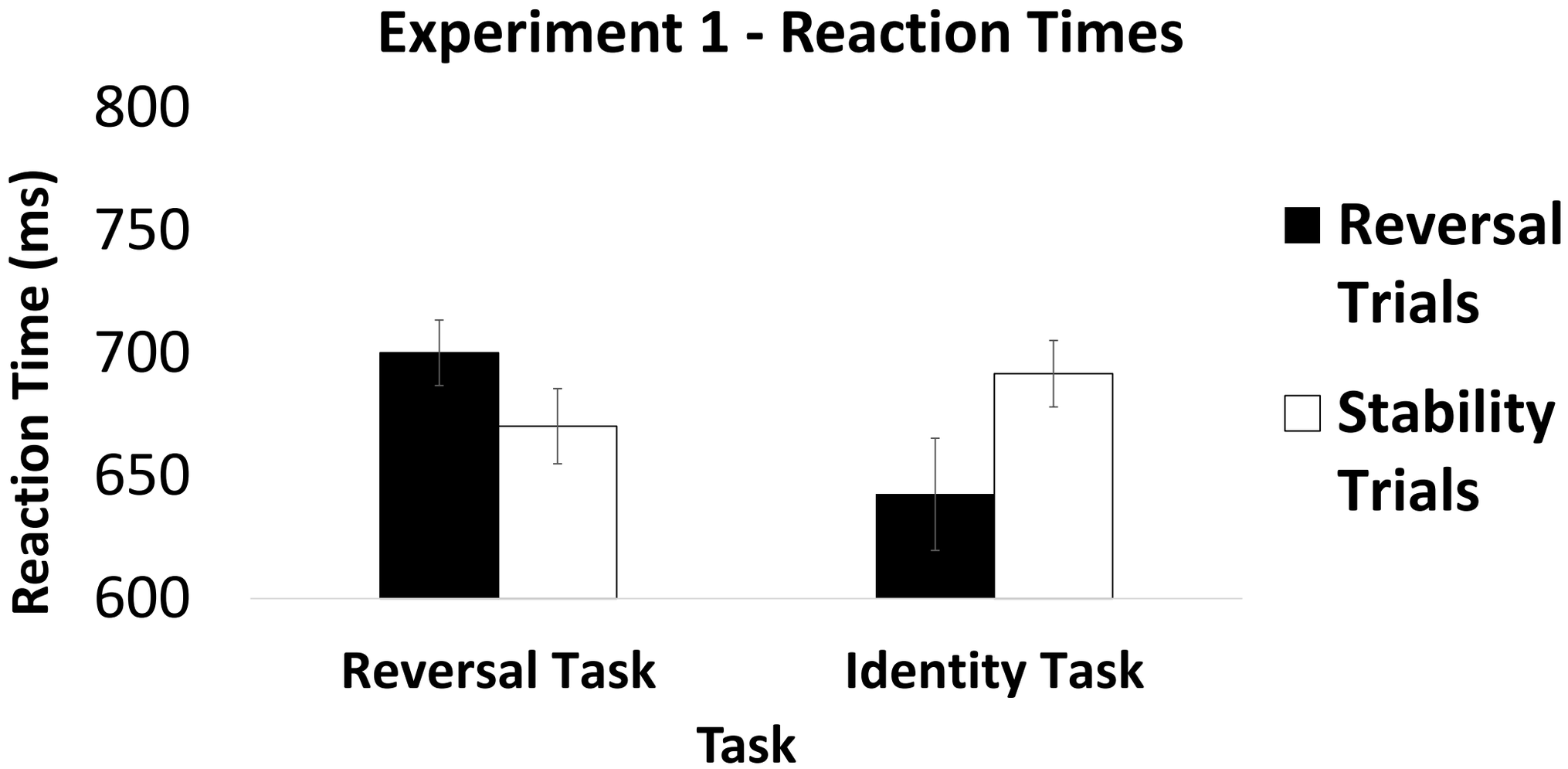
(A)



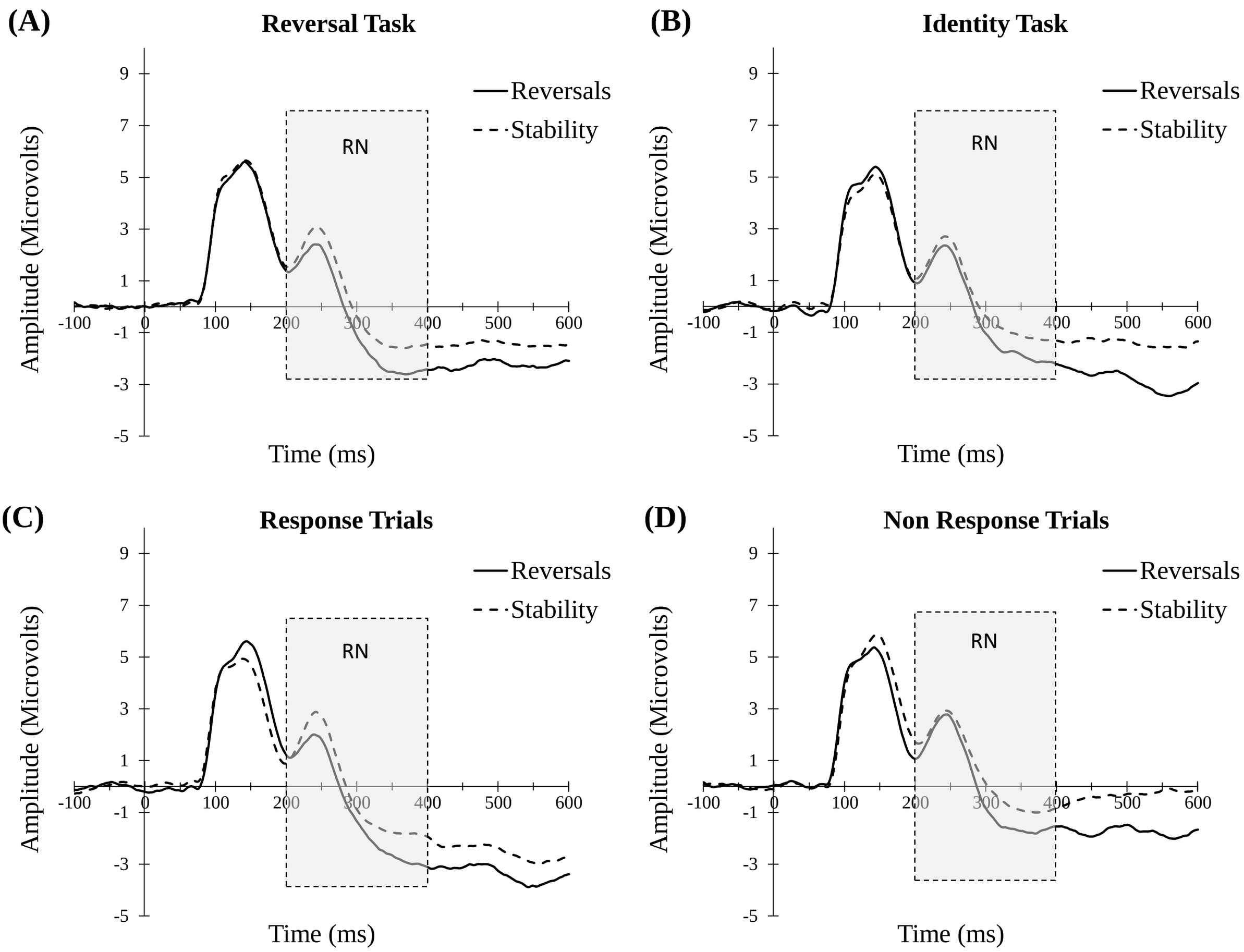

(B)

| Task & Response Style | Trial Number | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| | Participant's Subjective Perception | *Right-Facing* | *Right-Facing* | *Left-Facing* | *Left-Facing* |
| | Reversal Task Respond Reversal | *No Response* | *No Response* | *Respond* | *No Response* |
| | Reversal Task Respond Stable | *No Response* | *Respond* | *No Response* | *No Response* |
| | Identity Task Respond Left-Facing | *No Response* | *No Response* | *Respond* | *Respond* |
| | Identity Task Respond Right-Facing | *Respond* | *Respond* | *No Response* | *No Response* |

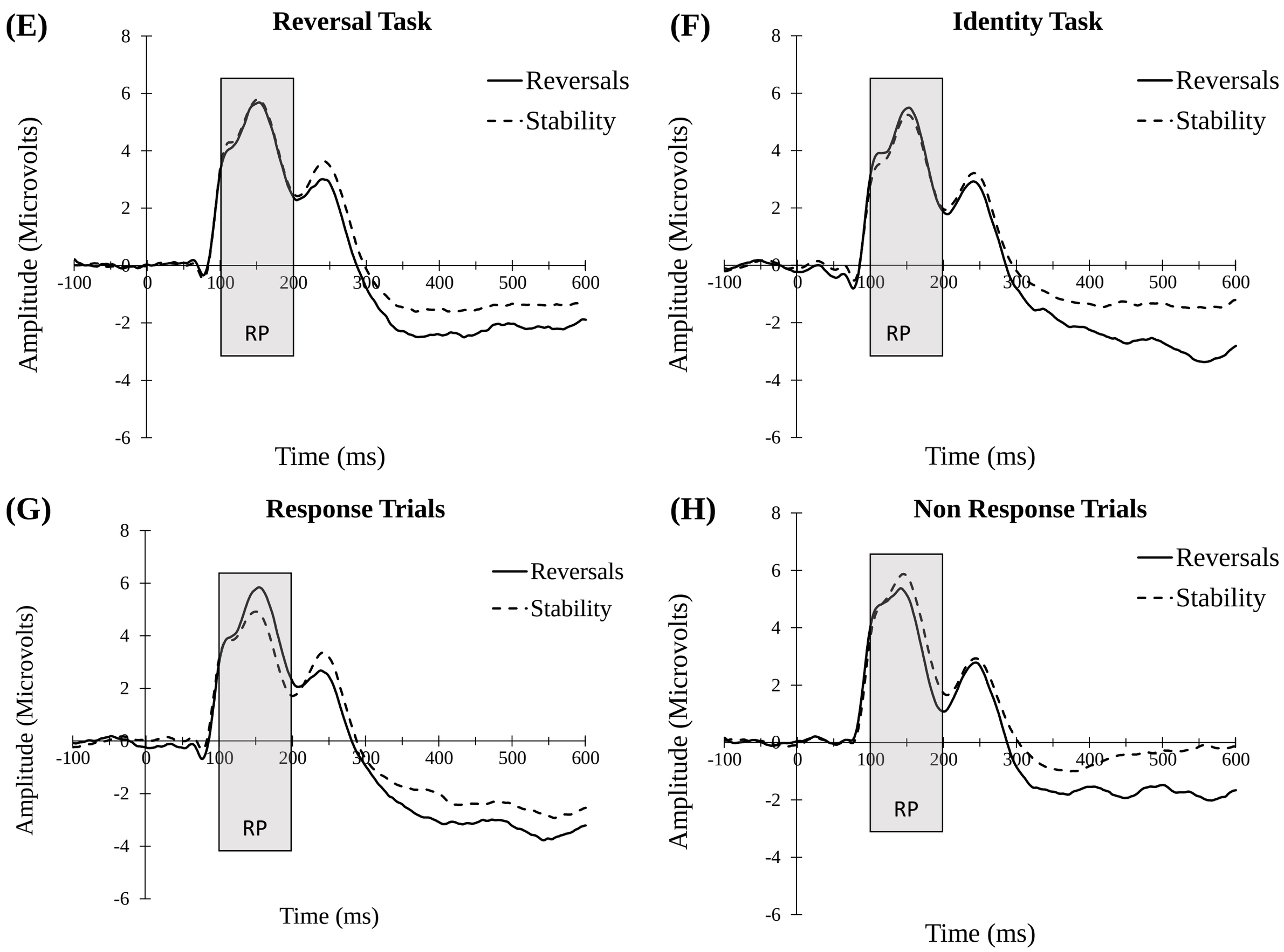
(C)



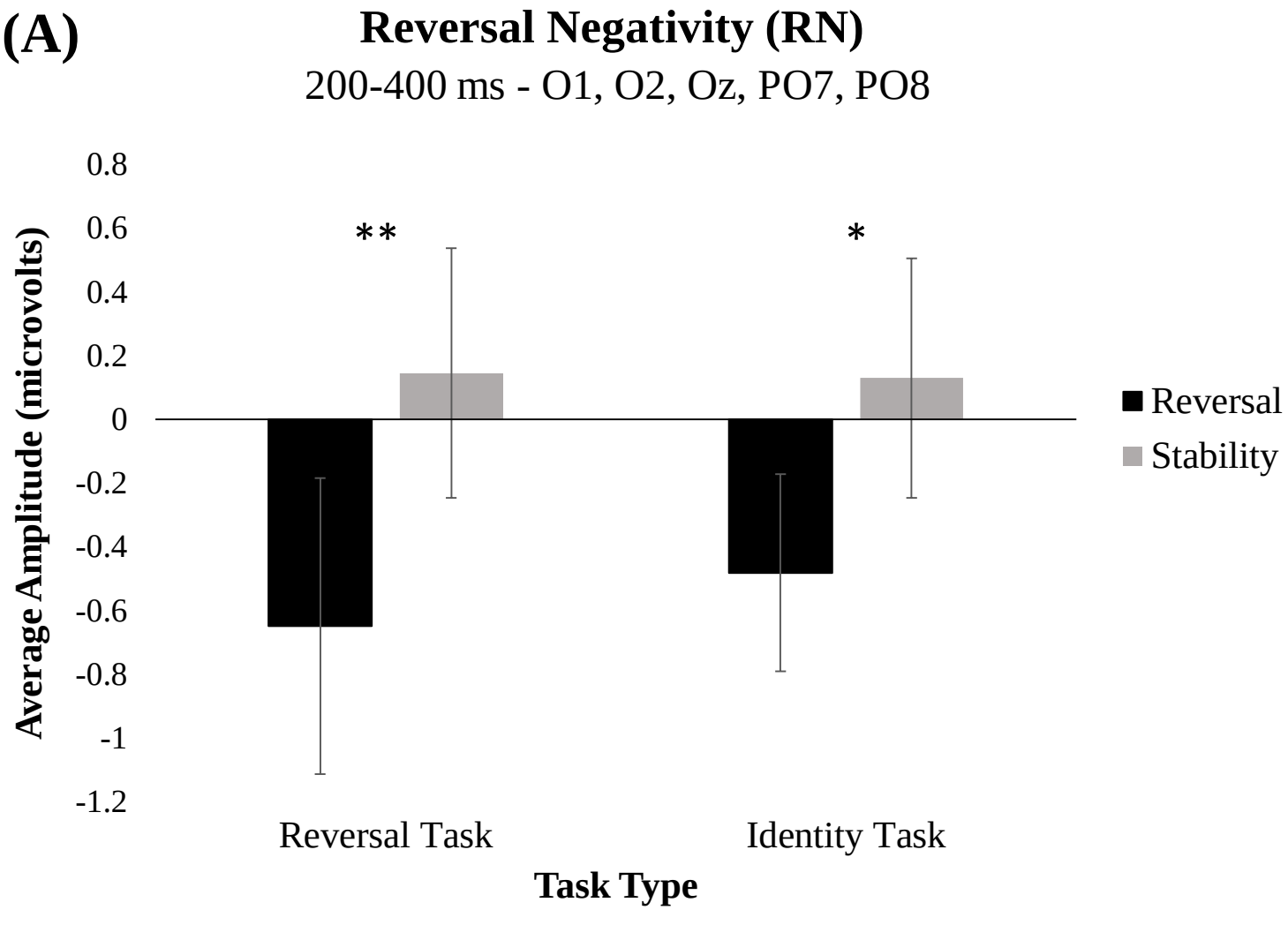
# RN ROI (O1, Oz, O2, PO7, PO8)



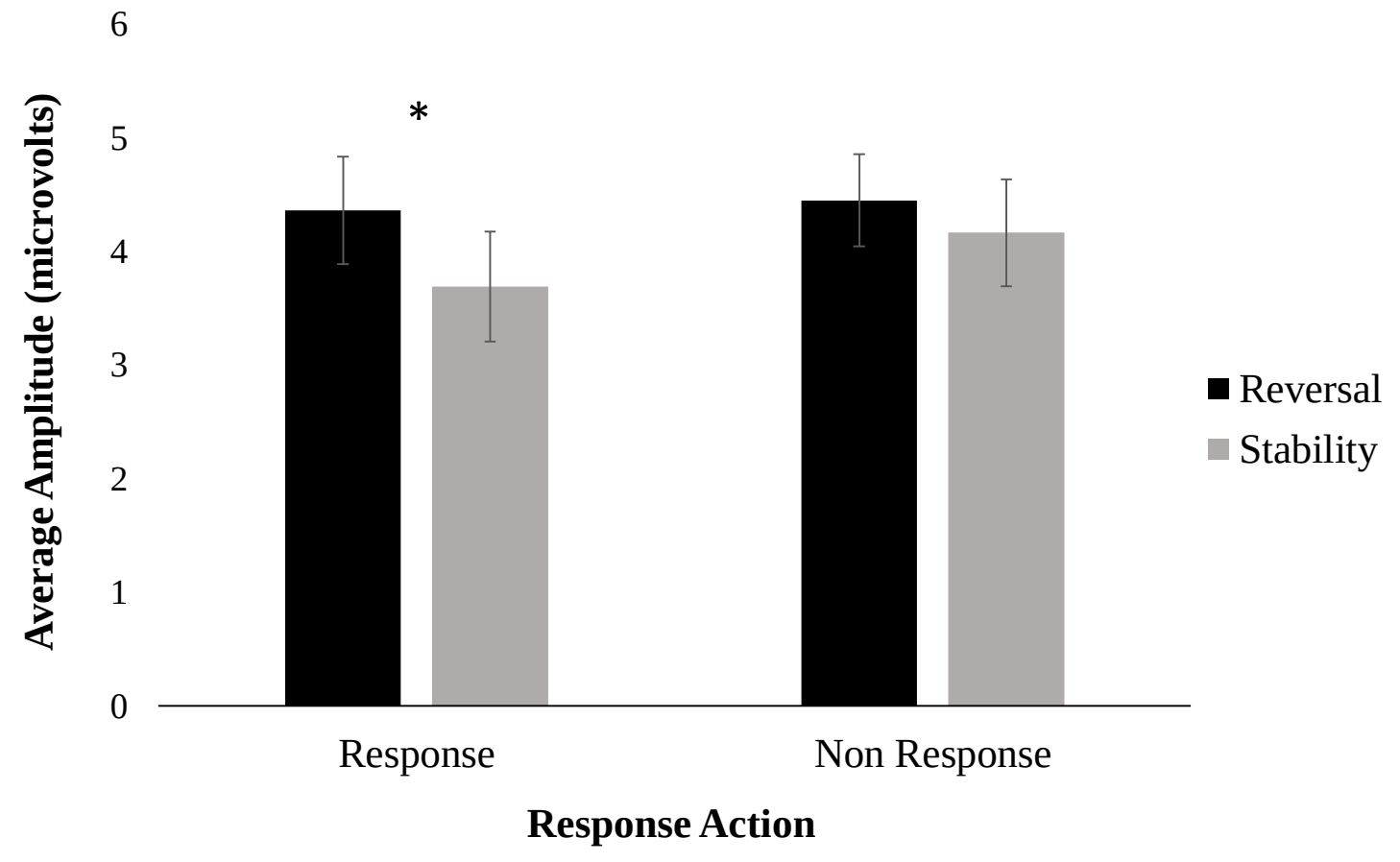

# RP ROI (O1, Oz, O2)



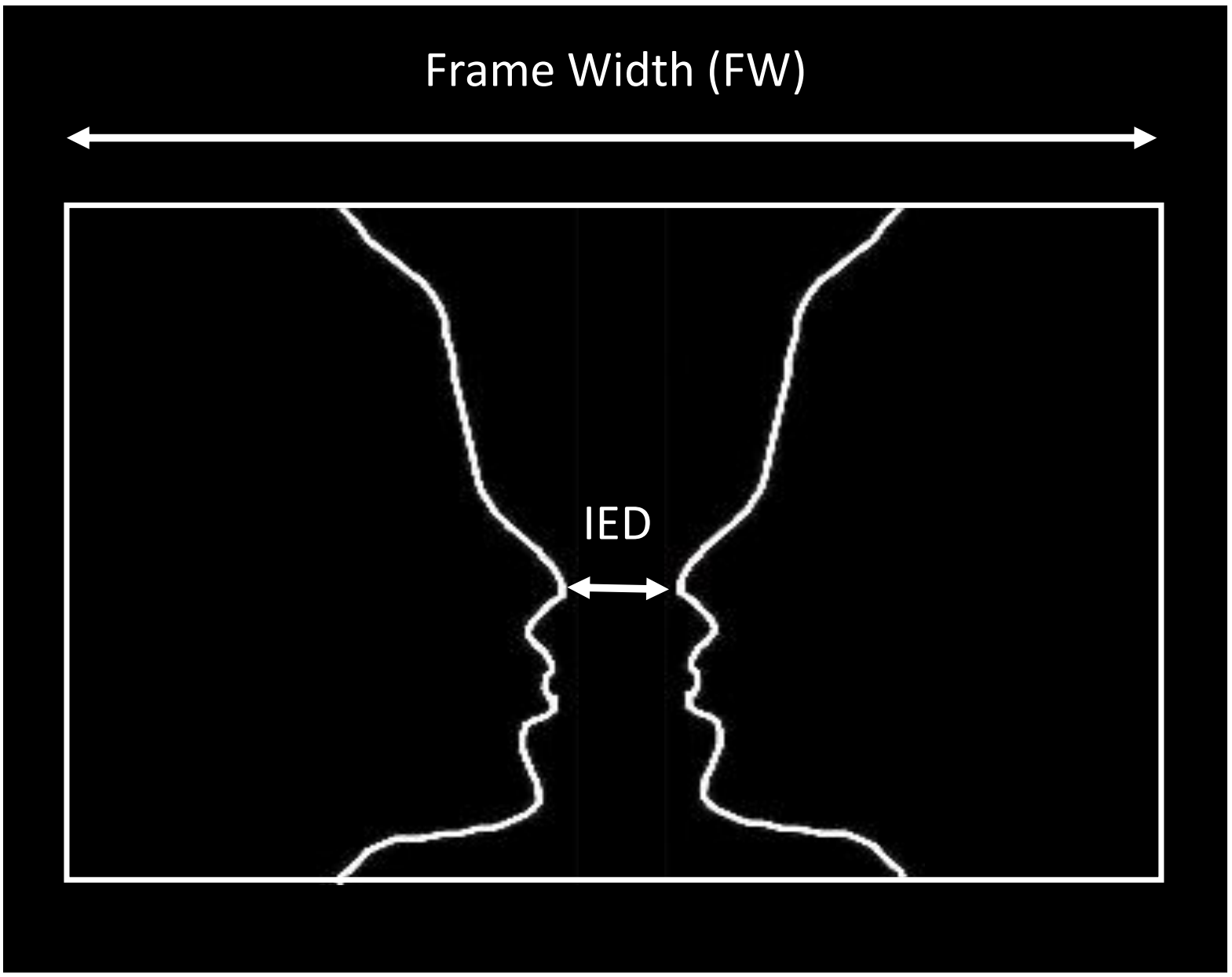
**(A)**


Reversal Negativity (RN)
200-400 ms - O1, O2, Oz, PO7, PO8
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
**
*
Average Amplitude (microvolts)
Reversal
Stability
Reversal Task
Identity Task
Task Type

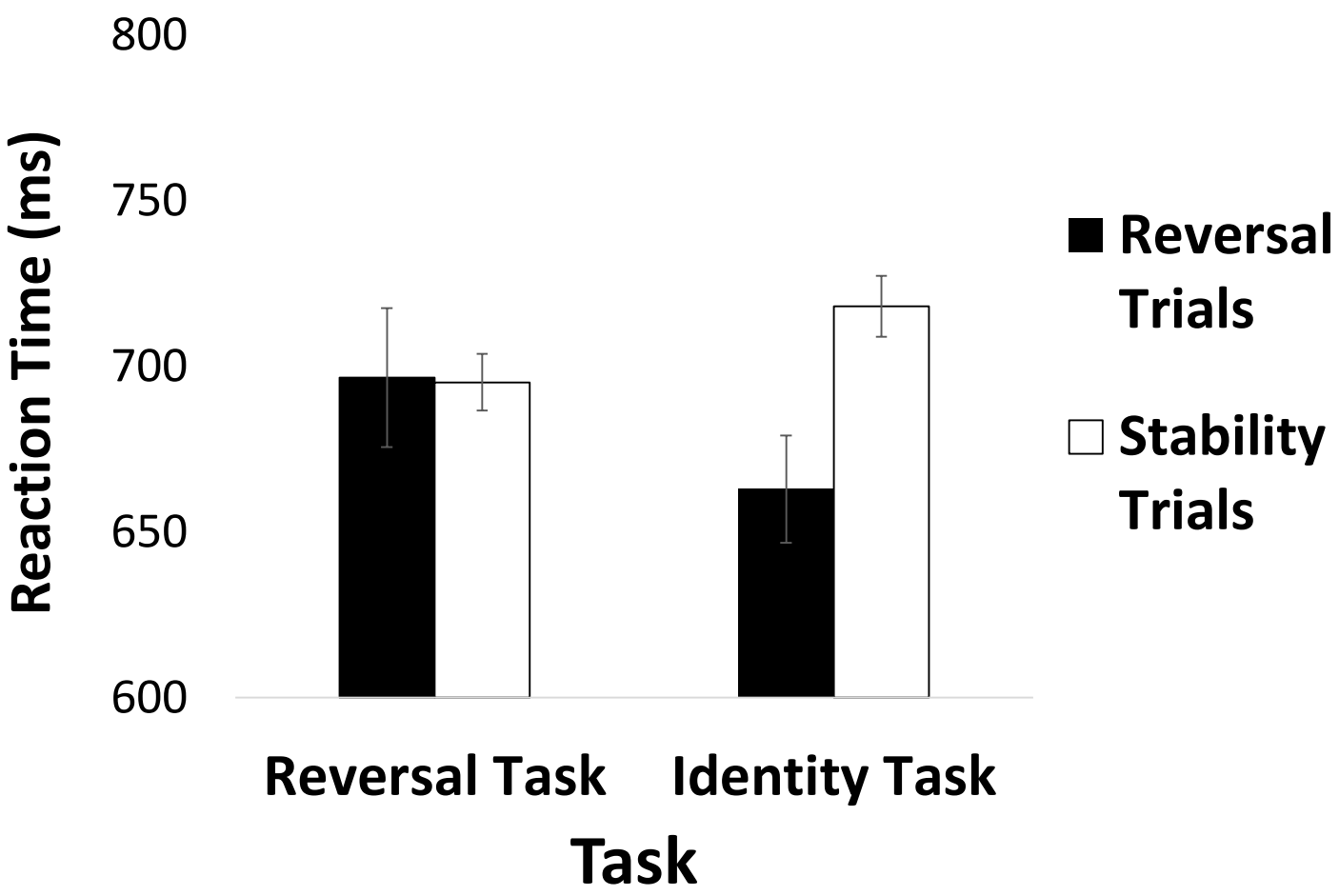
**(B)**


Reversal Positivity (RP)
100-200 ms - O1, O2, Oz
6
5
4
3
2
1
0
*
Average Amplitude (microvolts)
Reversal
Stability
Response
Non Response
Response Action

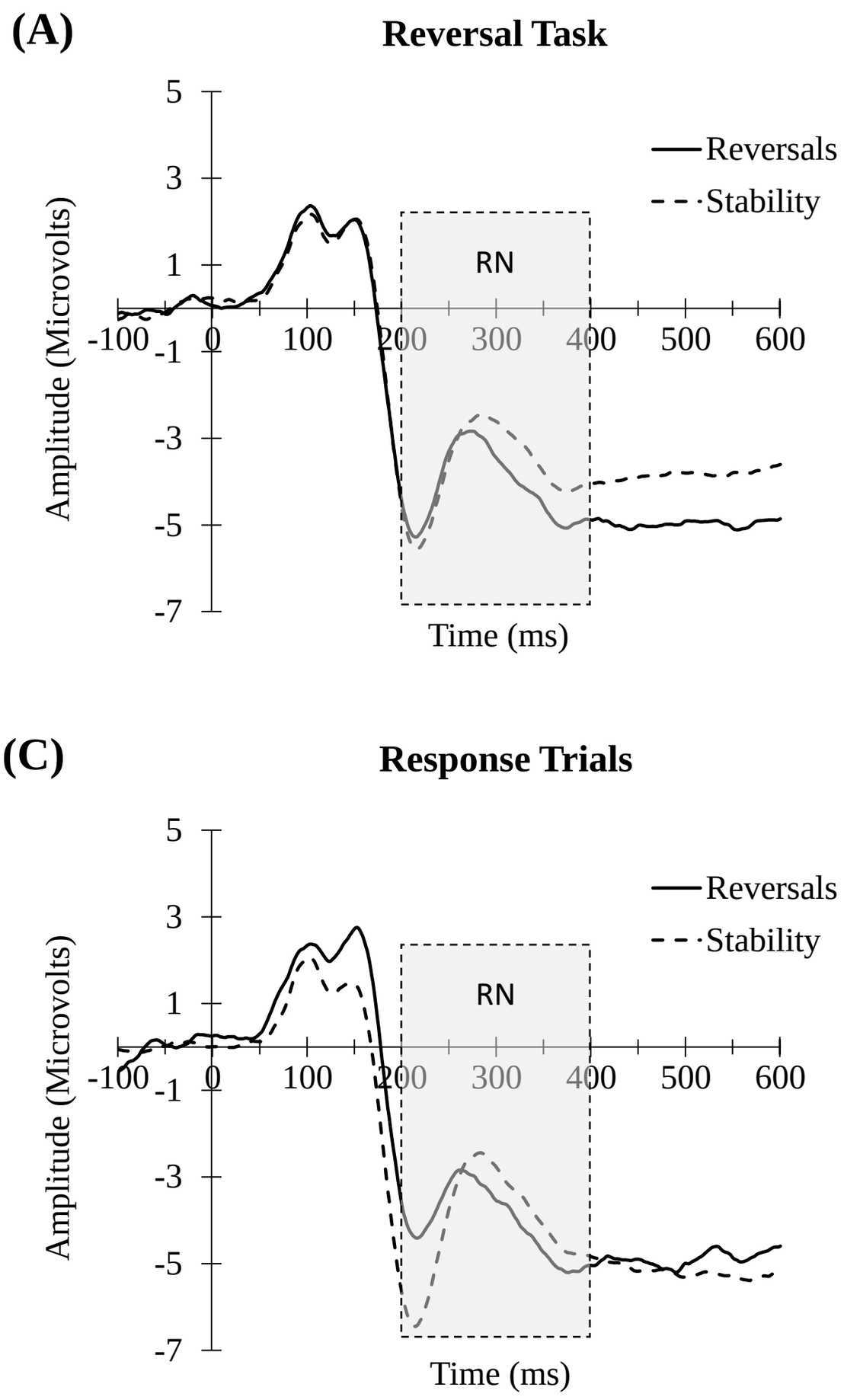
(A)

