The Feeling of Another's Knowing: How "Mixed Messages" in Speech Are Reconciled

Article in Journal of Experimental Psychology Human Perception & Performance · April 2016
Impact Factor: 3.36 · DOI: 10.1037/xhp0000240

2 authors:

Xiaoming Jiang
McGill University
24 PUBLICATIONS 156 CITATIONS

Marc D. Pell
McGill University
109 PUBLICATIONS 2,418 CITATIONS

Available from: Xiaoming Jiang
Retrieved on: 13 May 2016
The Feeling of Another’s Knowing: How “Mixed Messages” in Speech Are Reconciled
Xiaoming Jiang and Marc D. Pell
Online First Publication, April 28, 2016. http://dx.doi.org/10.1037/xhp0000240

CITATION
The Feeling of Another’s Knowing:  
How “Mixed Messages” in Speech Are Reconciled

Xiaoming Jiang and Marc D. Pell  
McGill University

Listeners often encounter conflicting verbal and vocal cues about the speaker’s feeling of knowing; these “mixed messages” can reflect online shifts in one’s mental state as they utter a statement, or serve different social-pragmatic goals of the speaker. Using a cross-splicing paradigm, we investigated how conflicting cues about a speaker’s feeling of (un)knowing change one’s perception. Listeners rated the confidence of speakers of utterances containing an initial verbal phrase congruent or incongruent with vocal cues in a subsequent statement, while their brain potentials were tracked. Different forms of conflicts modulated the perceived confidence of the speaker, the extent to which was stronger for female listeners. A confident phrase followed by an unconfident voice enlarged an anteriorly maximized negativity and late positivity for male listeners, suggesting that mental representations of another’s feeling of knowing in face of this conflict were hampered by increased demands of integration for males and increased demands on updating for females. An unconfident phrase followed by a confident voice elicited a delayed sustained positivity (from 900 ms) in female participants only, suggesting females generated inferences to moderate the conflicting message about speaker knowledge. We highlight ways that verbal and vocal cues are real-time integrated to access a speaker’s feeling of (un)knowing, while arguing that females are more sensitive to the social relevance of conflicting speaker cues.

Keywords: ERPs, mentalizing, metacognition, pragmatics, prosody
verbal description, when these cues were present in the same message (e.g., a doubtful text of legal argument produced in doubtful voice, Scherer et al., 1973; Walker, 1977). The interaction of expressed confidence encoded by vocal and verbal cues is especially relevant when these meanings appear to conflict, which can serve certain pragmatic functions. For example, an utterance produced in a relatively confident voice but accompanied by fillers (um; uh), or preceded by an unexpected verbal phrase with both verbal and vocal cues signaling lack of confidence (e.g., Maybe), could imply hesitation or a face-saving strategy by lowering the potential threat of an incorrect, confident statement (Bonnefon & Villejoubert, 2006); irrespective of their social function, these conflicting cues would modulate impressions of the speaker’s feeling of knowing. Similarly, verbal expressions of high confidence (Let’s be clear . . .) followed by vocal cues associated with uncertainty are likely to promote distinct interpretations about the speaker’s goals and true feeling of (un)knowing (Mol et al., 2013; Smith & Clark, 1993). Our study focused specifically on these conflict situations, in which multiple cues in the contextual phrase and main statement must be differentially weighed to arrive at a mental representation of the speaker’s feeling of (un)knowing, to further elucidate the neurocognitive system involved.

**Neural Evidence of Accessing Speaker’s Feeling of Knowing**

Studies illuminating the neurorecognition of ‘feeling of knowing’ are recent and relatively few. In terms of the underlying neural architecture, access to one’s own feeling of knowing or unknowing has been linked to different neural networks (White, Engen, Sørensen, Overgaard, & Shergill, 2014). In conditions of heightened uncertainty, increased activation has been found in the salience (anterior cingulate cortex and insula) and central executive (dorsolateral prefrontal and posterior parietal cortices) networks; in contrast, subjective confidence seems to engage a more ‘default’ network involving midline cortical and medial temporal lobe regions. Accessing another’s feeling of knowing recruits not only neural structures for encoding another’s behavioral cues (‘how’ a speaker is showing his feeling of knowing), but for attributing these expressions to mental states (‘why’ a speaker is showing his feeling of knowing, Kuhlen, Bogler, Swerts, & Haynes, 2015). In a functional MRI (fMRI) study in which participants viewed respondents answering questions in an unconfident manner (based on visual cues), stronger activation was observed in the mentalizing network (superior medial prefrontal cortex and bilateral temporal-parietal junction) as compared to when they viewed them responding with full confidence. These results argue that processes for mentalizing underpin pragmatic inferences about a speaker's feeling of (un)knowing (Kuhlen et al., 2015). These mentalizing networks seem to be activated more strongly when processing statements that contradict representations already formed about a speaker based on previous evidence; for example, this occurred when processing statements indicating that a character, who was described as being very friendly, “dared two strangers to fight” (Van Duynslaeger, Van Overwalle, & Verstraeten, 2007). These findings emphasize that pragmatic inferences are sometimes initiated to reconcile conflicting representations about another person.

With respect to the cognitive structure and time course of this processing, recent studies emphasize that the perceived confidence of a speaker is registered at distinct time points in the brain when this knowledge is revealed by vocal cues alone (Jiang & Pell, 2015) or by combined verbal and vocal information (Jiang & Pell, 2016). Using event-related potentials (ERPs), it was first demonstrated that linguistically identical utterances spoken in a confident and doubtful voice are differentiated at 200 ms following onset of the vocal expression (P200), whereas statements conveying an intermediate level of confidence (i.e., slight feeling of unknowing) or spoken in a neutral tone were differentiated from confident expressions at around 340 ms (late positivity) and 970 ms (delayed, sustained positivity), respectively (Jiang & Pell, 2015). Interestingly, the same confident/unconfident vocal expressions were differentiated earlier in the neural response (100 ms) when the statement is preceded by a meaning-congruent probability phrase indicating speaker confidence (I’m positive/ Maybe . . .); the presence of the probability phrase significantly reduced the amplitude of N1, P2, and N400 waves when processing the vocal expression of confidence in the following statement (Jiang & Pell, 2016). The comparison of the high versus low confidence expressions in these studies may be explained by low-level acoustic variation, because the confidence manipulation was marked by distinct vocal patterns for each expression (e.g., f0, speech rate, etc.). However, in keeping with three-stage models of vocal expression processing (e.g., Schirmer & Kotz, 2006), these data supply neurophysiological evidence that an utterance-initial phrase conveying speaker confidence in both lexical and vocal form creates a strong expectation that affects the processing of a meaning-congruent vocal stimulus in the subsequent statement; this biases listener responses at early stages of acoustic analysis (N1), coarse coding (‘salience detection,’ P2), and semantic access and elaboration (N400) of a speaker’s feeling of knowing. However, it is presently unclear how these operations are affected when the contextual phrase conflicts with vocal expressions of speaker confidence in the main statement as they sometimes do in daily interactions. Moreover, it is unclear how these contextual effects would vary as a function of the level of confidence encoded by vocal cues in the main statement.

