Brain and Cognition 129 (2019) 49–58

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

# Confidence tracks sensory- and decision-related ERP dynamics during auditory detection

Alexandria C. Zakrzewski[a,*], Matthew G. Wisniewski[a], Nandini Iyer[b], Brian D. Simpson[b]

[a] Department of Psychological Sciences, Kansas State University, Manhattan, KS 66506, USA
[b] U.S. Air Force Research Laboratory, Wright-Patterson Air Force Base, OH 45433, USA

ARTICLE INFO

*Keywords:*
Metacognition
Attention
Auditory-evoked potential
Confidence
Decision-making

ABSTRACT

Recent research has focused on measuring neural correlates of metacognitive judgments in decision and post-decision processes during memory retrieval and categorization. However, many tasks (e.g., stimulus detection) may require monitoring of earlier sensory processing. Here, participants indicated which of two intervals contained an 80-ms pure tone embedded in white noise. One frequency (e.g., 1000 Hz) was presented on ~80% of all trials (i.e., 'primary' trials). Another frequency (e.g., 2500 Hz) was presented on ~20% of trials (i.e., 'probe' trials). The event-related potential (ERP) was used to investigate the processing stages related to confidence. Tone-locked N1, P2, and P3 amplitudes were larger for trials rated with high than low confidence. Interestingly, a P3-like late positivity for the tone-absent interval showed high amplitude for low confidence. No 'primary' vs. 'probe' differences were found. However, confidence rating differences between primary and probe trials were correlated with N1 and tone-present P3 amplitude differences. We suggest that metacognitive judgments can track both sensory- and decision-related processes (indexed by the N1 and P3, respectively). The particular processes on which confidence judgments are based likely depend upon the task an individual is faced with and the information at hand (e.g., presence or absence of a signal).

## 1. Introduction

Metacognitive abilities are important for regulating one's performance. For instance, if an individual is uncertain about whether or not they are making the right decision, s/he may choose to obtain additional information or opt out of the decision altogether. This adaptive behavior can allow better decision accuracy and/or the avoidance of costly decision errors. Confidence ratings, uncertainty responses, judgments of learning (JOLs), feelings of knowing (FOKs), and tip-of-the-tongue (TOT) reports have been used extensively to explore the cognitive processes associated with metacognition (for a review, see Fleming & Dolan, 2012). Within this work, debate has arisen as to whether these behavioral measures track stimulus encoding, pre- and/or post-decision processes, or some combination thereof (Boldt & Yeung, 2015; Kao, Davis, & Gabrieli, 2005; King, Zechmeister, & Shaughnessy, 1980; Koriat, 1997; Paynter, Reder, & Kieffaber, 2009; Skavhaug, Wilding, & Donaldson, 2010, 2013).

Given the excellent temporal resolution of the electroencephalogram (EEG), several studies have used event-related potentials (ERPs) to investigate the time course of processing that relates to metacognitive judgments. Some have reported post-decision components in response-locked ERPs that correlate with metacognitive judgments. For example, the error positivity (Pe), a component typically elicited when participants make response errors, tends to be larger in amplitude for trials that are given low-confidence ratings than high-confidence ratings (Boldt & Yeung, 2015; also, see Scheffers & Coles, 2000). A number of studies have shown that within the time window of the classic P3 of the stimulus-locked ERP, the largest amplitudes are observed for high confidence (e.g., Boldt & Yeung, 2015; Curran, 2004), high JOL (Skavhaug et al., 2010, 2013), high FOK (Paynter et al., 2009), and successfully retrieved compared to TOT items (Díaz, Lindín, Galdo-Alvarez, Facal, & Juncos-Rabadán, 2007; Lindín & Díaz, 2010). Furthermore, Gherman and Philiastides (2015) were able to train a multivariate classifier to predict confidence ratings from single-trial EEGs within the P3 time window (after 300 ms and peaking around 600 ms post stimulus onset). Those authors concluded that confidence judgments are based on the accumulation of evidence used in the primary-task decision itself.

Though this recent work suggests relationships to metacognitive judgments in the primary-task decision process and post-decision monitoring, the earlier sensory processing on which many metacognitive judgments may be based remains largely unexplored. Understanding this

* Corresponding author.
*E-mail address:* aczakrzewski@ksu.edu (A.C. Zakrzewski).

https://doi.org/10.1016/j.bandc.2018.10.007
Received 11 May 2018; Received in revised form 18 October 2018; Accepted 18 October 2018
Available online 13 December 2018

relationship between metacognition and sensory processing is important for full characterization of metacognitive processes. Fatigue during long work hours can affect sensory processing of incoming information (e.g., Hillyard, Hink, Schwent, & Picton, 1973; Makeig, 1993; Strauß, Wöstmann, & Obleser, 2014; Wisniewski, 2017). The classic work of Hillyard et al. (1973) revealed how attention affects sensory responses in the ERP. Specifically, average N1 amplitude tracked attention as it shifted to target tones played in one ear and then the other (Hillyard et al., 1973). Utilizing selective attention for one signal can come at the expense of potentially relevant non-focused sources (e.g., an auditory warning indicating that fuel levels of an aircraft are low).

There are a great number of real-world tasks that contain significant variability in the processing of sensory signals which, in turn, can be used to adapt behavior after making a metacognitive judgment. However, the particular paradigms used in recent work typically involve higher-level categorization (e.g., "Is this image a face or a car?"; Gherman & Philiastides, 2015) and/or meta-memory judgments (e.g., "Have you solved this math problem before?"; Paynter et al., 2009) that are unlikely to fully characterize such scenarios. In contrast, in a seminal study by Squires, Hillyard, and Lindsay (1973), participants were tasked with determining when there was a 1000-Hz sinusoidal tone presented against a background of white noise. Participants made confidence ratings following each detection response. Both the auditory-evoked N1 and a later positivity systematically increased in amplitude as a function of confidence. This early work suggests that metacognitive judgments can track sensory-related processes (reflected by the N1) in addition to the relatively late decision-related and error-monitoring processes that have received recent attention in the metacognitive literature.

Similar to Squires et al. (1973), we asked whether confidence in auditory tone detection is related to sensory- and/or decision-related processes observable in the ERP. We made two key extensions on this early work. First, we used a detection paradigm (the *probe-signal method*) known to reveal systematic differences in detection sensitivity along the dimension of frequency (for review, see Scharf, 1998). Frequency probability (e.g., 80% of trials contain tones at 1000 Hz) was used to set up expectancy for a tone of specific frequency. In this paradigm, 'probe' tones at unexpected frequencies are generally detected with lower accuracy than 'primary' tones at the expected frequency. For instance, a tone may be detected ~90% of the time as the 'primary' but at chance level when it is the 'probe' (Scharf, 1998). Unlike Squires et al. (1973), who were forced to collect thousands of trials and rely solely on trial-by-trial variations in processing, we expected that this paradigm would generate those variations systematically. Second, whereas Squires et al. (1973) used a single-interval task, we used a two-interval, two-alternative forced-choice (2i-2afc) paradigm. Two consecutive intervals contained a noise masker, but only one also contained a tone. Participants' task was to indicate in which interval the tone occurred. This allowed us to assess relationships between confidence ratings and ERPs associated with both the presence and absence of a tone within a trial. Arguably, both types of information may be important for making a single confidence judgment.

We hypothesized that 'primary' tones would be detected with greater accuracy and would be rated with higher confidence than 'probe' tones. We planned our ERP analyses to look at components that have previously been related to metacognitive judgments: the auditory N1/P2 complex (Squires et al., 1973) and the P3 (e.g., Curran, 2004; Del Cul, Baillet, & Dehaene, 2007; Desender, Van Opstal, Hughes, & Van den Bussche, 2016; Díaz et al., 2007; Gherman & Philiastides, 2015; Hillyard, Squires, Bauer, & Lindsay, 1971; Kouider et al., 2013; Lindín & Díaz, 2010; Paul & Sutton, 1972, 1973; Paynter et al., 2009; Skavhaug et al. 2010, 2013; Squires et al., 1973; Squires, Squires, & Hillyard, 1975). While the N1/P2 complex is exogenous and does not necessarily require an individual's attention to a stimulus (though, as discussed earlier, we acknowledge that attention affects N1 amplitude; Hillyard et al., 1973; Okamoto, Stracke, Wolters, Schmael & Pantev, 2007), the P3 is endogenous and typically requires some decision to be made by the participant and/or attentive processing (for a review, see Donchin, 1981). In addition to comparing 'primary' versus 'probe' trials, we analyzed ERPs generated on trials given high- compared to low-confidence ratings. Finally, we assessed correlations between confidence ratings made on 'primary' and 'probe' trials and the amplitudes of ERP components generated under those conditions. This analysis assessed the hypothesis that individuals showing a larger effect in the confidence rating dependent variable also show larger effects in their ERPs. We expected that analyses would reveal at what stage or stages of processing differences between these conditions arise when both sensory and decision-related processes are task critical.