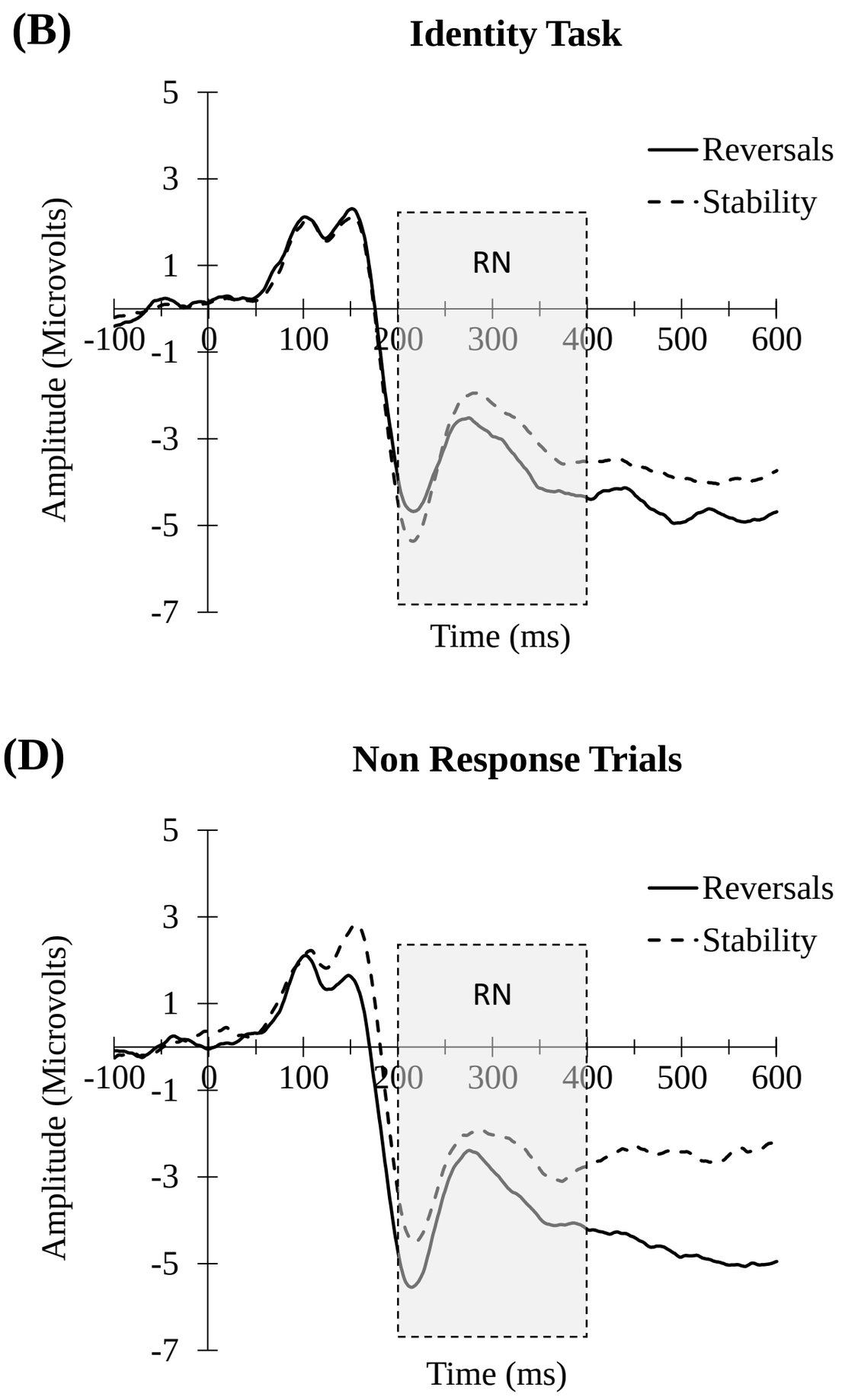

Frame Width (FW)
IED

(B)


Experiment 2 - Reaction Times
800
750
700
650
600
Reaction Time (ms)
Reversal Trials
Stability Trials
Reversal Task
Identity Task
Task

# RN ROI (O1, Oz, O2, PO7, PO8)



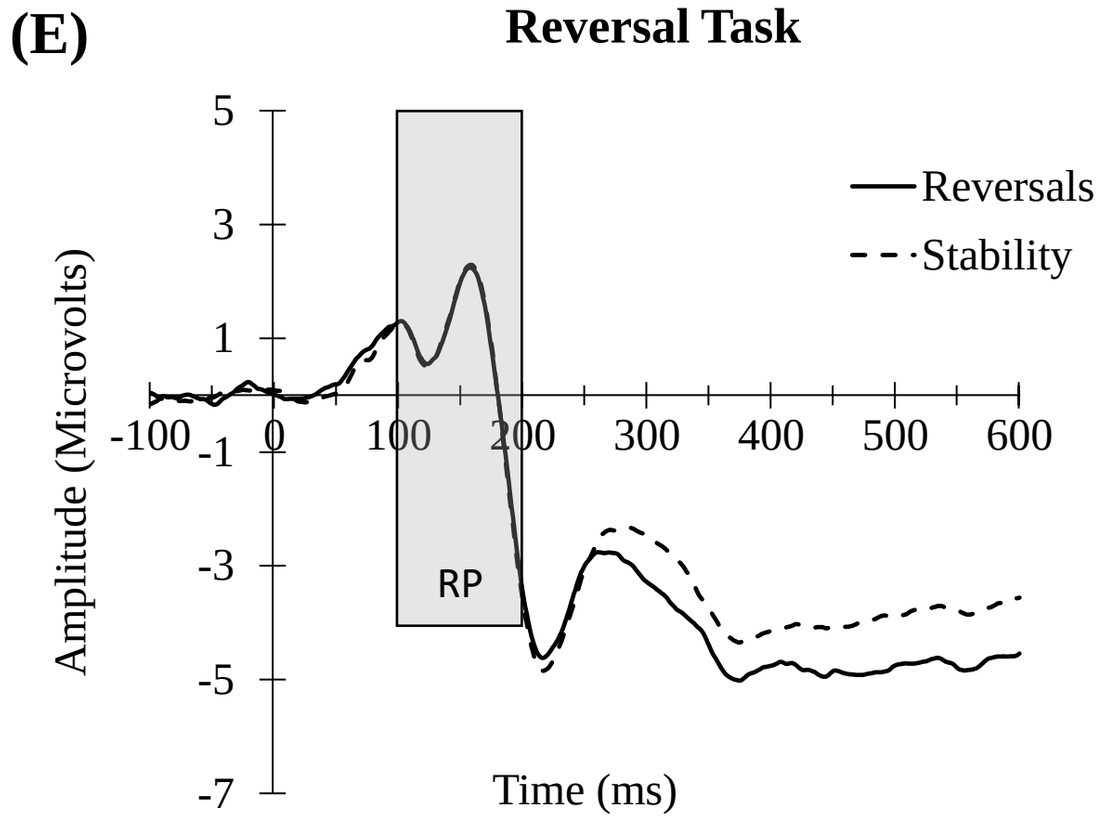



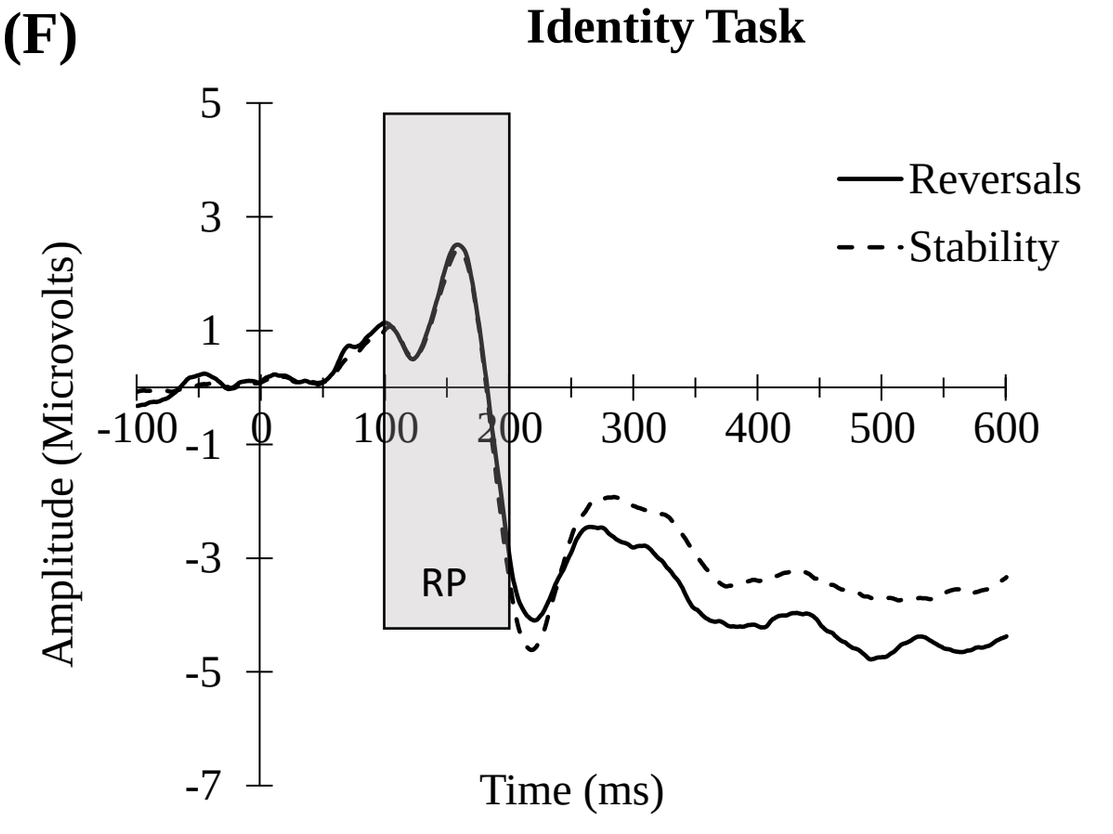


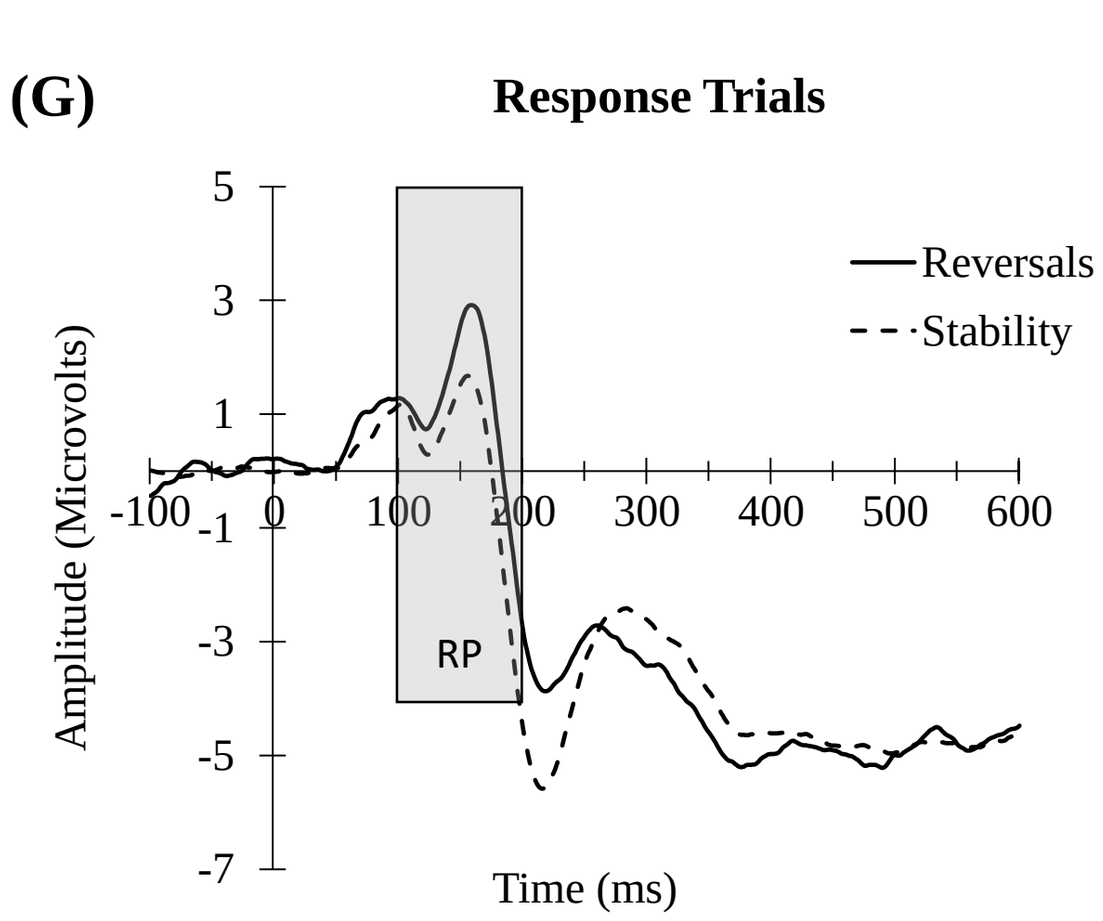


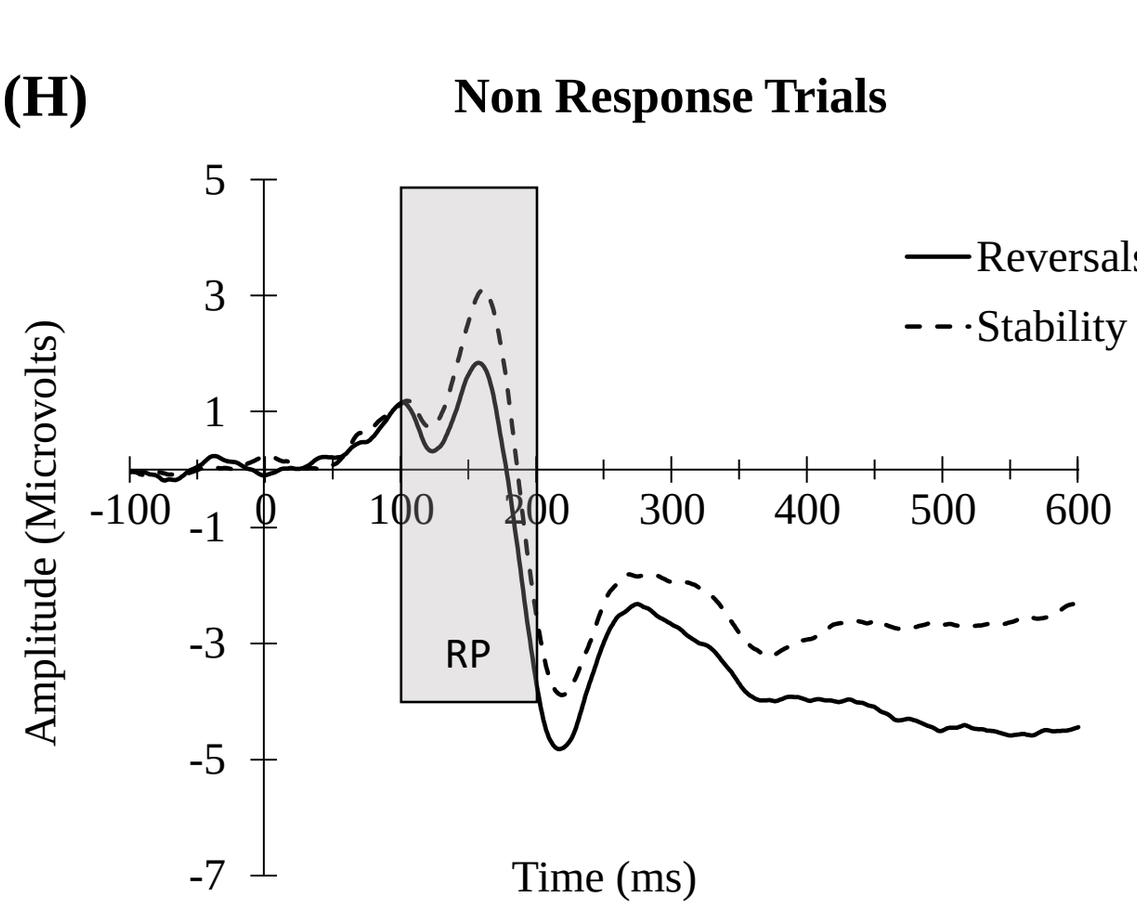


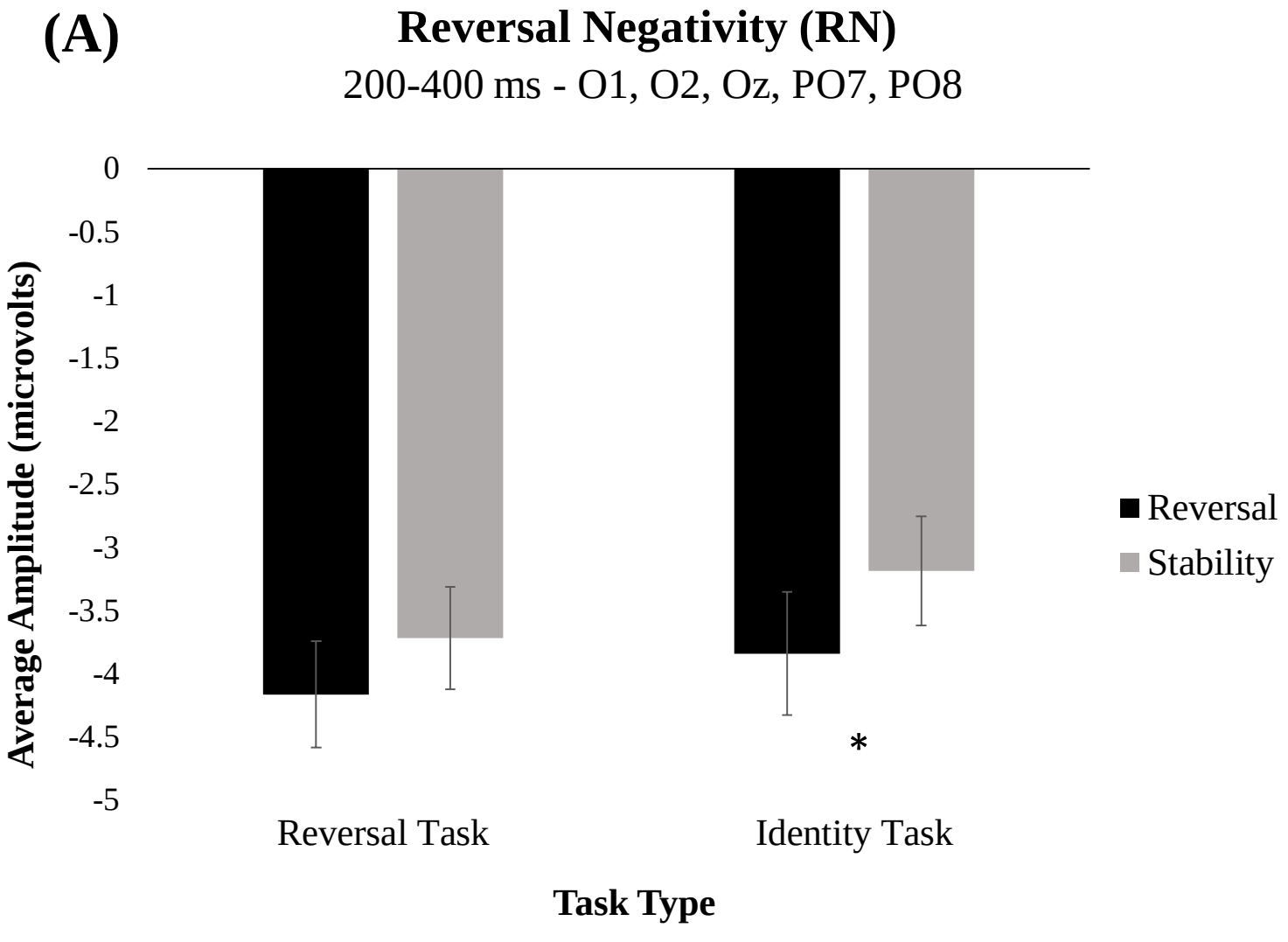

# RP ROI (O1, Oz, O2)

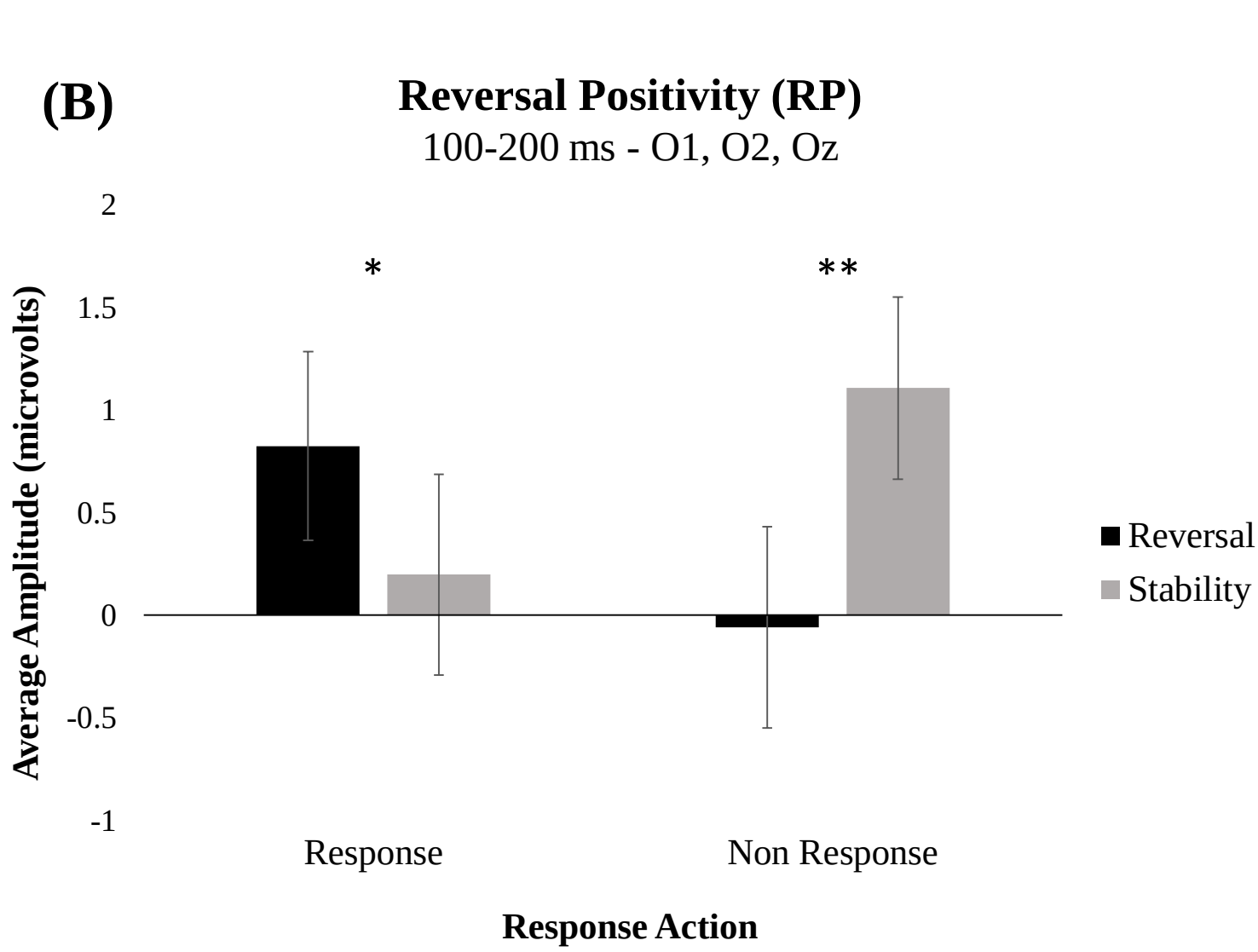



(A)
Reversal Negativity (RN)
200-400 ms - O1, O2, Oz, PO7, PO8
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
Average Amplitude (microvolts)
*
Reversal
Stability
Reversal Task
Identity Task
Task Type


(B)
Reversal Positivity (RP)
100-200 ms - O1, O2, Oz
2
1.5
1
0.5
0
-0.5
-1
Average Amplitude (microvolts)
*
**
Reversal
Stability
Response
Non Response
Response Action

## Supplementary Materials

**Table S1: ANOVA Table - Experiment 1 (Necker Lattice) - RN**

| *Factor/Interaction* | *df* | *F* | $\eta_p^2$ | *p* |
|---|---|---|---|---|
| Channels | 2.329 | 4.209 | 0.208 | 0.018 |
| Task | 1 | 0.042 | 0.003 | 0.84 |
| Trial Type | 1 | 18.272 | 0.533 | 0.001 |
| Response | 1 | 16.496 | 0.508 | 0.001 |
| Channels x Task | 2.692 | 1.005 | 0.059 | 0.393 |
| Channels x Trial Type | 2.427 | 4.346 | 0.214 | 0.015 |
| Task x Trial Type | 1 | 0.248 | 0.015 | 0.625 |
| Channels x Task x Trial Type | 1.951 | 0.201 | 0.012 | 0.814 |
| Channels x Response | 2.268 | 0.847 | 0.05 | 0.45 |
| Task x Response | 1 | 1.435 | 0.082 | 0.248 |
| Channels x Task x Response | 2.881 | 0.729 | 0.044 | 0.535 |
| Trial Type x Response | 1 | 0.068 | 0.004 | 0.798 |
| Channels x Trial Type x Response | 1.807 | 2.057 | 0.114 | 0.15 |
| Task x Trial Type x Response | 1 | 1.528 | 0.087 | 0.234 |
| Channels x Task x Trial Type x Response | 2.033 | 1.872 | 0.105 | 0.169 |

**Table S2: ANOVA Table - Experiment 1 (Necker Lattice) - RP**

| *Factor/Interaction* | *df* | *F* | $\eta_p^2$ | *p* |
|---|---|---|---|---|
| Channels | 1.375 | 2.386 | 0.13 | 0.129 |
| Task | 1 | 3.753 | 0.19 | 0.071 |
| Trial Type | 1 | 0.036 | 0.002 | 0.852 |
| Response | 1 | 2.724 | 0.145 | 0.118 |
| Channels x Task | 2 | 0.13 | 0.008 | 0.879 |
| Channels x Trial Type | 1.473 | 0.372 | 0.023 | 0.63 |
| Task x Trial Type | 1 | 0.421 | 0.026 | 0.525 |
| Channels x Task x Trial Type | 2 | 0.873 | 0.052 | 0.427 |
| Channels x Response | 1.579 | 0.003 | 0 | 0.991 |
| Task x Response | 1 | 1.563 | 0.089 | 0.229 |
| Channels x Task x Response | 1.253 | 0.971 | 0.057 | 0.357 |
| Trial Type x Response | 1 | 4.895 | 0.234 | 0.042 |
| Channels x Trial Type x Response | 2 | 2.126 | 0.117 | 0.136 |
| Task x Trial Type x Response | 1 | 0.609 | 0.037 | 0.447 |
| Channels x Task x Trial Type x Response | 2 | 0.566 | 0.034 | 0.573 |

**Table S3: ANOVA Table - Experiment 2 - RN Faces-Vase**

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Channels | 2.546 | 1.488 | 0.085 | 0.216 |
| Task | 1 | 0.469 | 0.028 | 0.503 |
| Trial Type | 1 | 15.461 | 0.491 | 0.001 |
| Response | 1 | 16.465 | 0.507 | 0.001 |
| Channels x Task | 1.369 | 12.395 | 0.437 | <0.001 |
| Channels x Trial Type | 2.020 | 0.318 | 0.019 | 0.865 |
| Task x Trial Type | 1 | 0.914 | 0.054 | 0.353 |
| Channels x Task x Trial Type | 2.25 | 1.470 | 0.084 | 0.243 |
| Channels x Response | 1.454 | 2.044 | 0.113 | 0.161 |
| Task x Response | 1 | 2.093 | 0.116 | 0.167 |
| Channels x Task x Response | 1.257 | 1.649 | 0.093 | 0.217 |
| Trial Type x Response | 1 | 0.634 | 0.038 | 0.437 |
| Channels x Trial Type x Response | 1.845 | 0.403 | 0.025 | 0.656 |
| Task x Trial Type x Response | 1 | 0.472 | 0.029 | 0.502 |
| Channels x Task x Trial Type x Response | 1.858 | 0.675 | 0.040 | 0.506 |

**Table S4: ANOVA Table - Experiment 2 - RP Faces-Vase**

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Channels | 1.529 | 7.477 | 0.318 | 0.005 |
| Task | 1 | 0.113 | 0.007 | 0.741 |
| Trial Type | 1 | 0.007 | 0.000 | 0.935 |
| Response | 1 | 3.464 | 0.178 | 0.081 |
| Channels x Task | 1.285 | 0.057 | 0.004 | 0.871 |
| Channels x Trial Type | 2 | 0.853 | 0.051 | 0.435 |
| Task x Trial Type | 1 | 0.074 | 0.005 | 0.788 |
| Channels x Task x Trial Type | 1.383 | 0.599 | 0.036 | 0.499 |
| Channels x Response | 1.258 | 3.686 | 0.187 | 0.061 |
| Task x Response | 1 | 0.148 | 0.009 | 0.705 |
| Channels x Task x Response | 1.5 | 0.997 | 0.059 | 0.362 |
| Trial Type x Response | 1 | 9.485 | 0.372 | 0.007 |
| Channels x Trial Type x Response | 1.595 | 2.165 | 0.119 | 0.143 |
| Task x Trial Type x Response | 1 | 0.073 | 0.005 | 0.791 |
| Channels x Task x Trial Type x Response | 1.874 | 0.819 | 0.049 | 0.444 |

**Table S5: ANOVA Table - Experiments 1 & 2 Combined – RN**

| *Factor/Interaction* | *df* | *F* | $\eta_p^2$ | *p* |
|---|---|---|---|---|
| Channels | 2.426 | 3.496 | 0.098 | 0.027 |
| Channels x Experiment | 2.426 | 4.029 | 0.112 | 0.016 |
| Task | 1 | 1.387 | 0.042 | 0.248 |
| Task x Experiment | 1 | 0.673 | 0.021 | 0.418 |
| Trial Type | 1 | 31.579 | 0.497 | <0.001 |
| Trial Type x Experiment | 1 | 0.449 | 0.014 | 0.507 |
| Response | 1 | 36.811 | 0.535 | <0.001 |
| Response x Experiment | 1 | 0.191 | 0.006 | 0.665 |
| Channels x Task | 3.181 | 1.352 | 0.041 | 0.261 |
| Channels x Task x Experiment | 3.181 | 0.886 | 0.027 | 0.456 |
| Channels x Trial Type | 2.749 | 3.139 | 0.089 | 0.033 |
| Channels x Trial Type x Experiment | 2.749 | 1.493 | 0.045 | 0.225 |
| Task x Trial Type | 1 | 0.003 | 0 | 0.959 |
| Task x Trial Type x Experiment | 1 | 0.759 | 0.023 | 0.39 |
| Channels x Task x Trial Type | 2.149 | 1.408 | 0.042 | 0.251 |
| Channels x Task x Trial Type x Experiment | 2.149 | 0.728 | 0.022 | 0.496 |
| Channels x Response | 2.322 | 4.448 | 0.122 | 0.011 |
| Channels x Response x Experiment | 2.322 | 0.776 | 0.024 | 0.481 |
| Task x Response | 1 | 0.07 | 0.002 | 0.793 |
| Task x Response x Experiment | 1 | 4.863 | 0.132 | 0.035 |
| Channels x Task x Response | 2.854 | 1.23 | 0.037 | 0.303 |
| Channels x Task x Response x Experiment | 2.854 | 4.167 | 0.115 | 0.009 |
| Trial Type x Response | 1 | 0.361 | 0.001 | 0.552 |
| Trial Type x Response x Experiment | 1 | 0.834 | 0.025 | 0.368 |
| Channels x Trial Type x Response | 2.337 | 0.188 | 0.006 | 0.861 |
| Channels x Trial Type x Response x Experiment | 2.337 | 2.64 | 0.076 | 0.069 |
| Task x Trial Type x Response | 1 | 1.871 | 0.055 | 0.181 |
| Task x Trial Type x Response x Experiment | 1 | 0.233 | 0.007 | 0.633 |
| Channels x Task x Trial Type x Response | 2.671 | 2.306 | 0.067 | 0.089 |
| Channels x Task x Trial Type x Response x Experiment | 2.671 | 0.981 | 0.03 | 0.399 |

**Table S6: ANOVA Table - Experiments 1 & 2 Combined – RP**

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Channels | 1.627 | 1.124 | 0.034 | 0.323 |
| Channels x Experiment | 1.627 | 5.874 | 0.155 | 0.008 |
| Task | 1 | 2.002 | 0.059 | 0.167 |
| Task x Experiment | 1 | 3.211 | 0.091 | 0.083 |
| Trial Type | 1 | 0.007 | 0 | 0.933 |
| Trial Type x Experiment | 1 | 0.038 | 0.001 | 0.846 |
| Response | 1 | 6.07 | 0.159 | 0.019 |
| Response x Experiment | 1 | 0.002 | 0 | 0.963 |
| Channels x Task | 1.72 | 0.058 | 0.002 | 0.922 |
| Channels x Task x Experiment | 1.72 | 0.137 | 0.004 | 0.841 |
| Channels x Trial Type | 1.612 | 0.367 | 0.011 | 0.649 |
| Channels x Trial Type x Experiment | 1.612 | 0.724 | 0.022 | 0.461 |
| Task x Trial Type | 1 | 0.219 | 0.007 | 0.643 |
| Task x Trial Type x Experiment | 1 | 0.495 | 0.015 | 0.487 |
| Channels x Task x Trial Type | 2 | 0.407 | 0.013 | 0.667 |
| Channels x Task x Trial Type x Experiment | 2.149 | 1.167 | 0.035 | 0.318 |
| Channels x Response | 1.603 | 0.927 | 0.028 | 0.383 |
| Channels x Response x Experiment | 1.603 | 1.087 | 0.033 | 0.333 |
| Task x Response | 1 | 0.641 | 0.02 | 0.429 |
| Task x Response x Experiment | 1 | 1.548 | 0.046 | 0.222 |
| Channels x Task x Response | 1.389 | 0.127 | 0.004 | 0.805 |
| Channels x Task x Response x Experiment | 1.389 | 1.834 | 0.054 | 0.18 |
| Trial Type x Response | 1 | 14.025 | 0.305 | 0.001 |
| Trial Type x Response x Experiment | 1 | 0.397 | 0.012 | 0.533 |
| Channels x Trial Type x Response | 1.644 | 0.337 | 0.01 | 0.673 |
| Channels x Trial Type x Response x Experiment | 1.644 | 3.969 | 0.11 | 0.032 |
| Task x Trial Type x Response | 1 | 0.606 | 0.019 | 0.442 |
| Task x Trial Type x Response x Experiment | 1 | 0.197 | 0.006 | 0.66 |
| Channels x Task x Trial Type x Response | 2 | 1.023 | 0.031 | 0.365 |
| Channels x Task x Trial Type x Response x Experiment | 2 | 0.335 | 0.01 | 0.717 |

**Table S7: ANOVA Table – Experiments 1&2 Combined - RN with Average Reference**

This supplementary analysis used the same data as Table S5 above except that the data were re-referenced to the average reference of 64 scalp EEG channels.