Similar studies comparing ERP responses to vocal cues conveying basic emotions, versus German statements with combined linguistic and vocal cues, have been performed (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008). Emotional utterances were “cross-spliced” in different ways to create a mismatch between an initial emotional interpretation conveyed vocally and the emotion conveyed by subsequent lexical/vocal information (e.g., She had[neutral] found the treasure[happy]). An N400-like effect (starting around 350 ms) was noted when both lexical and vocal information mismatched the emotional context, whereas a late positivity effect (starting around 600 ms) was found when only vocal cues mismatched the context. The late positivity was linked to increased effort to extract acoustic parameters needed to assign significance to these expressions (Jiang & Pell, 2015; Kotz & Paulmann, 2007; Rigoulot, Fish, & Pell, 2014; Schirmer et al., 2006). In contrast, the N400-like effect was attributed to differences in accessing (emotional) meaning, which showed facilitation when expectancies triggered by the initial prosodic context matched subsequent stimulus details (Kotz & Paulmann, 2007). This general approach can be usefully adapted to investigate how utterance-initial phrases encoding speaker confidence, both vocally and semantically, alter neural responses to vocal confidence expressions when the two
The Present Study

Here, we cross-spliced utterance-initial probability phrases that mark a speaker’s confidence level (I’m certain . . . / Maybe . . .) with statements produced in a confident or unconfident voice in a 2 x 2 factorial design as listeners’ electroencephalogram was recorded. Given evidence that listeners integrate verbal and vocal cues in the contextual phrase and vocal cues in the main statement in an ongoing manner to form a mental representation of a speaker’s feeling of knowing (Jiang & Pell, 2016), this paradigm allowed us to examine how listeners reconcile cross-channel conflicts during this form of mentalizing. Also, we could attempt to pinpoint what cognitive stage(s) this occurs at as listeners evaluate a speaker’s feeling of knowing, that is, causing difficulties in semantic access, continuous evaluation, or inferential processes in face of detected deviance. This paradigm has particular merits for examining how these processes vary with the level of confidence while controlling for the acoustic features that accompany the interpretation of speaker confidence (Jiang & Pell, 2015, 2016).

We hypothesized that phrases that conflict with subsequent vocal cues would attenuate perceptual impressions of confident vocal expressions but enhance unconfident vocal expressions when compared with conditions involving congruent cues. If listeners detect a deviance in the predicted acoustic pattern of utterances following the verbal expression, we expected an increased late positivity for incongruent versus congruent trials, and possibly a preceding N400 effect if the conflict promotes greater difficulty integrating vocal information into the contextual expectancy (Kotz & Paulmann, 2007). We also predicted that our two conflict situations (Utterance-initial Phrase[confident] Main Statement[unconfident] vs. Utterance-initial Phrase[unconfident] Main Statement[confident]) would have asymmetrical effects on the listener in their attempt to derive a coherent impression of the speaker’s feeling of knowing. For instance, a confident voice following a doubtful phrase could imply one’s hesitation, leading to pragmatic inferences about the speaker; in such contexts in which an expected way of speaking or acting conflicts with utterance contents, a delayed sustained positivity is often observed (Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009; Van Overwalle, Van den Eede, Baetens, & Vandekerckhove, 2009). In contrast, an unconfident voice following a confident phrase is less frequent in communication and more likely to engage a conflict control process (e.g., to inhibit information from one of the conflicting channels). Finally, given that male and female listeners were differentially sensitive to vocal cues about speaker confidence (Jiang & Pell, 2015) and in how they integrated an utterance-initial phrase with a congruent vocal expression (Jiang & Pell, 2016), we anticipated that participant sex would mediate certain neural responses to conflicting messages about a speakers’ feeling of knowing, especially at late (inferential) processing stages (where females tend to exhibit an enhanced response). The role of interpersonal sensitivity in how feeling of knowing is neurally registered was also examined in light of data linking this variable to differences in a sustained positivity associated with pragmatic language inferences (Jiang & Pell, 2016).

Method

Participants

Thirty university students consented to participate in the EEG experiment (15 female, mean age = 22.6 years, range = 18 to 30 years). All were right-handed, native Canadian English speakers who took part in a companion study focusing on vocal confidence perception (Jiang & Pell, 2015). None had experienced major psychiatric or neurological illness, speech, or hearing problems. At study onset, all participants were assessed on a measure of trait empathy/interpersonal sensitivity (IRI, Davis, 1983), gauging their ability to understand both the cognitive and emotional state of another person. At time of testing, female and male participants did not differ in mean age (Female: 22.2 ± 3.8, Male: 23.0 ± 3.7), years of education (Female: 15.7 ± 2.4, Male: 15.1 ± 2.1), handedness,1 (Female: .73 ± .21, Male: .67 ± .45), or interpersonal sensitivity (IRI total score: Female: 63.6 ± 10.5, Male: 59.3 ± 7.8). The research was carried out in accordance with the Declaration of Helsinki and ethically approved by the Faculty of Medicine Institutional Review Board at McGill University (Montréal, Canada).

Material and Design

Ninety-six utterance pairs were first selected from a database of vocal confidence recordings (Jiang & Pell, 2014). Each pair consisted of two variants of the same statement (She left the key by the fridge) recorded in a confident or unconfident tone of voice, preceded by a matching confident or unconfident probability phrase. These phrases varied in how likely what is addressed in the following statement agreed with the speaker’s knowledge or opinion. Confident utterances began with I’m certain/I’m positive/For sure/Definitely, and were meant to communicate absolute certainty that what the speaker says is true. Unconfident statements began with Perhaps/Maybe/There’s a chance/It’s possible and were expressed to communicate only a possibility that what the speaker says is true. Statements served three communicative functions: description of fact, making a judgment, or stating an intention (32 pairs/function, Table 1). Recordings were selected for 2 female and 2 male encoders who exhibited the most salient perceptual distinctions between confident and unconfident expressions in a validation study (Jiang & Pell, 2014, 2015). All vocal expressions were elicited by encoders in response to a personal question delivered face-to-face by the examiner (a native speaker), such as What did he do? - He left the key by the fridge. The encoders were encouraged to act out the intended level of confidence as naturally as possible and to avoid exaggeration. All sound recordings were mean-intensity normalized to 75dB for perceptual validation. On a 5-pt scale of perceived speaker confidence, items in the confident condition had a mean rating of 4.44 (SD = .74) and a mean rating of 1.86 (SD = .87) in the unconfident condition. Confident and unconfident stimuli displayed significant acoustic differences in both the main utterance (Jiang & Pell, 2015) and in the utterance-initial probability phrase (Jiang & Pell, 2016); relative to uncon-

1 The handedness measure was scored for each item and then summed for each hand according to Oldfield (1971). The final score was represented as (right − left)/(right + left).
fident expressions, confident expressions had a lower mean pitch, a higher mean and variation of intensity, and faster speaking rate.

Using these base recordings, two mismatch conditions were created from selected pairs using a cross-splicing method by extracting the initial phrase from each utterance and pairing it with the same statement conveying the opposite vocal meaning in the main utterance. During editing, the initial phrase was cut exactly before the onset of the main statement in the original recording. This yielded two types of incongruence between the initial phrase and main utterance: unconfident phrase + confident statement (Unconfident–Confident) and confident phrase + unconfident statement (Confident–Unconfident). When combined with the original congruent utterances (Confident–Confident, Unconfident–Unconfident), each statement appeared in four distinct presentation conditions of current interest (2 congruent, 2 incongruent). The initial phrase and the statement were cross-spliced between two critical sets of utterances of the same speaker and of the same communicative function, so that each phrase appeared equally often in the congruent and incongruent conditions for each communicative function. The edited recordings were evaluated by two research assistants (native English speakers) to eliminate any utterances that were perceived as unnatural. Experimental items were intermixed with stimuli produced by the same encoders that were not the specific focus of this report, including the following: statements conveying an intermediate level of confidence (“close-to-confident”) with unique probability phrases (e.g., *Most likely*, *n* = 96); prosodically neutral versions of the statements (*n* = 96); versions of the confident, unconfident, and close-to-confident statements without a preceding phrase (*n* = 288, Jiang & Pell, 2015); and cross-spliced filler versions of the close-to-confident statements with a conflicting confident/unconfident preceding phrase (*n* = 96). Two experimental lists ensured that the congruent and incongruent phrase paired with each vocal expression did not occur in the same list (48 items × 2 Confidencex Congruency/ list) and each listener was assigned to a different stimulus sequence.

