Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

On how the brain decodes vocal cues about speaker confidence

Xiaoming Jiang [*] *and Marc D. Pell*

School of Communication Sciences and Disorders and Center for Research on Brain, Language and Music, McGill University, Montréal, Canada

ARTICLE INFO

Article history:
Received 15 August 2014
Reviewed 27 November 2014
Revised 9 January 2015
Accepted 6 February 2015
Action editor Roberto Cubelli
Published online 21 February 2015

Keywords:
Feeling of knowing
Prosody
Expressed confidence
P200
Late positivity

ABSTRACT

In speech communication, listeners must accurately decode vocal cues that refer to the speaker's mental state, such as their confidence or 'feeling of knowing'. However, the time course and neural mechanisms associated with online inferences about speaker confidence are unclear. Here, we used event-related potentials (ERPs) to examine the temporal neural dynamics underlying a listener's ability to infer speaker confidence from vocal cues during speech processing. We recorded listeners' real-time brain responses while they evaluated statements wherein the speaker's tone of voice conveyed one of three levels of confidence (confident, close-to-confident, unconfident) or were spoken in a neutral manner. Neural responses time-locked to event onset show that the perceived level of speaker confidence could be differentiated at distinct time points during speech processing: *unconfident* expressions elicited a weaker P2 than all other expressions of confidence (or neutral-intending utterances), whereas *close-to-confident* expressions elicited a reduced negative response in the 330–500 msec and 550–740 msec time window. Neutral-intending expressions, which were also perceived as relatively confident, elicited a more delayed, larger sustained positivity than all other expressions in the 980–1270 msec window for this task. These findings provide the first piece of evidence of how quickly the brain responds to vocal cues signifying the extent of a speaker's confidence during online speech comprehension; first, a rough dissociation between unconfident and confident voices occurs as early as 200 msec after speech onset. At a later stage, further differentiation of the exact level of speaker confidence (i.e., close-to-confident, very confident) is evaluated via an inferential system to determine the speaker's meaning under current task settings. These findings extend three-stage models of how vocal emotion cues are processed in speech comprehension (e.g., Schirmer & Kotz, 2006) by revealing how a speaker's *mental state* (i.e., feeling of knowing) is simultaneously inferred from vocal expressions.

* *Corresponding author*. School of Communication Sciences and Disorders, 2001 McGill College, 8th Floor, Montréal, Canada.
E-mail addresses: xmjiang1983@gmail.com (X. Jiang), marc.pell@mcgill.ca (M.D. Pell).

http://dx.doi.org/10.1016/j.cortex.2015.02.002
0010-9452/

1. Introduction

Expressed confidence is a form of communication that records one's certainty or doubt about a fact or judgment; this serves as a device where a speaker's own evaluation of the evidence for his or her statement is revealed to the listener in human communication (Caffi & Janney, 1994). The act of expressing confidence is associated with certain metacognitive states (e.g., "feeling of knowing", Brennan & Williams, 1995; Swerts & Krahmer, 2005) as well as personal characteristics (such as credibility or persuasiveness, London, McSeveney, & Tropper, 1971; Maslow, Yoselson, London, 1971). For example, in the context of persuasion, expressed confidence can successfully lead to attitude change that affects decision-making (London, Meldman, Lanckton, 1970; London et al., 1971). Accurate decoding of confidence information from speech is also known to improve the efficiency of children's problem solving (Zimmerman & Ringle, 1981), word learning (Sabbagh & Baldwin, 2001), and interpersonal communication more broadly (Monetta, Cheang, & Pell, 2008; Pell, 2007). It is therefore of great relevance to understand the neurocognitive mechanisms underlying expressed confidence as one of the principle ways that speakers reveal their mental states, or 'feeling of knowing', to listeners, in addition to strategically using expressed confidence as a device to attain particular social-pragmatic goals (i.e., persuasion, to establish credibility).

Expressed confidence in speech communication has been studied in both verbal and non-verbal domains. For example, linguistic expressions such as, *I am positive that…*, have been associated with a speaker's knowledge level as well as professional task orientation (Scherer, London, Wolf, 1973). Speech-accompanying gestures, such as frequent self-touches or self-adapters, are associated with lack of social confidence or anxiety (Kimble & Seidel, 1991), while frequent eye contact and reduced blinking suggest convincing speech (DePaulo, Stone, & Lassiter, 1985; Hemsley, Doob, 1978; Swerts & Krahmer, 2005). When vocal expressions are specifically examined, it appears that a *confident* tone of voice is accompanied by increased speech rate and loudness, short and infrequent pauses, and sometimes, higher pitch (Barr, 2003; Kimble & Seidel, 1991; but see Apple, Streeter, Krauss, 1979). Intonational features of an utterance have also been associated with differences in perceived confidence, with pitch contours that fall at the end of an utterance signaling confidence and terminal rising contours signaling a lack of confidence (Bolinger, 1978). Confidence expressions may also be linked to other emotional attributes such as enthusiasm, activeness, relaxation, and perceived competence (Scherer et al., 1973).

Behavioral studies of how accurately vocal confidence is decoded from speech have investigated how speakers' own vocal expressions relate to their subjective 'feeling of knowing' when responding to trivia questions (Brennan & Williams, 1995; Kimble & Seidel, 1991; Smith and Clark, 1993), or how acted portrayals of confidence in spoken sentences (or pseudo-sentences) are judged by an independent group of listeners (Jiang & Pell, 2014; Monetta et al., 2008; Pell, 2007; Scherer et al., 1973). Typically, these studies include expressions of two opposing levels of intended speaker confidence (confident, unconfident) and sometimes a third intermediate level (close-to-confident, Monetta et al., 2008; Pell, 2007). Jiang and Pell (2014) also manipulated the communicative function of utterances bearing vocal confidence cues, which were statements of fact, intentions, or judgments. These studies show that *intended* levels of vocal confidence expressed by actors can be successfully differentiated by listeners to render graded judgments about speaker confidence (Jiang & Pell, 2014; Scherer et al., 1973). Interestingly, patients with focal right hemisphere lesions (Pell, 2007) or idiopathic Parkinson's disease (Monetta et al., 2008) show reduced sensitivity to different levels of confidence encoded by vocal cues when compared to healthy age-matched listeners; this suggests that vocal expressions of confidence, like vocal emotions, rely on distributed cortical-subcortical networks that attribute meaning to acoustic cue sequences in speech (Pell, 2006; Pell & Leonard, 2003; Schirmer & Kotz, 2006). However, these studies are limited by focusing on *offline* judgments in which the dynamic, real-time neurocognitive processes engaged as confidence cues in an utterance unfold cannot be measured.