## 2. Methods

### 2.1. Participants

Twenty-two individuals (9 females; ages 19–40) with normal audiometric thresholds (< 20 dB HL, 0.25–8 kHz) and experience in psychoacoustic tasks participated. All signed a U.S. Air Force Research Laboratory Institutional Review Board approved informed consent document. Two individuals, one from each counterbalance assignment (see below), were dropped from analysis because of excessive skin potentials in the data.

### 2.2. Apparatus

Sounds were presented diotically through Etymotic ER-2 insert earphones (Etymotic Research, Elk Grove Village, IL) at a fixed, comfortable listening level (< 81 dB SPL). Procedures were executed in MATLAB (Mathworks, Natick, MA). Event-code timing was controlled with a Tucker-Davis Technologies real-time processor (RP2.1; Tucker-Davis Technologies, Alachua, FL). Listeners sat in a sound-attenuating booth.

### 2.3. Stimuli

On each trial, two 2400 ms white noise sources (750 ms on- and off-ramps; Tukey window) were presented sequentially, with a 100-ms pause in between. Noise sources served as maskers to the tone. For simplicity, from henceforth we will refer to these sources as maskers even though technically one of the sources does not actually mask anything. The tone was an 80 ms sinusoidal tone at 1000 or 2500 Hz that was embedded in one of the noise maskers at 1160 ms relative to masker onset.

### 2.4. Tasks

A depiction of a typical trial is shown in Fig. 1.

#### 2.4.1. Detection task

On each trial, a tone was played in either the first or second masker interval with equal *a priori* probabilities. Participants' task was to indicate whether the tone occurred in the first or second masker interval using computer keys labeled "1" and "2".

Each participant's thresholds (~71% correct) for detection in this task (absent the metacognitive component) were estimated using 35-trial adaptive blocks in a pre-experimental session. Three blocks were run for each frequency starting at a signal-to-noise ratio of −10 dB, lowering the level of the target 1 dB after every correct response, and raising the level 2.43 dB after every incorrect response (Leek, Hanna, & Marshall, 1992).[1] The mean of the last 6 reversals in a block was taken

[1] A computer error left several participants with only 2 tracks per frequency. When this was the case, the median of the obtained tracks was used as the participant's threshold.

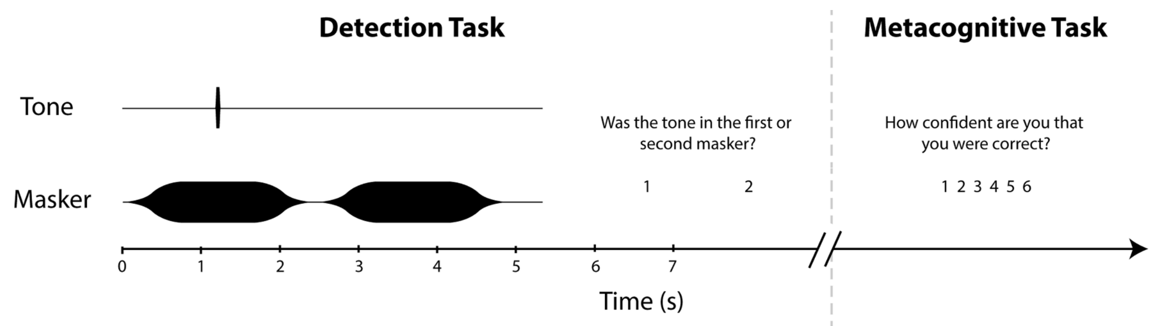


**Fig. 1.** Depiction of a typical trial in the experiment. A detection task and a metacognitive task occurred on each trial. Two consecutive noise maskers were presented. A tone was presented in one of the two masker intervals. The participant's task was to indicate which noise masker contained the tone. After the detection task, participants rated their confidence on a scale from 1 to 6.

as the threshold estimate for that block. The median of these threshold estimates was considered to be a participant's threshold. Thresholds were, on average, −21.86 dB (SD = 1.57) and −20.43 dB (SD = 2.07) for the 1000 Hz tone and the 2500 Hz tone, respectively.

During the experimental session, tones were presented 3 dB above each individual participant's thresholds (cf. Scharf, 1998). For half of participants, there was an *a priori* probability that the 1000 Hz ('primary') tone and the 2500 Hz ('probe') tone would occur 80% and 20% of the time, respectively. *A priori* probabilities of 80%/20% were reversed for the other half of participants. Participants were made aware that the tone could be one of two different frequencies. Five blocks of 65 trials were collected (325 trials total; ∼260 'primary' trials, ∼65 'probe' trials) per participant.

For both the pre-experimental and experimental sessions, participants were instructed to withhold responding until the end of sound playback. There was no response deadline. No feedback of correctness was provided; however, if an inappropriate key was accidentally pressed, the statement "you hit a wrong key!" was presented on the screen. These trials were discarded from analysis. Both correct and incorrect trials were included in all analyses reported in Section 3. *Results*. However, see *Supplementary Materials* for analyses with correct trials only as well as additional analyses that take into account the masker interval order (whether the tone was presented in the first or second masker).

#### 2.4.2. Metacognitive task.

Following each detection-task response in the experimental session, participants were presented with the prompt: "How confident are you that you were correct?" Printed below the prompt on the computer screen was a 6-point confidence scale. Participants were instructed to enter '1' when they had low confidence, and '6' when they were highly confident in the accuracy of their response. They were encouraged to use the entire scale. Ratings were made with the number keypad of the keyboard. There was no response deadline for confidence ratings.

### 2.5. Electrophysiology

Data were collected from a 38-channel array of electrodes with a BioSemi Active II system (BioSemi, Amsterdam, Netherlands), at a 2048 Hz sampling rate, and 24-bit A/D resolution. Data were referenced online to the Common-Mode-Sense/Driven-Right-Leg (CMS/DRL) reference of the BioSemi system (see www.biosemi.com). Thirty-two electrodes were fixed within a cap and arranged according to the international 10–20 system. Six additional electrodes were placed at the mastoids, and on lateral sides and below each eye. Electrode offsets relative to CMS/DRL were brought within 25 μV or else were rejected from analysis.

All offline analyses were performed using EEGLAB (Delorme & Makeig, 2004; http://sccn.ucsd.edu/eeglab) and custom MATLAB scripts/functions. The data were referenced to the average of mastoids, and resampled at 256 Hz (after applying a zero phase antialiasing filter). An additional bandpass zero-phase Hamming window sinc finite impulse response (FIR) filter (0.5 Hz and 40 Hz passband edges; filter order = 1536) was applied with the *pop_eegfiltnew()* function of the EEGLAB toolbox in order to remove slow-wave drifts and high-frequency activity that was not of interest for the planned ERP analyses. Based on visual inspection, continuous data and channels contaminated by excessive noise or movement artifacts were removed. Remaining data were submitted to independent components analysis (ICA). Independent components (ICs) identified as artifacts (e.g., eye blinks) were subsequently removed from the channel data (Jung et al., 2000; Wisniewski, Mercado, Church, Gramann, & Makeig, 2014). Following ICA, epochs were extracted from −200 to 1500 ms surrounding the tone onset for tone-present intervals. Similar epochs were also extracted for tone-absent intervals, time-locked to when the tone would have been presented. Mean baseline voltages were subtracted (−200 to 0 ms). ERPs were generated by averaging single-epoch voltage time courses. Statistics on trial counts per condition used in the main analyses were as follows: High confidence, M = 110.45, maximum = 233, minimum = 48; Low confidence, M = 80.45, maximum = 137, minimum = 21; Primary, M = 146.05, maximum = 223, minimum = 60; Probe M = 44.85, maximum = 89, minimum = 13.