### Table 1

<table>
<thead>
<tr>
<th>Communicative function</th>
<th>Phrase-statement congruence</th>
<th>Example statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact</td>
<td><strong>Definitely</strong></td>
<td><em>she has access to the building</em>&lt;sub&gt;confident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>Maybe</strong></td>
<td><em>she has access to the building</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>There's a chance</strong></td>
<td><em>she has access to the building</em>&lt;sub&gt;confident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>I'm certain</strong></td>
<td><em>she has access to the building</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td>Intention</td>
<td><strong>I'm positive</strong></td>
<td><em>she'll sell her car soon</em>&lt;sub&gt;confident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>Perhaps</strong></td>
<td><em>she'll sell her car soon</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>It's possible</strong></td>
<td><em>she'll sell her car soon</em>&lt;sub&gt;confident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>For sure</strong></td>
<td><em>she'll sell her car soon</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td>Judgment</td>
<td><strong>For sure</strong></td>
<td><em>she doesn't like sweet foods</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>It's possible</strong></td>
<td><em>she doesn't like sweet foods</em>&lt;sub&gt;confident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>Maybe</strong></td>
<td><em>she doesn't like sweet foods</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td><strong>Definitely</strong></td>
<td><em>she doesn't like sweet foods</em>&lt;sub&gt;unconfident&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

Note. The intended confidence of the main statement has been marked as subscript in square brackets, and the initial lexical phrase is displayed in bold. Event-related potentials were always measured from the onset of the main statement and the onset was marked as ‘T’.

Communicative function refers to the purpose of the statement in discourse: stating a past or present fact about a person or object (fact); describing one’s intention or willingness to do something (intention); and making comments or giving opinions about someone or something (judgment). Facts included an action verb in its past or present tense, intentions were expressed in the future tense and carried the contracted form of “will,” and judgments always contained a linguistic form expressing commentary (e.g., “good sense of humor,” “well enough”).

### Procedure

Participants were comfortably seated in an electrically shielded, sound-attenuated booth. Each trial began with a fixation cross, followed by a varied onset delay of the auditory event (presented by in-ear headphones). Participants were asked to listen to each statement and judge the level of the speaker’s confidence on a 5-pt rating scale, presented on a computer monitor 2000 ms post offset of the vocal stimuli (the direction of the scale was manipulated and counterbalanced across listener sex). Recordings of statements were normalized to a mean intensity level of 75dB. Participants completed 20 practice items before the critical trials. The electroencephalograms (EEGs) were recorded continuously from 64 electrodes using the ActiCap System (Brain Products, Germany) with a sampling rate of 500 Hz and a bandpass filter from 0.016 Hz to 100 Hz. The vertical and horizontal electrooculograms were recorded from independent electrodes. The recordings were online referenced to FCz and rereferenced offline to the mean of the bilateral mastoids.

### ERP Analysis

The preprocessing of EEG data was performed in EEGLAB. The continuous EEGs were visually inspected and signals with excessive movement artifacts, alpha activity or amplifier saturation were manually excluded. The subsequent EEGs were 30-Hz low-pass filtered and then decomposed with an ICA algorithm to remove ocular artifacts. The resulting datasets were segmented into epochs from 200 ms before to 1600 ms after the onset of the main statement (i.e., the acoustic onset of the first word of the vocal confidence expression after the initial phrase, which was identical for congruent/incongruent conditions). The ERPs were baseline corrected according to the mean EEG activity in the 100-ms prestimulus interval. Segments in which peak-to-peak voltage exceeded 70 μV within a 100-ms sliding window were automatically rejected. All artifact-free vocal expressions entered statistical analysis (Mean *N* = 37.6, 38.1, 36.7, 37.2 for
the Confident–Confident, Unconfident–Confident, Unconfident–
Unconfident, and Confident–Unconfident conditions, ps > .1).

Mean ERP amplitudes per subject and per condition, time
locked to the onset of the main statement, were entered into linear
mixed effects models (LMEM) to evaluate the significance of
experimental manipulations (Baayen, Davidson, & Bates, 2008;
Newman et al., 2012; Wilson et al., 2012) and their interaction
with participant variables such as sex and trait empathy (Jiang &
Pell, 2015). The LMEM can model fixed and random effects
simultaneously while maximizing the use of the random effect
structure by including both the random intercept and the random
slopes in the model (Barr, 2013; Barr, Levy, Scheepers, & Tily,
2013). Both continuous and discrete variables can be modeled
(Jiang & Pell, 2016; Payne, Lee, Federmeier, 2015). The LMEM
extends the repeated measures models in general linear model by
allowing unequal number of repetitions, and can explicitly model
the variance-covariance structure of the data, allowing for viola-
tions of sphericity and homogeneity of error variance (Snijders &
Bosker, 2011). The ERP signals for one subject were averaged
before entering the model because of the low signal-to-noise ratio
in the single-trial EEG (Newman et al., 2012; Payne, Lee, Feder-
meier, 2015) and that the current focus of interest was not to
examine the trial-level effect on an experimental or an individual-
level factor (Newman et al., 2012; Wilson et al., 2012). We built
LMEMs on the mean peak amplitude of N1 (70–160 ms) and P2
(180–250 ms) given the effect of verbal context on early ERP
responses in our previous report (Jiang & Pell, 2016). Based on
visual inspection of the grand average results (see Figure 1) and a
50-ms small-window analysis (Jiang & Pell, 2015), three addi-
tional time windows were selected to model mean amplitude
differences for the N400-like effect (400–550 ms), late positivity
(550–800 ms), and the sustained positivity (900–1550 ms). Anal-
yses focused on differences in the mean amplitude of the sustained
positivity (900–1550 ms) for the congruent/incongruent stimuli
with main statements expressed in a confident voice, and on the
N400-like effect (400–550 ms) and late positivity (550–800 ms)
for corresponding statements produced in an unconfident voice.
The LMEM included vocally expressed confidence level (confi-
dent vs. unconfident, reference = confident), phrase-statement
congruency (congruent vs. incongruent, reference = congruent),
speaker sex of voice (female vs. male, reference = female), and
topographic factors as fixed effects. Listener sex or interpersonal
sensitivity (IRI total) score was treated as an additional fixed factor
and included in separate models.

For all models, topographic factors included hemisphere and
region as fixed effects. Nine regions of interest (ROI) were
formed, each represented by 6 to 8 electrodes treated as repeated
measures: left anterior (AF3, FP1, F7, F5, F3, FT7, FC5, FC3), left
central (T7, C5, C3, TP7, CP5, CP3), left posterior (P7, P5, P3,
PO9, PO7, PO3), medial anterior (AF4, FP2, F1, FZ, F2, FC1,
FCZ, FC2), medial central (C1, CZ, C2, CP1, CPZ, CP2), medial
posterior (P1, PZ, P2, O1, POZ, O2), right anterior (F4, F6, F8,
FC4, FC6, FT8), right central (T8, C4, C6, TP8, CP4, CP6), and
right posterior (P4, P6, P8, PO4, PO6, PO10). Participants were
included as random intercept and random slopes to evaluate indi-
vidual adjustments in the magnitude of ERP responses as a func-
tion of fixed factors (Newman et al., 2012). To reduce the Type I
error, the full model was evaluated taking all fixed effects, by-
participant random intercepts, and random slopes for all within-
participant fixed factors and for all possible combinations of
higher-order interaction of these factors (Barr, 2013; Barr et al.,
2013). The variance explained by each fixed factor in the model

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Grand average waveforms showing the effect of an incongruent verbal phrase on vocal statements expressing confidence (A) and lack of confidence (B), each on nine representative electrodes time-locked to the onset of the main statement. See the online article for the color version of this figure.}
\end{figure}
was examined by way of $F$ tests for main effects and interactions, and $t$ tests for specific contrasts (reporting only lower-bound probability values, which was considered more conservative, for expository purposes). Given that differential ERP effects were shown for the congruency effect of the confident versus unconfident-intoned utterances, post hoc analysis was performed only when a significant interaction has been shown between the confidence level and phrase-statement congruency. All analyses were performed in R-studio (Version 3.1.0, http://cran.r-project.org) with lme4, lmerConvenienceFunctions, and lmerTest packages.