1.1. *Neurocognitive studies of vocal (emotion) processing in speech*

Our review highlights a number of unanswered questions: How is expressed confidence decoded from vocal cues in speech, devoid of other verbal and non-verbal cues for understanding speaker confidence, in real time? And how quickly are vocal cues signaling different levels of confidence differentiated and meaningfully elaborated in the neural responses of listeners as speech continues to unfold? To our knowledge, there has been no study investigating the temporal neural dynamics underlying the processing of vocal confidence expressions. In this experiment, we sought to answer these questions using event-related potentials (ERPs), which demonstrate high temporal resolution and good sensitivity to the cognitive processes that act on vocal cues during real-time speech comprehension (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006).

While data on expressed confidence are lacking, neurophysiological and imaging studies have shaped neurocognitive models highlighting how vocal information about basic *emotions* is extracted and mapped onto representation (see reviews in Belin, Fecteau, & Bédard, 2004; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), and how specific temporal predictive mechanisms could operate in vocal expression processing (Kotz & Schwartze, 2010; Schwartze & Kotz, 2013). Several influential models characterize the decoding of vocal emotion expressions as a multi-step process, with each step recruiting different (often lateralized) neural mechanisms (Belin et al., 2004; Kotz & Paulmann, 2011; Schwartze & Kotz, 2013; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). According to the multi-stage model of vocal emotion processing (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), an initial stage of extracting and analyzing acoustic properties (pitch, loudness, rhythm) in speech recruits bilateral primary and secondary auditory cortices and takes place early around 100 msec post-acoustic onset. At a second stage, auditory information is marked for its emotional relevance (e.g., valence, arousal, specific emotional qualities) after approximately 200 msec (involving right anterior superior temporal gyrus/sulcus). In the final stage, late occurring processes of higher level cognition—for integrating multiple cue sources (lexical/semantic and prosodic), information from the broader context (task, speech context), and world knowledge—takes place after 400 msec from onset of auditory presentation (involving bilateral inferior frontal and orbito-frontal cortex, Kotz & Paulmann, 2011). Kotz and Schwartze (2010) further separate temporal structure from the formal aspect (spectral content) of speech, and propose two parallel pathways that recruit subcortico-cortical networks in temporal processing; given that speech is highly time dependent, their dual-pathway architecture allows for the decoding of vocal behavior via specific prediction of temporal patterns with the unfolding of speech (see Schwartze & Kotz, 2013 for details). The extent to which these models fit data on the neurocognitive processing of *expressed confidence* in speech remains unclear.

Electrophysiological studies have contributed important details about early and late neurophysiological markers that index online perception of vocal emotion expressions in speech (and possibly expressed confidence), which roughly correspond to each hypothesized processing stage (Alain & Winkler, 2012; and see Kotz & Paulmann, 2011; Schirmer & Kotz, 2006 for reviews). The early fronto-central auditory N1 is known to respond to a wide range of auditory stimulus types (Näätänen and Picton, 1987) as a measure of perceptual-level processing. This component is generated in auditory cortex within Heschl's gyrus approximately 50 msec after the initial response of primary auditory cortex, and thus originates early in language processing. In vocal emotion processing, N1 has been linked to the extraction of acoustic cues that differentiate different types of vocal signals, for example frequency and intensity parameters (Pantev, Elbert, Ross, Eulitz, & Terhardt, 1996; Paulmann & Kotz, 2008; Pourtois, de Gelder, Vroomen, Rossion, Crommelinck, 2000), and is largely unaffected by differences in emotional meaning.

The fronto-central P200 has been associated with the early attentional allocation or evaluation of vocal signals, after initial analysis of acoustic information (Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), ensuring preferential processing of emotional stimuli ("early emotional significance detection", Kotz & Paulmann, 2011). Studies have demonstrated modulation of P200 amplitude differentiating emotional versus neutral speech, comparing a wide range of emotions (anger, fear, disgust, happiness, pleasant surprise, sadness, Paulmann & Kotz, 2008a). Recent work has also reported differentiation of the P200 amplitude as a function of basic emotion types and possibly general arousal features (Paulmann, Bleichner, & Kotz, 2013; Sauter & Eimer, 2010), suggesting that this component may reflect an early function of "tagging" emotional or motivationally significant stimuli. Despite some inconsistencies, the P2 amplitude tends to be greater for angry versus fearful vocal expressions (Jessen & Kotz, 2011; Paulmann et al., 2013; cf. Paulmann & Kotz, 2008a) and for non-linguistic vocalizations versus spectrally-rotated counterparts (Sauter & Eimer, 2010). This emotion-specific modulation occurs irrespective of speaker identity (e.g., gender, Paulmann & Kotz, 2008a), whether the speech is semantically meaningful or not (e.g., real-vs. pseudo-utterance, Paulmann & Kotz, 2008b), or the task relevance of vocal cues (Kotz & Paulmann, 2007). When the acoustic structure of vocal expressions was examined, a larger P200 tended to correlate with higher mean and range of F0, larger mean and range of speech amplitude, and slower speech rate (Paulmann et al., 2013), arguing that the early P200 modulation is partially explained by early meaning encoding and continued sensory processing (Schirmer, Chen, Ching, Tan, & Hong, 2013). The idea that other vocal cues, such as differences in emotion and speaker identity cues, are indexed by the P200 subsequent to sensory processes has also been proposed (Spreckelmeyer, Kutas, Urbach, Altenmüller, & Münte, 2009).