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Channel | 4 | 3.496 | 0.098 | 0.01 |
| Channel * Experiment | 4 | 4.029 | 0.112 | 0.004 |
| Task | 1 | 0.382 | 0.012 | 0.541 |
| Task * Experiment | 1 | 0.564 | 0.017 | 0.458 |
| TrialType | 1 | 16.707 | 0.343 | <0.001 |
| TrialType * Experiment | 1 | 0.983 | 0.030 | 0.329 |
| Response | 1 | 17.852 | 0.358 | <0.001 |
| Response * Experiment | 1 | 0.023 | 0.001 | 0.88 |
| Channel * Task | 4 | 1.352 | 0.041 | 0.254 |
| Channel * Task * Experiment | 4 | 0.886 | 0.027 | 0.474 |
| Channel * TrialType | 4 | 1.660 | 0.049 | 0.163 |
| Channel * TrialType * Experiment | 4 | 0.671 | 0.021 | 0.613 |
| Task * TrialType | 1 | 0.009 | 0.000 | 0.923 |
| Task * TrialType * Experiment | 1 | 0.434 | 0.013 | 0.515 |
| Channel * Task * TrialType | 4 | 0.964 | 0.029 | 0.43 |
| Channel * Task * TrialType * Experiment | 4 | 1.089 | 0.033 | 0.365 |
| Channel * Response | 4 | 4.448 | 0.122 | 0.002 |
| Channel * Response * Experiment | 4 | 0.776 | 0.024 | 0.543 |
| Task * Response | 1 | 0.019 | 0.001 | 0.891 |
| Task * Response * Experiment | 1 | 3.944 | 0.110 | 0.056 |
| Channel * Task * Response | 4 | 1.230 | 0.037 | 0.301 |
| Channel * Task * Response * Experiment | 4 | 4.167 | 0.115 | 0.003 |
| TrialType * Response | 1 | 0.573 | 0.018 | 0.455 |
| TrialType * Response * Experiment | 1 | 0.447 | 0.014 | 0.508 |
| Channel * TrialType * Response | 4 | 0.112 | 0.003 | 0.978 |
| Channel * TrialType * Response * Experiment | 4 | 1.089 | 0.033 | 0.365 |
| Task * TrialType * Response | 1 | 0.918 | 0.028 | 0.345 |
| Task * TrialType * Response * Experiment | 1 | 0.023 | 0.001 | 0.88 |
| Channel * Task * TrialType * Response | 4 | 1.472 | 0.032 | 0.148 |
| Channel * Task * TrialType * Response * Experiment | 4 | 1.138 | 0.034 | 0.342 |
| Experiment | 1 | 8.948 | 0.219 | 0.005 |

**Table S8: ANOVA Table – Experiments 1&2 Combined - RP with Average Reference**

This supplementary analysis used the same data as Table S6 above except that the data were re-referenced to the average reference of 64 scalp EEG channels.

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Channels | 2 | 1.136 | 0.034 | 0.328 |
| Channels x Experiment | 2 | 5.919 | 0.156 | 0.004 |
| Task | 1 | 2.517 | 0.073 | 0.122 |
| Task x Experiment | 1 | 1.244 | 0.037 | 0.273 |
| Trial Type | 1 | 0.210 | 0.007 | 0.650 |
| Trial Type x Experiment | 1 | 0.846 | 0.026 | 0.365 |
| Response | 1 | 4.683 | 0.128 | 0.038 |
| Response x Experiment | 1 | 2.042 | 0.060 | 0.163 |
| Channels x Task | 2 | 0.029 | 0.001 | 0.971 |
| Channels x Task x Experiment | 2 | 0.144 | 0.004 | 0.866 |
| Channels x Trial Type | 2 | 0.509 | 0.016 | 0.603 |
| Channels x Trial Type x Experiment | 2 | 0.649 | 0.020 | 0.526 |
| Task x Trial Type | 1 | 0.449 | 0.014 | 0.508 |
| Task x Trial Type x Experiment | 1 | 0.806 | 0.025 | 0.376 |
| Channels x Task x Trial Type | 2 | 0.291 | 0.009 | 0.749 |
| Channels x Task x Trial Type x Experiment | 2 | 1.218 | 0.037 | 0.302 |
| Channels x Response | 2 | 1.204 | 0.036 | 0.307 |
| Channels x Response x Experiment | 2 | 1.397 | 0.042 | 0.255 |
| Task x Response | 1 | 0.247 | 0.008 | 0.622 |
| Task x Response x Experiment | 1 | 0.729 | 0.022 | 0.400 |
| Channels x Task x Response | 2 | 0.149 | 0.005 | 0.861 |
| Channels x Task x Response x Experiment | 2 | 1.591 | 0.047 | 0.212 |
| Trial Type x Response | 1 | 15.052 | 0.320 | <0.001 |
| Trial Type x Response x Experiment | 1 | 0.619 | 0.019 | 0.437 |
| Channels x Trial Type x Response | 2 | 0.397 | 0.012 | 0.674 |
| Channels x Trial Type x Response x Experiment | 2 | 0.149 | 0.005 | 0.861 |
| Task x Trial Type x Response | 1 | 0.494 | 0.015 | 0.487 |
| Task x Trial Type x Response x Experiment | 1 | 0.000 | 0.000 | 0.993 |
| Channels x Task x Trial Type x Response | 2 | 1.048 | 0.032 | 0.357 |
| Channels x Task x Trial Type x Response x Experiment | 2 | 0.241 | 0.007 | 0.786 |
| Experiment (between-subjects) | 1 | 8.166 | 0.007 | 0.203 |

**List of Scalp Electrode Locations**

Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz (inion), Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, Afz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2

**Figure S1: RP Results with Average Reference Analysis – Experiments 1&2 Combined**

The following show the significant interaction of Trial Type and Response in the RP time window. The data in this analysis include participants from both Experiments 1 and 2 and represents the average amplitude at electrodes O1, Oz, and O2 from 100-200 ms.

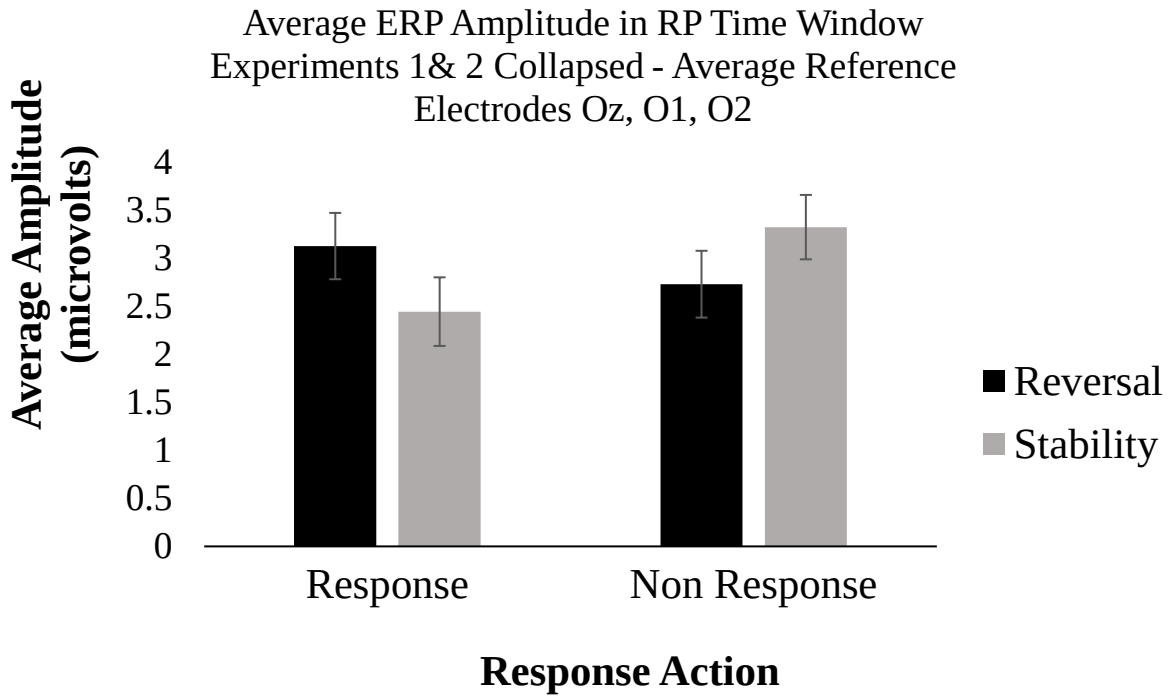
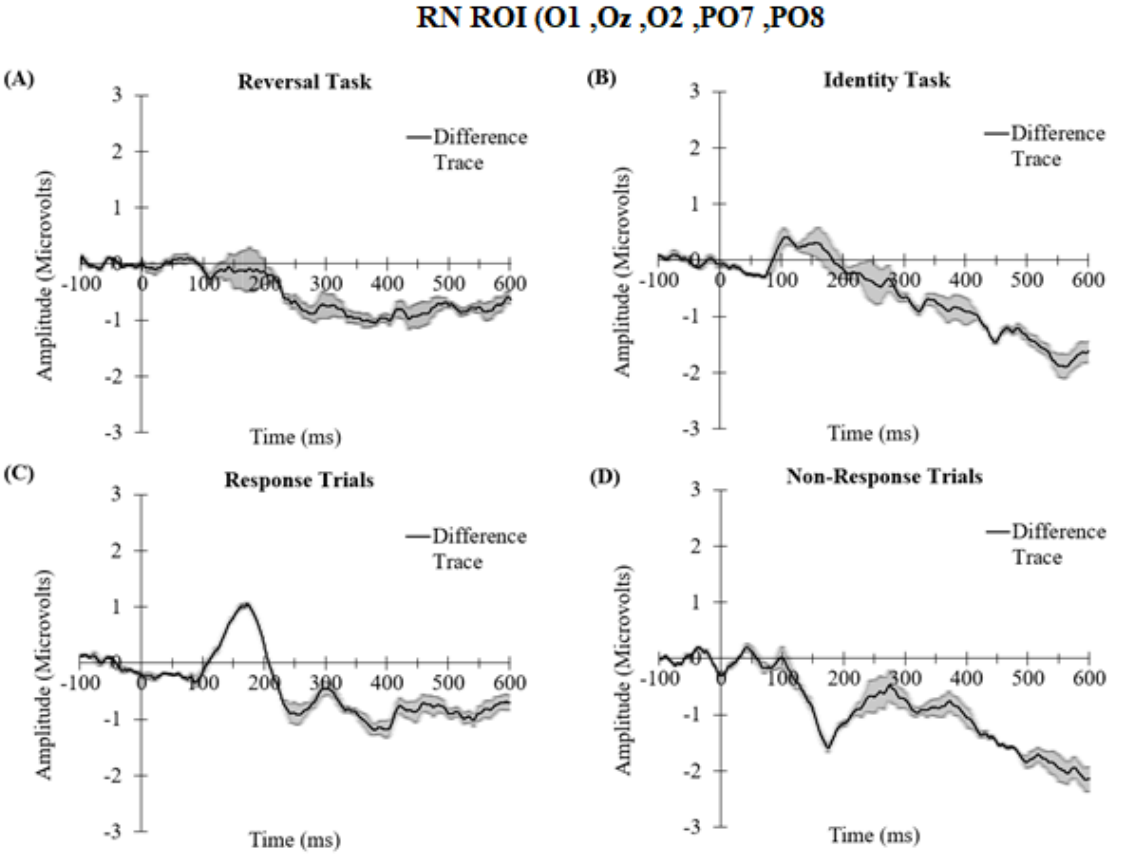



**Figure S2: Difference ERP Traces for Experiment 1 (Necker Lattice)**

Each panel plots the average difference of the ERP waveforms (dERP) for reversal and stability trials in Experiment 1 (Necker Lattice). There is a plot corresponding to each of the panels in Figure 3. Panels A-D are averaged over the RN ROI (O1, O2, Oz, PO7, & PO8). (A) Reversal Task (collapsed over response); (B) Identity Task (collapsed over response); (C) Response trials (collapsed over task); (D) Non-response trials (collapsed over task). Panels E-H are averaged over the RP ROI (O1, O2, & Oz). (E) Reversal Task (collapsed over response); (F) Identity Task (collapsed over response); (G) Response trials (collapsed over task); (H) Non-response trials (collapsed over task) z. The shaded area indicates the Standard Error of the Mean for the difference wave.


RN ROI (O1 ,Oz ,O2 ,PO7 ,PO8
(A)
Reversal Task
(B)
Identity Task
(C)
Response Trials
(D)
Non-Response Trials
Difference Trace
Amplitude (Microvolts)
Time (ms)

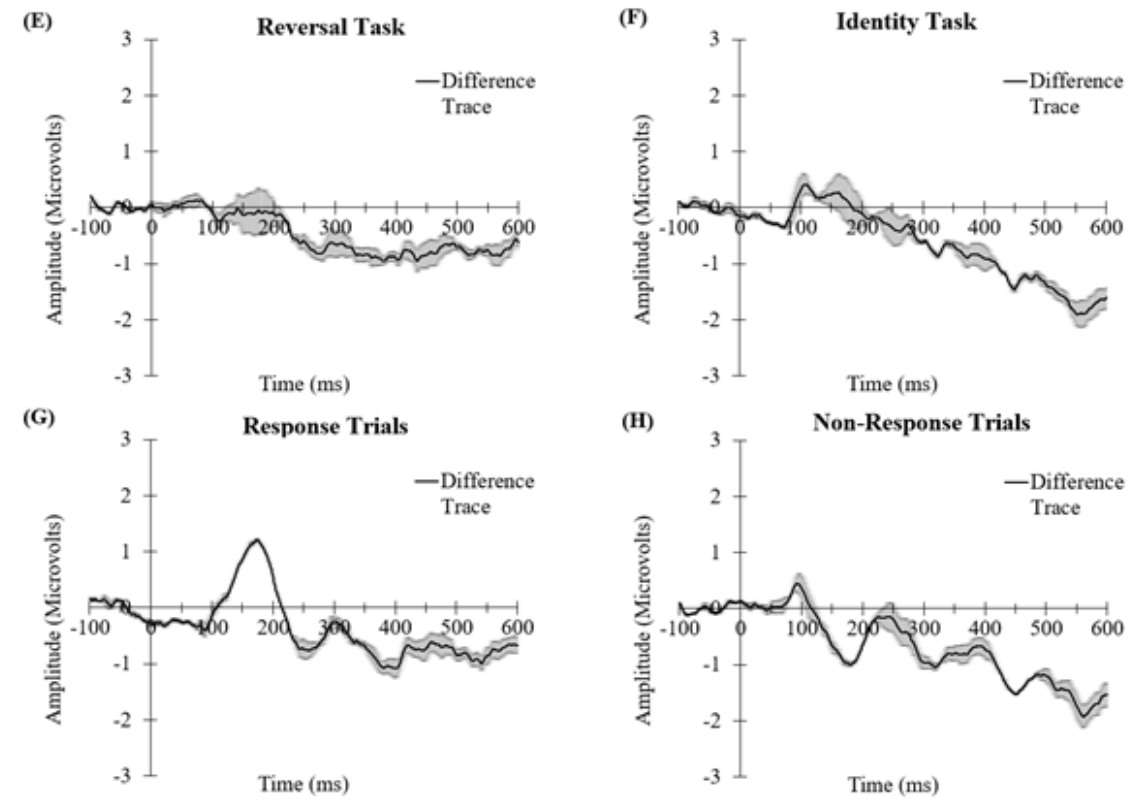
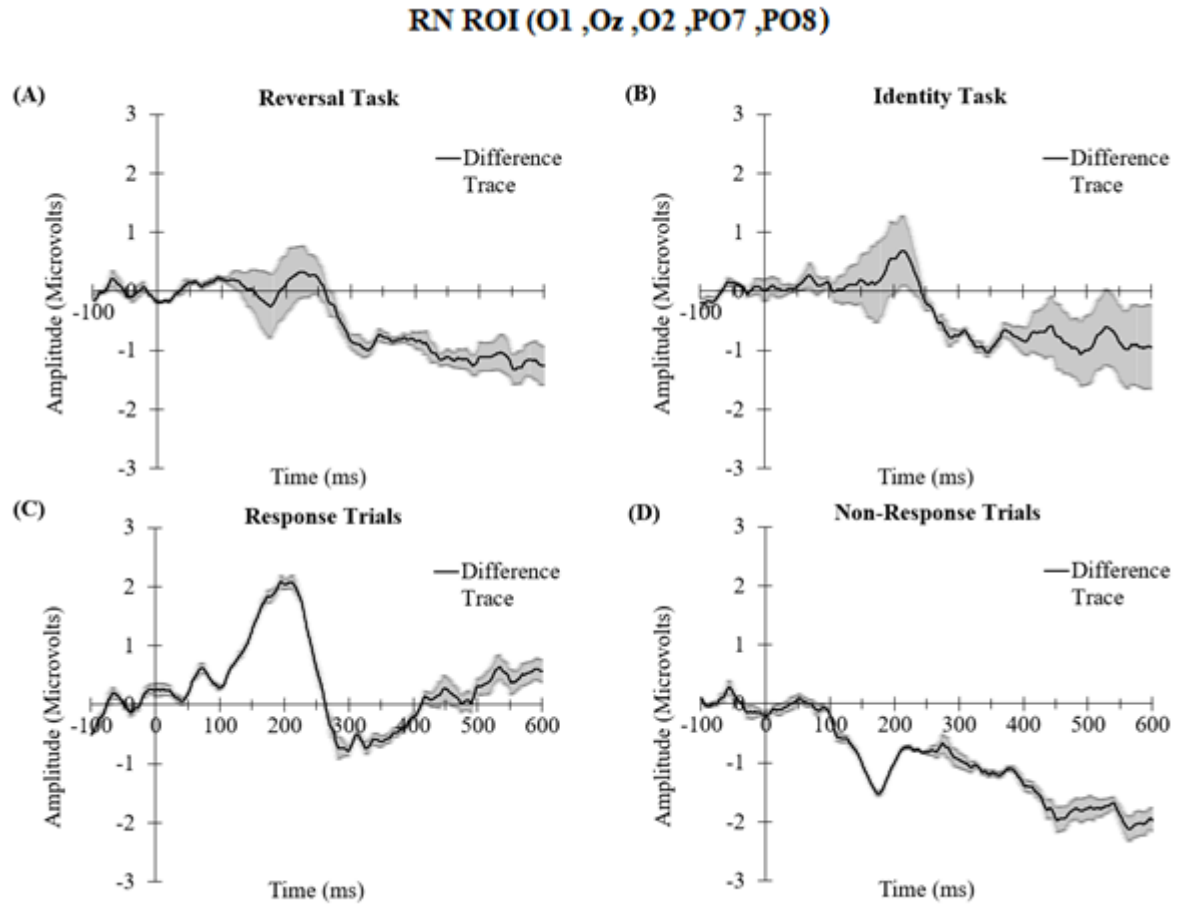

RP ROI (O1, O2, Oz)
(E)
Reversal Task
(F)
Identity Task
(G)
Response Trials
(H)
Non-Response Trials
Difference Trace
Amplitude (Microvolts)
Time (ms)

**Figure S3: Difference ERP Traces for Experiment 2 (Faces-Vase)**

Each panel plots the average difference of the ERP waveforms (dERP) for reversal and stability trials in Experiment 2 (Faces-Vase). There is a plot corresponding to each of the panels in Figure 6. Panels A-D are averaged over the RN ROI (O1, O2, Oz, PO7, & PO8). (A) Reversal Task (collapsed over response); (B) Identity Task (collapsed over response); (C) Response trials (collapsed over task); (D) Non-response trials (collapsed over task). Panels E-H are averaged over the RP ROI (O1, O2, & Oz). (E) Reversal Task (collapsed over response); (F) Identity Task (collapsed over response); (G) Response trials (collapsed over task); (H) Non-response trials (collapsed over task) z. The shaded area indicates the Standard Error of the Mean for the difference wave.



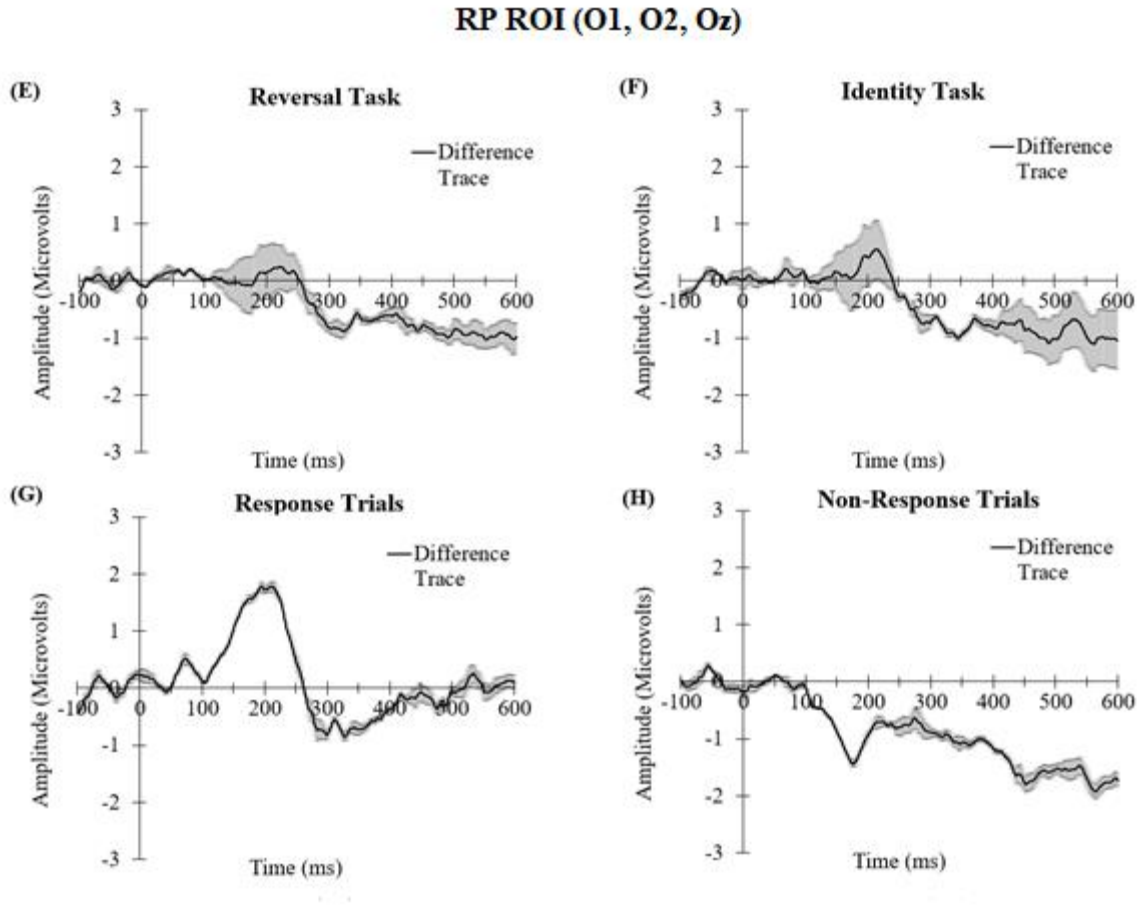

RP ROI (O1, O2, Oz)
(E)
Reversal Task
(F)
Identity Task
(G)
Response Trials
(H)
Non-Response Trials
Difference Trace
Amplitude (Microvolts)
Time (ms)
3
2
1
0
-1
-2
-3
-100
0
100
200
300
400
500
600

**Figure S4: Topographic Maps for the Reversal and Stability Trials for Experiment 1 (Necker Lattice)**

Each panel plots the grand average amplitude (across participants) of the difference between reversal and stability trials under the different response and task conditions covered in Figure 3. Panels A-D cover the RN time window (200-400 ms) and panels E-H cover the RP time window (100-200 ms): (A & E) in the reversal task (collapsed over response); (B & F) in the identity task (collapsed over response); (C &G) for the response trials (collapsed over task); (D & H) for the non-response trials (collapsed over task).

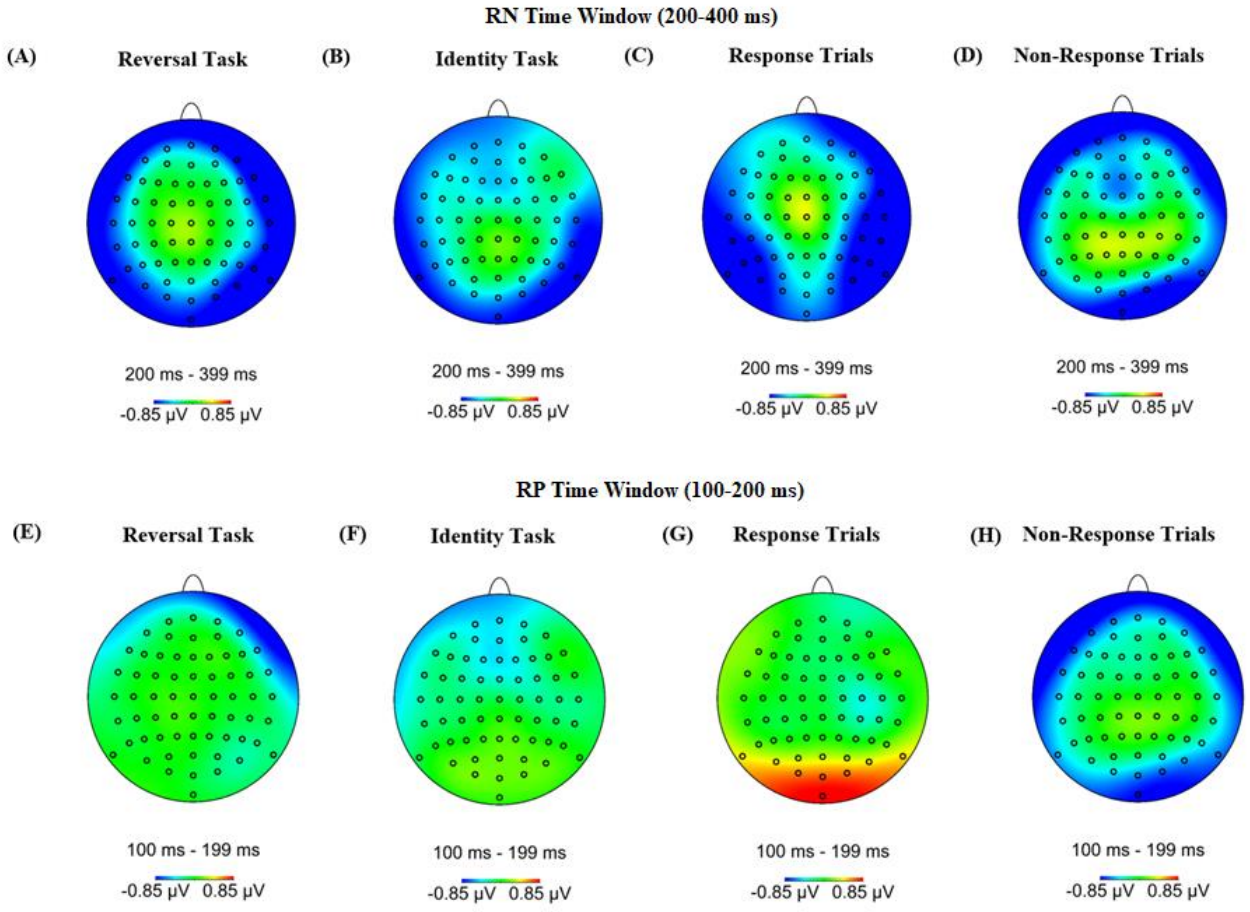


**Figure S5: Topographic Maps for the Reversal and Stability Trials for Experiment 2 (Faces-Vase)**

Each panel plots the grand average amplitude (across participants) of the difference between reversal and stability trials under the different response and task conditions covered in Figure 3. Panels A-D cover the RN time window (200-400 ms) and panels E-H cover the RP time window (100-200 ms): (A & E) in the reversal task (collapsed over response); (B & F) in the identity task (collapsed over response); (C &G) for the response trials (collapsed over task); (D & H) for the non-response trials (collapsed over task).