### 2.6. Statistics

Amplitudes of ERP components were considered to be the mean voltage within time-windows determined from the grand-average waveforms across all conditions. Mean amplitudes were averaged across groups of scalp locations encompassing frontal, central, posterior, left lateralized, and right lateralized scalp locations (see Figs. 2 and 4 in Section 3. *Results*). Time-windows for N1 and P2 were 140–180 ms and 260–340 ms, respectively. Note that latencies of these components are later than canonical latencies (e.g., ∼100 ms for N1). This was expected given that tone masking generally delays components of the auditory ERP (for a review, see Billings, Bennett, Molis, & Leek, 2011). Tone-present P3 amplitudes were considered to be mean voltage from 400 to 1000 ms. In addition to the P3 on tone-present intervals, we also analyzed the P3 on tone-absent intervals (400–1200 ms). This served to assess how late positivity on tone-absent intervals also related to the primary/probe manipulation and to participants' confidence (cf. Squires et al., 1973). See Figure S1 in *Supplementary Materials* for butterfly plots of grand-average waveforms and time-windows for computing mean amplitudes.

Mean amplitudes were analyzed with standard ANOVAs. All conducted post-hoc tests were interpreted with Bonferroni corrections (uncorrected *p*-values reported). Where ERP component amplitudes were related to behavior, Pearson's *r* was computed and assessed for significance in the same way (i.e., as post-hoc).

## 3. Results

All analyses were performed after collapsing across counterbalance assignments. Including tone frequency as a factor in any of the analyses added no significant main effects or interactions, *ps* > .05.

### 3.1. Behavioral data

Surprisingly, there was no significant difference in proportion correct between primary and probe trial types, $t < 1$. We believe that a lack of a significant effect might be due to a larger frequency separation between primary and probe tones than is typical (1500 Hz vs. ≤500 Hz; Scharf, 1998). Nevertheless, proportion correct was trending such that primary tone trials ($M$ = .86, $SE$ = .03) were more accurate than probe trials ($M$ = .81, $SE$ = .04).

Participants did respond with higher confidence ratings on primary ($M$ = 4.14, $SE$ = .27) than probe ($M$ = 3.83, $SE$ = .29) trials, however, this difference was not significant, $t < 1.1$. For a more detailed look at overall confidence resolution, Table 1 shows mean accuracy for high- and low-confidence levels on primary and probe trials. Individual participants had different biases in how they used the confidence rating scale. Using the midpoint of the scale (i.e., a rating of 3.5) as a separation between high and low confidence thus led to drastically unequal trial counts in these two categories across subjects. To control for bias, we labeled trials as high confidence and low confidence using individualized midpoints, derived from an individual's mean confidence rating. Trials with confidence ratings higher than the mean were considered to be high-confidence trials. The opposite was true for trials with confidence ratings below the mean. Participants' confidence ratings paralleled accuracy such that trials given high confidence showed a greater proportion correct than trials given low confidence for primary trials, $t(19)$ = 5.64, $p < .001$, Cohen's $d$ = 1.32, and probe trials, $t(19)$ = 5.15, $p < .001$, Cohen's $d$ = 1.14 (see Table 1). This shows that participants' confidence ratings did relate to task accuracy.

**Table 1**
Mean proportion correct by confidence level and trial type and statistical comparisons.

| Trial type | High confidence | Low confidence | *p* |
|---|---|---|---|
| Primary | .92 (.04) | .75 (.03) | < .001 |
| Probe | .90 (.04) | .70 (.04) | < .001 |

*Note.* Values in parentheses are standard errors of the mean.

### 3.2. ERP data – high confidence vs. low confidence

ERPs split up by high and low confidence based on participants' mean confidence as individualized midpoints (just described in Section 3.1 *Behavioral Data*), presence/absence of a tone, and electrode cluster are shown in Fig. 2. Scalp maps of ERP components are shown in Fig. 3.

When tones were present, they evoked an N1–P2 complex that was clear at fronto-central and left- and right-lateralized locations (see Section 2.6 *Statistics* for a discussion of component latency). At the posterior cluster of electrodes there was also an apparent P3 peaking between 500 ms and 1000 ms. This was expected (Donchin, 1981; Luck, 2005; Wisniewski et al., 2016). All three of these components appeared to be strongest for high-confidence trials. Interestingly, there was a comparable positivity for tone-absent intervals. This feature was also strongest at the posterior cluster. Squires et al. (1973) similarly observed a late positivity on tone-absent trials in their experiment that was assumed to be a P3. Due to its similarity in scalp map to the P3 locked to tone presentations (see Fig. 3), and the similarity to previous data, we will henceforth refer to this feature as a "tone-absent P3". With the possible exception of the N1, which showed relatively weak negativity on low-confidence trials, and a strong temporal projection (cf. Näätänen & Picton, 1987), scalp maps are consistent with their labeling. Note also that the local minimum at Cz observed for the N1 on high-confidence trials is likely due to the overlapping P2 component which shows a strong positivity at Cz (for review, see Luck, 2005).

Analyses employed 2 (Confidence: High, Low) × 5 (Electrode Cluster: Frontal, Central, Posterior, Left, Right) repeated-measures

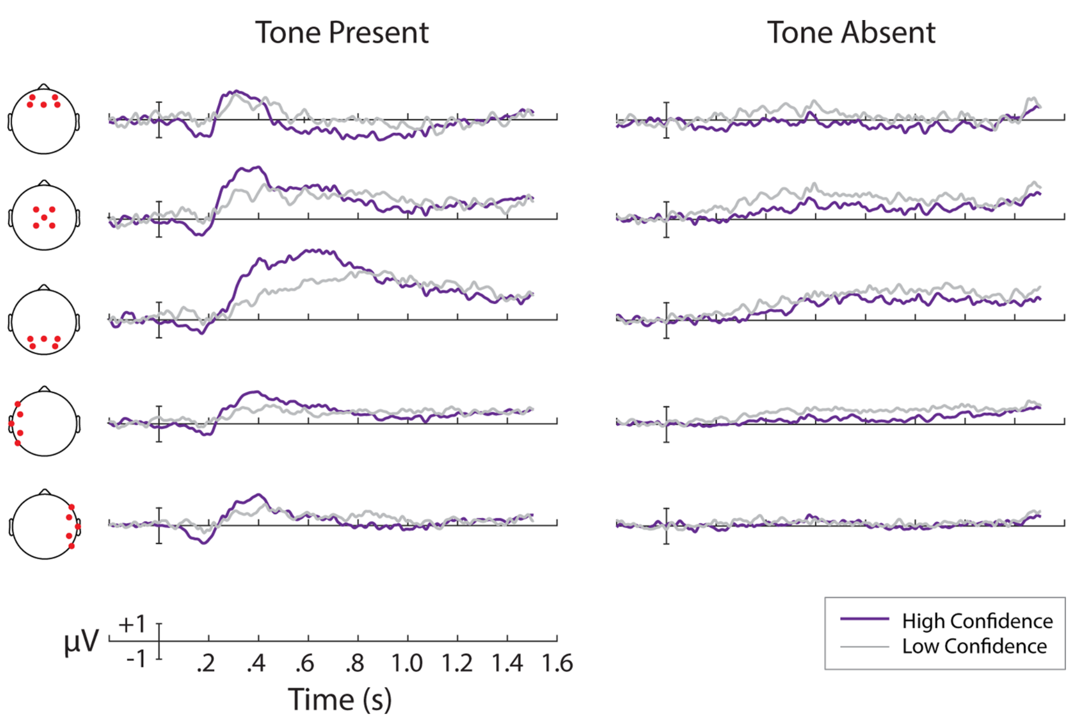


**Fig. 2.** ERPs at each electrode cluster for when the tone was present (left) and absent (right). ERPs corresponding to trials in which a participant gave high-confidence ratings are in purple. Low-confidence trials are depicted in grey. Red points in scalps to the left of each row represent locations within the electrode cluster for that row.