In recent work, a late centro-parietal positivity (also named LPC) evoked by vocal emotion expressions has been identified as a positive-going wave starting about 500 msec post-onset of the auditory stimulus (Paulmann et al., 2013; Schirmer et al., 2006), and sustained until 1000 msec depending on stimulus characteristics. In vocal emotion studies, the LPC is thought to involve multiple in-depth or second-pass cognitive processes such as evaluation (or re-evaluation) of the meaning of vocal emotional signals and/or access to emotional memory representations (Kotz & Paulmann, 2011; Paulmann et al., 2013; Schirmer et al., 2006). The LPC was found to be larger in emotional (especially high-arousing) vocal stimuli, leading to parametric differences in the LPC amplitude among basic emotion types (Paulmann et al., 2013); this suggests a more elaborate processing of vocal information at this stage (e.g., building-up of emotional representation and/or construction of intended speaker meaning). The late positivity effect has also been reported for a variety of processes in speech and language comprehension not specifically tied to vocal emotions ("P600", see Brouwer, Fitz, & Hoeks, 2012 for a review). For example, the positivity effect has been linked to ambiguity resolution (Kolk & Chwilla, 2007) or access to non-literal meanings signaled in part by vocal information (Regel, Gunter, Coulson, 2010; Regel, Gunter, & Friederici, 2011; Regel, Meyer, & Gunter, 2014; Rigoulot, Fish, & Pell, 2014; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013). Although different descriptions of the late positivity (LPC and P600) appear sensitive to different aspects of speech stimuli (e.g., emotional *vs* linguistic), they may reflect a task-relevant in-depth, second-pass evaluation (Schact & Sommer, 2009; Schirmer, Kotz, & Friederici, 2002; 2005; Schirmer et al., 2006) or task-irrelevant monitoring process (Kotz & Paulmann, 2007; Paulmann et al., 2013) which aims at constructing a mental representation of what is intended in the message by the speaker (Brouwer, Fitz & Hoeks, 2012; Brouwer & Hoeks, 2013). Besides, a more delayed sustained positivity may reflect a listener's attempt to infer the goal/intention of a speaker or third person, for example when an expected way of speaking or acting is mismatched in an utterance or sentence context (e.g., a positivity starting from 500 msec and sustaining for 1000 msec; Jiang, Li, & Zhou, 2013; Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009; Van Overwalle, Van den Eede, Baetens, & Vandekerckhove, 2009). Together, these ERP effects allow us to make predictions about the time course of neural responses to vocal expressions of expressed confidence in the current study.

1.2. Current approach and hypotheses

This study investigated how *and* when the brain responds to expressed confidence in the vocal channel of speech, to provide new data on how listeners infer key aspects of a speaker's mental state (i.e., 'feeling of knowing'). To this end, listeners judged a set of pre-validated utterances (selected from Jiang & Pell, 2014) reflecting the confident, close-to-confident, unconfident, or neutral disposition of the speaker while online neural responses (EEGs) were recorded. In light of existing data and multi-stage models of vocal *emotion* processing, we hypothesized that different levels of expressed confidence would register at different time points during vocal processing, reflected in distinct ERP effects. Resembling vocal emotions, vocal confidence expressions are encoded by an evolving constellation of acoustic parameters (differences in frequency, amplitude, and temporal properties) over the course of the utterance, and there is evidence from patients that the ability to decode vocal emotions and vocal confidence may rely on shared neural systems (Monetta et al., 2008). However, *different* from vocal emotions, expressed confidence encodes *relational* meanings that are socially relevant to the speaker – hearer in the interpersonal context in which they appear; in social-pragmatic terms, they mark the reliability or correctness of an assertive speech act along a continuum of confident/doubtful (Caffi & Janney, 1994). We predicted that the early extraction and analysis of vocal acoustic changes (such as mean f0 and amplitude) should lead to differentiation of early ERP responses, e.g., N100 and/or P200; in particular, we expected that *confident* utterances would elicit larger early components from stimulus onset when compared to *unconfident* statements due to their physical properties and/or motivational salience to listeners .[1] Although an unconfident utterance may bear some motivational salience in a certain case when it implies the lack of credibility of the speaker (Kimble & Seidel, 1991; Scherer et al., 1973), a confident voice is more likely to be evaluated as salient in a task when the speaker's confidence is the focus of processing. We also hypothesize the *close-to-confident* level to be differentiated from the confident and the unconfident level in these early responses, if the initial acoustic analysis led to a detection of some emotional/motivational significance of this intermediate level of confidence that can be differentiated from that of the other levels (e.g., lower than confident and higher than the unconfident). Moreover, an in-depth, second-pass evaluation of the speaker's confidence level should lead to different amplitude in a late positivity; in this later time window, we predicted that more subtle, intermediate levels of confidence (i.e., *close-to-confident* or "mildly unconfident" voices) would be differentiated from less ambiguous confident and unconfident signals registered by early components, reflecting processes of ambiguity resolution (Kolk & Chwilla, 2007) or pragmatic inferencing (Jiang et al., 2013; Van der Cruyssen et al., 2009; Van Overwalle, Van den Eede, Baetens, & Vandekerckhove, 2009), given that the correct evaluation of the relevant social-relational meaning depends on a fine-grained analysis/dissociation of subtle variations between speaker's confidence level. Finally, although no expressed confidence was intended by neutral utterances presented in the experiment, we anticipated that these stimuli might trigger an inferential process in light of our task demands to focus listeners' attention towards the speaker's confidence level; since listeners tend to rate intended neutral statements as different but relatively high in confidence (Jiang & Pell, 2014), sustained evaluation of these stimuli might lead to a delayed, sustained positivity for neutral expressions in the context of our experiment.

2. Methods

2.1. Participants

Thirty university students consented to participate in the EEG experiment (15 female/15 male, mean age = 22.6 years, range = 18–30 years). All were native Canadian English speakers who were right-handed. None had suffered from major psychiatric or neurological illness, or had speech or hearing problems, according to self-report, nor had any participated in our previous studies (e.g., Jiang & Pell, 2014). This study was carried out in accordance with the Declaration of Helsinki and was ethically approved by the Institutional Review Board of the Faculty of Medicine at McGill University (Montréal, Canada).