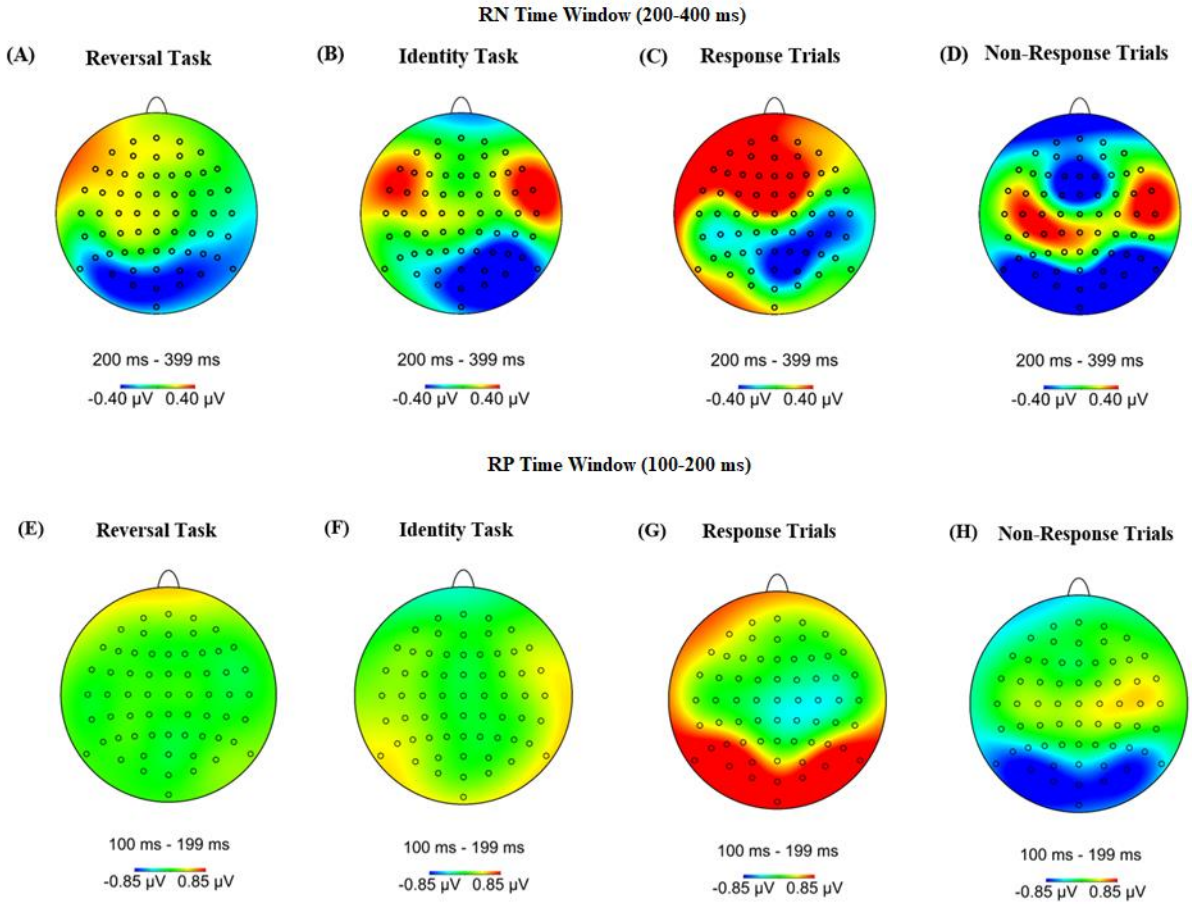
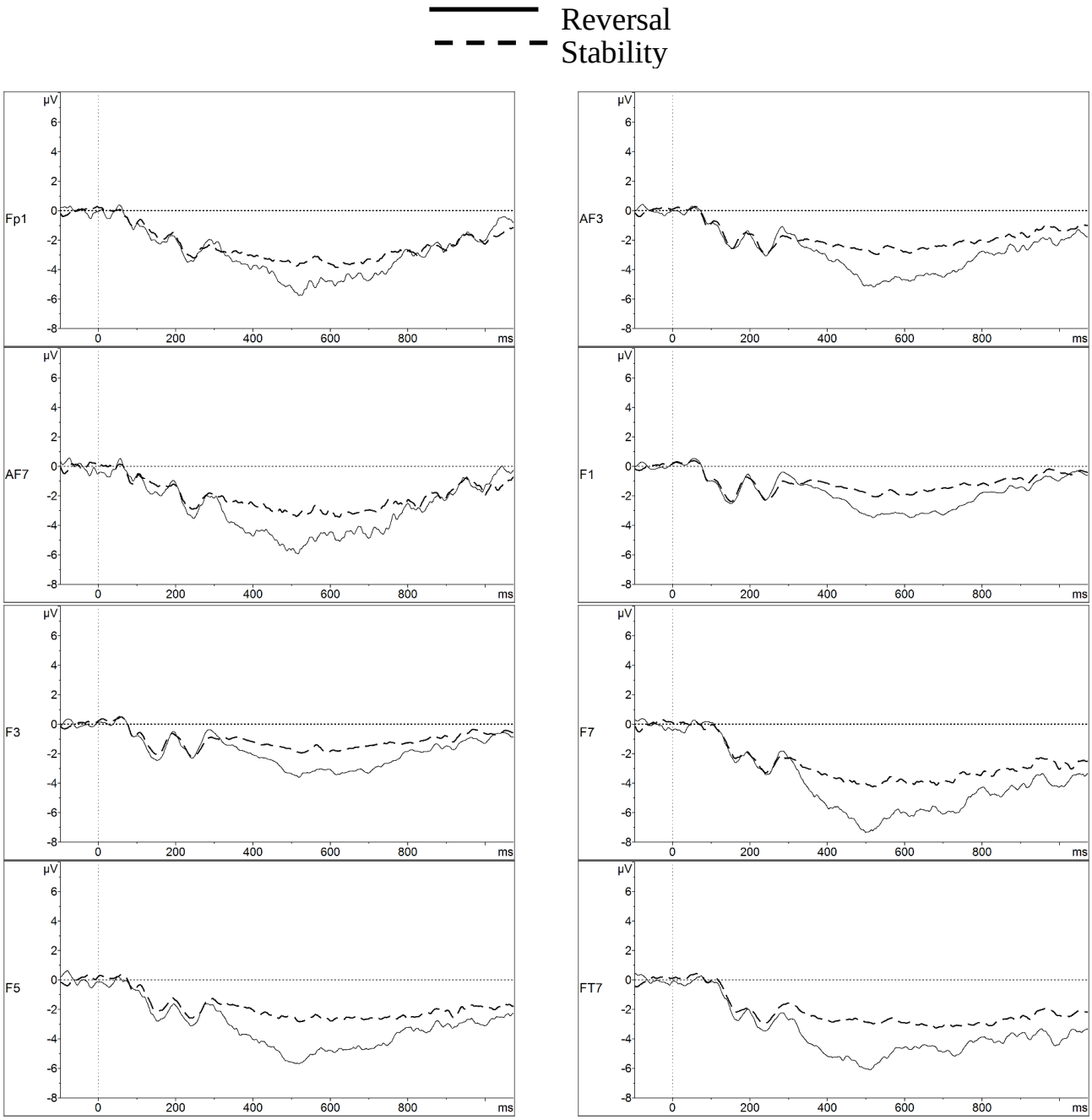
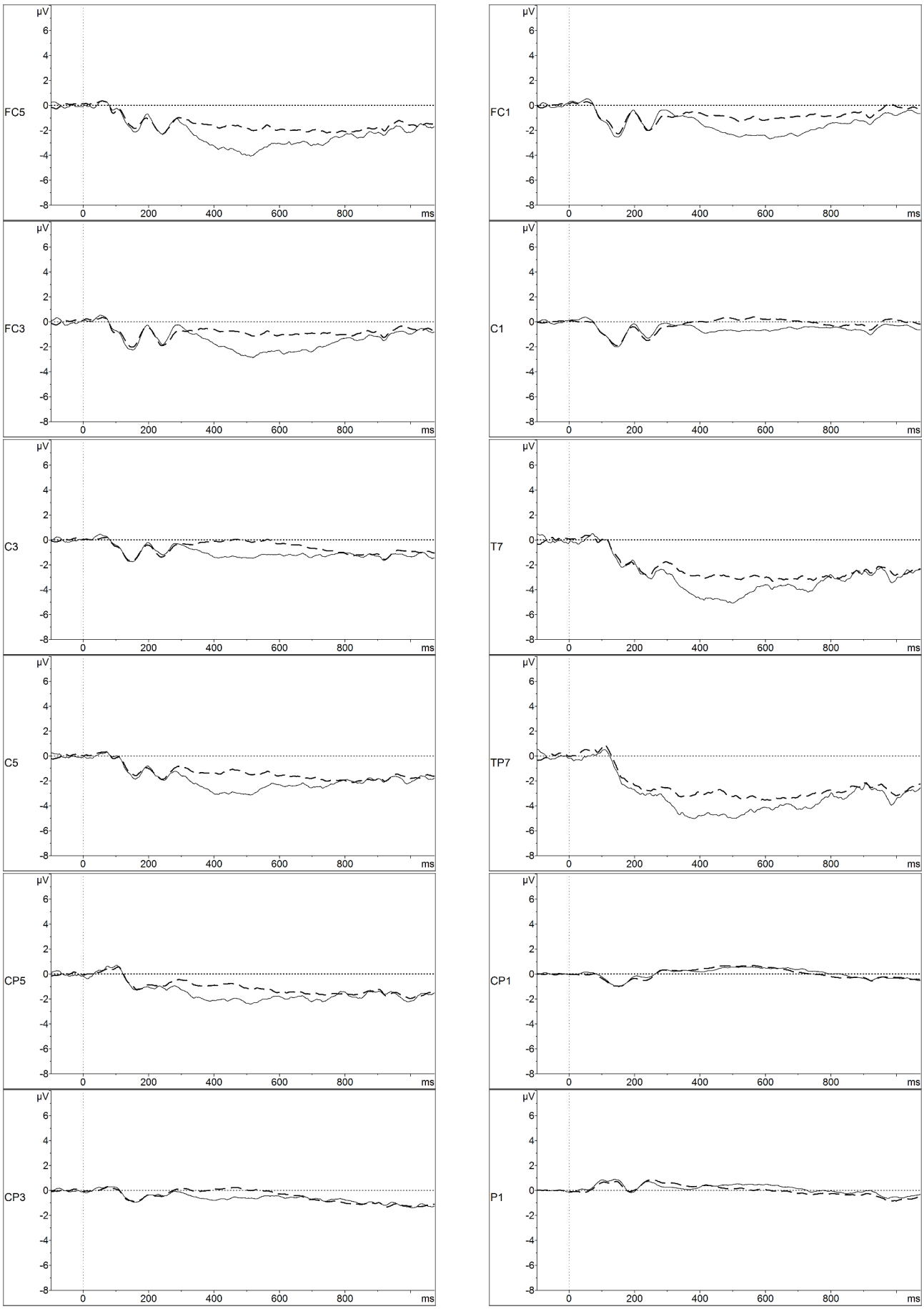


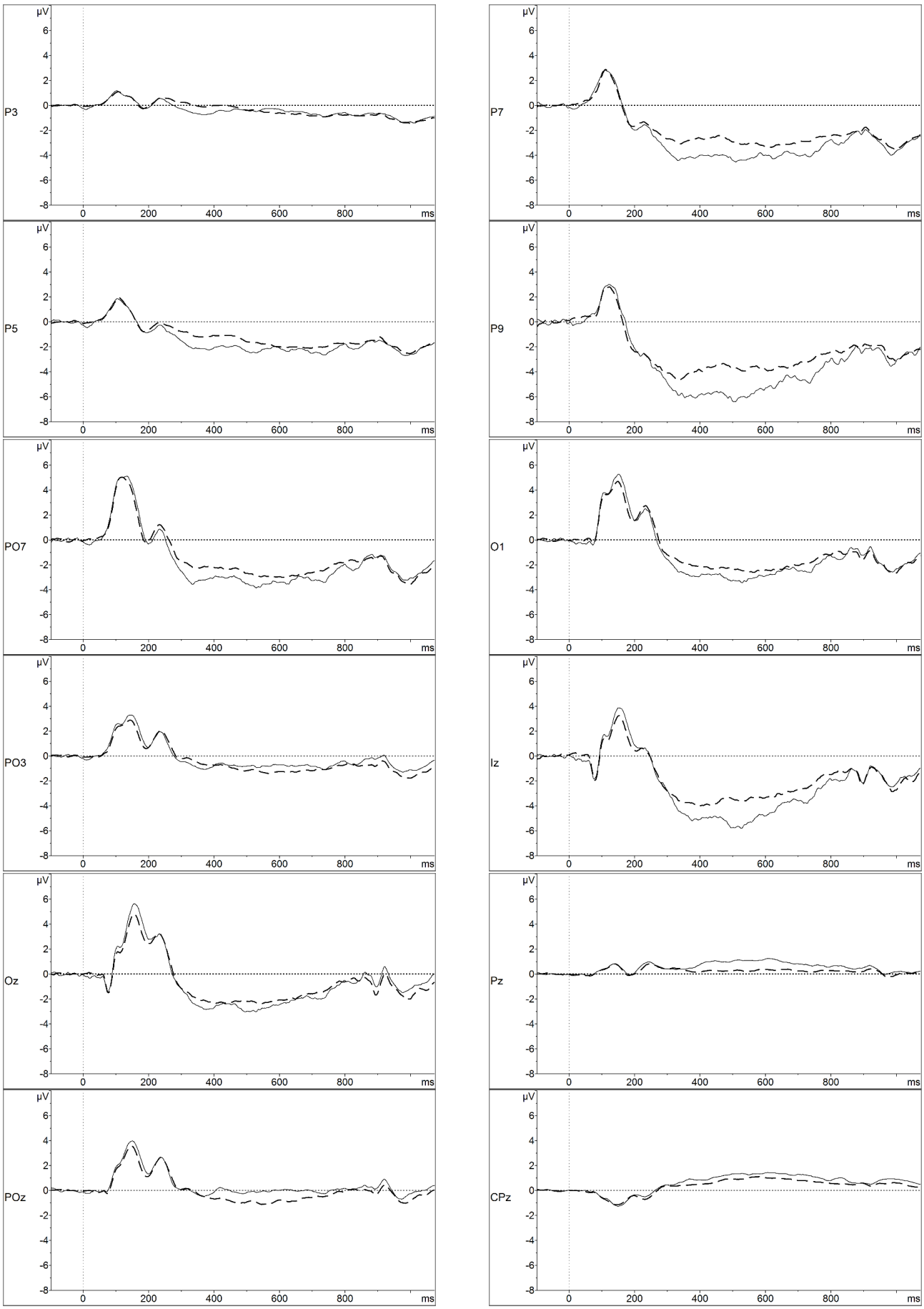
**Figure S6: Experiment 1 (Necker Lattice) Grand Average ERP plots for all electrodes comparing Reversal vs. Stable trials in the Reversal Task**

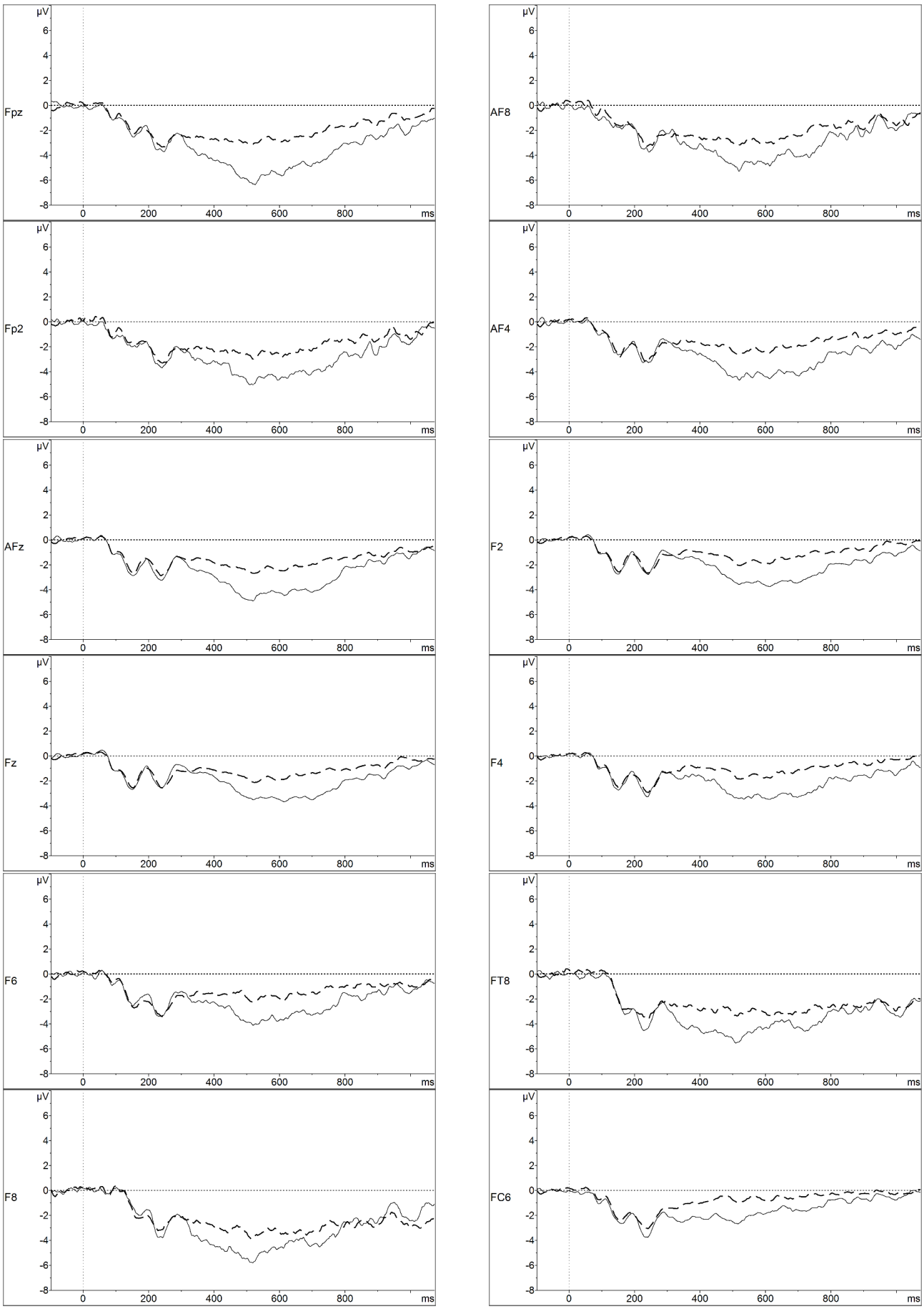
Each panel plots the grand mean ERP signals for reversal (solid line) and stability (dashed line) trials in Experiment 1 (Necker Lattice) from all recorded scalp locations. This is done separately for different conditions aligned to those presented in Figure 3: (a) ERP waveforms in the reversal task (collapsed over response); (b) ERP waveforms in the identity task (collapsed over response); (c) ERP waveforms for the response trials (collapsed over task); (d) ERP waveforms for the non-response trials (collapsed over task).

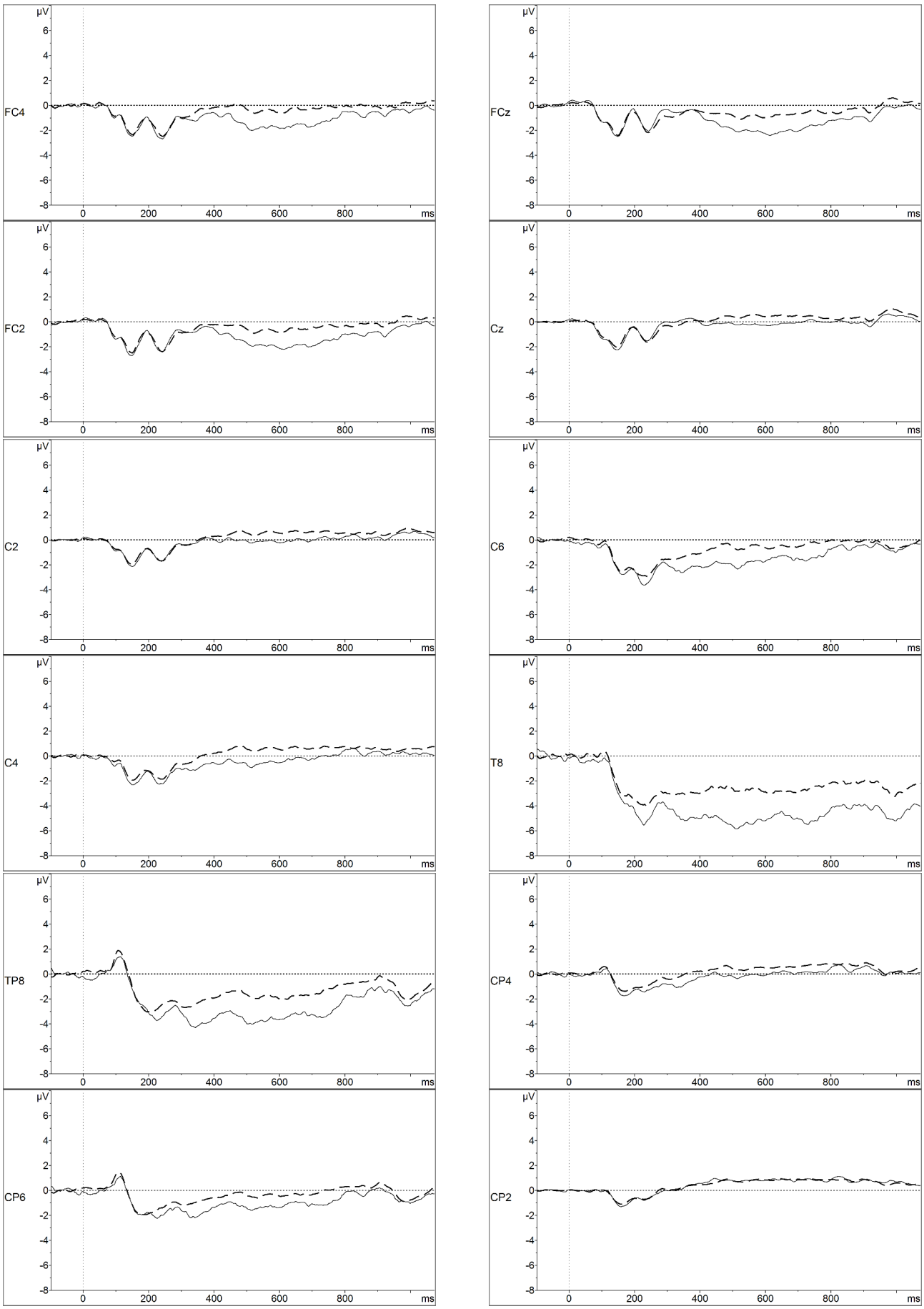
**(A) Reversal Task**

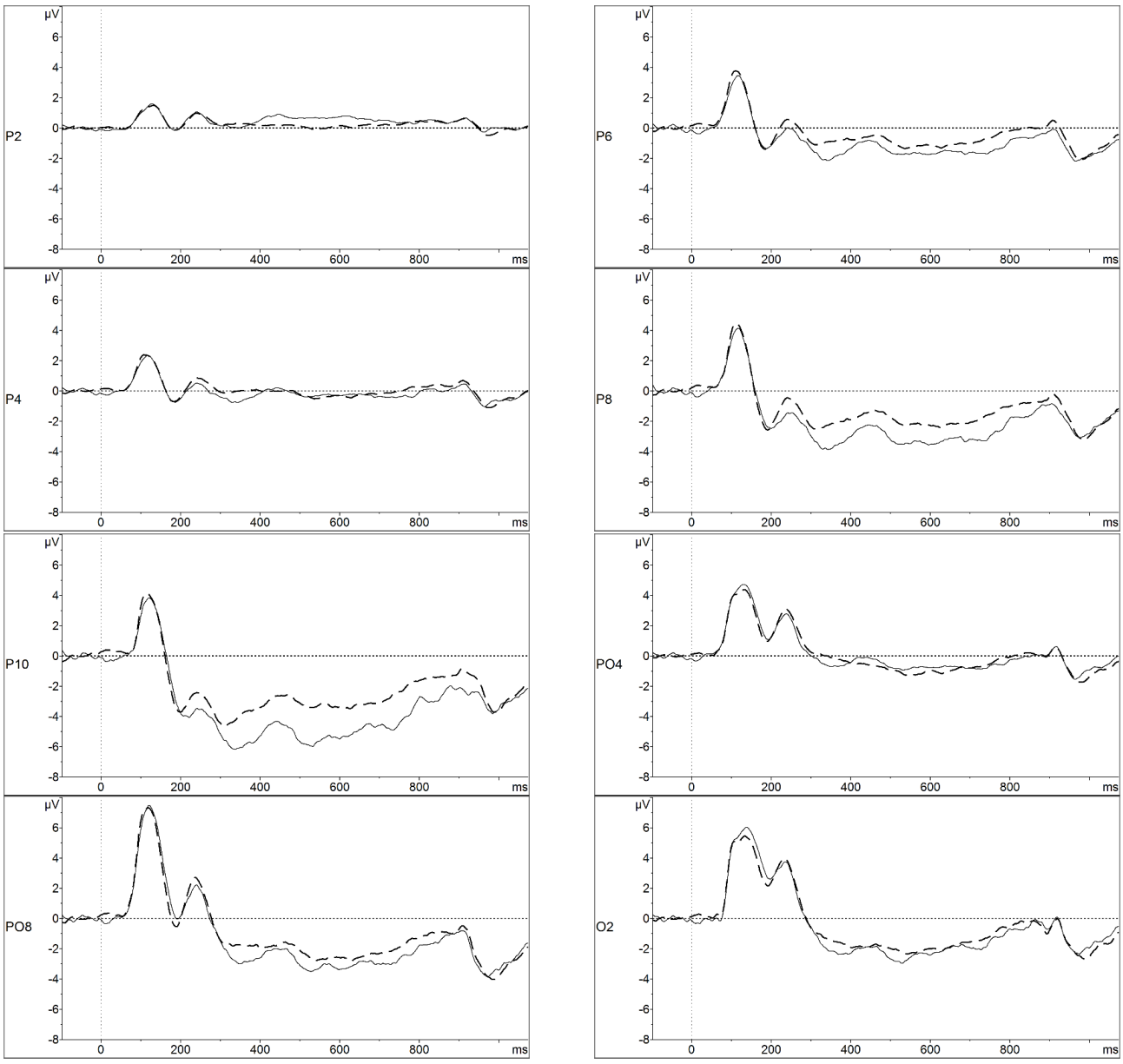



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

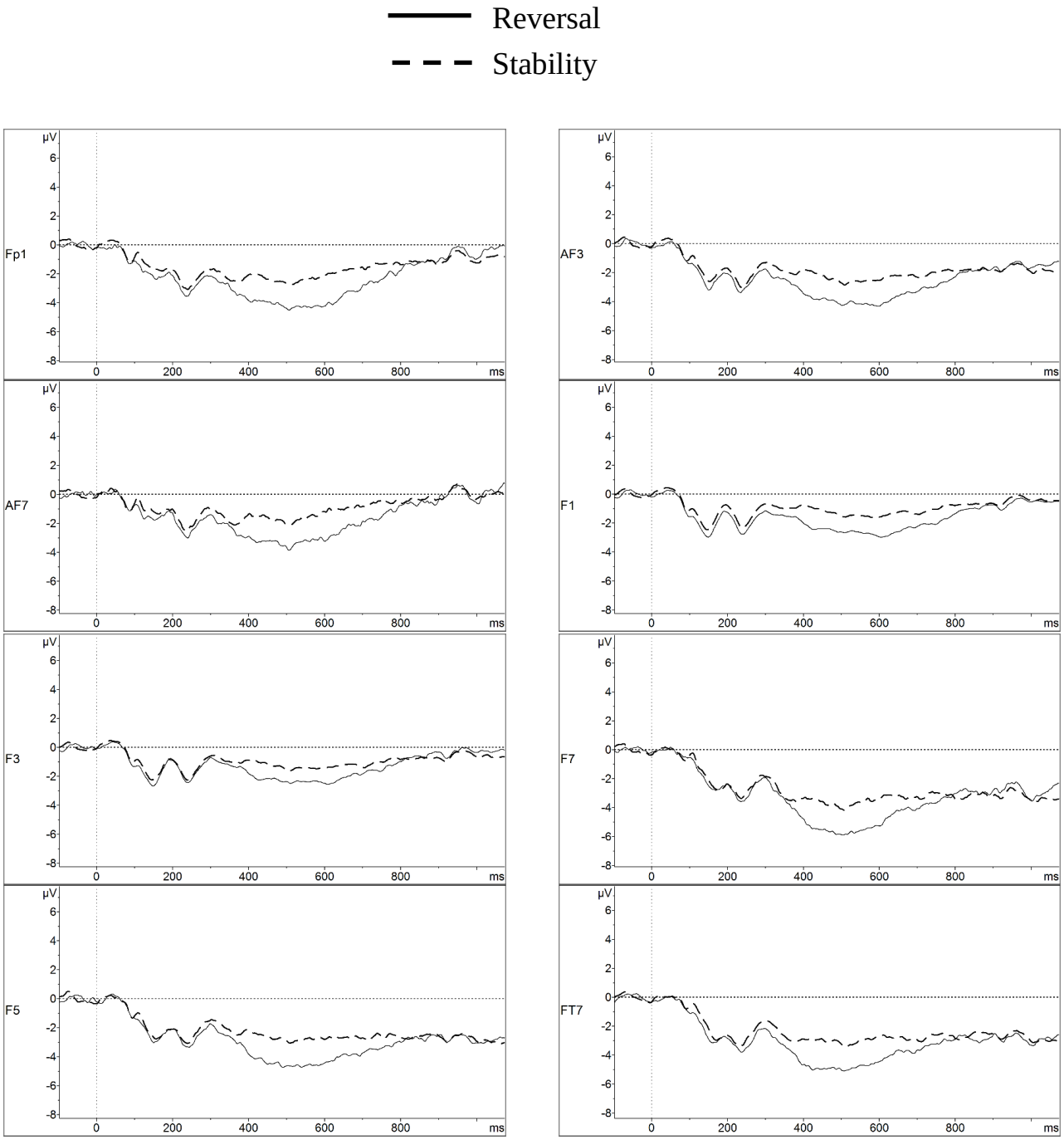

P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

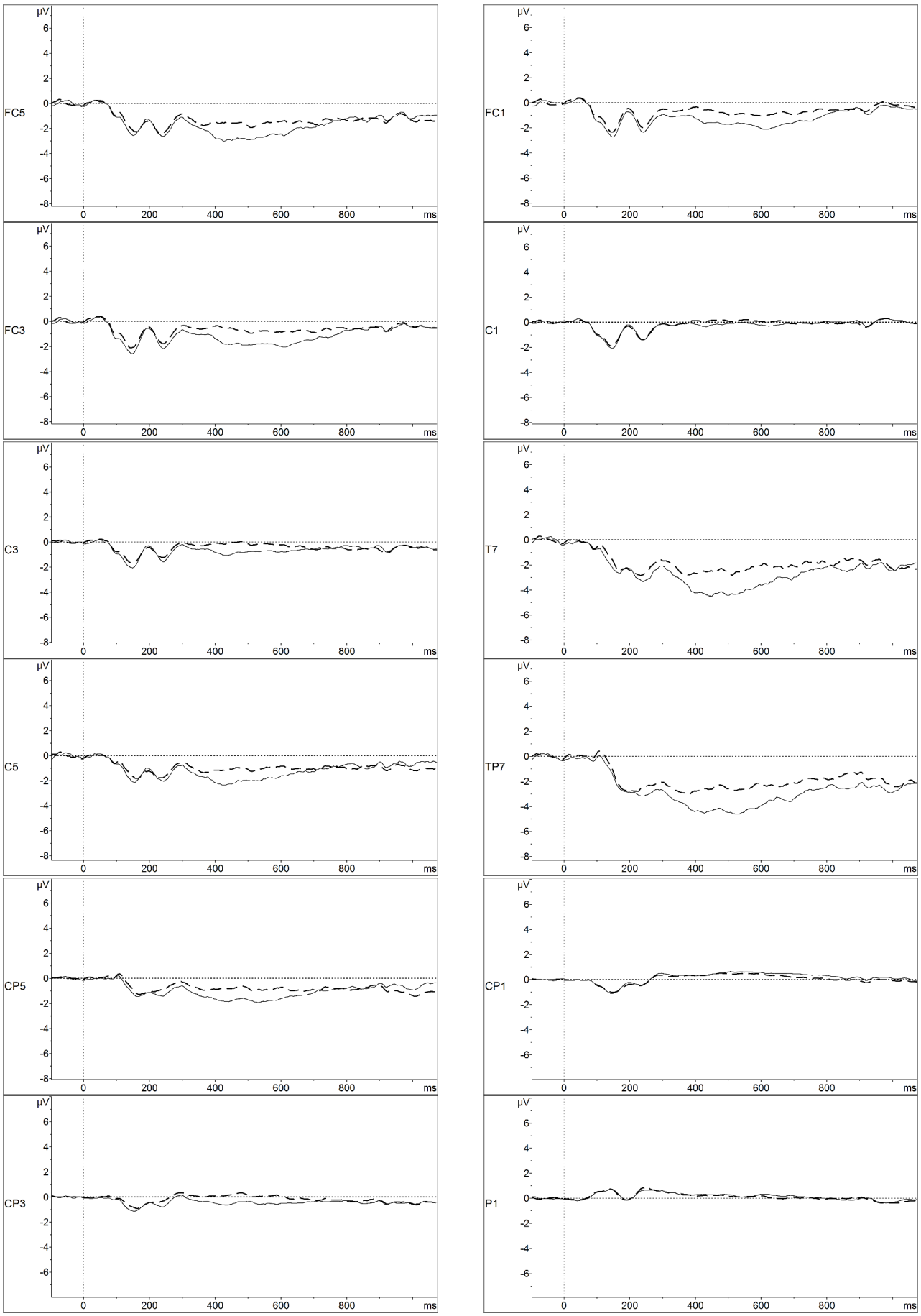

Fpz
AF8
Fp2
AF4
AFz
F2
Fz
F4
F6
FT8
F8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