At a final stage, sLORETA was performed on the scalp-recorded electrical activity to localize neural responses underlying ERP differences between incongruent and congruent vocal confidence expressions (Pascual-Marqui, 2002; Rigoulot et al., 2014) with 5 mm spatial resolution on a dense grid of 6239 voxels. The sLORETA is a functional imaging method that models the cortex as a collection of volume elements (voxels) in digitized MNI coordinates corrected to the Talairach coordinates (Pascual-Marqui, 2002). Here, the average waveforms for each condition and participant were converted for all 100 time-samples preonset and 800 time-samples postonset of the vocal expression. Paired $t$ tests were then computed between conditions for all time-samples (2 ms per sample) of the epoch using $t$-statistical nonparametric mapping (SnPM). The average reference was used and 5000 random permutations (i.e., bootstrapping) were performed. Levels of significance were corrected for multiple comparisons and false positives. The tests were performed on male and on female listeners separately while speaker sex of voice was collapsed. For each comparison, the peak sample which showed the highest significance was detected and the 5–10 time samples before and after the peak ($p < .05$, two-tailed) was selected for further analysis. The $t$-statistics generally replicated the LMEM performed on more generous time window for the ERP effects. sLORETA values for the selected time samples were computed for each condition and participant, respectively, followed by the statistical nonparametric mapping employing a log of $F$ ratio for average statistics for each analysis. The result of SnPM was displayed as $t$ values for each voxel with multiple comparisons ($p < .05$). We report the maximum significant differences between conditions at respective MNI coordinates and Brodmann areas (BA).

**Results**

**Effects of Mixed Messages on Perceptual Ratings of Speaker Confidence**

The LMEM included fixed factors of vocal Confidence level (2), phrase-statement Congruency (2), Speaker Sex of Voice (2), and Listener Sex (2), by-participant and by-item random intercepts, by-participant random slopes for all within-participant fixed factors and for all the higher-order interactions of these factors, as well as by-item random slopes for Listener Sex. Overall, statements produced in a confident voice were rated as more confident than statements in an unconfident voice, regardless of the congruency of the initial phrase. There was an interaction of Confidence level, Congruency, and Listener Sex, $F(1, 4983) = 20.16, p < .0001$. For both female and male listeners, statements produced in a confident tone were perceived as less confident following an incongruent versus congruent phrase, whereas unconfident statements were rated as more confident following an incongruent (confident) phrase. The congruency effects were larger in female than in male listeners, especially for the confident statement (Congruence × Listener Sex: $F(1, 2490) = 32.78, p < .0001$, for confident; $F(1, 2484) = 2.44, p = .05$, for unconfident; Table 2). These findings imply that the perceived confidence of the speaker based on their vocal cues in the main statement cannot be overridden by conflicting information in its preceding phrase; however, these phrases modulate impressions of a speaker’s feeling of knowing by attenuating the impact of confident phrases (Confident–Confident = 4.48, Unconfident–Confident = 3.17) and amplifying ratings of unconfident phrases (Unconfident–Unconfident = 2.05, Confident–Unconfident = 2.59). There was no interaction of Confidence Level, Congruency and Speaker Sex nor of Confidence Level, Congruency, Sex of Voice, and Listener Sex, $Fs < 1$; this suggests that the effect of incongruent phrases, and the larger congruency effects on female listeners, did not depend on the sex of the speaker’s voice. A subsequent model including Confidence Level, Congruency, and Interpersonal sensitivity did not uncover a relationship between individual IRI scores and the perceptual ratings.

**Effects of Mixed Messages on ERP Responses to the Vocal Expression**

Previous ERP data have established that confident and unconfident vocal expressions are differentiated in both early (N1, P2) and late (delayed positivity) processing time windows (Jiang & Pell, 2015). Of interest here, distinct effects of phrase congruency were also witnessed for responses to confident versus unconfident vocal expressions, as well as listener sex differences in how feeling of knowing was registered in mismatching conditions (see Figure 1; Table 3).

Focusing on effects informed by the congruency manipulation, linear mixed effects models on N1 and P2 did not reveal significant main effects or interactions of Congruency and vocal Confidence level on these components. A model run on the 300- to 450-ms time window produced a significant interaction of Confidence level, Congruency, Region, and Listener Sex, $F(2, 11706) = 5.00, p = .01$. Models to break down the interaction revealed a significant effect of phrase congruence when processing statements were produced in an unconfident tone, distributed in the central region for male listeners, $F(1, 768) = 8.02, p = .004$, and in the anterior and central regions for female listeners, $F(1, 1128) = 24.08, p < .0001; F(1, 768) = 6.10, p = .01$. When compared with utterances that consistently marked a speaker’s lack of confidence, hearing a confident phrase followed by a doubtful statement elicited an early, anteriorly maximized negativity (N400-like effect) in females $(-1.52 \mu V$ for anterior; $-1.04 \mu V$ for central; $-52 \mu V$ for posterior), whereas an early starting centrally distributed positivity was recorded in males $(60 \mu V$ for incongruent vs. congruent; Figure 2).

The interaction of Confidence level, Congruency, Listener Sex, and Speaker Sex of voice was also significant, $F(1, 11706) = 6.11, p = .01$, suggesting that the early effects of an incongruent phrase on the unconfident statement which differed by listener sex also depended on the sex of the speaker’s voice; more profound effects
were associated with male versus female voices (male vs. female: 2.30/V vs. .74/V for the negativity in the anterior region, 1.55/V vs. .53/V for the negativity in the central region; 1.12/V vs. .08/V for the central positivity). The interaction of Confidence level, Congruence, Region, and Listener Sex was also significant in the 450- to 800-ms time interval, F(2, 11706) = 3.22, p = .02. For male participants only, phrase-statement congruency continued to influence processing of unconfident vocal statements in central (1.20/V) and posterior (1.35/V) regions, with a more positive-going response in the incongruent condition, suggesting that this response began in the 300- to 450-ms window for male listeners (Figure 2A and 2B).

Table 2
Mean Speaker Confidence Rating in the Electroencephalographic Study According to the Vocal Confidence Level of the Main Statement, the Congruency of the Lexical Phrase and the Main Statement, and the Sex of the Speaker and Listener

<table>
<thead>
<tr>
<th>Level of confidence in main statement</th>
<th>Congruency between lexical phrase and main statement</th>
<th>Sex of voice</th>
<th>Perceptual rating (1–5)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Total</td>
</tr>
<tr>
<td>Confident</td>
<td>Congruent&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Female</td>
<td>4.78 (.26)</td>
<td>4.43 (.49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>4.61 (.32)</td>
<td>4.12 (.65)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>4.69 (.29)</td>
<td>4.27 (.55)</td>
</tr>
<tr>
<td>Incongruent</td>
<td></td>
<td>Female</td>
<td>3.30 (.48)</td>
<td>3.26 (.46)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>3.06 (.53)</td>
<td>3.06 (.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>3.18 (.47)</td>
<td>3.16 (.49)</td>
</tr>
<tr>
<td>Unconfident</td>
<td>Congruent&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Female</td>
<td>1.93 (.34)</td>
<td>2.30 (.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>1.86 (.25)</td>
<td>2.12 (.48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>1.89 (.26)</td>
<td>2.21 (.57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>2.49 (.62)</td>
<td>2.77 (.63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
<td>2.46 (.57)</td>
<td>2.63 (.53)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>2.47 (.58)</td>
<td>2.7 (.55)</td>
</tr>
</tbody>
</table>

Note. 1 = not at all confident, 5 = very confident. Standard deviation in parentheses.
<sup>a</sup> Data in the congruent condition, averaged across male and female speakers, were reported in Jiang & Pell (2016).  
<sup>b</sup> Mean congruency effect was calculated by subtracting the speaker rating in congruent statement from the incongruent statement per listener before averaging.