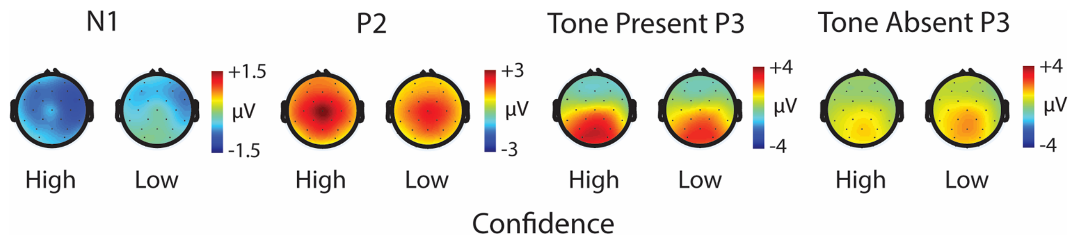


**Fig. 3.** Scalp maps of N1, P2, tone-present P3, and tone-absent P3 broken up by confidence level.

ANOVAs. With N1 amplitude as the dependent variable, we found a significant main effect of confidence, $F\,(1,19) = 7.40$, $p = .014$, $\eta_p^2 = .28$. Participants appeared to rate a trial with high confidence if a relatively large N1 was evoked, but low confidence if a relatively small N1 was evoked. Qualitatively, this effect is mirrored in scalp maps (Fig. 3), which show very little evidence of a N1 on low-confidence trials. No other main effects or interactions were found, $Fs < 2$.

For P2 amplitude, we found a significant main effect of electrode cluster, $F\,(4,76) = 4.66$, $p = .002$, $\eta_p^2 = .20$, likely reflecting the fact that P2 amplitude was larger at centrally-located electrodes (cf. Näätänen & Picton, 1987; Ball et al., 2017). The main effect of confidence was also significant, $F(1,19) = 4.72$, $p = .043$, $\eta_p^2 = .20$. Similar to N1, amplitudes were larger on high- compared to low-confidence trials. The interaction was not significant, $F < 2$.

For the tone-present P3, there was a main effect of electrode cluster, $F(4,76) = 27.25$, $p < .001$, $\eta_p^2 = .59$. As is typical, tone-present P3 amplitude was largest at the posterior electrode cluster. There was also a significant electrode cluster × confidence interaction, $F(4,76) = 14.67$, $p < .001$, $\eta_p^2 = .44$. The interaction likely stems from the fact that tone-present P3 amplitude was greater on high- compared to low-confidence trials, primarily at the posterior cluster of electrodes. Indeed, post-hoc repeated-measures *t*-tests (Bonferroni corrected; uncorrected *p*-values reported) testing the P3 amplitude difference between high- and low-confidence trials at each electrode cluster revealed a significant difference only for posterior electrode cluster, $t(19) = 3.42$, $p = .003$, Cohen's $d = 1.22$. The other comparisons were not significant, $ts < 2.81$.

For the tone-absent P3, we found a significant main effect of electrode cluster, $F(4,76) = 9.01$, $p < .001$, $\eta_p^2 = .32$, likely reflecting the strongest positivity being at posterior electrode sites (see Fig. 2). There was also a significant main effect of confidence, $F(1,19) = 6.35$, $p = .045$, $\eta_p^2 = .20$, such that tone-absent P3 amplitude was greater for low-confidence than high-confidence trials. Interestingly, the direction of this main effect was opposite of that found for the tone-present P3. The interaction was not significant, $F < 2$.

An additional control analysis was done using only correct high- and low-confidence trials. Significant ERP component amplitude differences between high- and low-confidence trials remained for the N1, $p = .02$, P2, $p = .034$, and tone-absent P3, $p = .017$, as well as a significant confidence by cluster interaction for tone-present P3, $p < .001$. Thus, these effects were not dependent upon differences in correctness. For more details, see *Supplementary Materials*.

### 3.3. ERP data – primary vs. probe.

Fig. 4 displays ERPs at all electrode clusters locked to the onset of the tone (tone present) and to when the tone would have been presented (tone absent). ERPs in each panel are separated by primary (cyan) and probe (grey) conditions. Scalp maps of components are shown in Fig. 5. There appeared to be some difference between primary

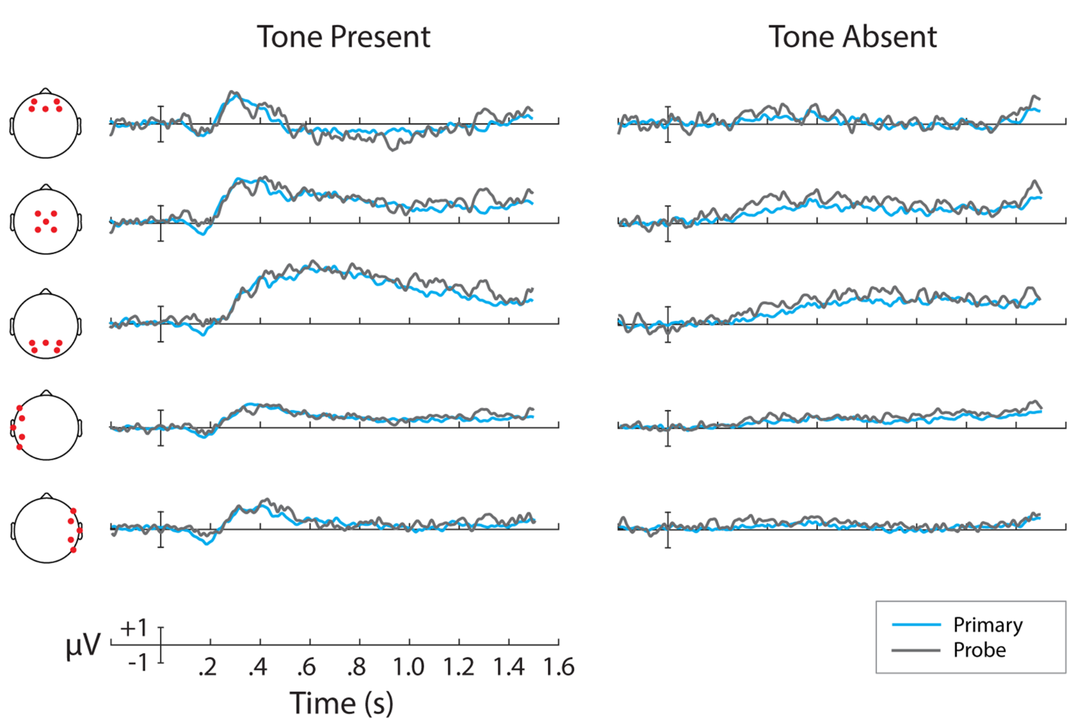


**Fig. 4.** ERPs at each electrode cluster for when the tone was present (left) and absent (right). ERPs corresponding to trials with primary tones are in cyan. Probe trials are depicted in grey. Red points in scalps to the left of each row represent locations within the electrode cluster for that row. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

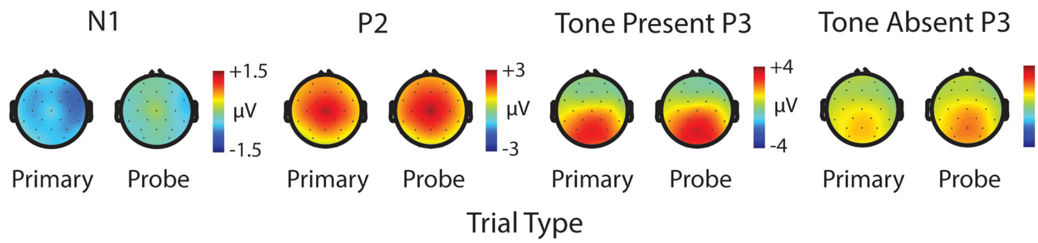


**Fig. 5.** Scalp maps of N1, P2, tone-present P3, and tone-absent P3 broken up by trial type.

and probe trials observable in the scalp maps of N1 and the tone-absent P3. However, unlike the differences between high- and low-confidence trials, the differences between primary and probe trials appeared minimal across all components.