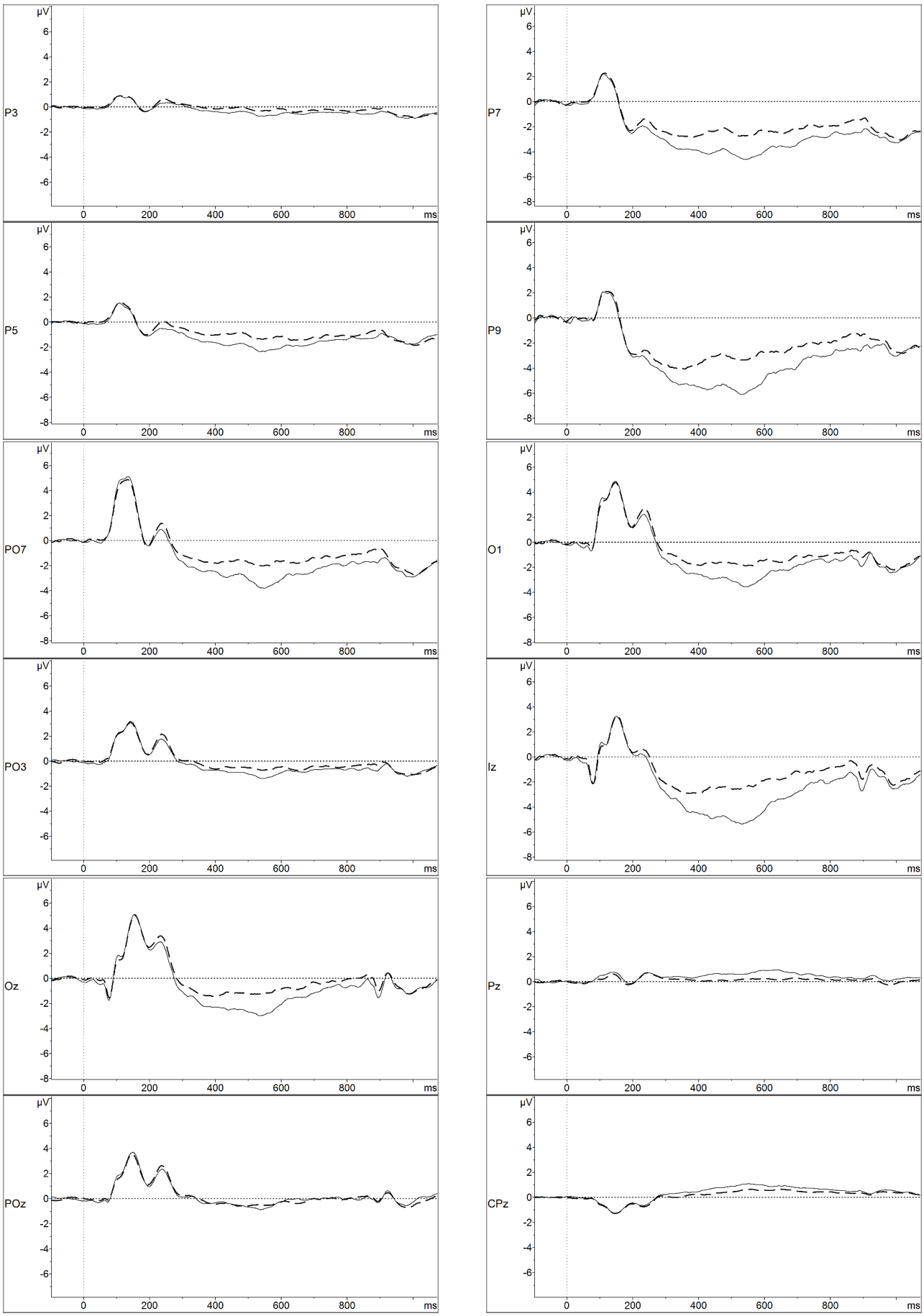

FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

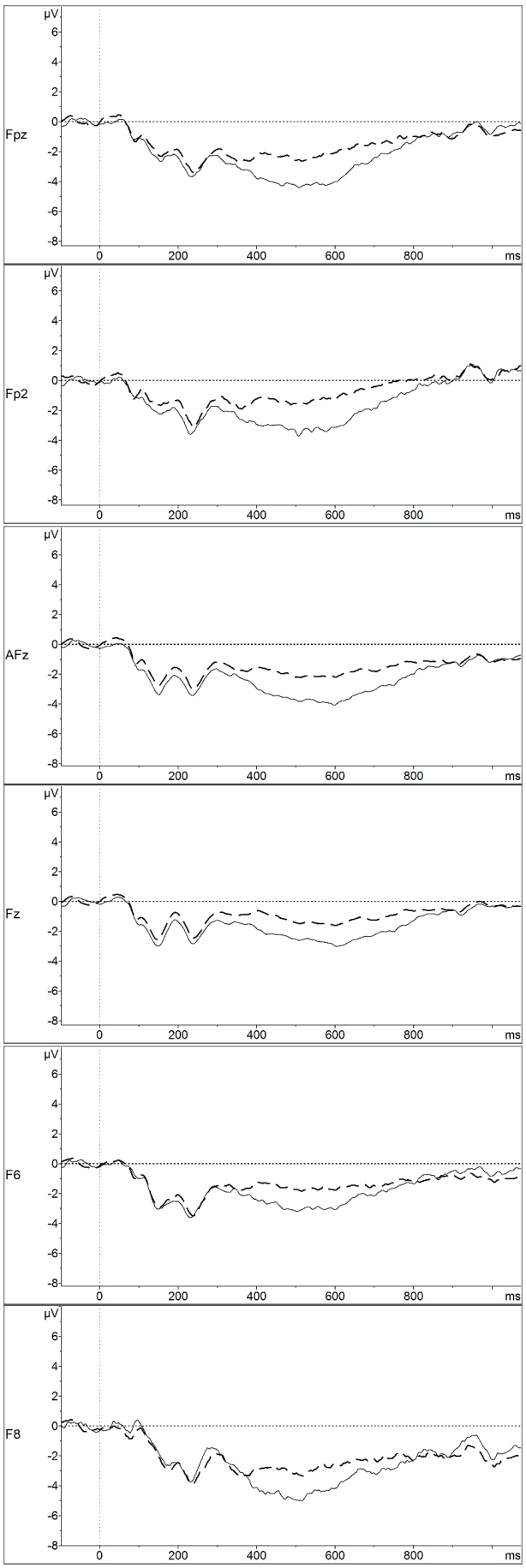

P2
P6
P4
P8
P10
PO4
PO8
O2
μV
ms
0
200
400
600
800
6
4
2
-2
-4
-6
-8

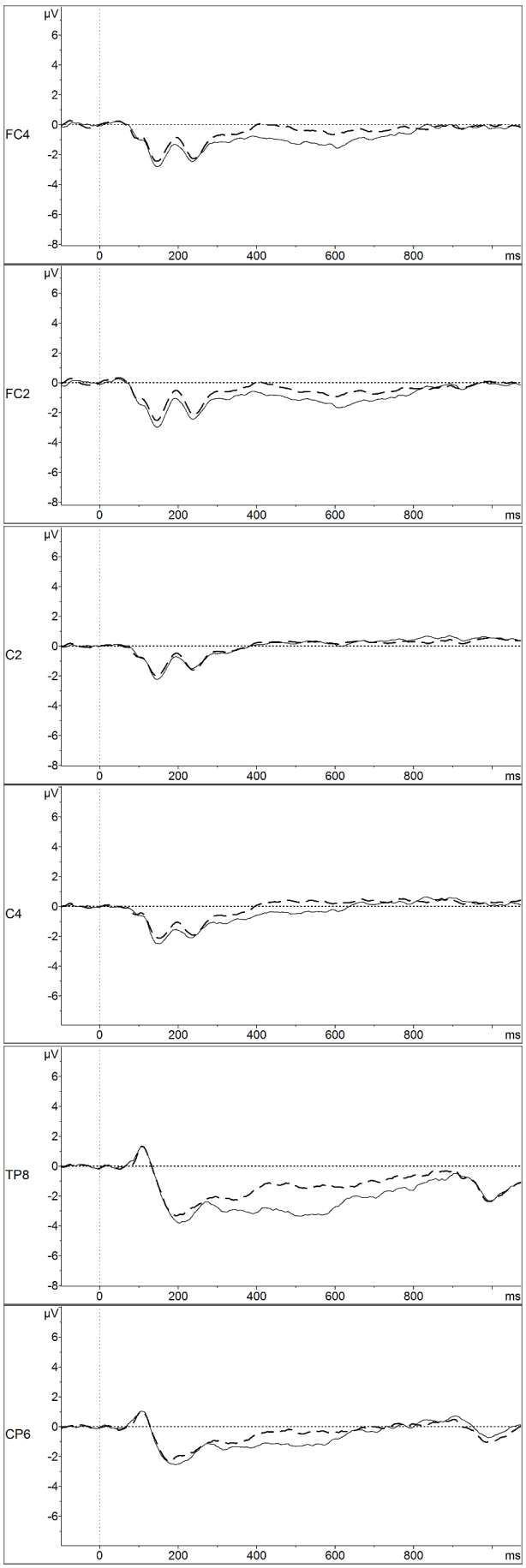
**(B) Identity Task**

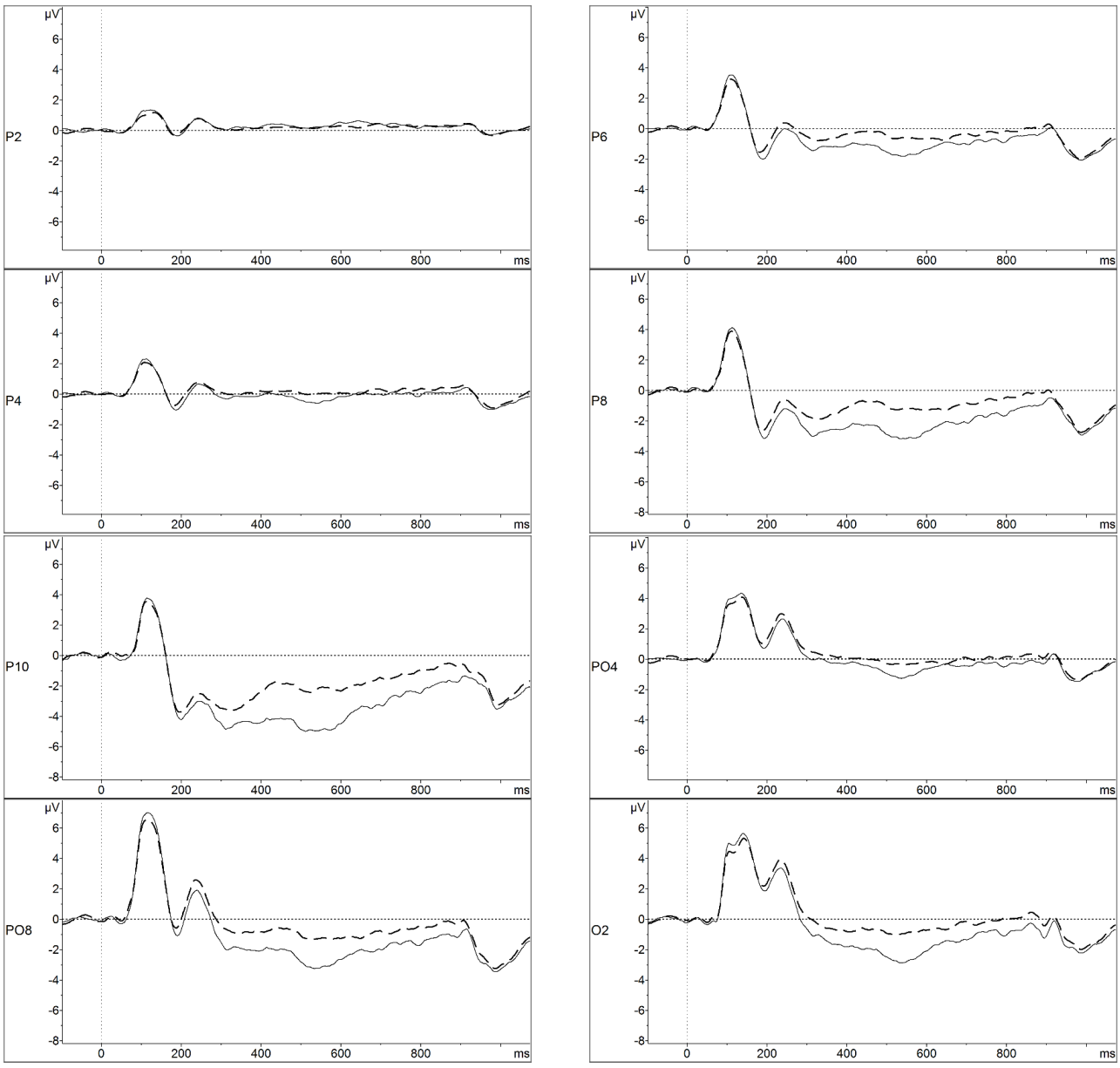



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

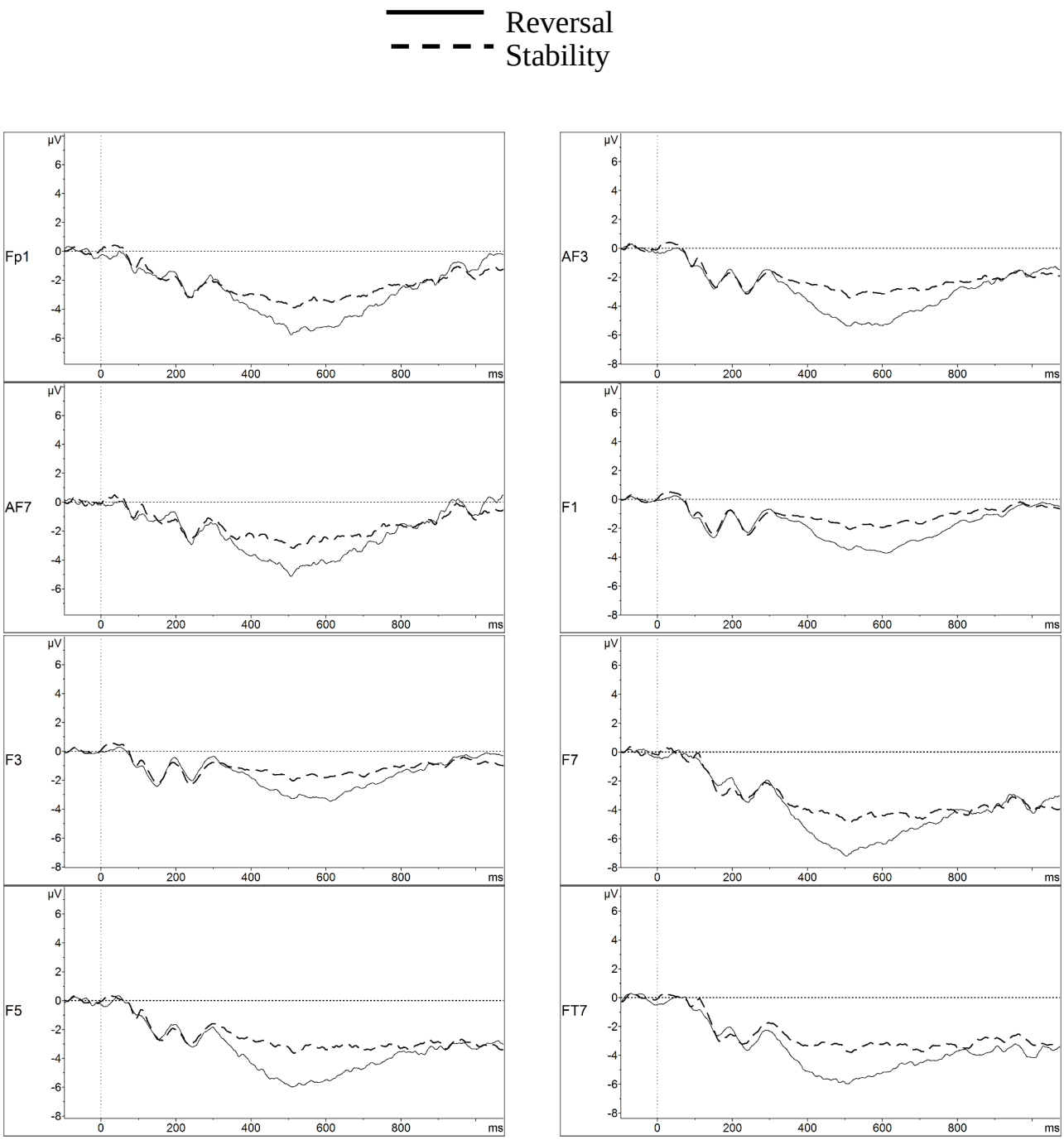

P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
μV
ms
0
200
400
600
800

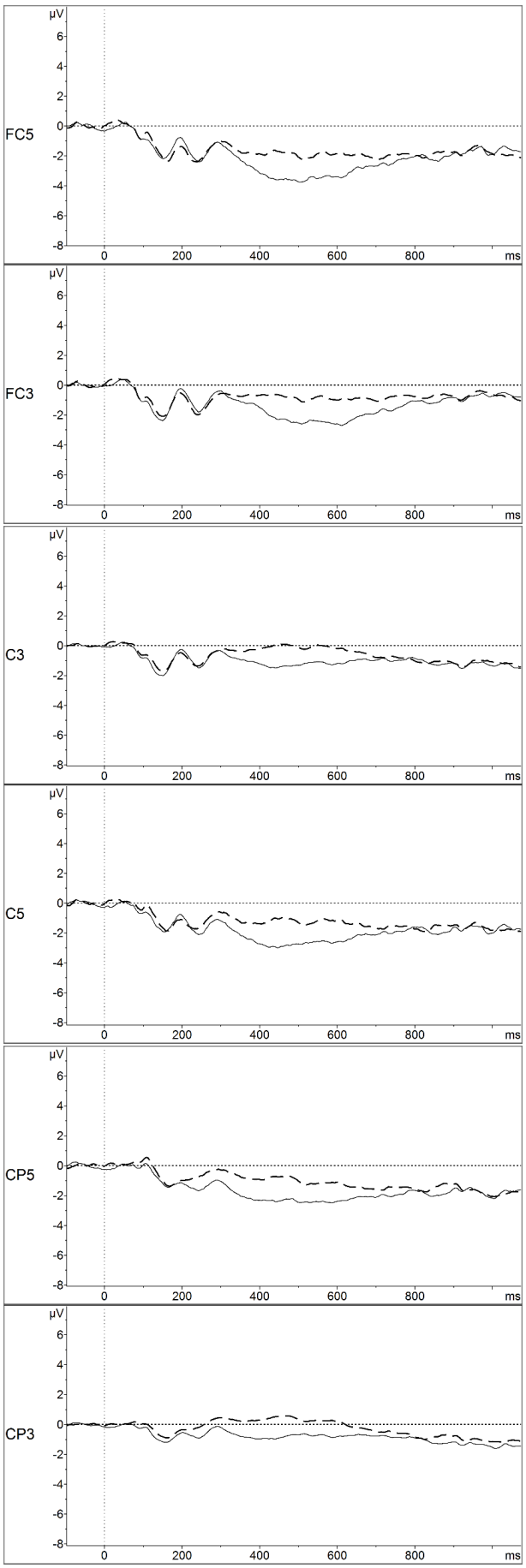

Fpz
Fp2
AFz
Fz
F6
F8
AF8
AF4
F2
F4
FT8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