Table 3
Overview of the ERP Results of Phrase-Statement Congruency in the Confident and Unconfident Statement per Listener Sex in All Types of Analysis

<table>
<thead>
<tr>
<th>Confidence in the main statement</th>
<th>Listener sex</th>
<th>ERP effects (incongruent vs. congruent)</th>
<th>Effect of IRI on ERP effects</th>
<th>Mediation of ERPs on sex difference</th>
<th>Source localization of ERP effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confident</td>
<td>Female</td>
<td>Delayed, sustained positivity (900–1550 ms)</td>
<td>The lower IRI total score, the larger positivity effect in the posterior region</td>
<td>Female listeners revealed a larger sustained positivity, which in turn resulted in a larger reduction of confidence rating in the statement preceded by unconfident phrase</td>
<td>bilateral PoCG</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Unconfident</td>
<td>Female</td>
<td>Anteriorly-maximized negativity (300–450 ms)</td>
<td>/</td>
<td>/</td>
<td>left pre-SMA</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>Centro-posterior positivity (300–800 ms)</td>
<td>/</td>
<td>/</td>
<td>left MFG, right anterior insula, left STG</td>
</tr>
</tbody>
</table>

Note. ERP = event-related potential; IRI = interpersonal reactivity index; PoCG = bilateral postcentral gyrus; SMA = supplementary motor area; MFG = middle frontal gyrus; STG = superior temporal gyrus.
action of Confidence level, Congruence, Listener Sex and Speaker Sex of voice was significant, $F(1, 11706) = 12.16, p = .001$, suggesting that the positivity specific to the male listeners was larger in response to male versus female voices ($1.99 \mu V$ vs. $.41 \mu V$ for the central and $1.76 \mu V$ vs. $.93 \mu V$ for the posterior region). In the 900- to 1550-ms analysis window, there was a significant interaction of Confidence level, Congruence, and Listener Sex, $F(1, 11706) = 66.49, p < .0001$. For statements produced in a confident voice, a conflicting phrase (*Maybe*) elicited a delayed, sustained positivity when compared with corresponding statements that were congruent. However, this response was only witnessed in female listeners (Figure 2C) irre-
spective of the sex of the speaker’s voice (Confidence level × Congruence × Listener Sex × Sex of Voice, $F(1, 11706) = 2.60$, $p = .11$, 1.50 μV vs. 1.48 μV).

To evaluate the relationship between a listener’s interpersonal sensitivity and ERP effects for resolving conflicts about a speaker’s feeling of knowing, LMEMs were built in each time window with significant congruence effects, for confident and unconfident vocal expressions separately, taking Congruence, IRI total score, and topographic variables as fixed effects. The positivity effect (900–1550 ms) elicited by confident statements following an unconfident hedge phrase was significantly larger in listeners with lower IRI scores (Congruence × IRI: $F(1, 6354) = 7.16, p = .008$, see Figure 4). An additional model including Listener Sex as a fixed factor indicated that effects of IRI on this positivity were more pronounced in anterior and central region for the male listener and more pronounced in posterior region for the female listener (Congruence × IRI × Region × Listener Sex: $F(1, 6318) = 5.97, p = .003$, and see Figure 3 for individual differences in the delayed positivity by region and listener sex). No modulation of IRI was found on the congruence effects when processing unconfident statements.

### Mediation Analysis

To this point, our analyses reveal that (a) female listeners exhibited larger behavioral adjustments when judging incongruent cues about a speaker’s feeling of knowing, and (b) female and male listeners varied in ERP responses while processing statements containing conflicting cues. To evaluate how the neural responses could mediate differences of listener sex in the perceptual adjustment to incongruent statements, we first calculated the difference in perceptual ratings between incongruent and congruent conditions for confident and unconfident statements, respectively, for each listener and then the difference in the neural responses between conditions per level of confidence per listener per time window$^2$. Then, we built two multiple mediation models, one for the confident and the other for the unconfident statements by deriving 95% bias-corrected confidence intervals from 5000 bootstrap estimates (Preacher & Hayes, 2004). The multiple mediation model estimates the global mediation effect containing conflicting cues. To evaluate how the neural activity (900–1550 ms) recorded in the posterior region for confident and unconfident statements, the difference was calculated based on the mean amplitudes in the 900- to 1550-ms window (delayed positivity); for unconfident statements, the difference was calculated based on the mean amplitudes in the 300- to 450-ms window (early negativity) and in the 900- to 1550-ms time window (delayed positivity).

Individual mediation effect was observed for the delayed positivity (900–1550 ms) recorded in the posterior region for confident statements. The global mediation with delayed positivity in all regions as a group was not significant. As illustrated in Figure 2, female listeners demonstrated a more positive-going delayed response to incongruent confident statements ($b = -1.89$, $t = -2.11, p = .04$), which in turn predicted a decrease in confidence ratings in that condition ($b = -0.14$, $t = -2.17, p = .04$). The effect of the direct path between listener sex and behavioral adjustment was reduced from being significant, $b = .40$, $t = 2.22$, $p = .03$, to becoming nonsignificant when the mediator was entered in the model, $b = .34$, $t = 1.49, p = .13$, suggesting that the delayed positivity in 900–1550 ms completely mediated the relationship between listener sex and behavioral adjustment (see Figure 4).

### Source Localization of ERP Effects

Analyses to localize our effects using sLORETA pointed to potentially different underlying neural mechanisms for resolving different forms of conflict about a speaker’s feeling of knowing (see Figure 5). When processing unconfident vocal expressions after a conflicting cue in the phrase, analysis of the early anterior-central negativity (410–450 ms) for female listeners revealed broadly increased activity in the left anterior supplementary motor area (pre-SMA, $[-5, 25, 60]$, Brodmann Area 8). For the same comparison in male listeners, the early positivity effect (426–450 ms) was associated with increased brain activity in the left middle frontal gyrus (MFG, $[-40, 20, 50]$, Brodmann Area 9), whereas the late positivity (744–784 ms) could be traced to the left MFG ($[-40, 20, 50]$, Brodmann Area 8), right anterior insula ($[40, 15, 5]$, Brodmann Area 48), and left superior temporal gyrus (STG, $[-35, -55, 25]$, Brodmann Area 39). When processing confident vocal expressions, analysis of the delayed positivity in female listeners corresponded to increased activity in the bilateral post-central gyrus (PoCG, $[20, -35, 70]$, Brodmann Area 4) following an incongruent versus congruent preceding phrase.

### Discussion

This study tested how inconsistent cues encountered in many social interactions modulate listener impressions of a speaker’s feeling of knowing when processing personal knowledge state-

---

2 For confident statements, the difference was calculated based on the mean amplitudes in the 900- to 1550-ms time window (delayed positivity); for unconfident statements, the difference was calculated based on the mean amplitudes in the 300- to 450-ms window (early negativity) and in the 550- to 800-ms window (LPC/P600).

3 The mediation analysis was conducted per region (anterior, central and posterior) based on the finding that (a) the congruency effects on the unconfident statement were region-specific in different listener sex (with male listener revealing central-posterior positivity and female listener revealing anteriorly maximized negativity), (b) the delayed positivity on the confident statement was modulated by IRI total score in different region for different listener sex, and (c) the female-specific delayed positivity effects previously reported (comparing neutral-intending and confidence-intending in Jiang & Pell, 2015 and comparing unconfident/close-to-confident and confident expression in Jiang & Pell, 2016) were region-sensitive (posterior in Jiang & Pell, 2015 and centro-posterior in Jiang & Pell, 2016).
ments that have no shared truth-value. We observed important distinctions in how the vocal cues of confident and unconfident statements were interpreted following an incongruent probability phrase (‘I’m sure, maybe’). Perceptually, incongruent phrases modulated ratings by reducing the perceived confidence of confident voices while amplifying the perceived confidence of doubtful voices, a behavioral pattern that was more pronounced in female than male listeners. At the neurocognitive level, responses to each “type” of mixed message were unique: we observed an anteriorly maximized negativity (N400-like effect, in female listeners) or a late positivity (LPC/P600, in male listeners) when processing unconfident vocal cues after a confident phrase, effects that were further modulated by the sex of the speaker’s voice. In contrast, a delayed positivity effect was observed only in female listeners when processing confident utterances after an unconfident phrase.