2.2. Design and materials

Ninety-six stimulus "quartets" were selected from a database of vocal expressions of confidence (see the details of the elicitation and selection of stimuli in 2.3 and in Jiang & Pell, 2014). Within each quartet, identical statements expressed either a confident, close-to-confident, unconfident, or neutral-intending message based only on changes in tone of voice. All statements were simple, canonical English utterances, consisting of a subject, verb, and complement with high frequency of use in spoken English (see Appendix A for exemplars). The 96 statements varied in communicative function: 32 were descriptions of fact (e.g., *She has access to the building*), 32 were statements of judgment (e.g., *He's too old to split the wood*), and 32 were statements of intention (e.g., *We'll help them with it*). All statements described a concrete event and were about an unspecified person or unknown object and none described an event or fact associated with strong expectancies of being true, or having a known truth value. All vocal materials were recorded in a separate expression elicitation study (Jiang & Pell, 2014; see below for details). The EEG study used recordings from two male and two female speakers who encoded the most salient distinctions in the low to high levels of confidence when judged by naïve listeners in the validation study (see Appendix A).

Altogether, 384 recordings (96 statements × 4 confidence levels) were chosen as the experimental stimuli. Another 480 utterances that contained an initial lexical phrase for inferring confidence (e.g., *I'm positive…*, *Maybe…*) were also presented for

[1] The acoustic profile of confident versus unconfident voices cannot lead to a uniform prediction of P200 patterns based on results reported by Paulmann et al. (2013), since confident expressions sound louder and faster, but are also lower in pitch on average, than *unconfident* statements (see also Kimble & Seidel, 1991; Pell, 2007; Scherer et al., 1973).

the purpose of a companion study investigating the effects of verbal cues on vocal confidence perception (for these items, sentences were identical to the critical stimuli but included an initial lexical phrase that was not removed prior to the EEG experiment). To reduce the effect of repetitive presentation of different variants of the stimuli, two experimental lists were generated so that half of the stimulus quartets ($n = 48$) were assigned to only one of the two lists, and the corresponding utterances that contained the lexical phrases were always assigned to a different list. Thus, the 384 critical utterances were divided into two experimental lists, with each list consisting of 192 critical statements with 48 of each expression type (confident, close-to-confident, unconfident, neutral-intending). The other utterances with different experimental purposes were also divided in the two lists, with each list consisting of 240 statements. Another 48 statements, with different sentence stems from the critical statements, were added as fillers and were repeated in the two lists. In total there were 480 statements (=192 + 240 + 48) in each list. Thirty different sequences were generated according to a pseudo-randomization procedure that ensured no more than five consecutive statements conveyed the same level of confidence, and no more than five consecutive utterances were produced by the same speaker. Each participant was assigned randomly to one sequence. An equal number of male and female participants were presented each list.

2.3. *Expression elicitation, validation, and stimulus selection*

Full details on the construction and validation of the confidence recordings used in this study are reported in Jiang and Pell (2014). Briefly, all recordings were elicited from six native English speakers with experience in professional acting and/or public speaking, from whom two female and two male speakers (mean age = 21.8 years, range = 19–25 years) were selected for the EEG experiment. Each speaker produced 151 statements, each with three levels of confidence and in a neutral manner. All utterances were recorded to digital media using a high quality head-mounted microphone attached to a professional Tascam Recorder, with a sampling rate of 16 bit, 44.1 kHz. During the recording session, speakers were coached to adopt a mental state corresponding to the intended level of confidence; for the confident condition, they were told to express absolute confidence that what is described in the sentence is true, and to aid naturalistic productions, confident statements began with either *I'm certain/I'm positive/For sure/Definitely*. Statements in the close-to-confident condition were intended to encode a near certainty about what is described in the sentence and began with *Most likely/I'm almost certain/I think*/or *I'm pretty sure*. The statements in the unconfident level began with *Perhaps/Maybe/There's a chance/It's possible* and were intended to encode only a possibility that the speaker thinks information in the sentence is true. In the neutral statements, there was no preceding phrase and the speaker was told they were simply reporting this information to another person in a non-emotive manner. Recordings were generated in blocks according to the intended confidence level, randomized across speakers. The naturalness of each expression was further facilitated by having the speaker express each statement as an answer to a question posed by the examiner, as if they were a real conversational partner.

Each recording was edited to generate two versions: a with-phrase version (the linguistic phrase was maintained) and a no-phrase version (the linguistic phrase was removed). The edited no-phrase recordings (altogether 3624 tokens, which were composed of 151 sentences by 4 speakers in 4 expression types) along with the with-phrase recordings were separated into 6 experimental lists to ensure minimum repetition of the same sentence by a speaker. They were entered into a perceptual study involving 60 native English speakers (10 judges for each list) who listened to each individual statement and judged: 1) if it conveyed some level of confidence (yes or no); and 2) if yes, how confident the speaker sounded on a 1–5 scale of increasing confidence (see Jiang & Pell, 2014 for detailed report of the results).

These perceptual data served to identify well-controlled stimulus quartets for the EEG study ensuring that: 1) all confident, close-to-confident, and unconfident stimuli were positively identified on the first question as *intending* to convey some level of confidence (minimum 70% agreement); 2) for neutral-intending statements, at most 70% participants positively agreed the speaker means to convey some level of confidence; and 3) all confident items had a mean rating on the second question above 4, all close-to-confident items had a mean rating between 2.9 and 4, and unconfident items had a mean rating below 2.8 (see Table 1). The stimuli were selected by quartets to ensure that psycholinguistic properties of the stimuli (e.g., the lexical choice) in each category were always the same. Twenty-four quartets were selected for each speaker. The mean duration of utterances naturally differed between conditions: 1615 msec ($SD = 371$ msec) for confident expressions, 1589 msec ($SD = 330$ msec) for close-to confident, 1791 msec ($SD = 463$ msec) for unconfident, and 1325 msec ($SD = 248$ msec) for neutral-intending expressions.