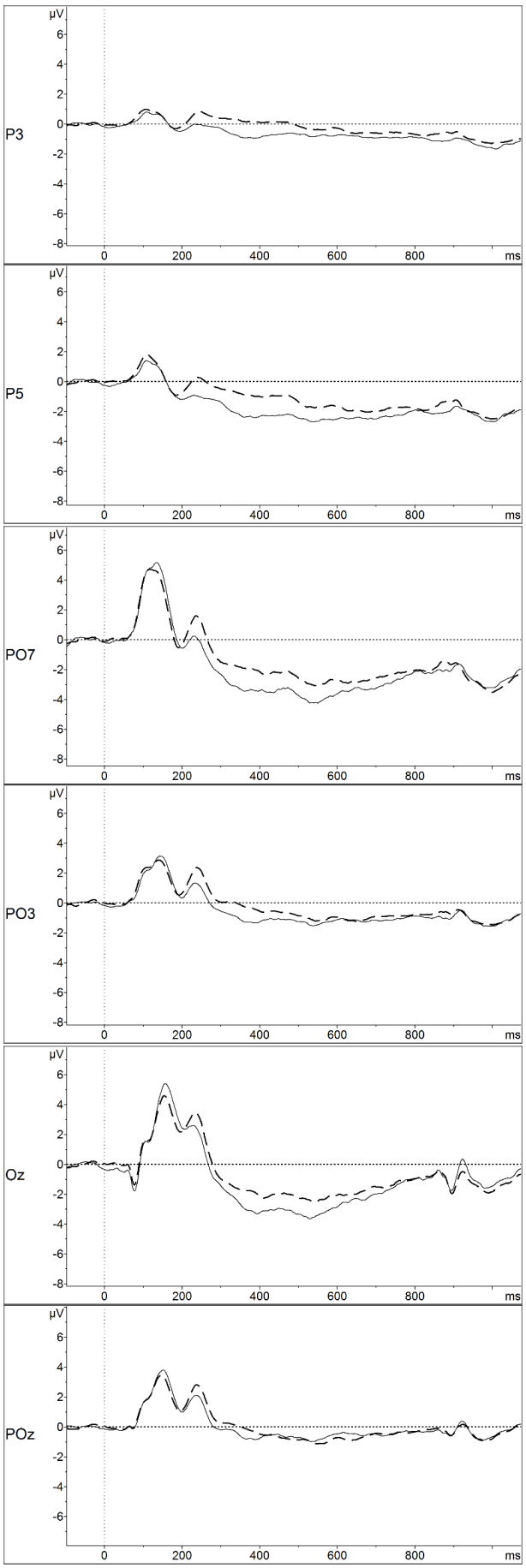

FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
μV
ms
0
200
400
600
800

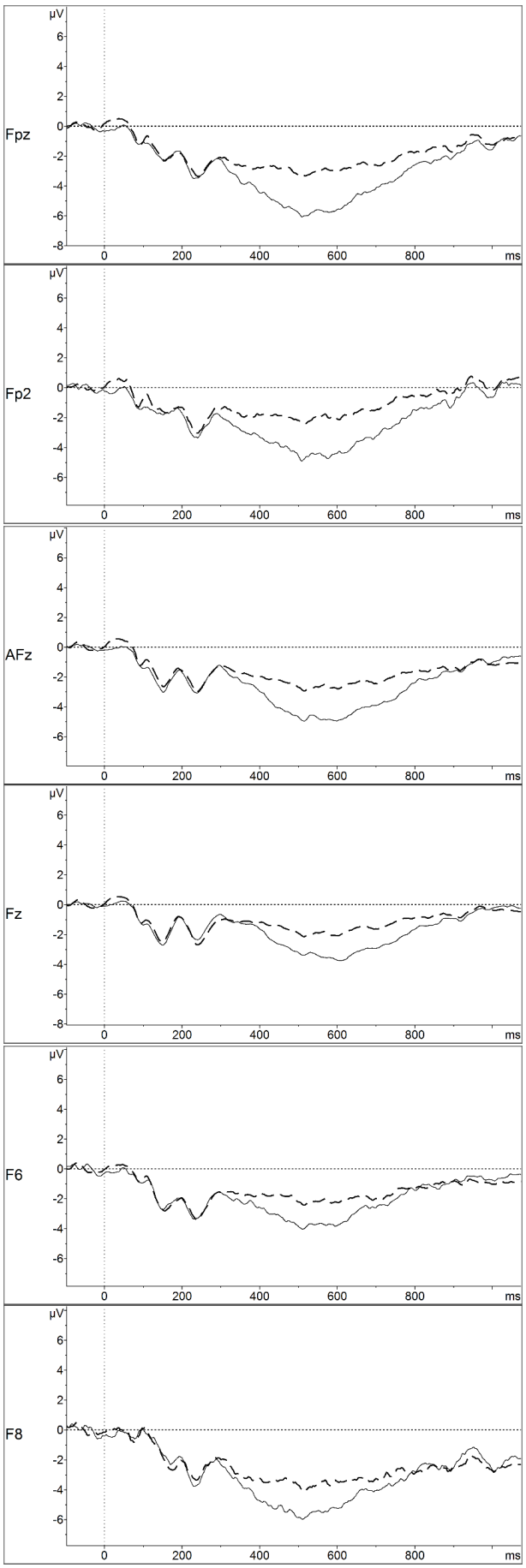

P2
P6
P4
P8
P10
PO4
PO8
O2
μV
ms
0
200
400
600
800

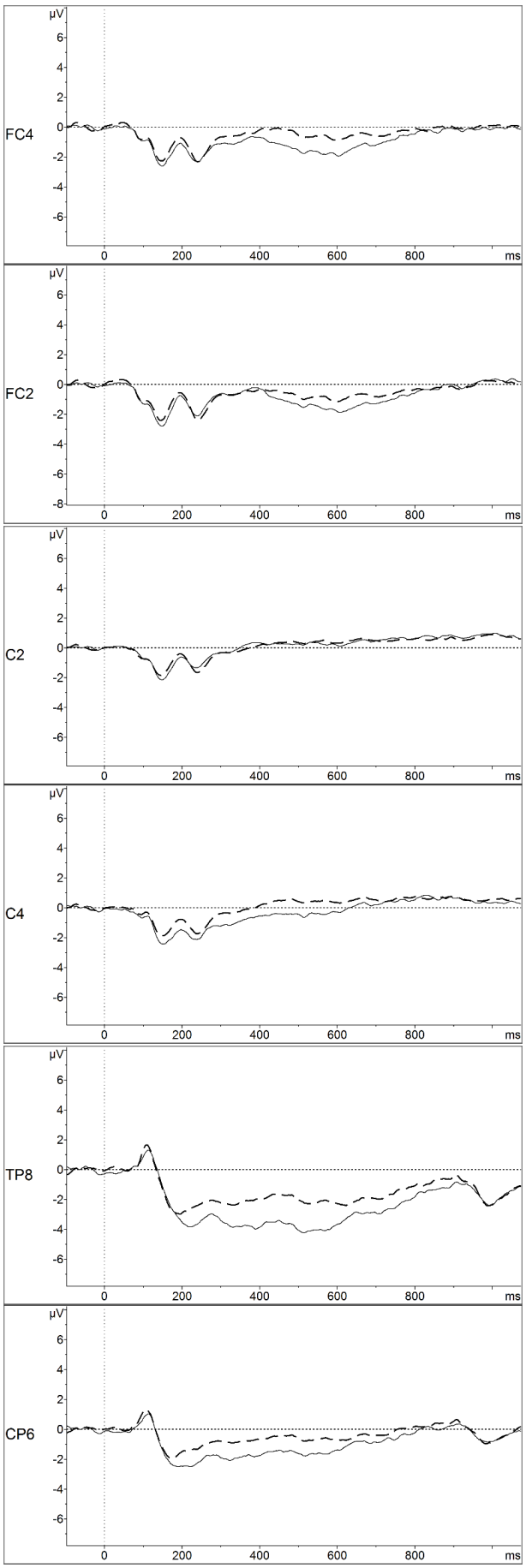
**(C) Response Trials**

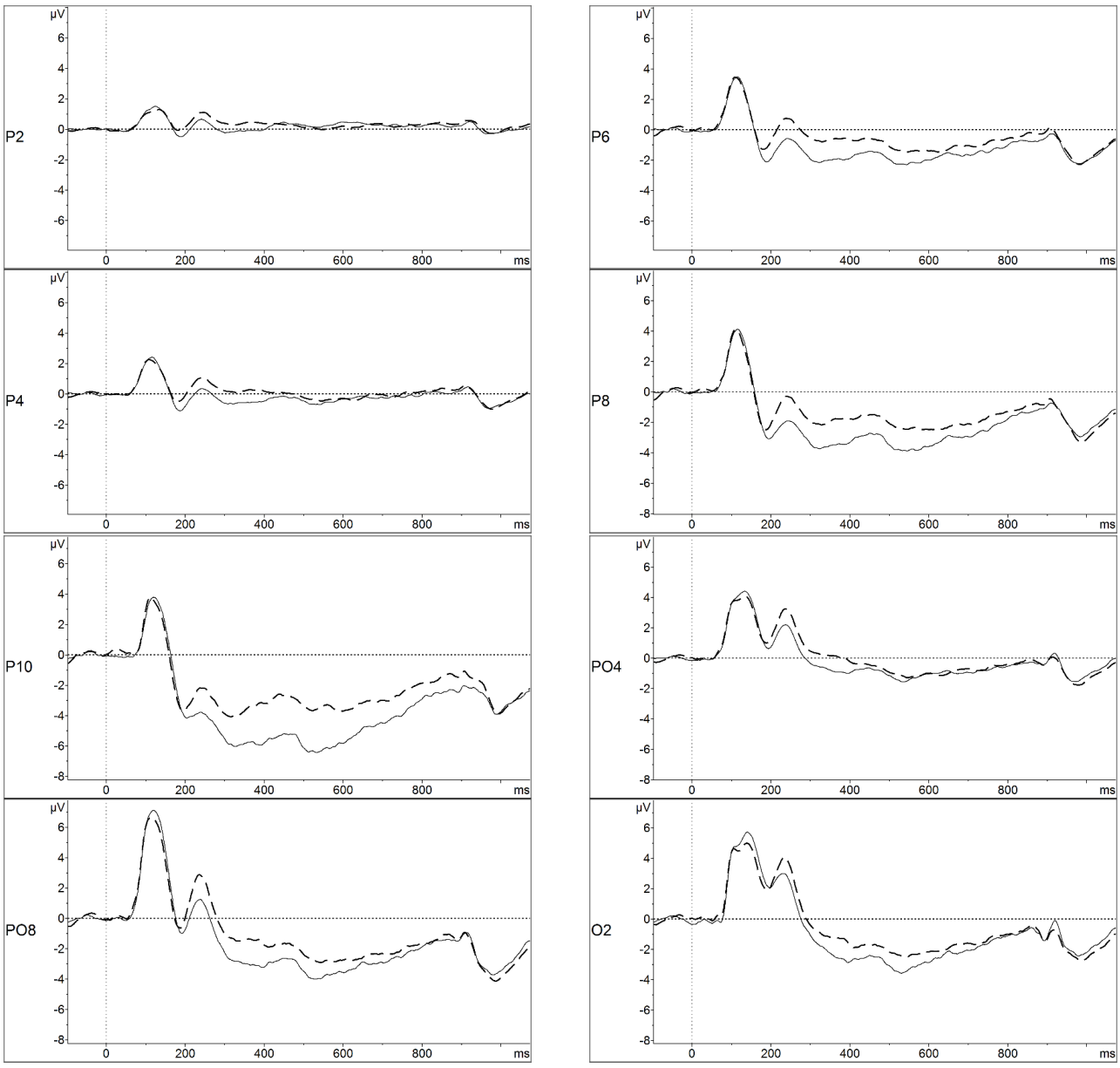



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

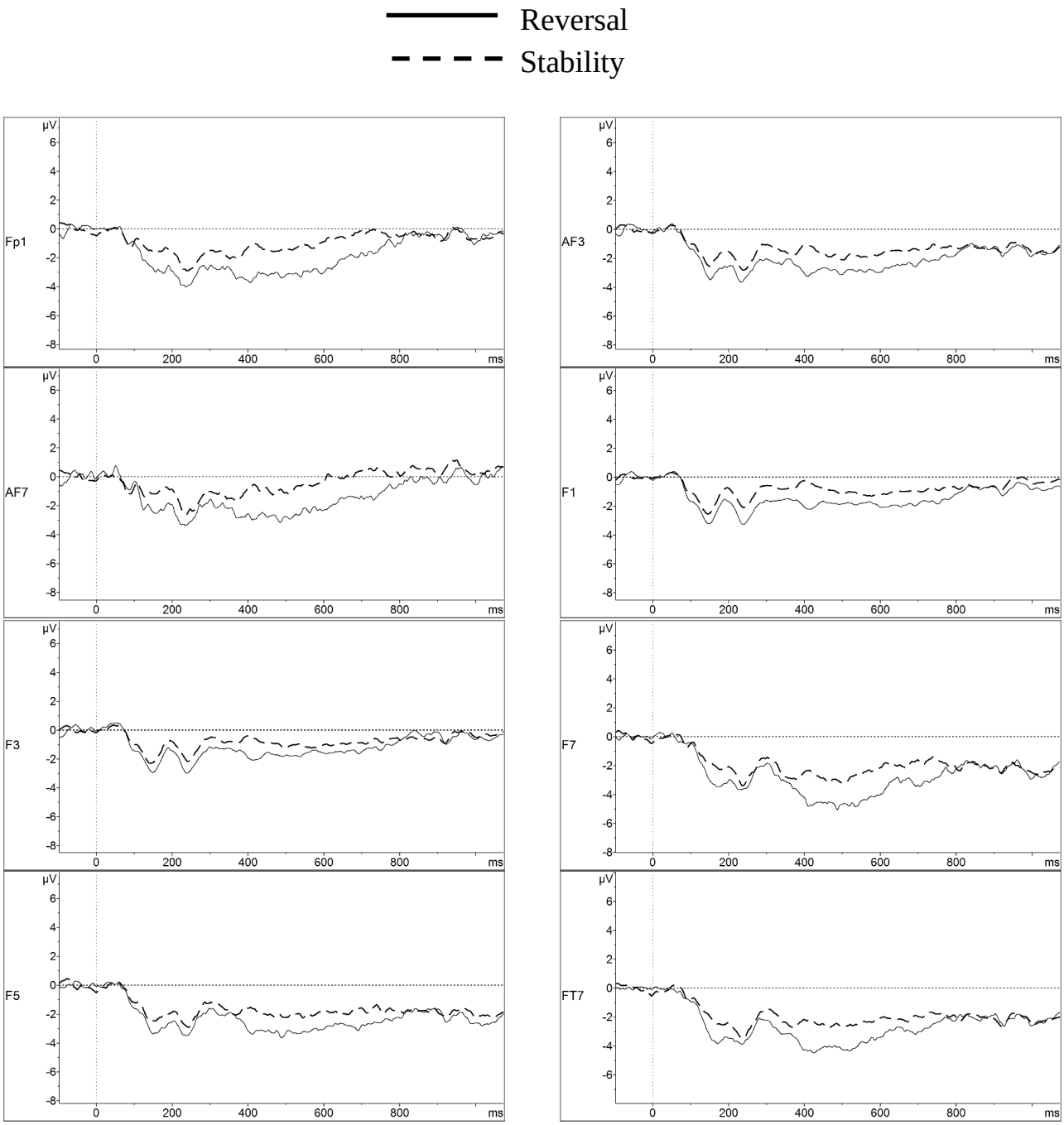

P3
P5
PO7
PO3
Oz
POz
P7
P9
O1
Iz
Pz
CPz
µV
ms
0
200
400
600
800

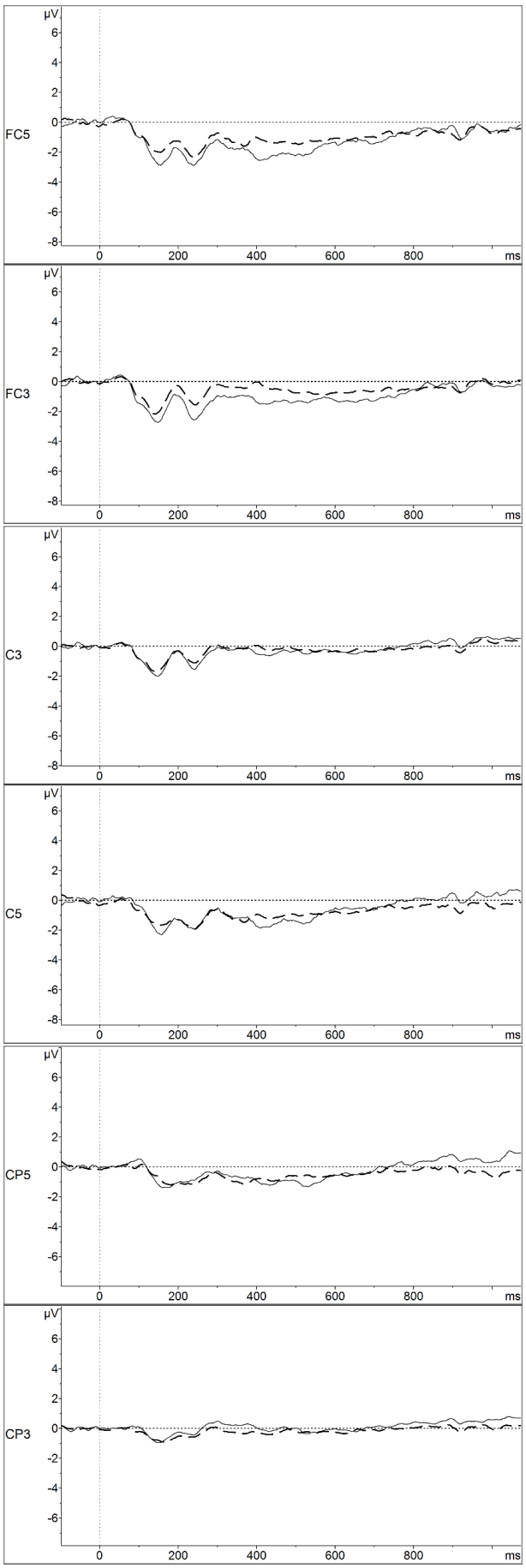

Fpz
AF8
Fp2
AF4
AFz
F2
Fz
F4
F6
FT8
F8
FC6
μV
ms
0
200
400
600
800

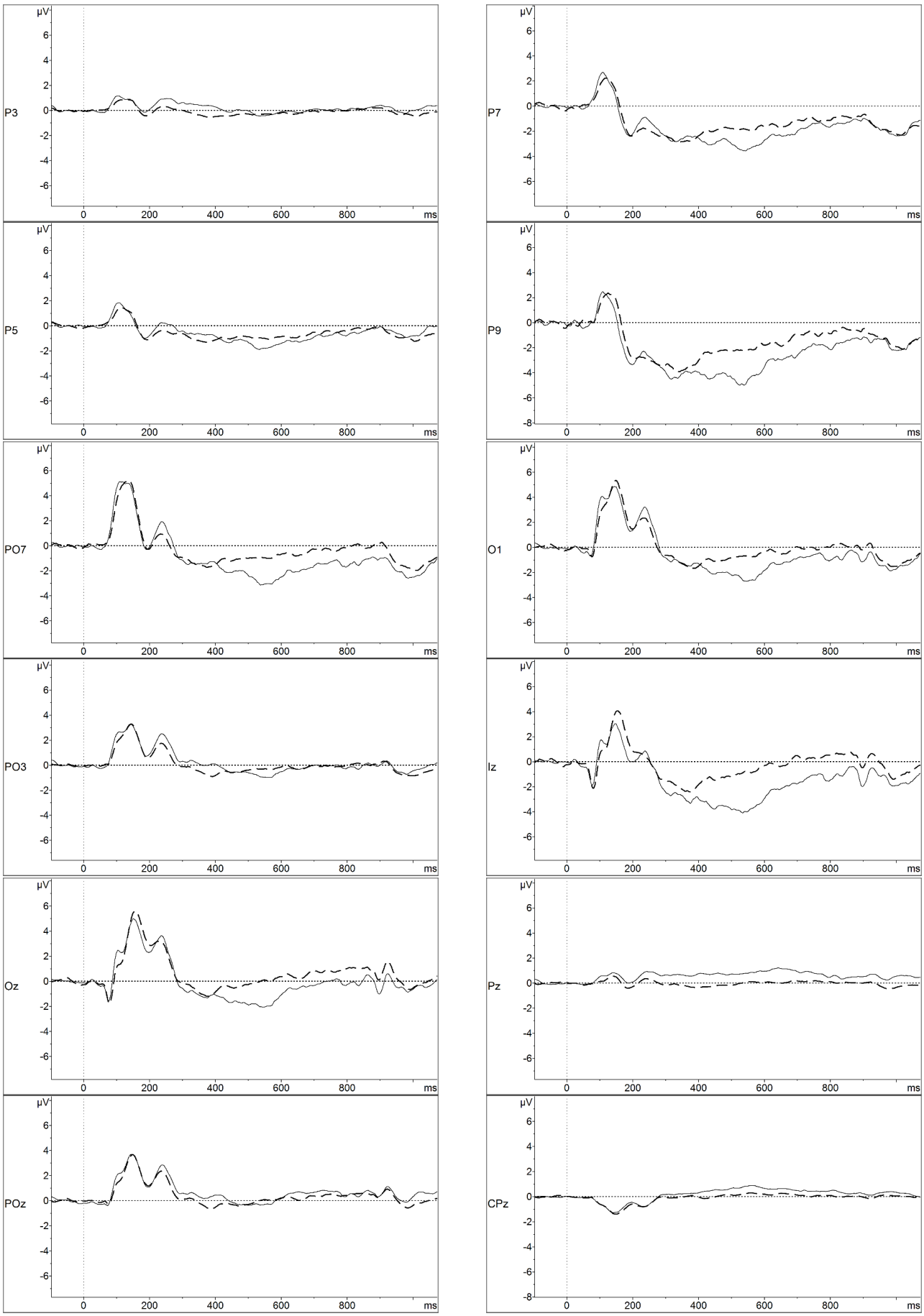

FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
μV
ms
0
200
400
600
800

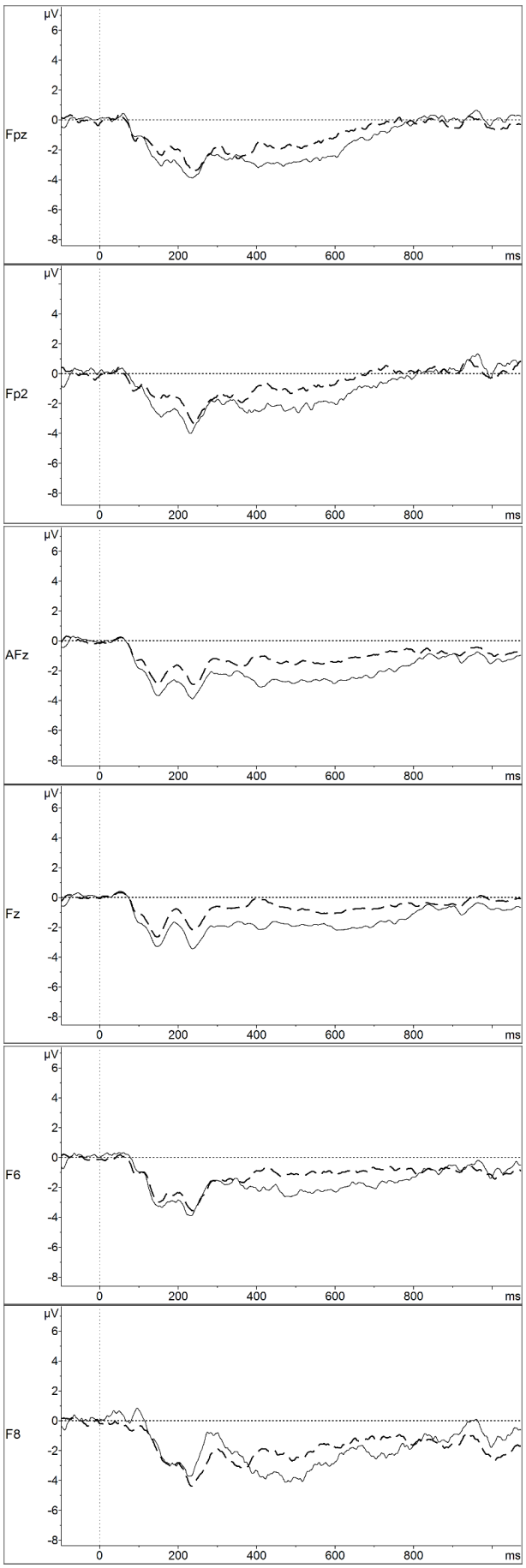

P2
P6
P4
P8
P10
PO4
PO8
O2
μV
ms
0
200
400
600
800

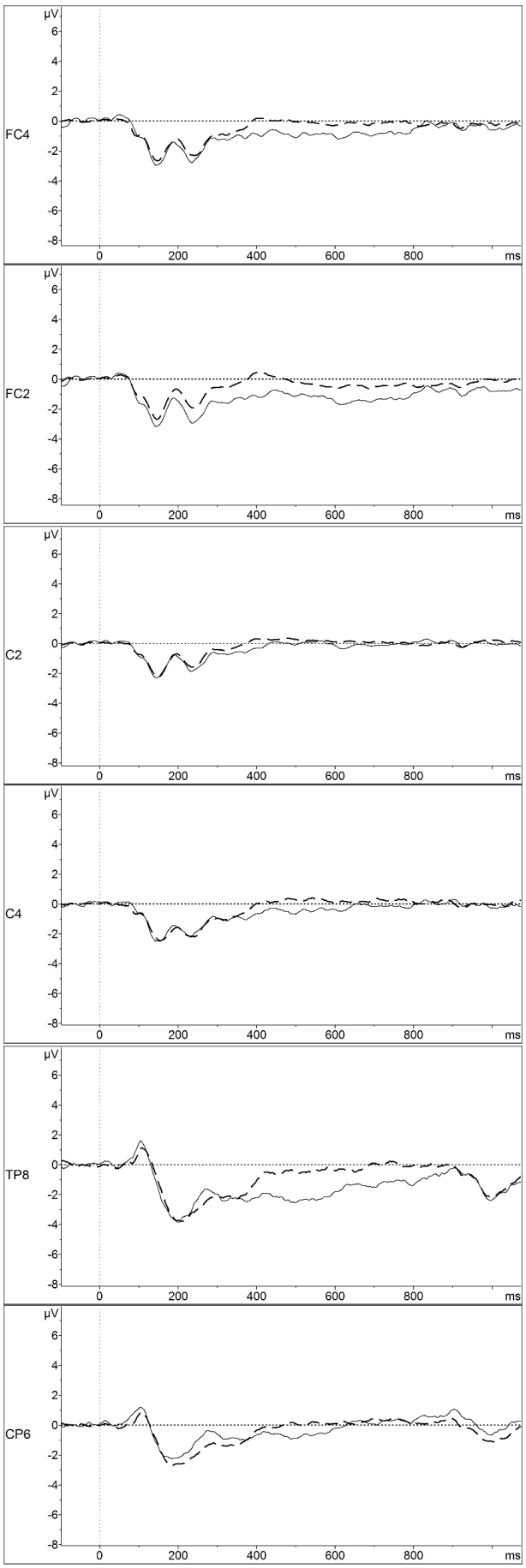
**(D) Non-Response Trials**

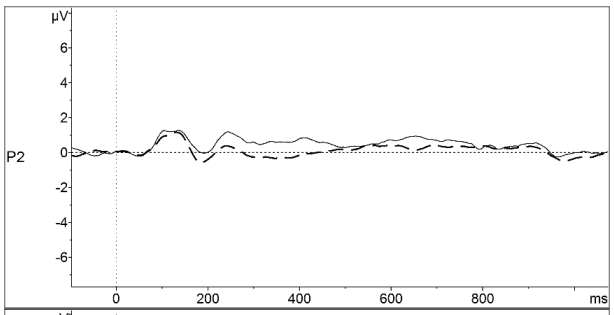


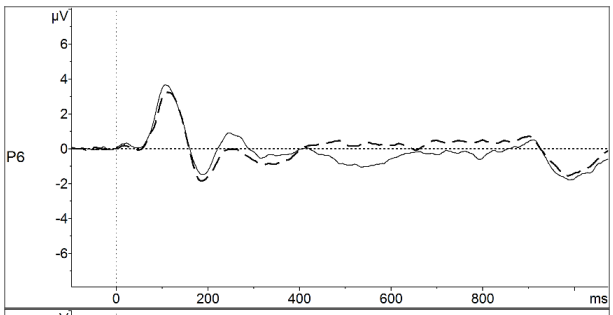

FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

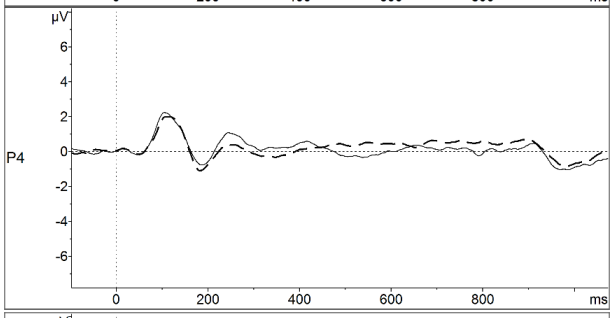

P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
μV
ms
0
200
400
600
800

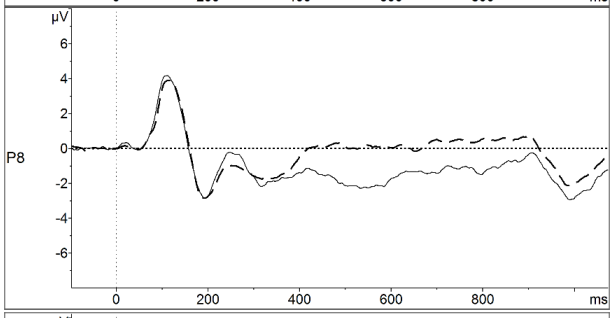

Fpz
AF8
Fp2
AF4
AFz
F2
Fz
F4
F6
FT8
F8
FC6
µV
ms
0
200
400
600
800

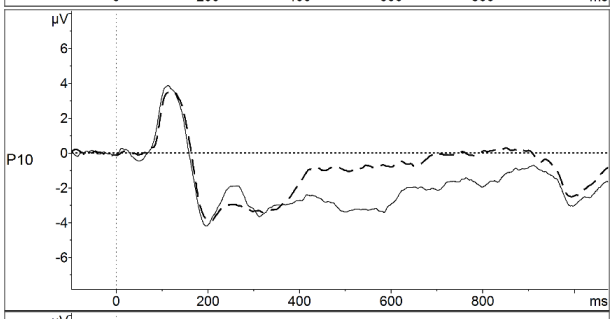

FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

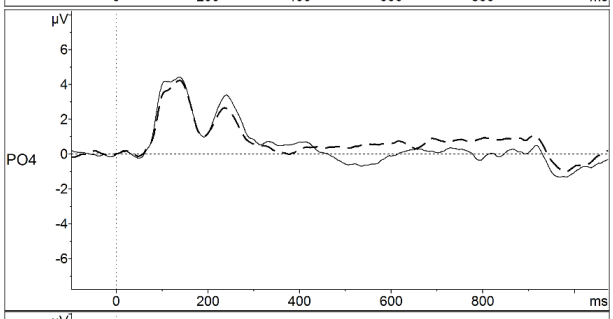

P2
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms

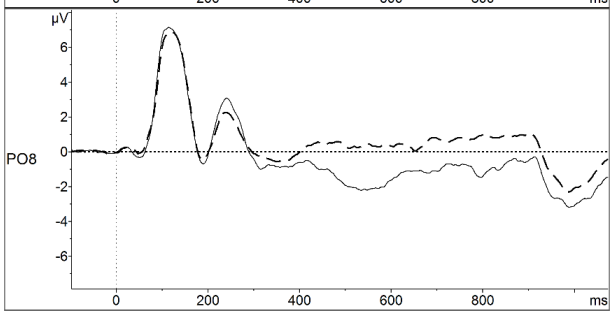

P6
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms


P4
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms


P8
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms


P10
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms


PO4
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms


PO8
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms

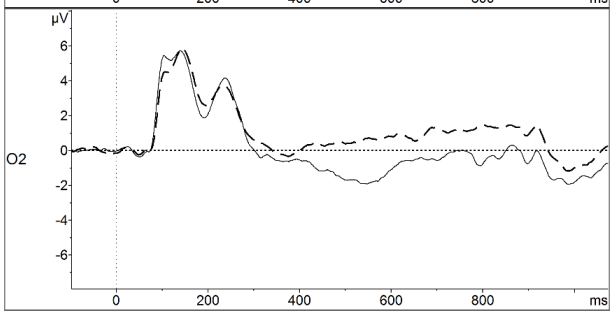
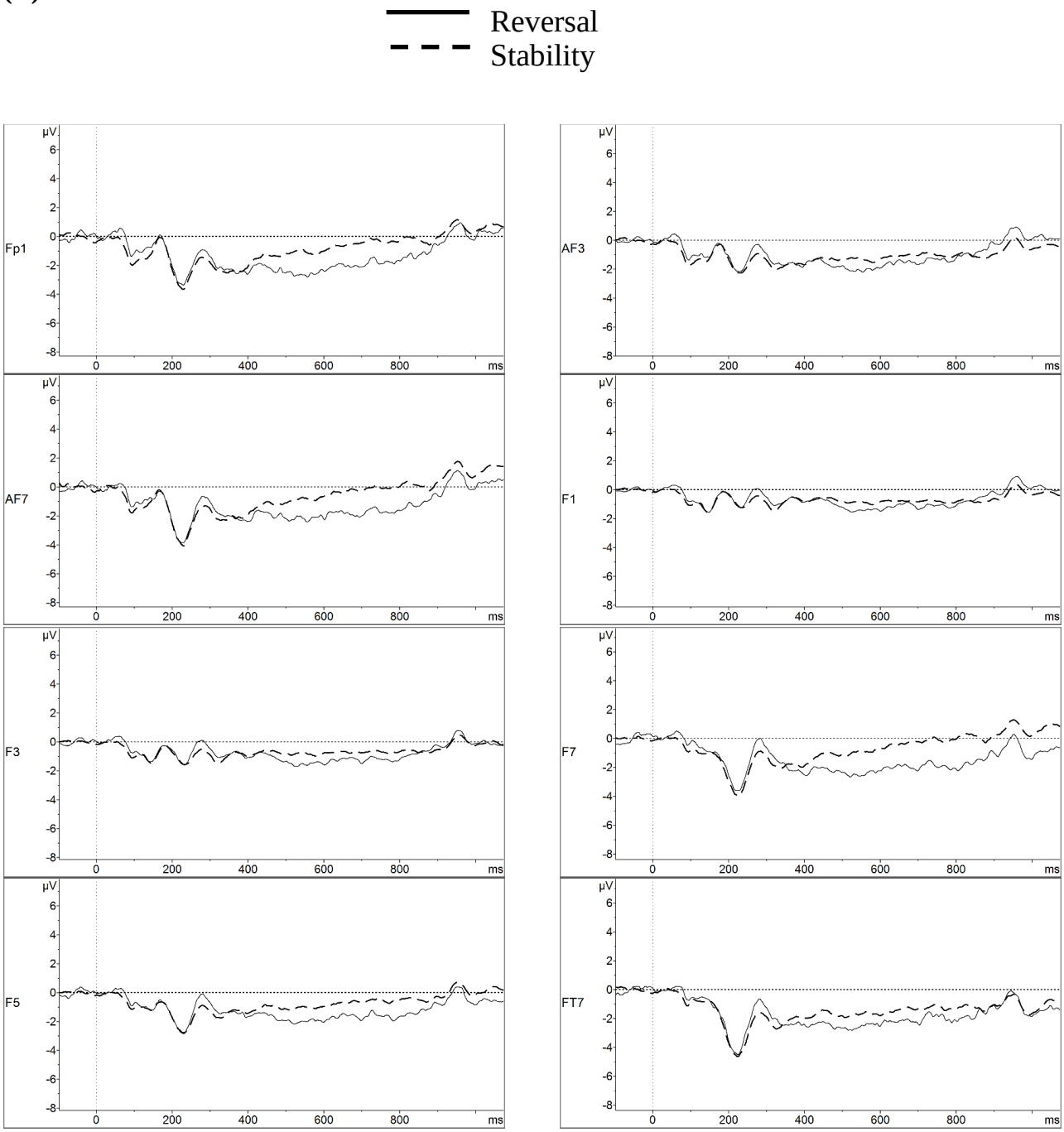
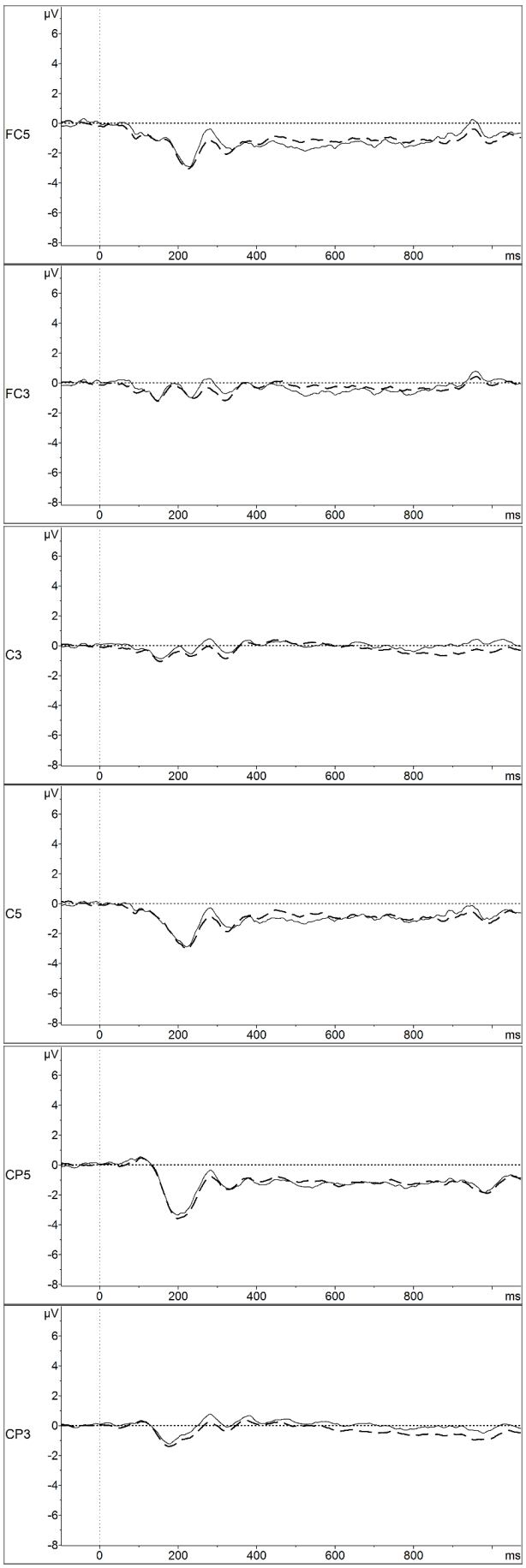

O2
µV
6
4
2
0
-2
-4
-6
0
200
400
600
800
ms

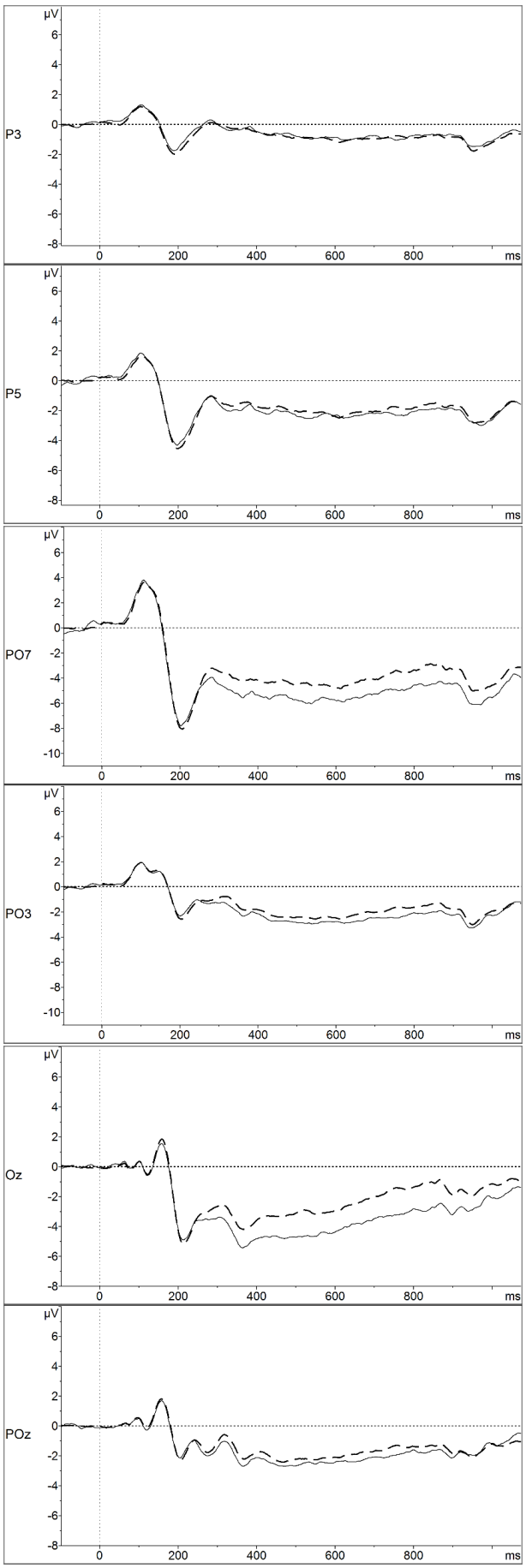
**Figure S7: Experiment 2 (Faces-Vase) Grand Average ERP plots for all electrodes comparing Reversal vs. Stable trials in the Reversal Task**

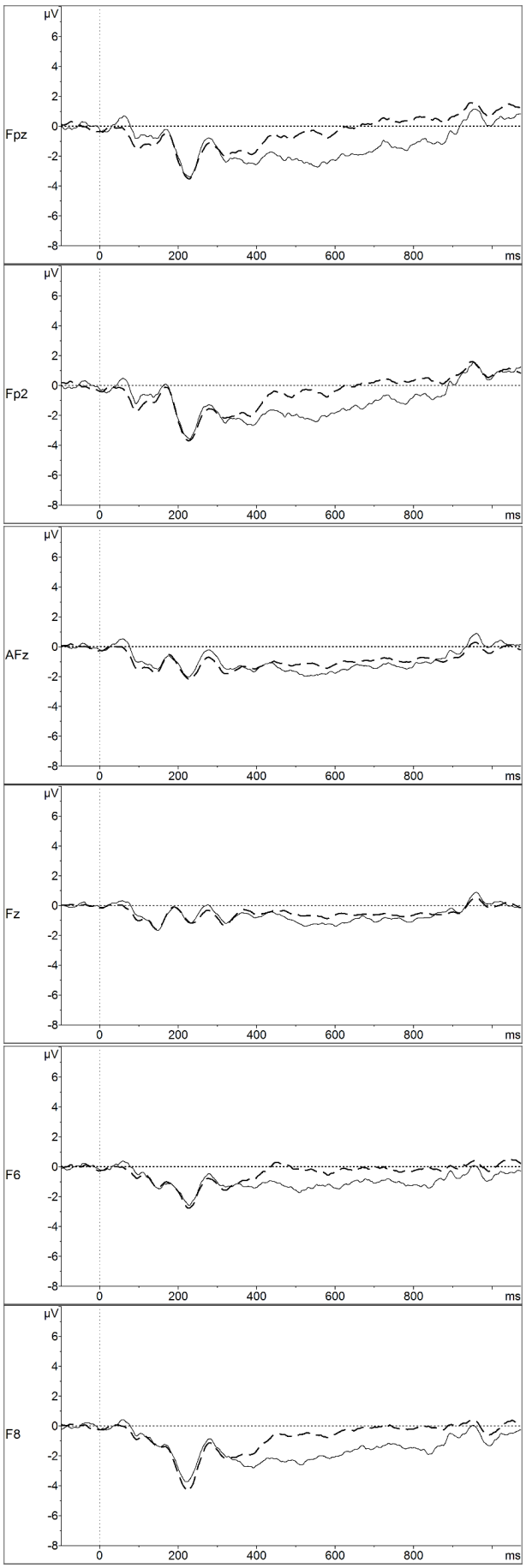
Each panel plots the grand mean ERP signals for reversal (solid line) and stability (dashed line) trials in Experiment 2 (Faces-Vase) from all recorded scalp locations. This is done separately for different conditions aligned to those presented in Figure 6: (a) ERP waveforms in the reversal task (collapsed over response); (b) ERP waveforms in the identity task (collapsed over response); (c) ERP waveforms for the response trials (collapsed over task); (d) ERP waveforms for the non-response trials (collapsed over task).