As the main utterance in the congruent/incongruent conditions was identical for each type of vocal confidence expression, these patterns cannot be explained by known acoustic differences for expressing low versus high feeling of knowing (e.g., changes in speech rate, pitch, etc., Jiang & Pell, 2015). Rather, they shed light on operations for forming an online mental representation of a speaker’s feeling of knowing in the face of disparate evidence registered in the utterance-initial probability phrase and vocal characteristics of the main statement, while illuminating how individual differences such as listener sex inform these operations. We begin by discussing the implications of encountering each type of “mixed message” about another’s feeling of knowing in turn.

Case 1: Confident Probability Phrase—Unconfident Voice

At the onset of the main statement, vocal cues that refer to the speaker’s feeling of knowing must be somehow integrated into the speaker representation built from verbal and vocal cues in the phrasal context. Often, when listeners are confronted with conflicting meanings in the verbal and vocal channels of a single input, or meanings must be accessed following contextual violations in speech, an increased N400 is reported (see Kutas & Federmeier, 2011 for a review). The N400 has been linked to increased difficulty in accessing a communicative meaning of vocal expressions that are facilitated by linguistic and/or speaker information in the context (Brouwer, Fitz, & Hoeks, 2012; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). For example, using a Stroop paradigm, a larger N400 was elicited by emotional words produced with an incongruent versus congruent prosody (e.g., “happy” spoken in an angry voice, Schirmer & Kotz, 2003).

Figure 3. The association between interpersonal reactivity index (IRI) total score and mean amplitude difference between incongruent and congruent statement on the confident voice in the 900- to 1550-ms time window for each of the region. Listener sex is color-shape coded. Level of fitness of the regression line (whether significant or not has been coded as solid or dashed line) to the actual data was included per region per listener sex. *p < .05, ***p < .001. See the online article for the color version of this figure.

Figure 4. Mediation model showing the delayed positivity effect to confident vocal statements in 900- to 1550-ms window mediating the relationship between listener sex and behavioral ratings following incongruent versus congruent phrases. *p < .05. The positivity effect was calculated as the mean of the event-related potential difference in the posterior region between the incongruent versus congruent condition in the confident expression. For both panels, solid lines are significant and dashed lines represent the nonsignificant direct path when the mediator is entered in the model.
Figure 5. Source localization images (sLORETA) showing statistical differences (Log of ratio of averages) between incongruent and congruent conditions in the early negativity, late positivity, and delayed, sustained positivity. See the online article for the color version of this figure.
Although the N400 is typically maximal at centro-posterior sites, an anterior N400 has been reported when comparing vocal expressions conveying different basic emotions, or between individuals with different musical or language expertise as they categorized different sound types (Elmer et al., 2014; Grossmann et al., 2013). This more anterior response could reflect semantic-level processes for evaluating vocal expressions (Kotz & Paulmann, 2007; Schirmer & Kotz, 2006; Schirmer et al., 2006).

As listeners explicitly attended to the speaker’s expressed confidence level, we observed an early anteriorly maximized negativity in the 300- to 450-ms time window post onset of the vocal expression, which emerged only when unconfident vocal cues violated expectancies built up by a confident phrase (and not vice versa) and only for female listeners. Given that confident and unconfident voices begin to be differentiated after 200 ms of acoustic input (differences in P2 amplitude, Jiang & Pell, 2015), this negativity could register the early detection of vocal cues in our main utterance that deviated from the semantic expectancy created by the initial probability phrase (Paulmann & Kotz, 2008). This deviation may have placed demands on accessing the speaker’s mental state with enlarged N400, extending previous evidence on contextual effects on the N400-like effect to how listeners access representations about a speaker’s feeling of knowing during online speech processing.

Interestingly, this effect had a source in the left pre-SMA, a region associated with increased demands for accessing the meaning or abstract rule of an object in semantic memory (e.g., Hart et al., 2013). As the source of this N400 effect was less typical, it could be treated as a left anterior negativity (LAN) typically observed when syntactic or predictive rules are not satisfied during speech and language comprehension (e.g., when anomalous speech input violates grammatical agreement independent of semantic content, Tanner & Van Hell, 2014). Other work has differentiated the LAN in response to linguistic rule violations from an anterior negativity that responds to mismatches in vocal regularity (e.g., patterned syllable stress, Schmidt-Kassow & Kotz, 2008). In our study, the unconfident voice in the main statement (with higher pitch, reduced intensity, and slower speech rate) mismatched the expected vocal pattern for signaling high confidence from the phrase, and as such, could render an anterior negativity that reflects the detection of these acoustic irregularities. Moreover, the overlap of this anterior effect with the typical N400 may drive the anterior-maximum of the negativity effect observed in the Confident-Unconfident condition for female listeners.

In male listeners, a confident phrase followed by a statement expressing doubt elicited a positive-going response in the same time interval, one that persisted as unconfident vocal cues continued to unfold (450–800 ms). Elsewhere, it has been shown that when German listeners attended to emotional features of cross-spliced utterances (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008), processing contextually incongruent vocal cues elicited an enhanced late positivity (LPC/P600 response); this occurred irrespective of whether the mismatch in emotional meaning was driven solely by vocal expression information or combined vocal-verbal cues in the preceding context. In our previous work on vocal confidence, differences in the LPC were observed in a condition when listeners heard voices that are ambiguous in signaling the speaker’s level of confidence (e.g., when hearing close-to-confident vs. very confident expressions, Jiang & Pell, 2015). Given that LPC amplitudes in that study correlated with the time needed to rate vocally expressed speaker confidence, it was argued that vocal meanings that are underspecified promote increased attentional analysis in the 450- to 800-ms time window with concomitant increases in the LPC (Jiang & Pell, 2015; see also Pell et al., 2015 for a related view).

Our behavioral measures underscore that ambiguity about the speaker’s feeling of knowing was introduced not solely by features of the voice, but by the need to integrate divergent properties of the vocal expression into a mental representation formed by the preceding verbal expression of probability. It seems that dealing with “mixed messages” of this nature is associated with a late positivity response, reflecting ongoing attempts of the listener to disambiguate or recover the underspecified meaning about the speaker’s confidence level, at least for male listeners. Hypothetically, male listeners could treat the ambiguous cue as the speaker’s unintentional mistake and make further attempts to extract relevant information—vocal cues from the main statement or combined vocal and verbal cues from the leading phrase—to achieve a fine-grained (re-)analysis of the speaker’s feeling of knowing. The late positivity could also be linked to a P600-like effect, reported in previous studies of speech processing in which listeners had to comparatively analyze verbal and vocal cues to arrive at intended speaker meanings (Regel, Gunter, & Friederici, 2011; Rigoulot et al., 2014). It has been said that the P600 indexes the relative demands of updating representations of what the speaker is communicating (Brouwer et al., 2012; Brouwer, Fitz, & Hoeks, 2012), possibly in a task-irrelevant manner (Van de Meerendonk, Indefrey, Chwilla, & Kolinsky, 2011). For example, P600 effects can emerge when linguistic input is inconsistent with a highly constrained discourse context (Kuperberg, 2007) or when an alternative nonliteral interpretation can be easily accessed (Jiang, Li, & Zhou, 2013a; Regel et al., 2011). In our paradigm, exposure to vocal patterns in the main statement that failed to confirm meanings established by the probability phrase could instigate a process to update representations of the speaker’s mental state (true feeling of knowing), promoting a P600-like response in certain participants.