To understand differences in perceived confidence among our four conditions, a linear mixed effect model (LMEM) was first built on the percentage of response of conveying some level of confidence to each selected item ($n = 384$), including level of confidence and list for EEG experiment as fixed effects, and by-item intercepts as a random effect. The results revealed a significant main effect of level of confidence, $F(3, 328) = 68.62$, $p < .0001$. Post-hoc comparisons revealed that the neutral-intending statements were less likely to be judged as conveying some level of confidence than the confident (neutral > confident: $\beta = -.18$, $t = -3.50$, $p = .0005$), the close-to-confident (neutral > close-to-confident: $\beta = -.16$, $t = -2.98$, $p = .0031$) and the unconfident statements (neutral > unconfident: $\beta = -.16$, $t = -3.01$, $p = .0028$). No differences were found between the other levels of confidence ($ps > .5676$). Neither the main effect of list nor the interaction between list and level of confidence reached significance ($Fs < 1$). A second model was built on the confidence rating of each selected item by each rater ($n = 3840$), including level of confidence and EEG list as fixed effects, and by-rater and by-item intercepts as random effects. The results revealed a significant main effect of level of confidence, $F(3, 2581) = 962.51$, $p < .0001$. Post-hoc comparisons revealed that perceived speaker confidence gradually decreased from *confident* statements (confident > unconfident: $\beta = 2.15$, $t = 49.97$,

Table 1 – Means (standard deviations) of the behavioral validations for each level of confidence across the selected stimuli in the perceptual study.

Intended level of confidence	Percentage of "Yes" response in judging whether the speaker conveyed some level of confidence (%)[a]			Speaker confidence rating (1–5)[b]		
	List 1	List 2	Total	List 1	List 2	Total
Confident	89.9 (9.2)	90.5 (9.4)	90.2 (9.3)	4.28 (.30)	4.26 (.27)	4.27 (.28)
Close-to-confident	86.0 (9.8)	85.5 (10.2)	85.8 (10.0)	3.45 (.49)	3.50 (.55)	3.48 (.52)
Unconfident	85.8 (8.0)	84.8 (7.9)	85.3 (7.9)	2.16 (.57)	2.06 (.66)	2.11 (.62)
Neutral	42.1 (7.5)	43.2 (6.2)	42.7 (6.8)	4.26 (.40)	3.93 (.44)	3.99 (.42)

[a] The mean percentage was calculated for each item as the number of "Yes" response divided by total number of response ($n = 10$); the standard deviation reveals inter-item variation.
[b] The mean speaker confidence rating was calculated as the mean rating of each item that has been judged as conveying some level of confidence ($n = 10$); the standard deviation reveals inter-item variation.

$p < .0001$; confident > close-to-confident: $\beta = .59$, $t = 13.77$, $p < .0001$; confident > neutral: $\beta = .29$, $t = 6.33$, $p < .0001$) to *neutral-intending* statements (neutral > close-to-confident: $\beta = .29$, $t = 6.45$, $p < .0001$; neutral > unconfident: $\beta = 1.86$, $t = 39.90$, $p < .0001$), to *close-to-confident* expressions (close-to-confident > unconfident: $\beta = 1.56$, $t = 35.89$, $p < .0001$), with unconfident expressions receiving the lowest ratings. Neither the main effect of list nor the interaction between list [$F(1, 2581) = 1.99$, $p = .1589$] and level of confidence [$F(3, 2581) = 1.60$, $p = .1880$] reached significance. These findings confirmed our hypothesis that 1) the intended levels of confidence were well perceived based on vocal cues; and 2) that neutral-intending utterances, while not meant to express confidence, are inferred as relatively high in confidence under a task demand that emphasized the confidence judgment (greater than close-to-confident but less than confident; see also Appendix A for the results of full perceptual experiment).

Acoustic specification of the critical statements considered five acoustic measures commonly reported in studies of vocal emotion expression: mean and range of the speaker's fundamental frequency (in Hz, normalized); speech rate (syllables/second); and the mean and range of amplitude (in dB, normalized).[2] The LMEM taking intended level of confidence as a fixed effect, and by-item intercepts and slopes for the fixed effect revealed a significant effect of level of confidence on mean speaker f0, $F(3, 284) = 63.55$, $p = .0001$, mean amplitude, $F(3, 284) = 6.62$, $p = .0002$, amplitude range, $F(3, 284) = 25.66$, $p = .0001$, and speech rate, $F(3, 284) = 60.24$, $p = .0001$. The acoustic data and results showing which confidence conditions were associated with significant differences for each measure are presented in Table 2 and illustrated in Fig. 1 (for pitch contour). In general, mean f0 was higher when a speaker was unconfident than in all other conditions; also, close-to-confident and confident expressions (which did not differ) displayed a higher mean f0 than neutral-intending utterances. For amplitude, the speaker mean tended to be highest in neutral-intending and confident expressions (which did not differ) and lowest in close-to-confident expressions; confident expressions also displayed a larger range in amplitude than all other expression types (and unconfident expressions exhibited more amplitude variation than close-to-confident or neutral-intending stimuli). Speech rate was significantly faster for neutral-intending expressions than all three types of confidence expressions, and notably slower when speakers were unconfident than confident or close-to-confident. The analysis of f0 range did not yield significant differences across conditions, $F < 1$, although unconfident expressions tended to display the largest f0 range. These findings verify that intended confidence levels can be differentiated by a minimal set of acoustic parameters and that neutral-intending statements also demonstrated important distinctions from expressions associated with different confidence levels (see Table 2).

2.4. Testing and EEG recording procedure

Testing was conducted in an electrically shielded, sound-treated booth. Participants were seated in a comfortable chair and were instructed to listen carefully to a series of statements spoken by different speakers in different ways (stimuli were normalized to a peak intensity of 70 dB prior to presentation). Each trial began with a fixation point followed by a jittered delay (mean = 700 msec across conditions, range = 500–900 msec) of the onset of the vocal stimulus. Participants were told to look at the fixation point as they listened to each statement and then judge the level of the speaker's confidence on a 5-point rating scale, which appeared on the screen 2000 msec after the end of the auditory stimulus. Responses were recorded via a response box (Cedrus, USA) with five buttons programmed to correspond to each of the five points shown on the screen (from "not at all" to "very much"). Two scales, one with the leftmost and one with the rightmost point as the most confident, were counterbalanced across participants of each gender. The scale disappeared when a response was made or after 5 s without a response. The next trial started after a 1500-msec blank screen. Prior to the formal session, 20 practice statements were presented to each participant. The whole session lasted about 2.5 h including electrode preparation.