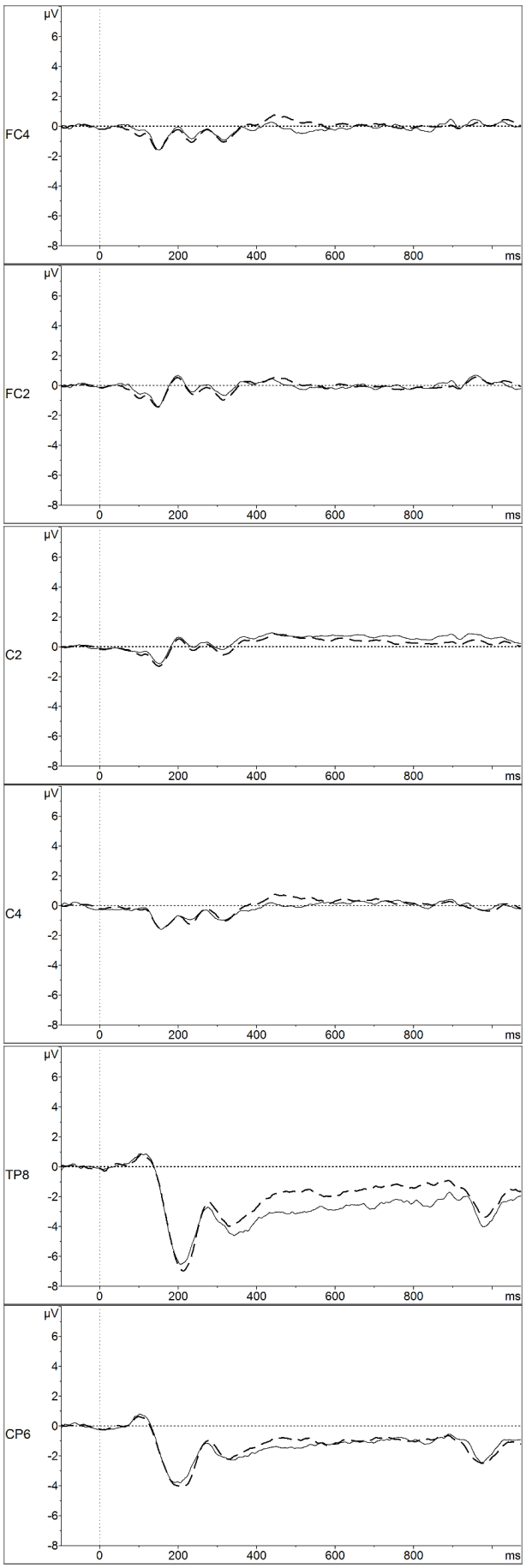
**(A) Reversal Task**

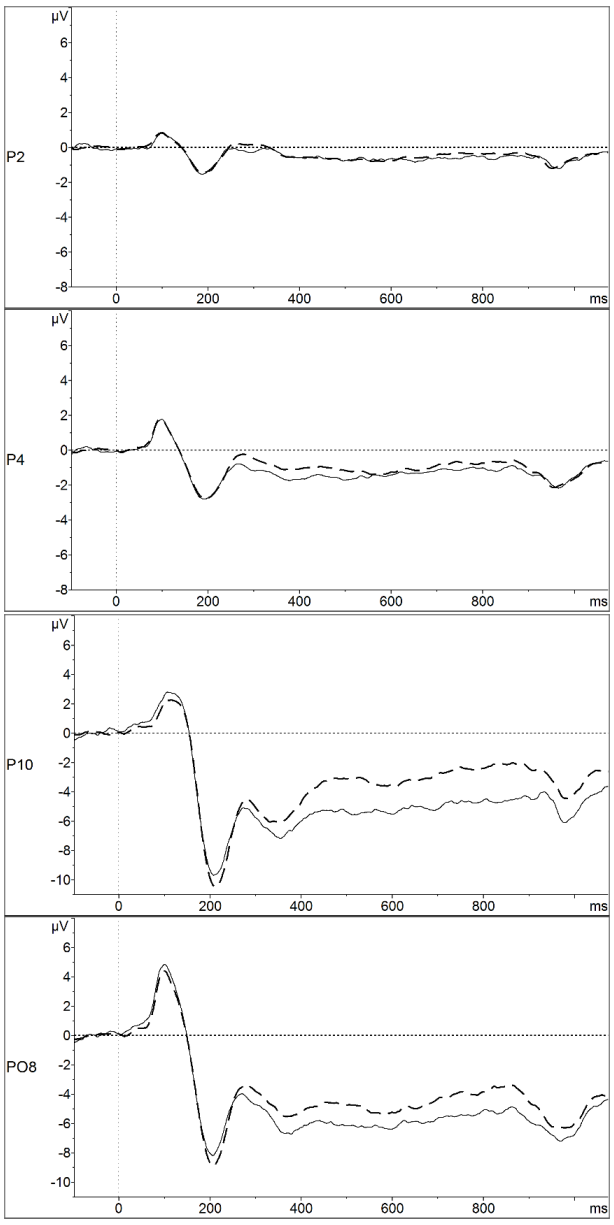



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

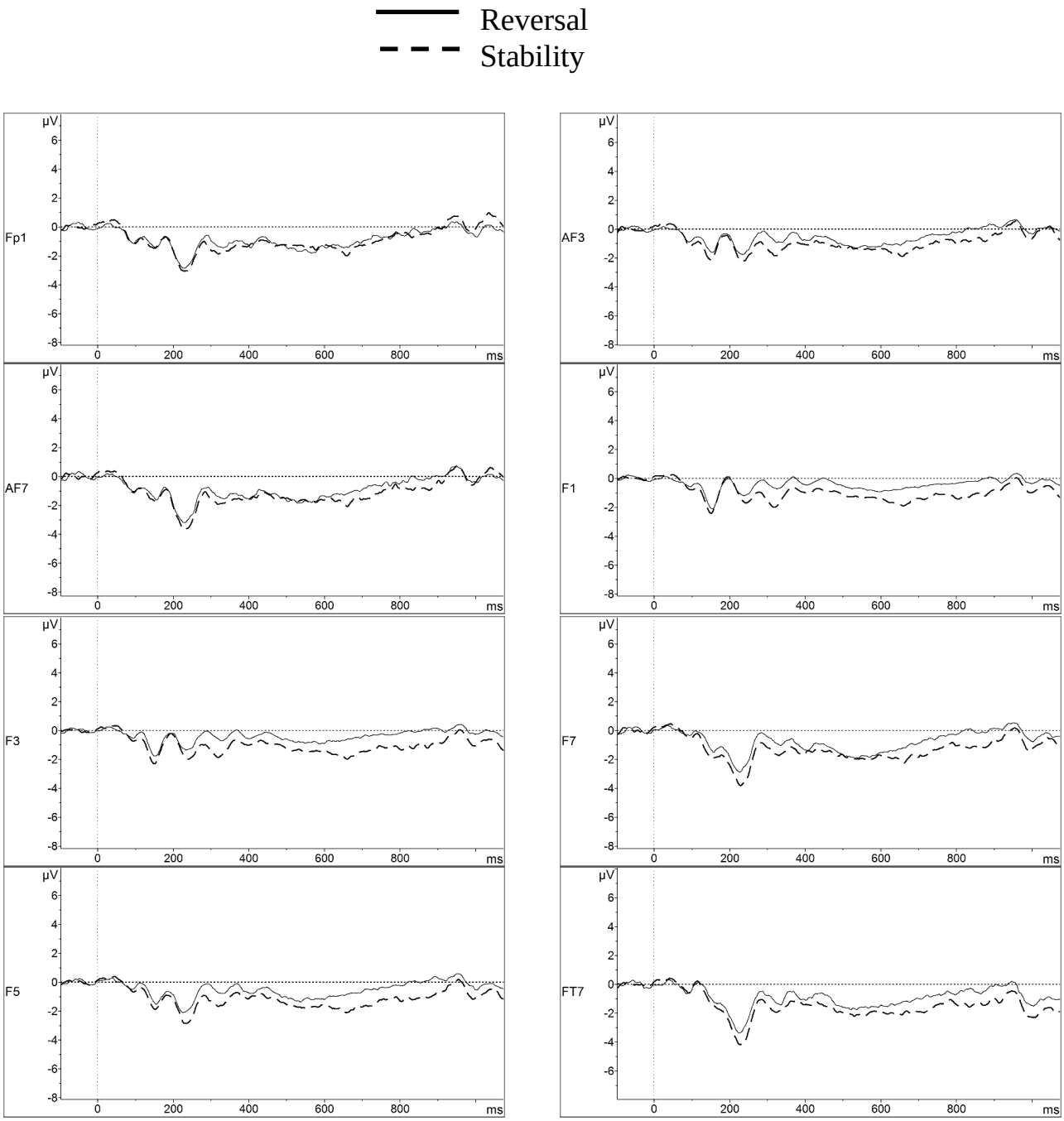

P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
µV
ms
0
200
400
600
800

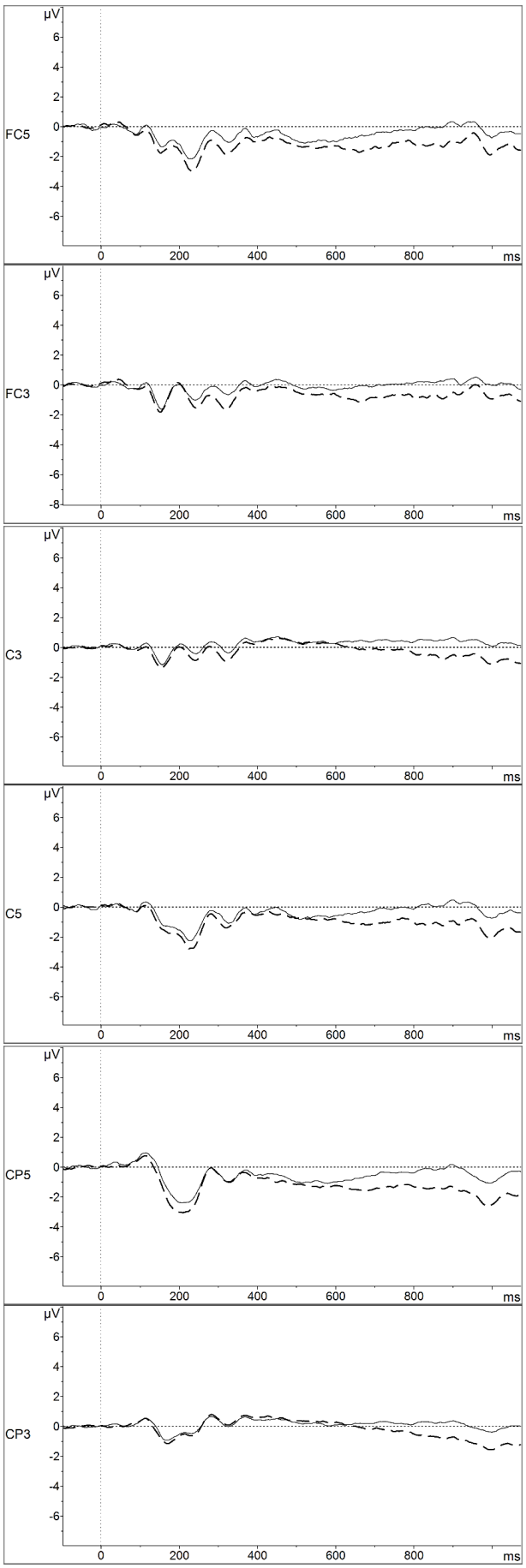

Fpz
Fp2
AFz
Fz
F6
F8
AF8
AF4
F2
F4
FT8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

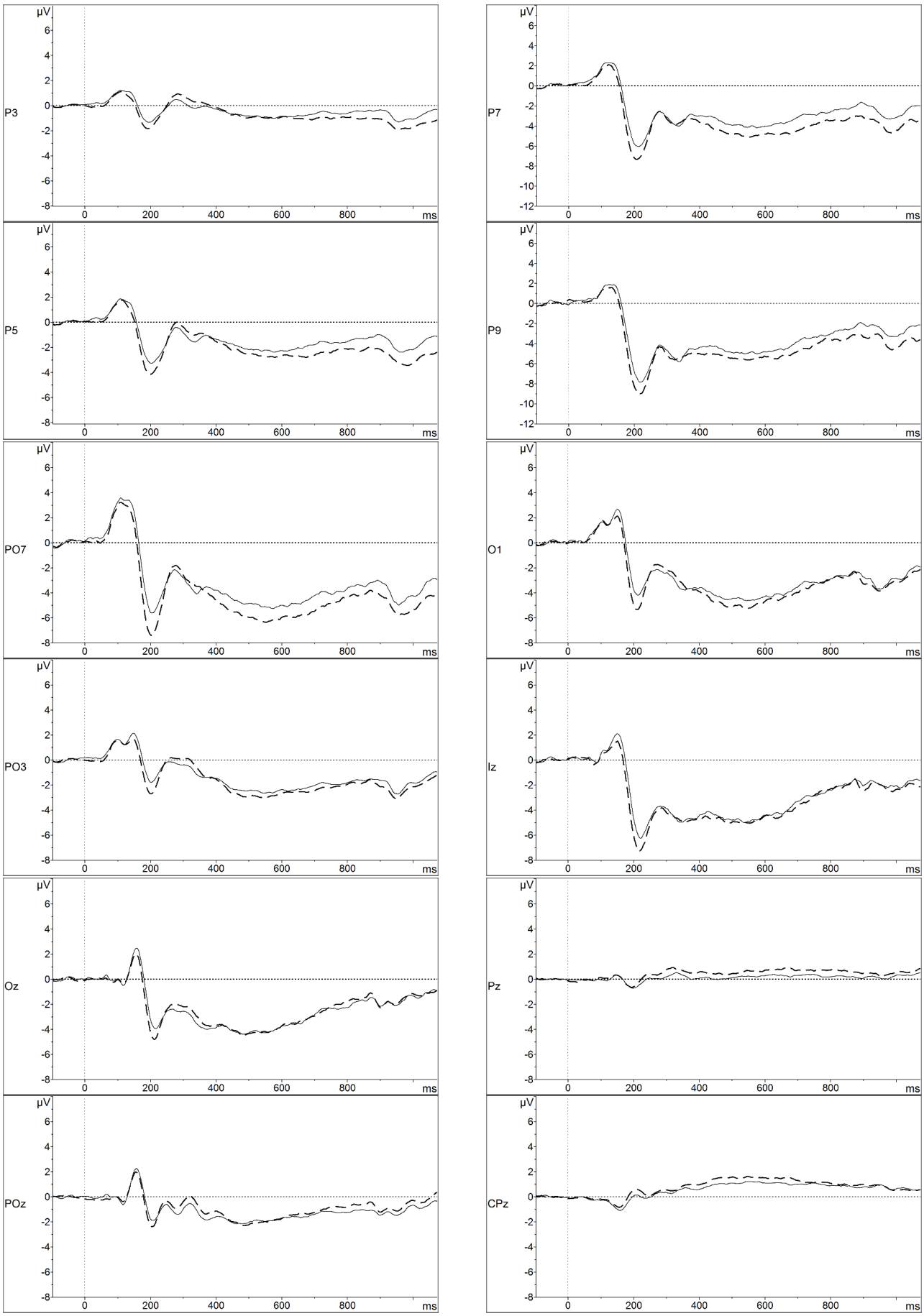

FC4
FC2
C2
C4
TP8
CP6
FCz
Cz
C6
T8
CP4
CP2
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

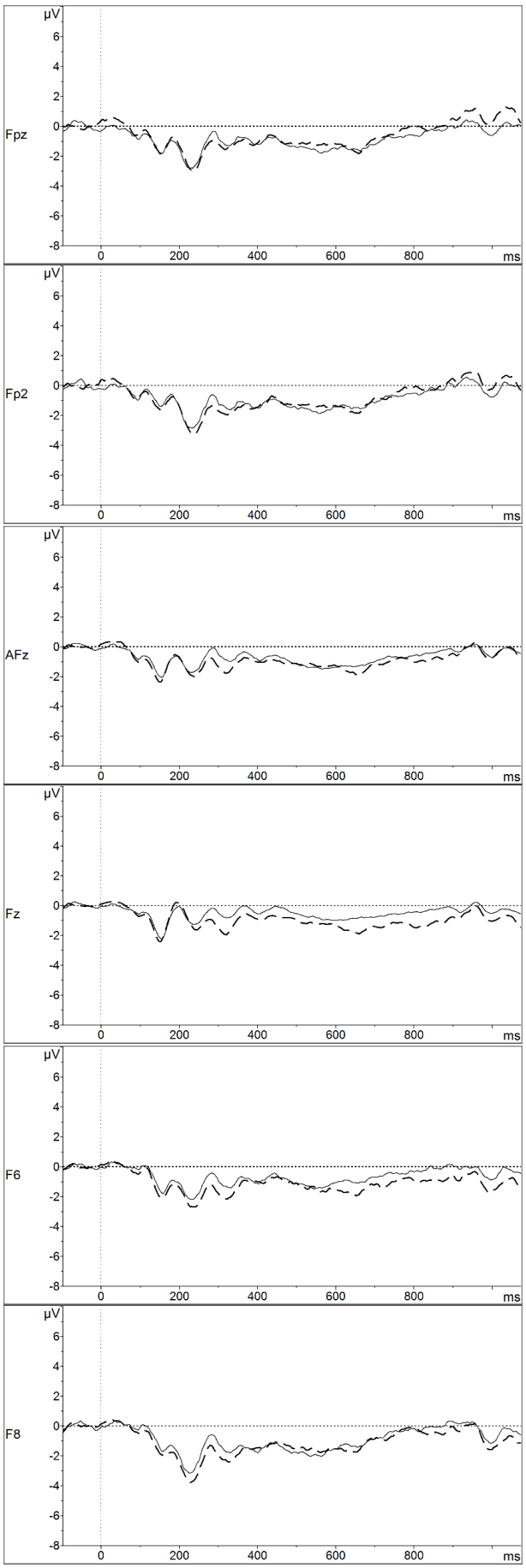

P2
P4
P10
PO8
P6
P8
PO4
O2
μV
ms
0
200
400
600
800

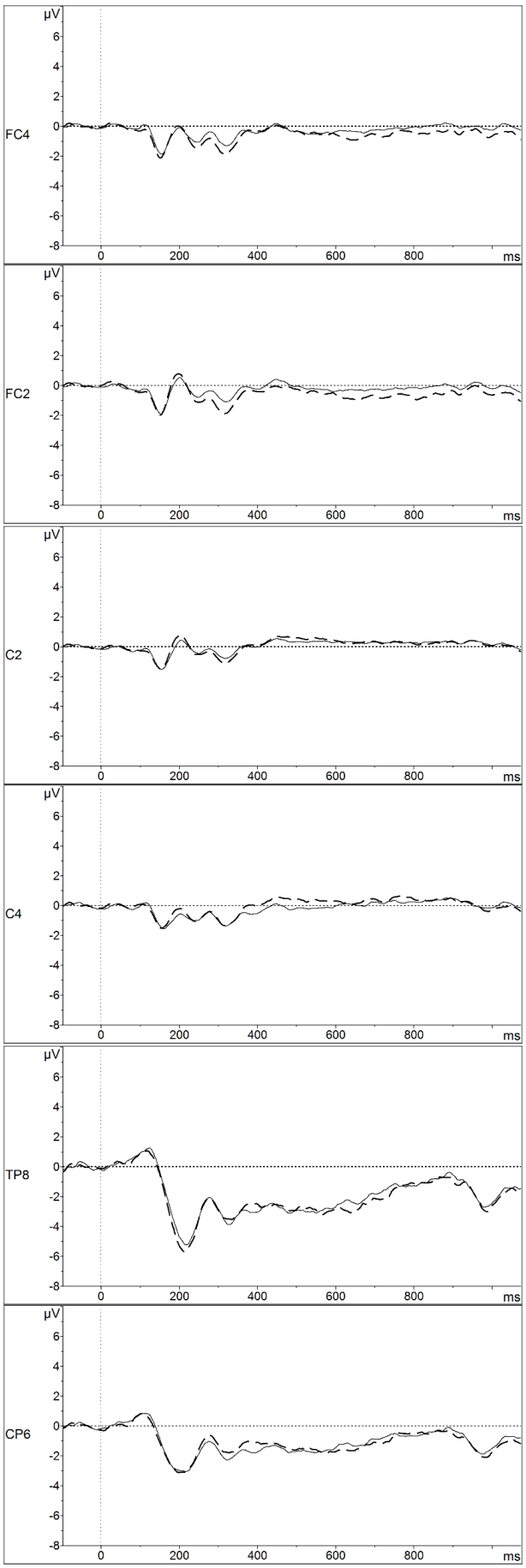
**(B) Identity Task**

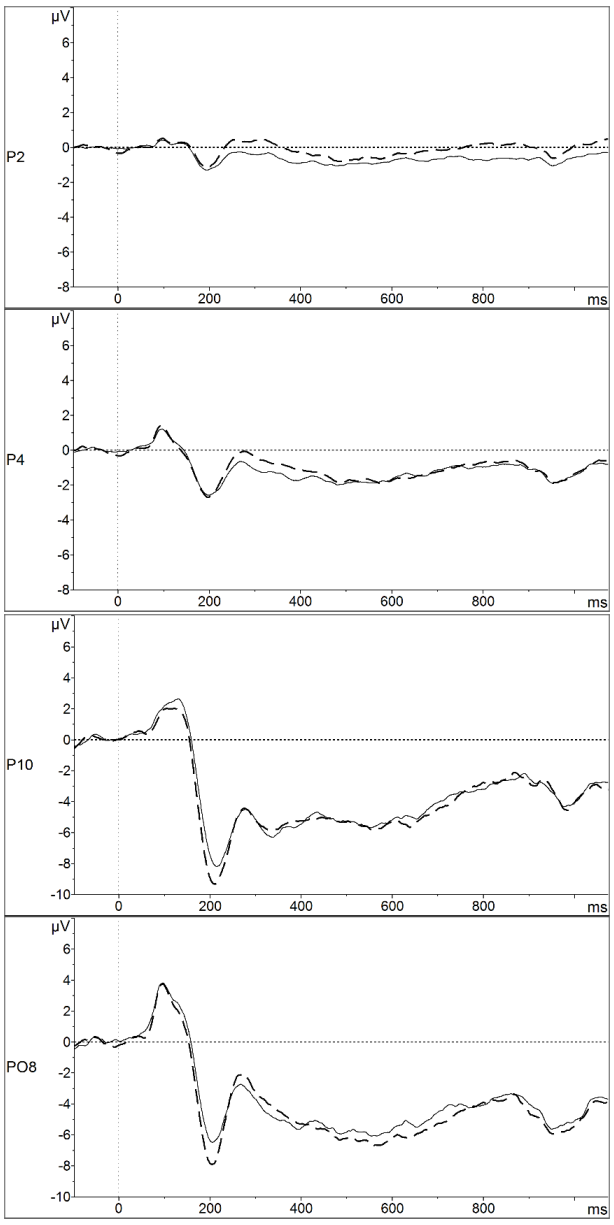



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
ms
0
200
400
600
800

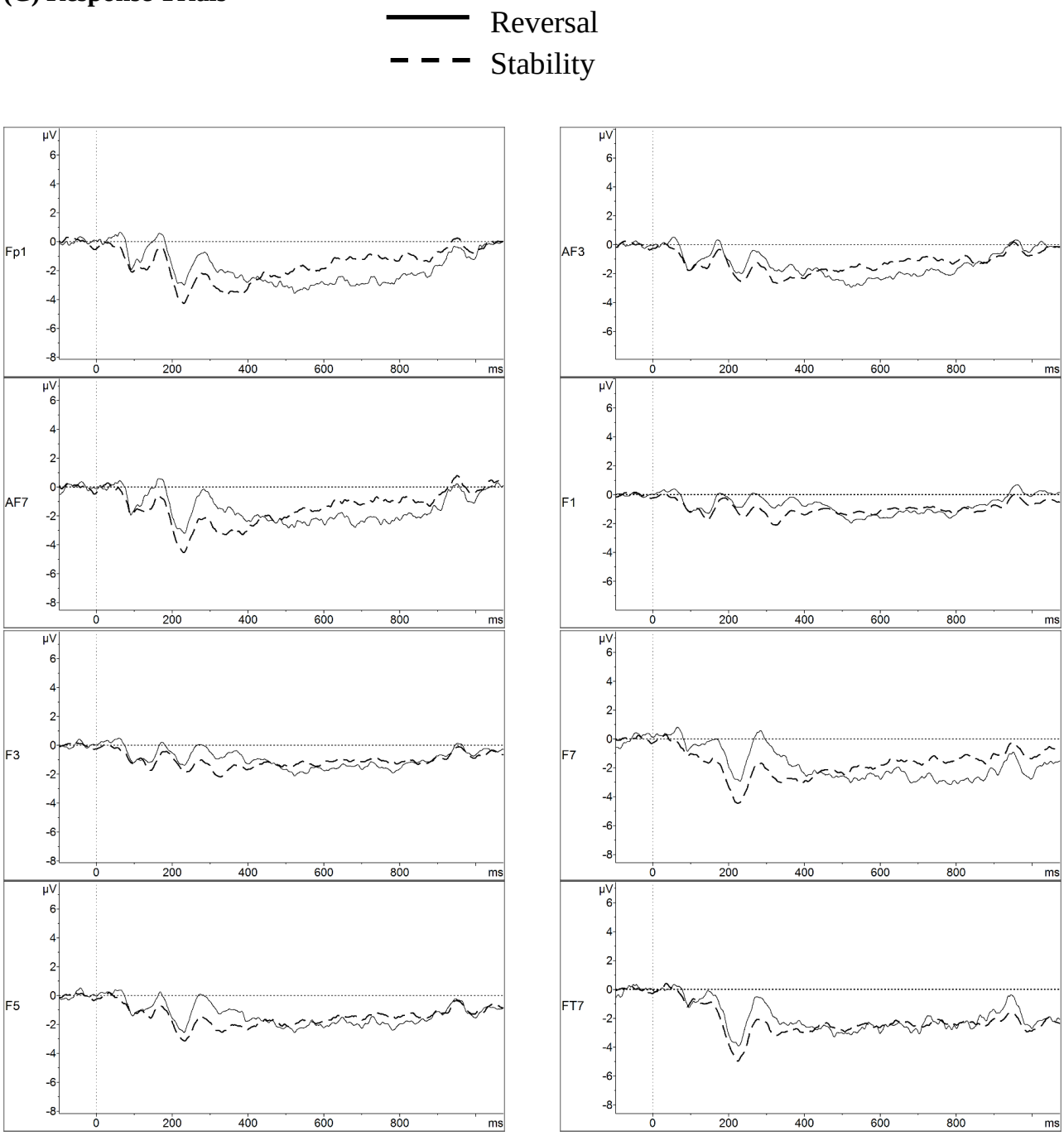

P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
μV
ms
0
200
400
600
800

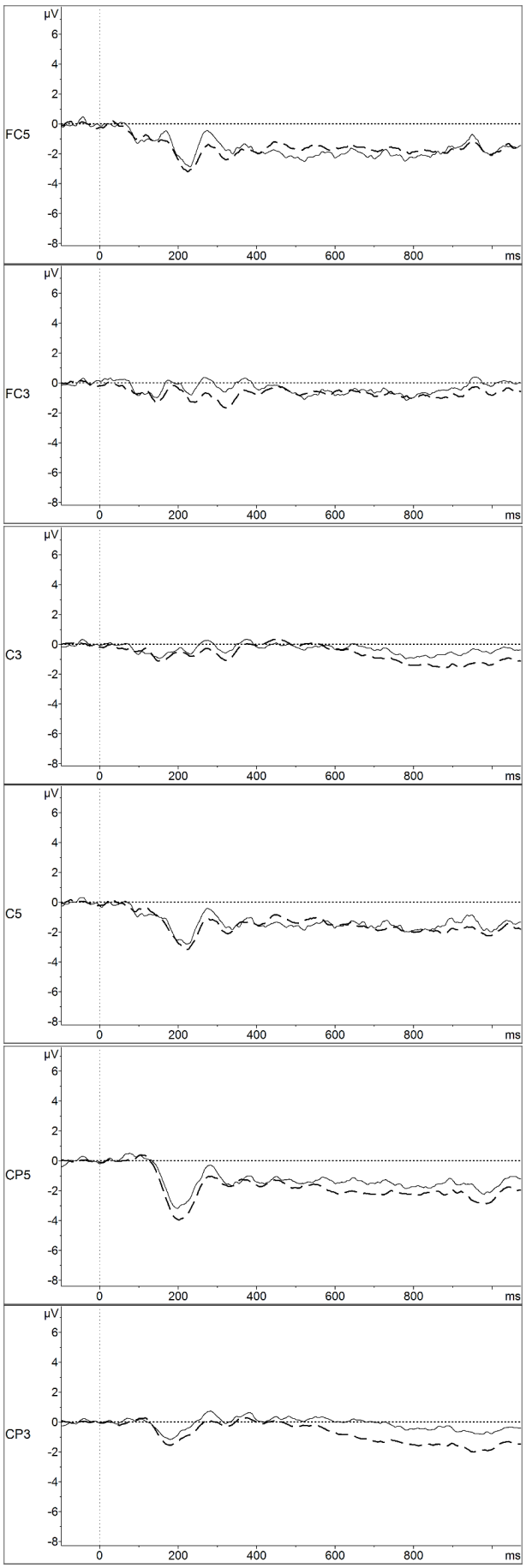

Fpz
Fp2
AFz
Fz
F6
F8
AF8
AF4
F2
F4
FT8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