Source analysis implied that the positivity effects in male participants emanated from brain networks that were different from those subserving the early negativity in female participants: the left middle frontal gyrus (MFG), responsible for attentional control (associated with the early positivity); and the left superior temporal gyrus (STG) and right insula, involved in vocal and socioemotional processing (associated with the late positivity). Left MFG involvement has been previously reported in tasks demanding conflict control (e.g., Kouneiher, Charron, & Koechlin, 2009) and inhibitory control at the “perceptual level” (Milham et al., 2001). Here, processing an unconfident voice after hearing a confident phrase may have triggered an early process for resolving perceptual conflict in the input to adjust for the perceived confidence of the speaker, a control process that was slightly different for male versus female listeners. The source of the late positivity was then traced to the left STG, a region involved when making evaluative judgments of vocal expressions embedded in meaningful semantic content (Frühholz & Grandjean, 2013), and the (right) insula, a structure tied to the ability to make socioemotional judgments and social inferences (Craig, 2009; Lamm & Singer, 2010; Rigoulot et al., 2014).
et al., 2014). These neural regions appear to be important for males to update mental representations about the speaker’s feeling of knowing when confronted with an unconfident voice after hearing a conflicting phrase in the context.

**Case 2: Unconfident Probability Phrase—Confident Voice**

Our ERP data furnish strong evidence that hearing an unconfident hedge (Maybe) followed by a confident vocal statement does not produce strong contradictions in how speaker confidence is initially assigned; rather, this condition was marked by a qualitatively distinct sustained positivity from 900 to 1550 ms postonset of the vocal expression. The delayed positivity effect, which was unique to female listeners in our sample, may be partly attributed to conditions that promoted pragmatic inferences in these individuals. For instance, a similarly temporally delayed, sustained positivity has been observed when use of the polite form of the second-person pronoun conflicts with the social status of the speaker, promoting an ironic interpretation of the pronoun (Jiang et al., 2013a; Regel et al., 2011), or after words that violate implied characteristics of a protagonist described in a context (Van Overwalle et al., 2009). These patterns suggest that the delayed positivity is linked to a late inferential process for deriving contextually appropriate nonliteral interpretations (Van der Cruyssen et al., 2009).

One reason that the transition from an unconfident phrase to a confident statement (Maybe | he’s a good leader [confident tone]) does not hamper online processes for integrating and/or updating the mental representation of the speakers’ feeling of knowing is that early registration of the confident vocal statement easily overrides expectancies created by the preceding phrase (see discussion in next section). Nonetheless, this form of deviance does promote social inferences about the speaker. Our data suggest that females disproportionately engaged in late processes for assigning nonliteral or implied social meanings to these utterances (e.g., that contrasting cues about speaker confidence were used intentionally to express irony, or might reflect the speaker’s initial indecision or hesitation to express the main statement). Increases in the sustained positivity effect might therefore reflect the relative effort associated with this inferential process instigated by females in the Unconfident-Confident condition (Frank, Baron-Cohen, & Ganzel, 2015). Consistent with this argument, the delayed positivity effect appeared to be generated by mechanisms in the left postcentral gyrus, which is known to be activated during tasks for recognizing the mismatch between what is said and what is fact, and has been associated with lie perception (Wu, Loke, Xu, & Lee, 2011). We also found that individuals with reduced interpersonal sensitivity (IRI total score) displayed a larger sustained positivity effect in the same condition, meaning that it was more effortful for them to make pragmatic inferences (Li, Jiang, Yu, & Zhou, 2014). Interestingly, this latter association was evident in different scalp regions depending on whether individual differences were assessed in female versus male listeners, emphasizing that both listener sex and an individual’s interpersonal sensitivity influence pragmatic inferences about the feeling of another’s knowing (Jiang & Pell, 2016).

### Effects of Task, Speaker, and Listener Characteristics

Several factors could explain why each type of mixed message had unique effects on neural responses to the statements of confidence as described above. The fact that we observed N400- and P600-like context effects in only one of our two mismatch conditions—when a confident phrase preceded an unconfident statement—may be partly explained by the relevance of particular information to the task. Listeners were instructed to judge “how confident the speaker is,” emphasizing the positive end of the speaker’s feeling of knowing, which likely rendered confident cues more salient than unconfident ones in our paradigm. In previous work, confident voices elicited enhanced P200 amplitudes over unconfident voices when listeners also explicitly attended to the speaker’s confidence level (Jiang & Pell, 2015). This argues that speaker cues demonstrating a high feeling of knowing lead to preferential deployment of attention at early stages for (rough) semantic encoding of the stimulus (Jiang & Pell, 2015; Pell et al., 2015). Interestingly, we found that ERP responses to an unconfident voice following a confident phrase were amplified when listening to male versus female speaking voices, suggesting that this conflict engaged even more attention when communicated by male speakers. When producing the confident phrase, male voices displayed a lower pitch and greater intensity than female voices, some of the key perceptual determinants of a speaker’s feeling of knowing (Jiang & Pell, 2015). Assuming that confident cues in our task are quickly marked as task-relevant and attract preferential attention, it can be said that the demands of integrating or updating representations about a speaker’s feeling of knowing are more pronounced after a confident phrase is disconfirmed, leading to a more elaborate/effortful process in this condition, especially for male speaking voices.

According to the multistage model of vocal expression processing, the consolidation of multiple channels of information at later stages is achieved by operations that are sensitive to individual differences (e.g., as reflected by the N400, LPC, etc. Paulmann & Kotz, 2008; Schirmer & Kotz, 2006). This raises the question of how individual differences influence processing of expressed confidence in speech and which operations are affected. As emphasized above, a major outcome of this work is that neurophysiological responses to utterances expressing feeling of knowing often vary for male and female listeners, and that female listeners display a larger perceptual adjustment to contextual cues than males (assigning lower ratings to confident vocal expressions and higher ratings to unconfident voices after a conflicting verbal phrase). Although there is ample evidence of listener sex asymmetries in the decoding of a speaker’s vocal emotion state (Besson, Magne, & Schön, 2002; Schirmer, Chen, Ching, Tan, & Hong, 2013; Schirmer, Kotz, & Friederici, 2005; Schirmer et al., 2006; Schirmer, Kotz, & Friederici, 2005), our data underscore that sex differences in the decoder further characterize how vocal cues are used to attribute mental states to other speakers (Frank et al., 2015; Rothermich & Pell, 2015).

Of note, when unconfident statements followed a confident phrase, the grand average across participants (Figure 1B) highlights a biphasic pattern composed of an early negativity–late positivity; however, closer analysis indicated that these late neural processing stages were modulated by biological sex, with the N400-like effect predominantly observed in female listeners and
the LPC/P600 effect in males. Similar heterogeneous ERP effects in the face of certain types of linguistic conflict have been reported elsewhere (e.g., to the violation of grammatical rules, Tanner & Van Hell, 2014, or to the mismatch of semantic constraints, Kos, Van den Brink, & Hagoort, 2012). In these studies, a global biphasic (anterior) negativity–positivity could be explained by the “mix” of brains showing ERPs at different points of a continuum between a monophasic (anterior) negativity and a monophasic late positivity: listeners who displayed a larger negativity also exhibited a smaller positivity, suggesting that a common neural mechanism manifested in different forms on the scalp.

In our study, there was no apparent association between the congruency effect on the anterior negativity in 300–450 ms and the centro-posterior positivity in 450–800 ms in either male or female listeners. The fact that female participants displayed an N400-like effect in the confident-unconfident condition, whereas male participants displayed an LPC/P600, has precedence when male and female listeners were presented utterances with speaker identity-content mismatches (Van den Brink et al., 2012). These effects could exemplify sex-specific processing strategies at the stage of coordinating multiple sources of vocal and verbal information, extending current models of vocal processing (Schirmer & Kotz, 2006). For example, females may attempt to integrate vocal information with the phrasal context with varying levels of difficulty, promoting differences in the N400, whereas males engage processes to resolve conflicting messages about speaker meaning (as reflected in the LPC/P600). Further research will be needed to verify these claims.