We conducted 2 (Trial Type: Primary, Probe) × 5 (Electrode Cluster: Frontal, Central, Posterior, Left, Right) repeated measures ANOVAs on component amplitudes. For N1, there were no significant effects found, $Fs < 2.26$. For P2, there was a significant main effect of electrode cluster, $F(4,76) = 4.98$, $p = .001$, $\eta_p^2 = .21$ (cf. Section 3.2), however, the main effect of trial type and trial type × cluster interaction were not significant, $Fs < 2$. There were significant main effects of electrode cluster for the tone-present P3, $F(4,76) = 29.20$, $p < .001$, $\eta_p^2 = .61$, and tone-absent P3, $F(4,76) = 9.98$, $p < .001$, $\eta_p^2 = .34$. However, no other main effects or interactions were found, $Fs < 2$. Thus, similar to the behavioral data reported in Section 3.1, there was little evidence for any ERP differences between primary and probe trials.

### *3.4. ERP/behavior relationships*

It was unexpected that the primary/probe manipulation would reveal little difference in behavior and ERPs. Rather, the largest differences reported thus far were between trials given high- compared to low-confidence ratings. We thus conducted analyses to inquire into the relationship between confidence on primary and probe trials and the ERP data. Of particular interest to us was whether or not participants' differences in ERPs between primary and probe trials correlated with differences in confidence ratings made on primary and probe trials. In order to characterize differences in ERPs, we calculated a difference measure for the N1, P2, tone-present P3, and tone-absent P3. For each ERP component, the difference between primary and probe trials was characterized by the primary-minus-probe amplitude difference. Similarly, the behavioral difference between primary and probe trials was considered to be average confidence for primary trials minus average confidence for probe trials.

Fig. 6 shows scatterplots of each individual's difference in average confidence ratings for primary and probe trials (y-axes) as a function of their difference in component amplitude for primary and probe trials (x-axes). The relationship was significant for the N1, $r(18) = -.61$, $p = .005$, and tone-present P3, $r(18) = .60$, $p = .006$. Both significant relationships trended such that there were larger amplitudes on primary compared to probe trials when confidence was greater on primary compared to probe trials. Correlations were not significant for the P2 and tone-absent P3.

## 4. Discussion

In this study, participants were asked to provide confidence ratings regarding the accuracy of their responses in a 2i-2afc auditory detection task. The probabilities of a tone presented at 1000 Hz and 2500 Hz were manipulated in an attempt to systematically induce different accuracies and confidence ratings (Scharf, 1998). The intent was to examine, with ERPs, how metacognitive judgments relate to several different stages of task-related processing when detectability was similar across trials (3 dB above ~71% correct thresholds). Surprisingly, the manipulation of tone probability (i.e., the primary vs. probe manipulation) failed to yield any significant effects in either behavior or ERPs. However, participants' confidence ratings paralleled accuracy such that trials given high confidence showed a greater proportion correct than trials given low confidence. Further, after splitting trials up into high- and low-confidence trials and computing ERPs, several ERP components were clearly related to confidence. Correlational analyses of confidence ratings and ERP amplitudes between primary and probe trials also revealed that individual differences in the N1 and P3 to primary and probe presentations were related to differences in confidence ratings.

The finding that N1 was larger in amplitude for high- compared to low-confidence trials is unlike many of the recent ERP results employing non-auditory tasks, but similar to early auditory work (Squires et al., 1973). The result is also consistent with the work of one group who recently found neural responses to be related to confidence early in visual perception. Graziano, Parra, and Sigman (2015) sought neural markers of confidence at the single trial level. Participants were asked to indicate the identity of the letter in an array at a cued location and rate confidence on each trial. They found confidence was partially indexed by an early stage during the initial processing of the stimulus (overlapped with the P1 and P2 components or the visual-evoked

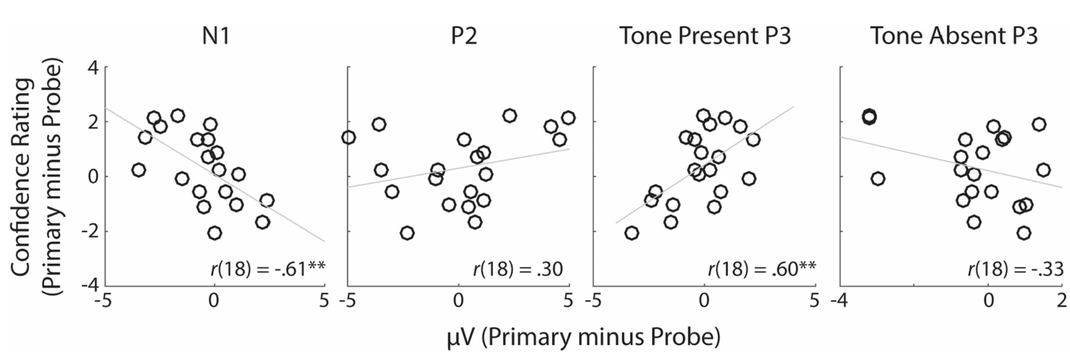


**Fig. 6.** Scatterplots of primary minus probe N1, P2, tone-present P3, and tone-absent P3 amplitudes as they relate to primary minus probe confidence ratings. Pearson's *r* is shown for each correlation. Lines represent best linear fits of the data. Asterisks highlight significant correlations ($^{**}p < .01$).

potential), suggesting that relationships between metacognitive judgments and early sensory processing are not restricted to the auditory modality.

P2 amplitudes were also larger on high- compared to low-confidence trials. Our ERP data revealed a clear N1/P2 complex at frontocentral and left- and right-lateralized locations when a tone was present (see Fig. 2, left panel). Indeed, the N1 and P2 commonly co-occur in the auditory evoked potential (for review, see Billings et al., 2011; Näätänen & Picton, 1987). It is not all too surprising that both N1 and P2 behaved similarly. It might also be the case that the P2 amplitude tracked confidence in Squires et al.'s (1973) study. Though Squires et al. (1973) presumed the positive peak on high-confidence trials to be a P3, they also acknowledged that the P2 component may have contributed to the effects (p. 267). Because they only measured ERP amplitudes from a single vertex electrode, it was difficult to determine which component their results reflected. Our results, which contained data from an array of electrodes covering the scalp, showed that both P2 *and* P3 (discussed next) showed significant differences between confidence levels. Thus, confidence did not parallel either, but instead paralleled both. That the amplitudes of components in the N1/P2 complex paralleled confidence, even when the stimuli presented were matched for detectability, suggests that trial-by-trial variations in cortical sensory processing contribute to an individual's confidence.

For tone-present P3 amplitudes, we found a confidence by electrode cluster interaction indicating that amplitude was larger on trials given high- compared to low-confidence ratings primarily at the posterior cluster of electrodes. This is consistent with the known scalp topography of the P3 component (for review, see Luck, 2005; Donchin, 1981), and repeated findings that the amplitude of the P3 component can increase with increased confidence in auditory detection tasks (e.g., Hillyard et al., 1971; Paul & Sutton, 1972, 1973; Squires et al., 1973, 1975). Recent work continues to examine the P3 in the context of metacognition, but primarily in visual tasks (e.g., Curran, 2004; Del Cul et al., 2007; Kouider et al., 2013). In Desender et al. (2016), participants decided whether a target arrow was pointed left or right. An arrow was presented just prior to the target arrow presentation as a prime. Primes were either congruent (pointed in same direction as the target) or incongruent (pointed in opposite direction as the target). As a consequence, congruency varied metacognitive experiences of trial-by-trial difficulty. Participants tended to judge congruent trials as "easy" and incongruent trials as "hard". Desender et al.'s (2016) ERP results showed that the stimulus-locked P3 component was modulated by congruency but that this modulation interacted with metacognitive ("easy" vs. "hard") judgments. Our results in an auditory detection task fit well with this ERP work. That is, both studies employed manipulations meant to facilitate or hinder performance, and subsequently found that trials separated by metacognitive ratings showed different ERPs.