The electroencephalograms (EEGs) were recorded continuously from 64 electrodes using the ActiCap System (Brain Product, Germany). The vertical electrooculograms (VEOG) were recorded from above and below the right eye and the horizontal electrooculograms (HEOG) were recorded from the

[2] Frequency and amplitude values were normalized per statement per speaker as their proportional distance in reference to the speaker's resting frequency/amplitude (i.e., the mean of each speaker's minimum frequency/amplitude in neutral statement; see Liu & Pell, 2012; Pell, Paulmann, Dara, Alasseri, Kotz, 2012).

Table 2 – Means (standard deviations) of the behavioral responses in the EEG study, and of the acoustic measures in each level of confidence and the relative ranking of different confidence expressions by acoustic parameter.

Intended level of confidence	Perceptual rating (1–5)[a]			Acoustic measures									
				Mean f0		Range f0		Speech rate		MeanAmp[c]		RangeAmp[d]	
	Female mean (SD)	Male mean (SD)	Total mean (SD)	Mean (SD)	Ranking[b]	Mean (SD)	Ranking[b]	Mean (SD)	Ranking[b]	Mean (SD)	Ranking[b]	Mean (SD)	Ranking[b]
Confident	4.65 (.33)	4.28 (.60)	4.48 (.47)	.31 (.22)	2	1.60 (1.30)	1	4.49 (1.00)	2	.54 (.08)	1	1.03 (.10)	1
Close-to-confident	4.12 (.70)	3.88 (.72)	3.99 (.73)	.33 (.18)	2	1.56 (1.28)	1	4.52 (.87)	2	.50 (.13)	3	.920020(.10)	3
Unconfident	2.28 (.86)	2.56 (.87)	2.37 (.86)	.55 (.22)	1	1.67 (1.01)	1	4.10 (1.04)	3	.52 (.11)	2	.98 (.09)	2
Neutral	4.40 (.52)	4.12 (.66)	4.24 (.62)	.25 (.14)	3	1.55 (1.40)	1	5.37 (.89)	1	.55 (.11)	1	.90 (.09)	3

[a] 1 – "not at all confident 5 – means very confident". In the EEG recording session, participants were asked to listen carefully to each statement and then rate "how confident you think the speaker is" on a 5-point scale. The standard deviation reveals inter-item variation.
[b] 1 – Highest ranking, 4 – Lowest ranking. The ranking indicates significant differences in post-hoc comparisons (Markov Chain Monte Carlo sampling tests).
[c] Mean amplitude.
[d] Range of amplitude.

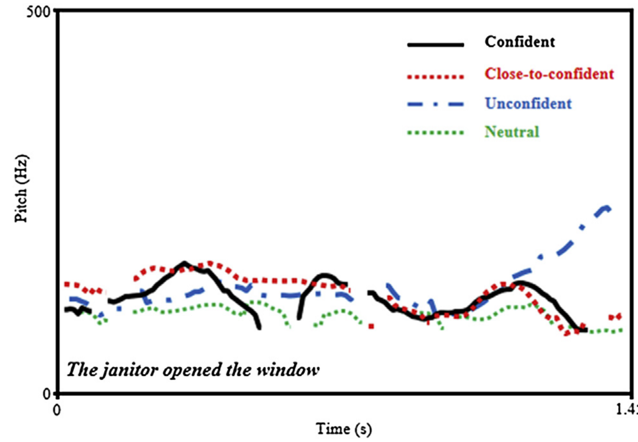

Fig. 1 – Example of pitch contours for the four expressions of confidence for one speaker.

outer canthus of both eyes. The recordings were online referenced to FCz and re-referenced offline to the bilateral mastoids. The EEGs were digitized at 500 Hz and filtered with a band-pass from .016 Hz to 100 Hz.

2.5. *EEG analysis*

All pre-processing procedures were performed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014). The continuous EEGs were first visually inspected and signals with excessive movement artifact, alpha activity or amplifier saturation were manually excluded from the analysis. The subsequent EEGs were filtered using a 30 Hz low-pass Butterworth of the sixth order with zero-lag and then decomposed with an ICA algorithm (Makeig, Bell, Jung, Sejnowski, 1996) to remove ocular artifacts. Given that speaker confidence is sometimes conveyed by a falling versus rising pitch contour toward the end of an utterance as it unfolds (Bolinger, 1978), and given that the mean duration of our items by condition (ranging from 1325 to 1791 msec) allowed us to capture the late and sustained ERP effects, the resulting datasets were segmented into an epoch starting 200 msec before and 1300 msec after the onset of each stimulus, covering the majority of the utterance. The epoch was baseline corrected based on the mean EEG activity in the 100-msec pre-stimulus interval. Segments with signal peak-to-peak voltage exceeding 70 μV within a 100-msec sliding window were automatically rejected.

The ERP response was time-locked to the onset of the vocal stimuli. Linear mixed effects models (LMEM) were used to fit the mean ERP amplitude that was calculated per subject per condition, and to statistically evaluate the difference between conditions. The LMEM allows us to evaluate the effects of interest in the experimental manipulation as fixed factors while considering the individual difference across subjects in its intercept and its interaction with the main effect of interest as random factors (Baayen, Davidson, Bates, 2008; Newman, Tremblay, Nichols, Neville, & Ullman, 2012). Moreover, it extends the repeated measures models in general linear model by allowing unequal number of repetitions. The ERP signals within one subject were averaged before being entering into the model to ensure signal-noise ratio (see Newman et al., 2012).