The idea that females selectively (or more robustly) engage in processes for inferring a speaker’s communicative intentions, as implied by the delayed positivity effect in our unconfident-confident condition, also merits further commentary. In past studies that presented a mismatch of the emotional meaning encoded by verbal and vocal cues, a female bias (or ‘advantage’) appeared primarily when the task discouraged listeners from attending to the congruence between channels (e.g., using an emotional Stroop or oddball paradigm, Ishii, Kobayashi, & Kitayama, 2010; Schirmer & Kotz, 2003; Schirmer et al., 2005, 2006). One account attributes these differences to higher estrogen levels in females that increase sensitivity to socially relevant information, such as vocal emotion expressions (Schirmer et al., 2013; Schirmer et al., 2008). Along similar lines, when a mismatch in cues expressing a speaker’s feeling of knowing was implicitly detected in our task, female listeners displayed a unique pattern of neural responses that could reflect heightened sensitivity to the potential social relevance of the mismatch. According to a mediation analysis, shifts in how female listeners rated speaker confidence when faced with incongruent information were directly predicted by the strength of the delayed positivity effect (i.e., inferential process). This finding argues that the delayed response allowed females to recalibrate their initial impression of the mixed expression using what they inferred about the speaker (e.g., the intention to express some level of hesitance). Mediation did not occur either when speakers expressed confidence consistently between channels (where differences in the N1, associated with acoustic differentiation of vocal signals, predicted females’ behavioral judgments). These data underscore that female listeners are more likely to use context to engage a late, sustaining mechanism that taps into pragmatic knowledge to conceive of the speaker’s feeling of knowing, implying that they are more sensitive to the social relevance of vocal information (Hall, 1978; Schirmer et al., 2013).

Mediation of confidence ratings did not occur for the early anterior negativity effect in the confident-unconfident condition in female listeners, despite sex differences in speaker ratings and in the neural response. This implies that the early negativity, rather than indexing the outcome of perceptual evaluation processes, may simply reflect difficulty dealing with conflicting messages. Elsewhere, the magnitude of the anterior negativity (300–450 ms) toward phrases that violate pragmatic constraints of how likely an event is (e.g., even that rich person can afford an expensive house) was negatively associated with the difficulty of the response, with slower responses eliciting a larger negativity (Jiang, Li, Zhou, 2013b). Mediation was not shown for the LPC/P600 effect in male listeners either, even though the LPC/P600 in the posterior region was positively associated with speaker ratings when averaged across participants (mediation was not guaranteed for this analysis because female listeners often demonstrated an opposite pattern to males). Thus, neural responses that are likely to mediate behavioral adjustments in perceived speaker confidence in the confident–unconfident condition are less clear for our data.

Limitations and Future Directions

Like most studies in this literature (e.g., Frühholz & Grandjean, 2013; Pell, Paulmann, Dara, Allasseri, & Kotz, 2009; Scherer, 2013), the interpretation of our data may be restricted by the fact that vocal expressions were portrayed by speakers trained in acting or public speaking, allowing the possibility that our confidence stimuli sometimes reflected prototypical or slightly exaggerated realizations of speaker confidence. Simulated expressions were necessary to tightly control our stimuli for the social function of the main utterance, recording quality, and to mitigate effects of other mental and emotional states of the speaker on the expression of feeling of knowing; nonetheless, our acoustic data provide similar evidence of how vocal confidence is expressed as reported following other types of induction procedures (Scherer et al., 1973; Swerts & Krahmer, 2005), suggesting that our results can be generalized to more naturalistic forms of communication.

The explicit demand of evaluating a speaker’s feeling of knowing throughout our task could also limit the generalizability of these findings to everyday social situations in which expressions of confidence are irrelevant to the topic but unconsciously attended. For instance, we implicitly process confident vocal cues when others are trying to persuade us (e.g., in marketing behavior, Smith, De Houwer, & Nosek, 2013) and detect vocal cues of uncertainty when a speaker attempts to mitigate a face-threatening situation to the listener (e.g., in doctor-patient communication, Bonnefon & Villejoubert, 2006). Research on emotion shows that effects of integrating and continuously evaluating verbal and vocal information in conflicting statements are greater when vocal cues to the intended speaker meaning are task-relevant, yielding more extensive ERP effects when vocal cues are explicitly attended to (producing a larger N400 effect in the explicit than the implicit condition but a late positivity only in the explicit condition, see Kotz & Paulmann, 2007 for details). One can therefore speculate that if listeners implicitly attended to the speaker’s feeling of knowing in our task, we would see a reduction of the negativity in the confident-unconfident condition, and/or the absence of the
delayed positivity (900–1550 ms) in female listeners in the unconfident-confident condition. This work represents an important direction for future projects.

In natural discourse, expressions that reveal a person’s confidence level do shift in real time with the speaker’s evolving mental state, or display obvious inconsistencies as one of the social goals of the speaker. By focusing on “mixed messages” about what a speaker knows, our data exemplify that listeners are exquisitely sensitive to multiple cues that encode a feeling of another’s knowing in the contextual phrase and the main statement during online speech processing, and rapidly *integrate* these cues to form a mental representation of the utterance and of the speaker. However, each context places unique demands on the neurocognitive system (and on male and female listeners).

Recent neurocognitive models of language processing posit the existence of multiple processing mechanisms when building a mental representation, with one relying on heuristics and priming between meaningful inputs regardless of how these inputs are organized, and the other relying on a combinatorial and integrative process where all constraints are considered to form a coherent representation (e.g., Brouwer et al., 2012; Kuperberg, 2007). Different patterns of ERPs are elicited when these processing streams conflict, or either of the streams fails, yielding a monophasic N400, monophasic positivity, or biphasic N400-late positivity response. The functional interpretation of these ERP effects is still a matter of significant debate; one view regards the N400 as difficulty in semantic integration/unification (Friederici, 2011), whereas the P600 reflects a second-pass or continued analysis due to initial processing failure (Kuperberg, 2007). This view assumes that the difficulty underlying N400 and P600 is interdependent. A second view interprets the N400 as difficulty in semantic access/retrieval (Kutas & Federmeier, 2011; Lau, Phillips, Pöppel, 2008) or in implicit memory formation driven by implicit prediction error (Laszlo & Pfaut, 2012; Rabovsky & McRae, 2014), whereas the P600 reflects an integrative process that aims to build a coherent representation taking into account all possible cues in the context (Brouwer et al., 2012). According to this view, the occurrence of N400 does not depend on the occurrence of P600. Our findings align more closely with the latter view by demonstrating a case of conflict (confident phrase—unconfident statement) that evokes negativity and positivity responses in different listener groups, effects that were uncorrelated in their magnitude across listeners. Moreover, we highlight that vocal characteristics referring to speaker identity (e.g., speaker sex) matter in these judgments by influencing the level of demands when processing incongruity between multiple sources of speaker information, revealing an immediate use of vocal identity information in integrating and updating mental representations (Foucart et al., 2015; Hanulková & Carreiras, 2015; Van Berkum et al., 2008; Van den Brink et al., 2012).

Additional research is needed to validate the patterns we observed and to extend our ideas to new contexts that progressively mirror the complexity of real social interactions. For example, given our focus on the processing of personal knowledge statements, it will be useful to explore how representations about a speaker’s feeling of knowing derived from online speech are integrated with *shared knowledge* about the truth value of the statement and to examine what types of pragmatic inferences are generated. Alternatively, the simultaneous impact of vocal features marking a speaker’s social identity (e.g., linguistic or geographic background, race, etc.) on vocal confidence processing and the perceived *believability* of the speaker would be highly instructive.

**References**


Received October 9, 2015
Revision received March 15, 2016
Accepted March 18, 2016