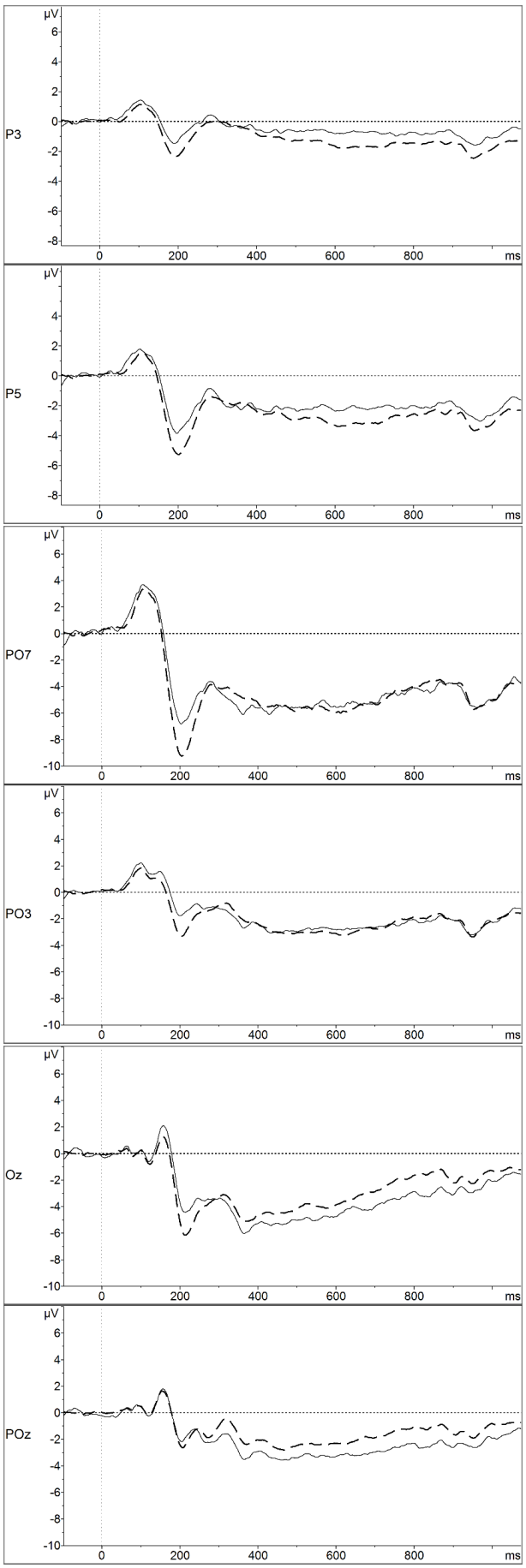

FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

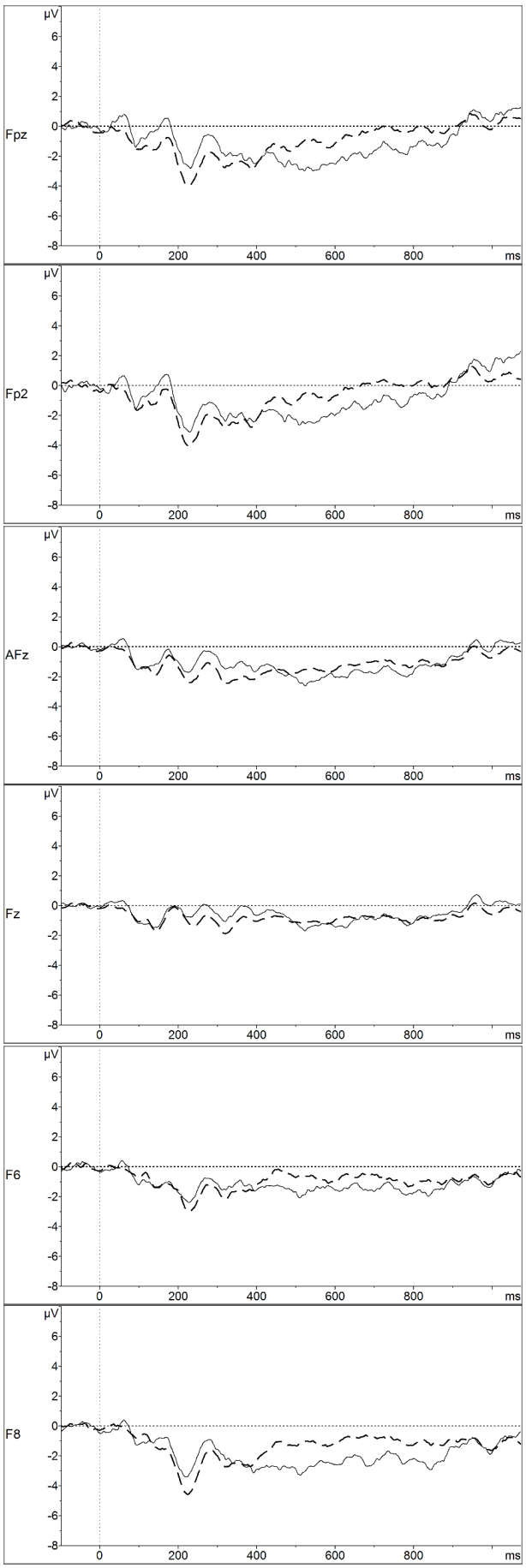

P2
P4
P10
PO8
P6
P8
PO4
O2
μV
ms
0
200
400
600
800

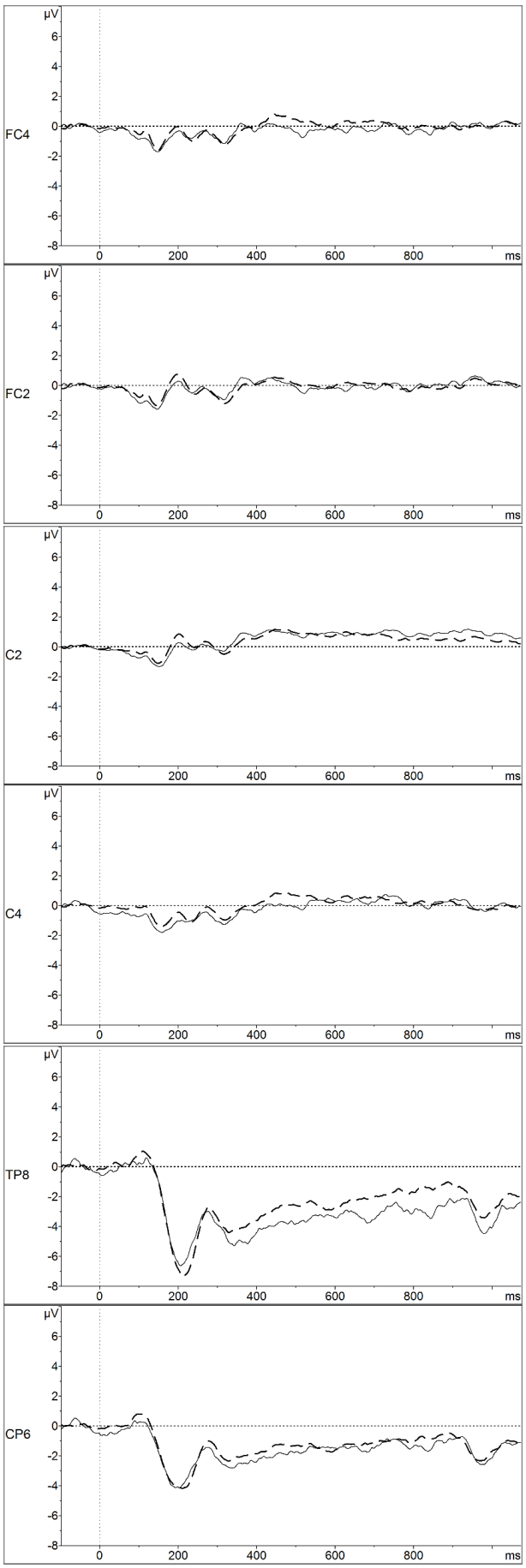
**(C) Response Trials**

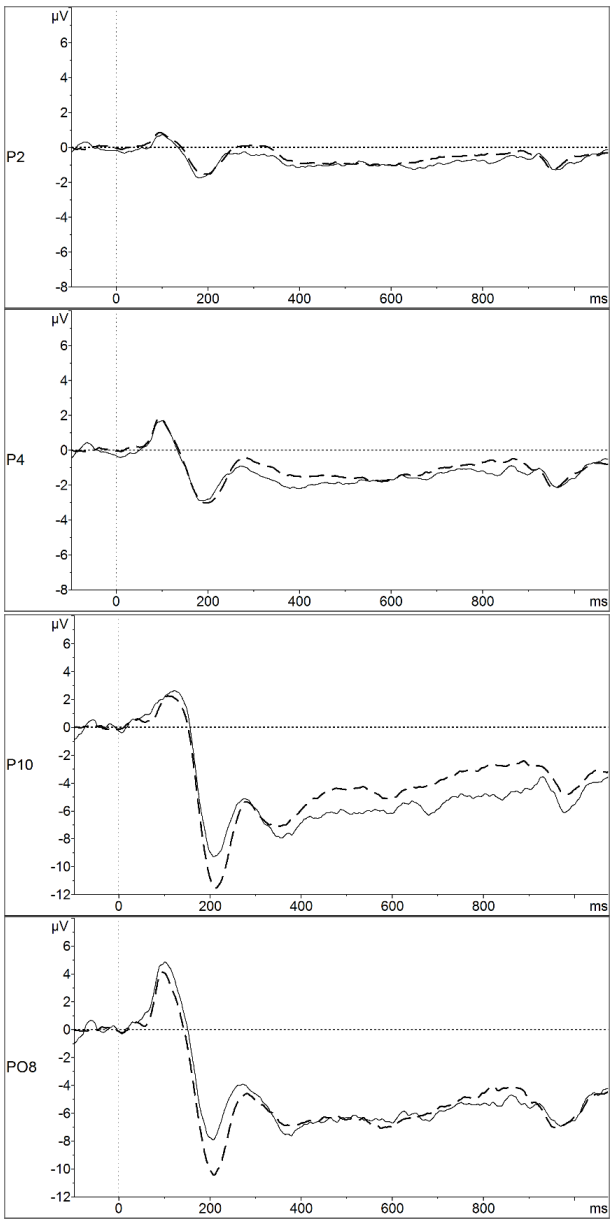



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

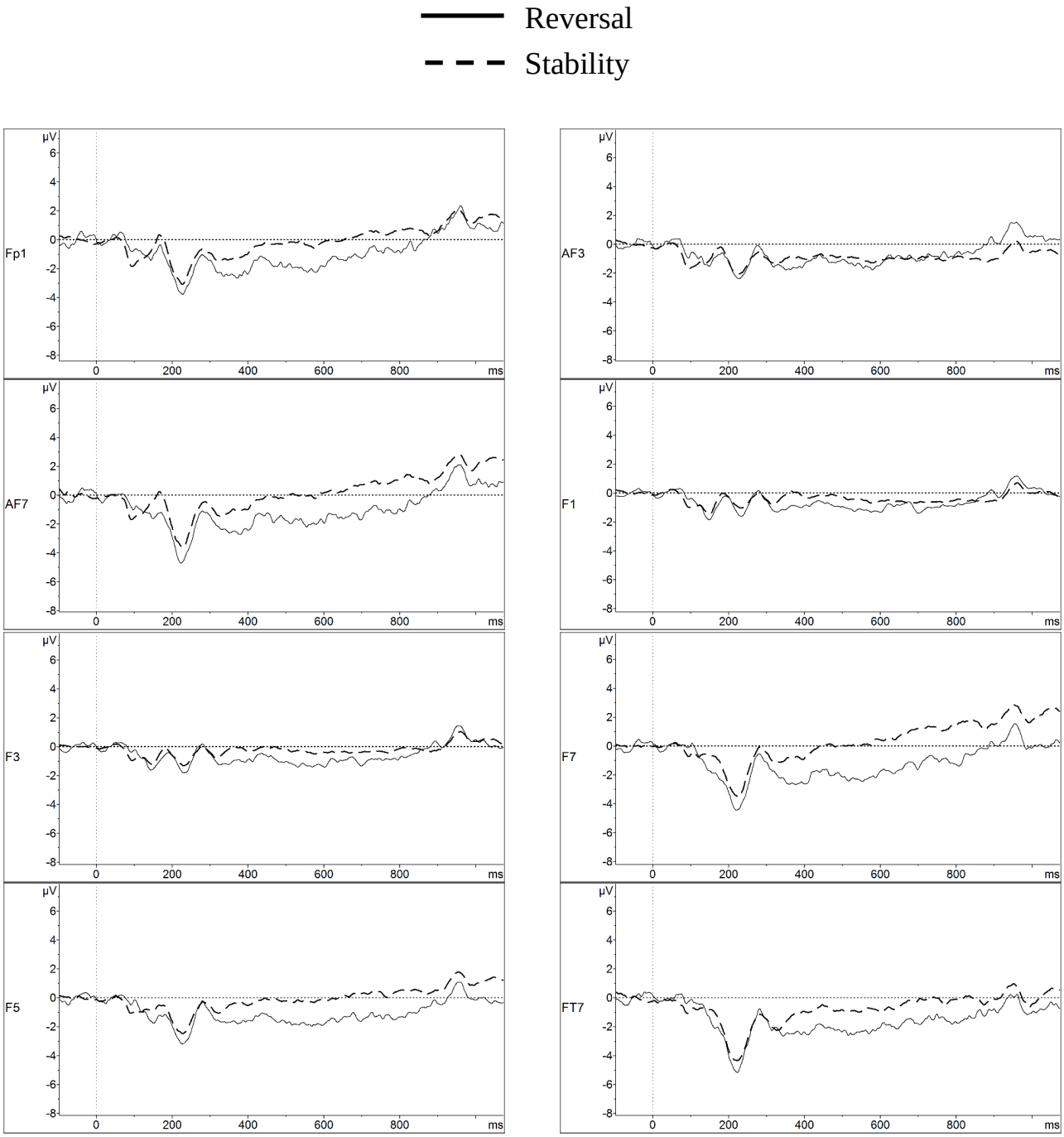

P3
P5
PO7
PO3
Oz
POz
P7
P9
O1
Iz
Pz
CPz
μV
ms
0
200
400
600
800

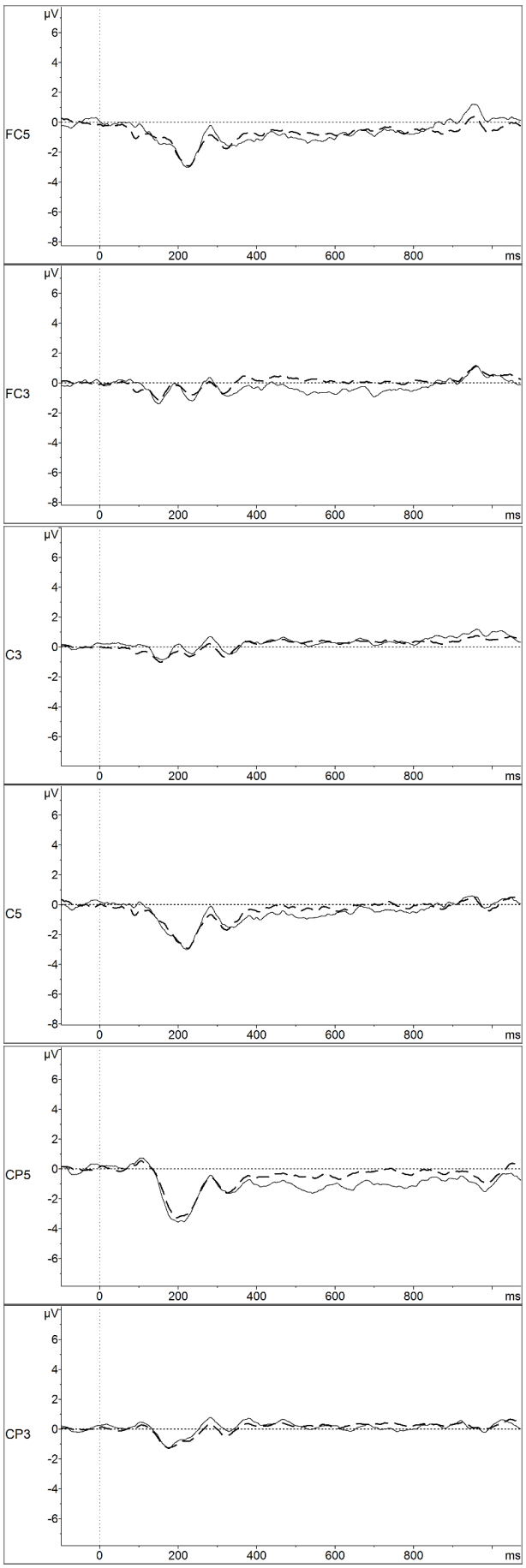

Fpz
Fp2
AFz
Fz
F6
F8
AF8
AF4
F2
F4
FT8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

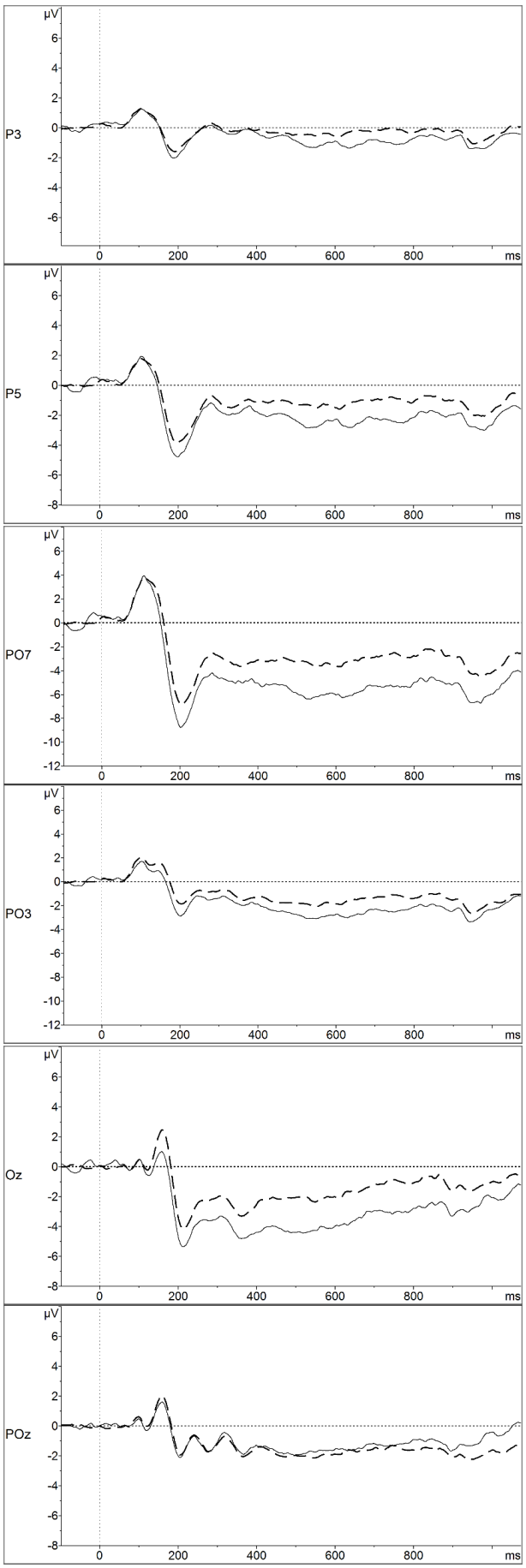

FC4
FC2
C2
C4
TP8
CP6
FCz
Cz
C6
T8
CP4
CP2
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms

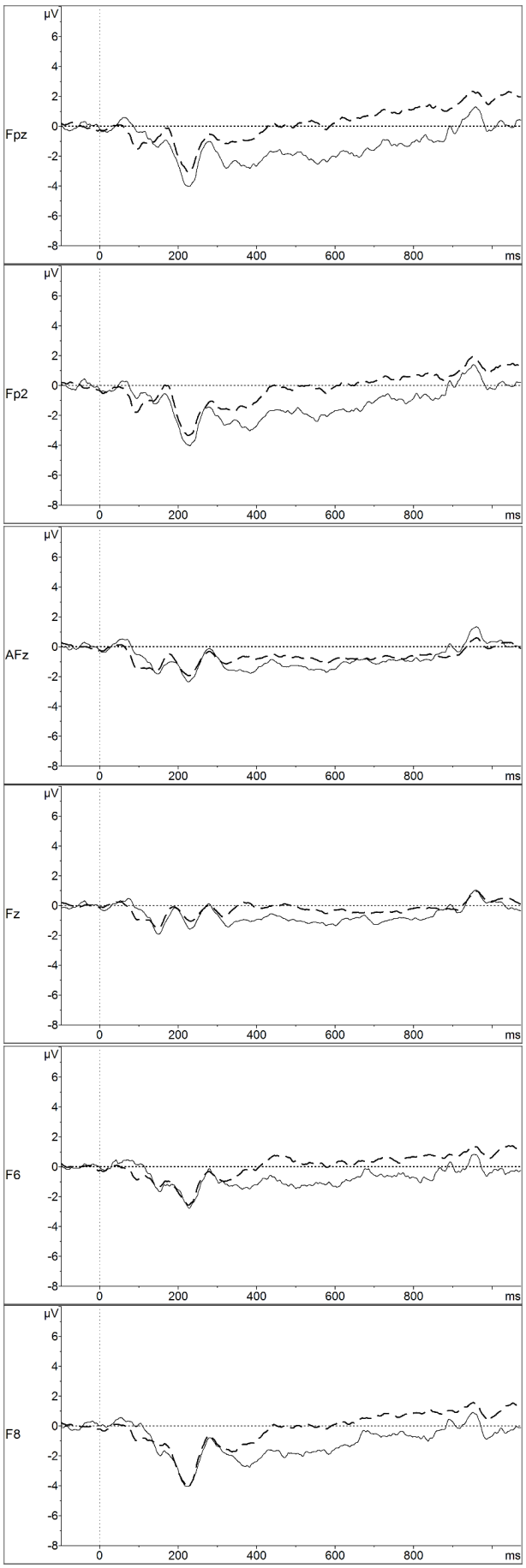

P2
P4
P10
PO8
P6
P8
PO4
O2
μV
ms
0
200
400
600
800

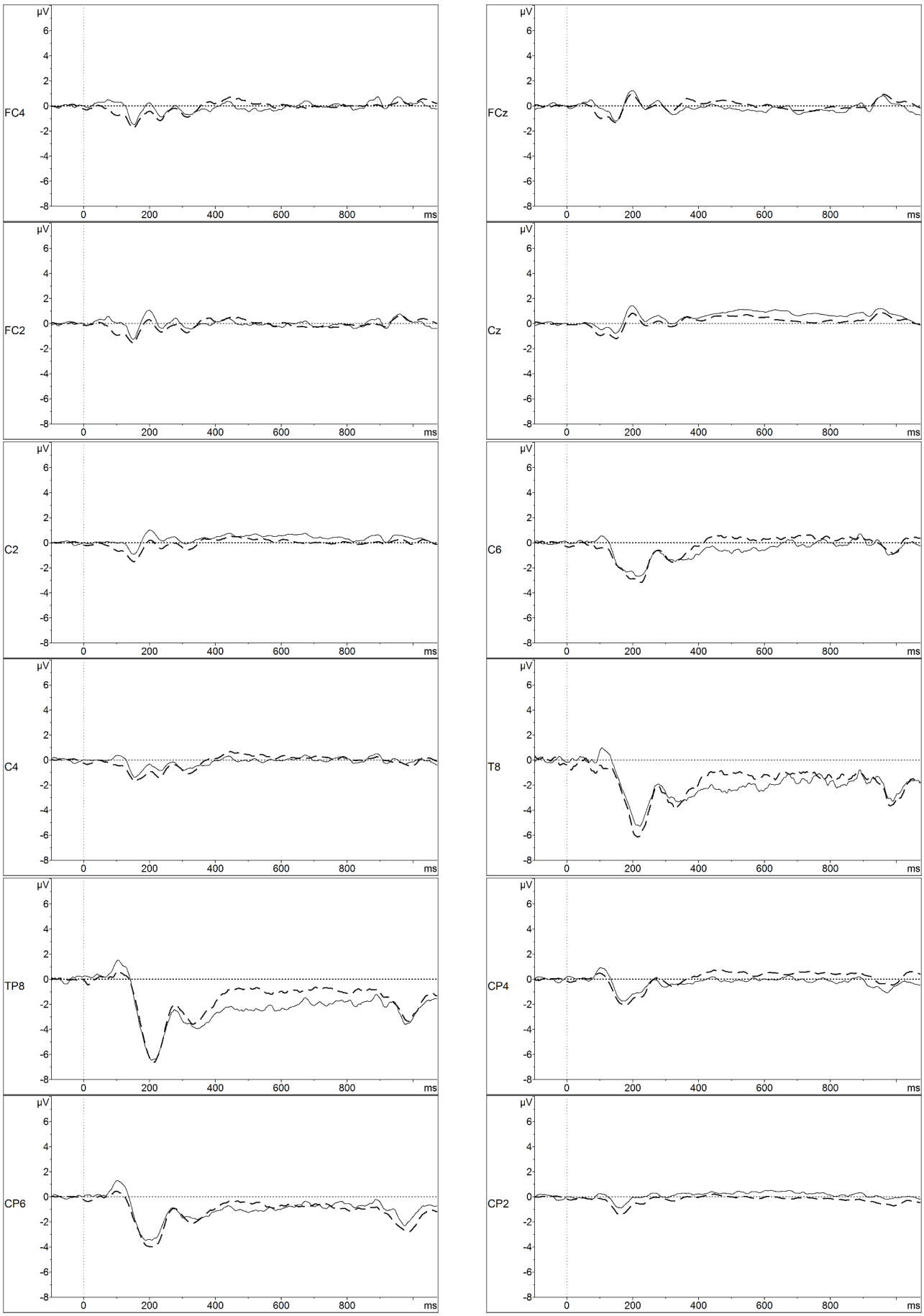
**(D) Non-Response Trials**

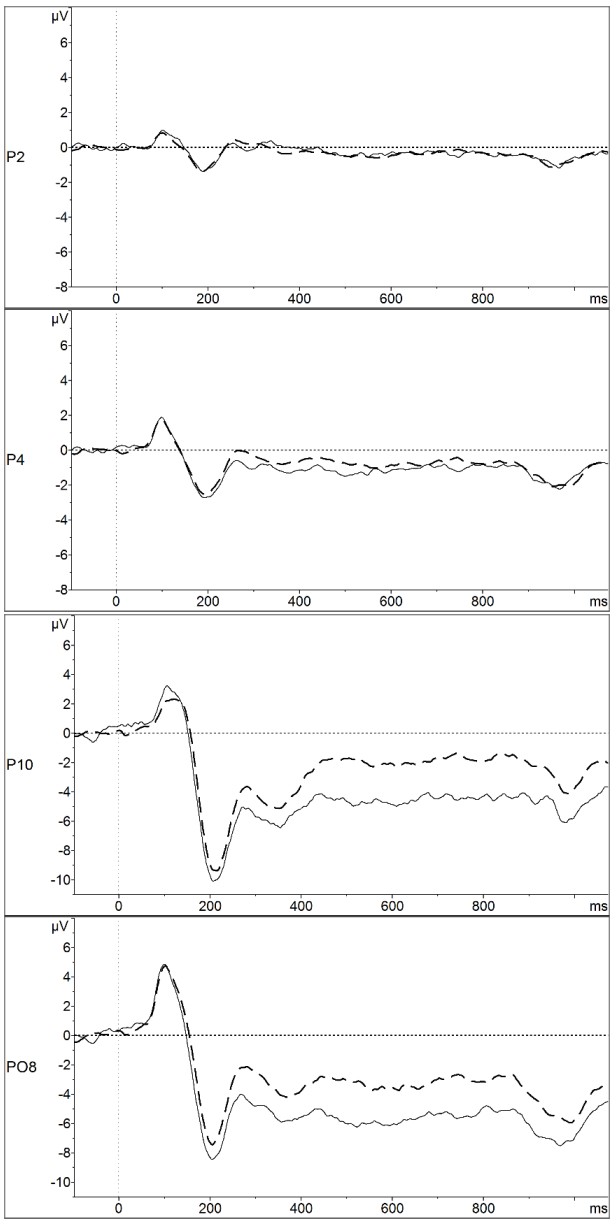



FC5
FC1
FC3
C1
C3
T7
C5
TP7
CP5
CP1
CP3
P1
μV
ms
0
200
400
600
800


P3
P7
P5
P9
PO7
O1
PO3
Iz
Oz
Pz
POz
CPz
μV
ms
0
200
400
600
800


Fpz
Fp2
AFz
Fz
F6
F8
AF8
AF4
F2
F4
FT8
FC6
µV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms


FC4
FCz
FC2
Cz
C2
C6
C4
T8
TP8
CP4
CP6
CP2
μV
6
4
2
0
-2
-4
-6
-8
0
200
400
600
800
ms


P2
P4
P10
PO8
P6
P8
PO4
O2
μV
ms
0
200
400
600
800

**Supplementary Analysis A: No-Go Negativity**

A reviewer suggested that non-response trials may elicit a no-go negativity at central/frontal locations (150-250 ms) and that this no-go effect might be affecting measurement of the RP at posterior channels in the non-response condition (which had an inverted, negative going RP). It is important to point out here that our inverted RP effect involved comparing reversal non-response and stability non-response ERPs. Thus, both of these were equated for non-response. So, it should have still been possible to measure the RP unless any no-go negativity differed between reversal and stable trials.

To assess the reviewer's proposal, we computed the no-go effect (go vs. no-go) separately on reversal and stable trials at fronto/central electrodes Fz, F1, F2, and FCz (as used in Pitts, et al., 2007). The analysis combined data from Experiments 1 and 2 in a 2x2x2 ANOVA with trial type (reversal vs. stable) and response (go vs. no-go) as within-subjects factors and Experiment as a between-subjects factor. Our results did not find a significant effect of go vs. no-go ($p = .400$) nor a trial type x response interaction ($p = .296$). Our analysis suggests that a no-go negativity is not affecting our results. Full results are presented in the table below.

| ***Factor/Interaction*** | ***df*** | ***F*** | $\eta_p^2$ | ***p*** |
|---|---|---|---|---|
| Trial Type | 1,32 | 1.710 | 0.051 | 0.200 |
| Trial Type x Experiment | 1,32 | 0.306 | 0.009 | 0.584 |
| Response (go vs. no-go) | 1,32 | 0.728 | 0.022 | 0.400 |
| Response x Experiment | 1,32 | 0.777 | 0.024 | 0.385 |
| Trial Type x Response | 1,32 | 1.128 | 0.034 | 0.296 |
| Trial Type x Response x Experiment | 1,32 | 0.072 | 0.002 | 0.790 |