To tap into early operations associated with the analysis of acoustic and motivational information in the vocal signal (Paulmann et al., 2013; Paulmann & Kotz, 2008a; Paulmann, Ott, & Kotz, 2011; Schirmer et al., 2013; Schirmer & Kotz, 2006), we calculated mean averaged amplitude[3] in the 70–160 msec window for N1 (mean latency = 119 msec, $SD = 21$ msec) and in the 150–280 msec window for P2 (mean latency = 213 msec, $SD = 28$ msec) per participant per condition, time-locked to statement onset.

In order to tap into the late continuous evaluation (Kotz & Paulmann, 2011; Paulmann et al., 2013; Regel et al., 2010; Regel et al., 2011; Regel et al., 2014; Rigoulot et al., 2014; Schirmer et al., 2006; Spotorno et al., 2013) and/or pragmatic inference (Jiang et al., 2013; Van der Cruyssen et al., 2009; Van Overwalle et al., 2009), we defined the time windows based on two approaches. We first performed a 10 msec-interval analysis on the mean ERPs and larger time windows with consecutive differences across multiple small windows between conditions were determined for further analysis[4] (see Strauß, Kotz, & Obleser, 2013; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010). To provide converging evidence on the temporal data patterns that were obtained in the mean amplitude analysis, a two-step principal component analysis (temporal-spatial PCA) was subsequently performed to decompose the time/space-locked correlational structure of the ERP amplitudes (ERP PCA Toolkit, Version 2.33, Dien, 2010; 2012). The PCA provides a data-driven approach that is free of the theoretical preconceptions that are implied by standard ERP component labeling (e.g., P200, N400, LPC/P600). This approach allows the separation of temporally and/or spatially overlapping components (e.g., N400 *vs* P300, Molinaro & Carreiras, 2010; Vespignani et al., 2010) and the confirmation of which independent spatial component has revealed the effect of the experimental factor within a certain time window. Mean amplitude of these time windows was calculated per participant per condition using ERPLAB. A 10-Hz low-pass filter was applied for the purpose of displaying figures.

Although our selection of vocal stimuli ensured that each level of confidence was distinct from each other in the acoustic-perceptual profile, there could still exist inconsistent ratings between the perceptual pretest and the EEG task. Firstly, the selection in the pretest was mainly based on the mean rating of each individual item averaged across all raters. For example, for a close-to-confident expression, all raters may give a 3 response or they may give responses varying from 1–5.[5] Secondly, vocal signals are generally more ambiguous than the linguistic signal (Wilson & Wharton, 2006) and there is inconsistency between the speaker's intention and the listener's perception of the vocal confidence (Pon-Barry & Shieber, 2010) as well as variations in the perceived confidence between listeners on a certain category (Pon-Barry & Shieber, 2010). In order to reveal online neural responses that were qualified by its behavioral response, we regrouped each confidence-intended statement and compared ERP responses for items reassigned according to their *perceived* level of confidence in the experiment based on perceptual ratings in the 5-point scale per subject. On a scale from 1 to 5, we assigned 5 to confident, 3 and 4 to close-to-confident and 1 and 2 to unconfident level (see Kimble & Seidel, 1991; Monetta et al., 2008; Wang, Zhu, Bastiaansen, Hagoort, & Yang, 2013 for a similar procedure). Such procedure ensures the assigned category was based on individual response to each item. Among all trials regrouped to be confident, 61.4% were statements intended to be confident and 34.0% were intended to be close-to-confident; among those regrouped to be close-to-confident, 44.1% were statement intended to be close-to-confident, 27.0% were intended to be confident and 28.8% were intended to be unconfident; among those regrouped to be unconfident, 82.1% were statements intended to be unconfident and 13.2% were intended to be close-to-confident. The mean confidence rating of each regrouped level of confidence was 5, 3.71 and 1.64 from highest to lowest. We did not regroup neutral-intending statements because there was evidence they were produced in a qualitatively different way and were judged less frequently to convey some level of confidence in the elicitation and validation study (Jiang & Pell, 2014).

For the ERP analyses, the LMEM included fixed effects of level of confidence (confident *vs* close-to-confident *vs* unconfident *vs* neutral, reference = confident[6]) and topographic factors (see below). The mean number of trials accepted after artifact rejection was 36 (confident), 42 (close-to-confident), 26 (unconfident), and 38 (neutral) in each reassigned confidence condition. The unconfident condition contained less trials than the other conditions, $F(3, 86) = 6.43$, $p = .0006$ (unconfident > confident: $\beta = -11.97$, $t = -2.95$, $p = .0022$; unconfident > close-to-confident: $\beta = -16.87$, $t = -4.30$, $p = .0001$; unconfident > neutral-intending: $\beta = -12.73$, $t = 3.25$, $p = .0015$; comparisons between confident, close-to-confident and neutral-intending, $ts < 1.82$). The incongruent number of trials is apparently due to a listener bias to not use the low (unconfident) end of the confidence rating scale (with the mean number of reassigned trials in each confidence category 43, 49 and 32 from highest to lowest). Therefore we performed an exploratory analysis of ERP patterns on a randomly-selected subset of trials with equal numbers in each category for each participant (ranging from 20 to 32 across participants). The resulting ERP pattern

[3] The mean number of accepted trials in unconfident expression was slightly lower than in other category, leading to a larger noise level than the other expression types. Given that the mean amplitude is not biased by the noise level (Luck, 2010), statistical analysis were performed on the mean amplitude.

[4] The LMEM was performed on the mean ERPs every 10 msec starting from 250 msec, comparing each level of confidence against each other. The onset time of the effect was defined as the time where at least five subsequent tests were significant.

[5] The standard deviation of rater's response to each selected item in the perceptual study was .71 ± .22 for the confident, .79 ± .25 for the close-to-confident, .74 ± .29 for the unconfident, and .77 ± .25 for the neutral-intending category. The standard deviation of the rater's responses in the EEG study was .78 ± .21 for the confident, 1.03 ± .28 for the close-to-confident, 1.18 ± .23 for the unconfident, and .92 ± .22 for the neutral-intending category.