On tone-absent intervals in our 2i-2afc task, there was a late positivity that resembled a P3. Interestingly, this tone-absent P3 was larger for low-confidence than high-confidence trials. That is, participants made low-confidence judgments on trials for which there was a larger P3 during the tone-absent interval. This supports the relationship between confidence and P3 even when there is no tone. Participants may use information, even in tone-absent intervals, to make confidence judgments. Similar conclusions have been made regarding high P3 amplitudes on false alarm trials in memory paradigms (Chen, Voss, & Guo, 2012). Although most work has associated greater amplitude P3s with higher confidence, it is interesting that large P3s for tone-absent intervals precedes low-confidence ratings. Our tone-absent result is consistent with the balance-of-evidence hypothesis, which predicts confidence as a function of the difference in sensory evidence accumulation between the selected and non-selected choice (Vickers, 1979). A large P3 amplitude during a tone-absent interval conflicts with any sensory evidence during the tone-present interval, contributing to low confidence in the accuracy of one's response. However, others have challenged this position, showing that conflicting information (e.g., evidence supporting the non-selected choice) does not contribute to confidence judgments (Zylberberg, Barttfeld, & Sigman, 2012). Our data suggest that it is too simple a story to say that larger P3 amplitudes mean greater confidence. The relationship between P3 amplitude and confidence depends upon the task at hand and the information being processed (e.g., presence vs. absence of a tone).

The probability manipulation (i.e., primary vs. probe) was expected to yield detection performance differences between 'primary' and 'probe' trials, but it did not. Indeed, unlike many previous studies (e.g., Scharf, 1998), detectability in our study was not significantly greater for a 'primary' tone than for a 'probe' tone, nor was confidence. ERPs were also comparable between primary and probe trials. As mentioned earlier, we believe that a lack of significant effects here might be due to a large frequency separation between primary and probe tones. Attentional filters set up by probability may have had excitatory and inhibitory regions that were outside the range of probe tone frequencies. Fritz, Shamma, Elhilali, & Klein (2003) found in ferrets that attention can facilitate neural responses to a target frequency while at the same time suppressing responses from neurons coding for frequencies immediately adjacent to the target. In our case, perhaps the probe frequency was too far removed (1500 Hz) from the target to be suppressed enough to see effects. A second possibility is that the probe, being 1500 Hz distant from the target frequency elicited a stronger response upon presentation due to novelty (for review, see Donchin, 1981). Still more, it is also possible that because listeners had knowledge that there were two different frequencies, they did not develop a consistent sensory perceptual template for one frequency over the other (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). For instance, some have proposed that distant frequency probes may lead listeners to adopt a multi-focus strategy by dividing attention among multiple frequency bands (Dai, Scharf, & Buus, 1991).

Even though primary vs. probe differences were not apparent in the grand-average data, our novel brain-behavior correlational analyses of primary and probe trials demonstrated in a new way the degree to which ERP dynamics parallel confidence ratings. We tested the correlation between primary minus probe confidence and primary minus probe ERP component amplitude. We found a significant correlation for N1 and tone-present P3. Individuals with larger differences between confidence on primary and probe trials displayed larger difference between primary and probe N1 and tone-present P3 component amplitude. These results demonstrate that the difference in an individual's ERP response to primary and probe tones predict confidence ratings made later on in the trial. They go further in demonstrating that the N1 *and* P3 are related to confidence on an individual differences level (cf. fMRI data; McCurdy et al., 2013). It is also worthwhile to note that both the correlational analyses and the grouped-level analyses convincingly demonstrate that ERP effects are related more clearly to confidence than stimulus frequency characteristics when sounds are detected at similar levels of accuracy (3 dB above ~71% correct thresholds).

Based on our data, and the existing literature, we believe that metacognitive judgments may be made using information at several stages of processing: sensory inputs, decision variables, predictions, decision processes, and a post-decision evaluation (Meyniel, Sigman, & Mainen, 2015). Metacognition may be task dependent with metacognitive judgments in different types of tasks relying disproportionally on one stage of processing over another. Tasks relying heavily on the quality of sensory information may show metacognitive judgments that are at least partially based on the quality of an internal representation of the stimulus, which can vary from trial-to-trial. In contrast, in tasks requiring the use of conceptual information (e.g., stimulus category or familiarity), or the monitoring of ongoing errors (e.g., mistakes in speeded decisions), metacognitive judgments may rely on later processing stages. Curran (2004) provides support for this position. He found parietal ERP (400–800 ms) amplitudes increased with confidence in recognizing old (but not new) items in a word list remember/know task. This finding was interpreted to mean that the parietal old/new effect—an effect which co-occurs with the P300 component (Bentin & McCarthy, 1994; Spencer, Vila Abad, & Donchin, 2000)—is not explained by generic decision processes producing confidence differences. Rather, it may be related to recollection process engaged with studied (old) items. Recent neuroimaging work is also supportive of this view. For example, Fleming, Ryu, Golfinos, and Blackmon (2014) found that lesions to the anterior prefrontal cortex (aPFC) impaired metacognitive ability (how well one's metacognitive ratings, such as confidence, discriminate between his or her correct and incorrect responses) in perceptual but not memory-related tasks. Additionally, McCurdy et al. (2013) found that grey matter volume in the aPFC predicted individual differences in visual metacognition while the medial parietal cortex grey matter volume predicted individual differences in memory metacognition (also, see Baird, Smallwood, Gorgolewski, & Margulies, 2013).

Not all neuroimaging work has supported the idea that metacognition processes are task-dependent. Recently, Lemaitre, Herbet, Duffau, and Lafargue (2018) demonstrated preserved perceptual metacognitive ability following brain damage and removal of Brodmann area 10, the anterior-most portion of the prefrontal cortex. This result conflicts with Fleming et al.'s (2014) aPFC lesions result. It is potentially useful to see how metacognition does and does not vary as a function of task type (de Gardelle & Mamassian, 2014; Song et al., 2011; Zakrzewski, 2016) and sensory modality (Ais, Zylberberg, Barttfeld, & Sigman, 2016; de Gardelle, Le Corre, & Mamassian, 2016; Faivre, Filevich, Solovey, Kühn, & Blanke, 2018). Metacognitive ability may generalize across sensory modalities when tasks in each modality are similar. For example, Faivre et al. (2018) found metacognitive efficiency to correlate across visual, auditory, and tactile tasks involving stimulus intensity judgments (in experiment 1) and congruency judgments (in experiment 2) but did not test supramodality across tasks that differed in paradigm (e.g., stimulus detection vs. stimulus categorization). Relatedly, McCurdy et al. (2013) found a behavioral correlation of metacognitive efficiency across memory and perceptual tasks, however, both used a 2AFC paradigm. In contrast, Baird et al. (2013) revealed no behavioral correlation across perceptual and memory tasks when the perceptual task involved an oddity task procedure while their memory task used a two-choice old/new discrimination procedure. In general, we believe that our work in the context of the existing literature supports a task-specific view of metacognition. That is, metacognitive processes vary depending upon the task at hand.

### 4.1. Considerations and future directions

It is important to note that it is not the aim of our study to identify the neural correlates of confidence alone. As Pouget, Drugowitsch, and Kepecs (2016) point out, confidence ratings may also depend on behavioral cues, context, bias, and so on. Therefore, any correlations between confidence ratings and neural activity may be the result of more than just internal monitoring of performance accuracy. We acknowledge this limitation and do not claim that our measure of ERP amplitudes index metacognition. For instance, the N1 is an exogenous ERP component that is elicited even in the absence of attention directed towards the sounds. It is plausible that the N1 reflects the quality of stimulus representations used as input in the metacognitive process. That quality may differ from trial to trial either because of the waxing and waning of attentional states (e.g., Boksem, Meijman, & Lorist, 2005), cortical state at stimulus onset (e.g., Hermann, Henry, Haegens, & Obleser, 2016), or experience-related changes in representations themselves (e.g., Budd, Barry, Gordon, Rennie, & Michie, 1998; Wisniewski, Radell, Guillette, Sturdy, & Mercado, 2012). The same could be said for the P3, although this is more debatable (see Gherman & Philiastides, 2015).

It should be further mentioned that other tasks might yield correlates of metacognitive judgments in other EEG features. Recently, Wöstmann, Herrmann, Wilsch, and Obleser (2015) found that 7–13 Hz alpha power during a difficult auditory number comparison task was related to confidence ratings. Lower alpha power between number presentations was associated with higher confidence. Based on theories wherein increases in alpha power reflect inhibition of task-irrelevant brain regions (e.g., Jensen & Mazaheri, 2010; also, see Wisniewski, Thompson, & Iyer, 2017), those authors interpreted increases of alpha power as indicative of more effortful attentional processes at play on low-confidence trials. In other words, early inhibition of task-irrelevant brain regions as well as listening effort and attention may contribute to subsequent confidence in the accuracy of one's response. Relatedly, Kubanek, Hill, Snyder, and Schalk (2015) found alpha activity following the presentation of an auditory stimulus predicted whether participants were "sure" or "unsure" of their choice. Additional tasks that may show an association between metacognitive judgments and sensory and/or stimulus-processing related features of EEG (besides N1) include those in which stimulus rate is important. These tasks might show confidence judgments that track variability in the brain's ability to synchronize to stimulus presentation rate (indexed by steady-state responses; Luck, 2005). Infrequent stimuli presented among frequently-presented standard stimuli are well known to generate a mismatch-negativity (MMN; Näätänen, Paavilainen, Rinne, & Alho, 2007). Other tasks in which the recognition in a change from frequent input may show a parallel of the MMN with confidence ratings. Our results are not at odds with these possibilities. The purpose of our study is to understand what the relationship between stimulus-locked ERP components (related to early and late stages of processing) and confidence might tell us about the stages of metacognition.

Although the present study was designed to examine the neural correlates of human metacognition, our findings have important implications for non-human animal (hereafter, animal) research as well. Researchers have developed ways to measure metacognition without verbal report, revealing similar behavioral patterns when humans and animals face difficulty (e.g., Kornell, Son, & Terrace, 2007; Smith, 2009; Smith, Beran, & Couchman, 2012). However, there is a debate over whether or not uncertainty responses (the inverse of confidence; Meyniel et al., 2015) are driven by associative, reinforcement learning processes (e.g., Basile, Schroeder, Brown, Templar, & Hampton, 2015; Carruthers, 2008; Jozefowiez, Staddon, & Cerutti, 2009; Le Pelley, 2012; Smith, Zakrzewski, & Church, 2016). This disagreement extends to neuroscience results (e.g., Kiani & Shadlen, 2009; Middlebrooks & Sommer, 2012; Paul et al., 2015). Even though our task gave no feedback, we replicated the N1, tone-present P3, and tone-absent P3 confidence results for correct-only trials (see *Supplementary Materials*). The present findings show a pattern of neural activity that is related to confidence under the same (internal) reinforcement conditions, challenging associative-based accounts of metacognition (e.g., Le Pelley, 2012).

Some researchers have already taken a crucial next step, asking whether or not neural features predict behaviors related to confidence in animals (e.g., Kepecs, Uchida, Zariwala, & Mainen, 2008; Kiani & Shadlen, 2009; Komura, Nikkuni, Hirashima, Uetake, & Miyamoto, 2013; Middlebrooks & Sommer, 2012). For example, Kiani and Shadlen (2009) showed neural responses in monkeys' parietal cortex related to certainty in an opt-out task. However, this and other opt-out tasks (e.g., Komura et al., 2013) still provide a small reward for opt-out responses, making behavioral interpretations less clear. Additionally, many studies vary stimulus characteristics across a continuum (e.g., color, motion, density, etc.), producing metacognitive responses that may relate to changes in the stimulus. The present study shows metacognitive responses relating to an internal response while the stimulus (tone) remains essentially constant on all trials. Future behavioral work controlling stimulus characteristics, coupled with neuroimaging, such as EEG, might allow a robust extension of the present ERP results to animals. Such work might reveal the time-course of confidence or uncertainty in animals as well.

### 4.2. Conclusions

We used the ERP method to examine processing that relates to metacognitive judgments in a simple auditory detection task. Auditory N1, P2, and P3 amplitudes paralleled ratings of confidence. Additionally, differences in confidence on two trial types ('primary' and 'probe') correlated with differences in N1 and tone-present P3 amplitude on those two trial types. The current results in the context of the literature suggest a need to examine how metacognition in different types of tasks may rely on different processes. They also open up possibilities to study metacognitive issues in new ways.

## Acknowledgements

Thank you to members of the Battlespace Acoustics branch of the 711th Human Performance Wing at the U.S. Air Force Research Laboratory for helpful discussions on this data and topic.

## Funding

This research was performed while A.C.Z. and M.G.W. held an Oak Ridge Institute for Science and Education Internship and Postdoctoral Fellowship at the U.S. Air Force Research Laboratory.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandc.2018.10.007.

## References

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition, 146*, 377–386. https://doi.org/10.1016/j.cognition.2015.10.006.

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *Journal of Neuroscience, 33*(42), 16657–16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013.

Ball, N. J., Wisniewski, M. G., Zakrzewski, A. C., Iyer, N., Simpson, B. D., Seccia, A., ... Spencer, N. (2017). Learning-related improvements in auditory detection sensitivities correlate with neural changes observable during active and passive sound processing. *Proceedings of Meetings on Acoustics, 30*, 050012. https://doi.org/10.1121/2.0000630.

Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology: General, 144*(1), 85–102. https://doi.org/10.1037/xge0000031.

Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(1), 130–149. https://doi.org/10.1037/0278-7393.20.1.130.

Billings, C. J., Bennett, K. O., Molis, M. R., & Leek, M. R. (2011). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear and Hearing, 32*(1), 53–60. https://doi.org/10.1097/AUD.0b013e3181ec5c46.

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research, 25*, 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience, 35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31*, 51–68. https://doi.org/10.1016/S0167-8760(98)00040-3.

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language, 23*(1), 58–89. https://doi.org/10.1111/j.1468-0017.2007.00329.x.

Chen, H., Voss, J. L., & Guo, C. (2012). Event-related brain potentials that distinguish false memory for events that occurred only seconds in the past. *Behavioral and Brain Functions, 8*, 36. https://doi.org/10.1186/1744-9081-8-36.

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42*(8), 1088–1106. https://doi.org/10.1016/j.neuropsychologia.2003.12.011.

Dai, H., Scharf, B., & Buus, S. R. (1991). Effective attenuation of signals in noise under focused attention. *The Journal of the Acoustical Society of America, 89*(6), 2837–2842. https://doi.org/10.1121/1.400721.

de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE, 11*(1), e0147901. https://doi.org/10.1371/journal.pone.0147901.

de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science, 25*(6), 1286–1288. https://doi.org/10.1177/0956797614528956.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(10), 2408–2423. https://doi.org/10.1371/journal.pbio.0050260.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Desender, K., Van Opstal, F., Hughes, G., & Van den Bussche, E. (2016). The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes. *Neuropsychologia, 82*, 54–64. https://doi.org/10.1016/j.neuropsychologia.2016.01.003.

Díaz, F., Lindín, M., Galdo-Alvarez, S., Facal, D., & Juncos-Rabadán, O. (2007). An event-related potentials study of face identification and naming: The tip-of-the-tongue state. *Psychophysiology, 44*(1), 50–68. https://doi.org/10.1111/j.1469-8986.2006.00483.x.

Donchin, E. (1981). Surprise!... surprise? *Psychophysiology, 18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x.

Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition. *The Journal of Neuroscience, 38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B Biological Sciences, 367*, 1338–1349. https://doi.org/10.1098/rstb.2011.0417.

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain, 137*(10), 2811–2822. https://doi.org/10.1093/brain/awu221.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6*, 1216–1223. https://doi.org/10.1038/nn1141.

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage, 106*, 134–143. https://doi.org/10.1016/j.neuroimage.2014.11.036.

Graziano, M., Parra, L. C., & Sigman, M. (2015). Neural correlates of perceived confidence in a partial report paradigm. *Journal of Cognitive Neuroscience, 27*(6), 1090–1103. https://doi.org/10.1162/jocn_a_00759.

Hermann, B., Henry, M. J., Haegens, S., & Obleser, J. (2016). Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. *Neuroimage, 124*, 487–497. https://doi.org/10.1016/j.neuroimage.2015.09.019.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177.

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science, 172*(3990), 1357–1360. https://doi.org/10.1126/science.172.3990.1357.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience, 4*, 186. https://doi.org/10.3389/fnhum.2010.00186.

Jozefowiez, J., Staddon, J. E. R., & Cerutti, D. T. (2009). The behavioral economics of choice and interval timing. *Psychological Review, 116*(3), 519–539. https://doi.org/10.1037/a0016171.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163–178. https://doi.org/10.1017/S0048577200980259.

Kao, Y. C., Davis, E. S., & Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience, 8*(12), 1776–1783. https://doi.org/10.1038/nn1595.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227–231. https://doi.org/10.1038/nature07200.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759–764. https://doi.org/10.1126/science.1169405.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*(2), 329–343. https://doi.org/10.2307/1422236.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., & Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nature Neuroscience, 16*(6), 749–755. https://doi.org/10.1038/nn.3393.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349.

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*(1), 64–71. https://doi.org/10.1111/j.1467-9280.2007.01850.x.

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., De Gardelle, V., ... Dehaene-Lambertz, G. (2013). A neural marker of perceptual consciousness in infants. *Science, 340*(6130), 376–380. https://doi.org/10.1126/science.1232509.

Kubanek, J., Hill, J., Snyder, L. H., & Schalk, G. (2015). Cortical alpha activity predicts the confidence in an impending action. *Frontiers in Neuroscience, 9*, 243. https://doi.org/10.3389/fnins.2015.00243.

Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning Memory and Cognition, 38*(3), 686–708. https://doi.org/10.1037/a0026478.

Leek, M. R., Hanna, T. E., & Marshall, L. (1992). Estimation of psychometric functions from adaptive tracking procedures. *Perception & Psychophysics, 51*(3), 247–256. https://doi.org/10.3758/BF03212251.

Lemaitre, A. L., Herbet, G., Duffau, H., & Lafargue, G. (2018). Preserved metacognitive ability despite unilateral or bilateral anterior prefrontal resection. *Brain and Cognition, 120*, 48–57. https://doi.org/10.1016/j.bandc.2017.10.004.

Lindín, M., & Díaz, F. (2010). Event-related potentials in face naming and tip-of-the-tongue state: Further results. *International Journal of Psychophysiology, 77*(1), 53–58. https://doi.org/10.1016/j.ijpsycho.2010.04.002.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, MA: MIT Press.

Makeig, S. (1993). Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. *Electroencephalography and Clinical Neurophysiology, 86*(4), 283–293. https://doi.org/10.1016/0013-4694(93)90110-H.

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience, 33*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013.

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron, 88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039.

Middlebrooks, P. G., & Sommer, M. A. (2012). Neuronal correlates of meta- cognition in primate frontal cortex. *Neuron, 75*, 517–530. https://doi.org/10.1016/j.neuron.2012.05.028.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375–425.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences, 24*, 283–288. https://doi.org/10.1016/S0166-2236(00)01790-2.

Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., & Pantev, C. (2007). Attention improves population-level frequency tuning in human auditory cortex. *Journal of Neuroscience, 27*(39), 10383–10390. https://doi.org/10.1523/JNEUROSCI.2963-07.2007.

Paul, E. J., Smith, J. D., Valentin, V. V., Turner, B. O., Barbey, A. K., & Ashby, F. G. (2015). Neural networks underlying the metacognitive uncertainty response. *Cortex, 71*, 306–322. https://doi.org/10.1016/j.cortex.2015.07.028.

Paul, D. D., & Sutton, S. (1972). Evoked potential correlates of response criterion in auditory signal detection. *Science, 177*(4046), 362–364. https://doi.org/10.1126/science.177.4046.362.

Paul, D. D., & Sutton, S. (1973). Evoked potential correlates of psychophysical judgments: The threshold problem. A new reply to Clark, Butler, and Rosner. *Behavioral Biology, 9*(4), 421–433. https://doi.org/10.1016/S0091-6773(73)80062-8.

Paynter, C. A., Reder, L. M., & Kieffaber, P. D. (2009). Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia, 47*(3), 796–803. https://doi.org/10.1016/j.neuropsychologia.2008.12.009.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*(3), 366–374. https://doi.org/10.1038/nn.4240.

Scharf, B. (1998). Auditory attention: The psychoacoustic approach. In H. Pashler (Ed.). *Attention* (pp. 75–117). UK: Psychology Press East Sussex.

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141.

Skavhaug, I. M., Wilding, E. L., & Donaldson, D. I. (2010). Judgments of learning do not reduce to memory encoding operations: Event-related potential evidence for distinct metacognitive processes. *Brain Research, 1318*, 87–95. https://doi.org/10.1016/j.brainres.2009.11.047.

Skavhaug, I. M., Wilding, E. L., & Donaldson, D. I. (2013). Immediate judgments of learning predict subsequent recollection: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning Memory and Cognition, 39*(1), 159–166. https://doi.org/10.1037/a0028885.

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences, 13*(9), 389–396. https://doi.org/10.1016/j.tics.2009.06.009.

Smith, J. D., Beran, M. J., & Couchman, J. J. (2012). Animal metacognition. In T. Zentall, & E. Wasserman (Eds.). *Comparative Cognition: Experimental Explorations of Animal Intelligence* (pp. 282–304). Oxford, UK: Oxford University Press.

Smith, J. D., Zakrzewski, A. C., & Church, B. A. (2016). Formal models in animal-metacognition research: The problem of interpreting animals' behavior. *Psychonomic Bulletin and Review, 23*(5), 1341–1353. https://doi.org/10.3758/s13423-015-0985-2.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition, 20*(4), 1787–1792. https://doi.org/10.1016/j.concog.2010.12.011.

Spencer, K. M., Vila Abad, E., & Donchin, E. (2000). On the search for the neurophysiological manifestation of recollective experience. *Psychophysiology, 37*(4), 494–506. https://doi.org/10.1017/S0048577200982210.

Squires, K. C., Hillyard, S. A., & Lindsay, P. H. (1973). Vertex potentials evoked during auditory signal detection: Relation to decision criteria. *Perception & Psychophysics, 14*(2), 265–272. https://doi.org/10.3758/BF03212388.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued observation intervals. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 268–279. https://doi.org/10.1037/0096-1523.1.3.268.

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience, 8*, 350. https://doi.org/10.3389/fnhum.2014.00350.

Vickers, D. (1979). *Decision Processes in Visual Perception.* New York, NY: Academic Press.

Wisniewski, M. G. (2017). Indices of effortful listening can be mined from existing electroencephalographic data. *Ear and Hearing, 38*(1), e69–e73. https://doi.org/10.1097/AUD.0000000000000354.

Wisniewski, M. G., Mercado, E., III, Church, B. A., Gramann, K., & Makeig, S. (2014). Brain dynamics that correlate with effects of learning on auditory distance perception. *Frontiers in Neuroscience, 8*, 396. https://doi.org/10.3389/fnins.2014.00396.

Wisniewski, M. G., Radell, M. L., Guillette, L. M., Sturdy, C. B., & Mercado, E., III (2012). Predicting shifts in generalization gradients with perceptrons. *Learning & Behavior, 40*, 128–144. https://doi.org/10.3758/s13420-011-0050-6.

Wisniewski, M. G., Romigh, G. D., Kenzig, S. M., Iyer, N., Simpson, B. D., Thompson, E. R., & Rothwell, C. D. (2016). Enhanced auditory spatial performance using individualized head-related transfer functions: An event-related potential study. *Journal of the Acoustical Society of America, 140*, EL539–EL544. https://doi.org/10.1121/1.4972301.

Wisniewski, M. G., Thompson, E. R., & Iyer, N. (2017). Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult. *Psychophysiology, 54*, 1916–1928. https://doi.org/10.1111/psyp.12968.

Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience, 35*(4), 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015.

Zakrzewski, A. C. (2016). Testing generalization of metacognitive accuracy (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (Accession Order No. 10127630).

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience, 6*, 79. https://doi.org/10.3389/fnint.2012.00079.