[6] Each level of the fixed factors was dummy-coded such that one level of factor was assigned as reference and compared to all other levels when the effect of the factor was evaluated by the LMEM. The selection of the reference does not affect the *F*-tests and the coefficients in the pairwise *t*-tests differ in the positive or negative sign depending on which of the levels is treated as the reference.

was essentially the same as the full dataset and therefore we report the results from the untrimmed data.

The topographic factors included hemisphere (medial *vs* left *vs* right, reference = medial) and region (anterior *vs* central *vs* posterior, reference = anterior) and were included as fixed factors. Nine regions of interest (ROI) were formed, with each represented by 6–8 electrodes: 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 (F1, FZ, F2, FC1, FCZ, FC2), medial central (C1, CZ, C2, CP1, CPZ, CP2), medial posterior (P1, PZ, P2, O1, POZ, O2), right anterior (AF4, FP2, F4, F6, F8, FC4, FC6, FT8), right central (C4, C6, T8, CP4, CP6, TP8), and right posterior (P4, P6, P8, PO4, PO8, PO10). The electrodes were treated as repeated measures. Listener sex was included as another fixed effect in order to evaluate any potential sex effect on vocal confidence processing, as shown in related studies of emotion (Schirmer, Kotz, & Friederici, 2003; Schirmer et al., 2005). Participants were included as random intercept as well as random slopes in order to evaluate the individual adjustments in the magnitude of ERP responses as a function of fixed factors (Newman et al., 2012). The identification of the optimal mixed-effects model was performed for each window of interest by comparing a full model taking all fixed effects and progressive simpler models using log-likelihood ratio tests (Baayen et al., 2008; Tremblay, 2014). This procedure allows removal of factors and interactions that do not explain significant amounts of variance. The model was then refit, and variance explained by each factor in the optimal model was examined by way of *F* tests for main effects and interactions and *t* tests for specific contrasts. Both upper-bound (liberal) and lower-bound (conservative) probability values were considered, although only lower-bound values are reported for expository purposes (Newman et al., 2012).[7] All analyses were performed in R-studio (Version 3.1.0, http://cran.r-project.org) with *lme4* and *lmerConvenienceFunctions* packages.

3. Results

3.1. Behavior

Behavioral ratings show that listeners in the EEG experiment perceived vocal expressions in each intended confidence condition (prior to item reassignment) in a distinct manner, consistent with the validation study. The LMEM including intended level of confidence and listener sex as the fixed effects, and by-participant and by-item intercepts and by-participant and by-item slopes for level of confidence as random effects, revealed a significant effect of confidence level, $F(3, 5368) = 142.08$, $p < .0001$. Speaker confidence ratings were significantly higher for confident expressions (confident > neutral: $\beta = .09$, $t = 2.42$, $p = .0155$; confident > close-to-confident: $\beta = 1.05$, $t = 2.75$, $p = .0060$; confident > unconfident: $\beta = 2.61$, $t = 6.83$, $p < .0001$), followed by neutral-intending and close-to-confident expressions which did not differ from one another (neutral > close-to-confident: $\beta = .13$, $t = .33$, $p = .7434$) and were the lowest for unconfident expressions (neutral > unconfident: $\beta = 1.69$, $t = 4.42$, $p < .0001$; close-to-confident > unconfident: $\beta = 1.56$, $t = 4.09$, $p < .0001$; see Table 2). The model also revealed a significant interaction between confidence and sex, $F(6, 5368) = 21.52$, $p < .0001$. Post-hoc comparisons on each sex type revealed that the four expressions were significantly differentiated from each other only in female listeners, while confident, close-to-confident and unconfident expressions were undifferentiated in male listeners (see Table 2 and Appendix A).

3.2. ERPs

Fig. 2 shows the grand average waveforms for vocal expressions in each reassigned level of confidence starting from the onset of the statement. The vocal expressions revealed a temporal dynamics which started with an N1–P2 complex maximized in anterior regions, followed by a centro-posteriorly prominent negative-going wave peaking at around 450 msec, which shifted towards positive voltage until the end of the epoch in the posterior regions. The 10-ms window analysis revealed the effect of speaker's confidence at 183 msec, where both confident and close-to-confident expressions elicited a larger P200 in the anterior regions, as compared with unconfident vocal expressions. The close-to-confident and neutral-intending expressions elicited less negative-going waves broadly-distributed from 330 to 500 msec. The close-to-confident expressions also elicited a less negative-going wave broadly-distributed from 550 to 740 msec than all other expression types. Finally, neutral-intending expressions elicited a larger, sustained posterior positivity (starting from 980 msec and sustained until 1270 msec) than stimuli in the other conditions. These patterns were later confirmed by linear mixed effects modeling; significant effects emerging from these analyses are displayed in Fig. 3 and statistical results are reported in full in the appendices. The PCA analysis revealed 11 main temporal-spatial components that lie in six temporal factors (see Appendix G and Fig. 4). With early peak latency, TF6SF1/TF6SF2 (118 msec) and TF3SF1/TF3SF2 (212 msec) correspond to the N100 and P200 (e.g., Fig. 4a). The TF2SF2 (peaking at 438 msec) corresponds to a centro-posterior negativity although the same temporal factor revealed also a frontal positivity (TF2SF1). The two factors are likely to reflect the two field sides of the same electrical dipole (Fig. 4b). The TF4SF1 (peaking at 610 msec) was temporally close and had a similar spatial distribution to TF2SF1. It is possible that both reflect similar patterns of a positivity statistically split into two factors (Vespignani et al., 2010). The TF5SF1/TF5SF2 (peaking at 848 msec) and TF1SF1/TF1SF2 (peaking at 1298 msec) revealed later peaks and represented the components after 800 msec (e.g., Fig. 4c).

[7] The upper-bound *p* value was calculated using the difference between the number of data points and the number of *df* used up by the fixed effects. The lower-bound *p* value was calculated using the difference between the number of data points and the number of *df* used up by both the fixed effects and the number of random intercepts and slopes (e.g., *df* used up by individual adjustments for the level of confidence plus *df* used up for individual adjustments for topographic factors, see also Newman et al., 2012).

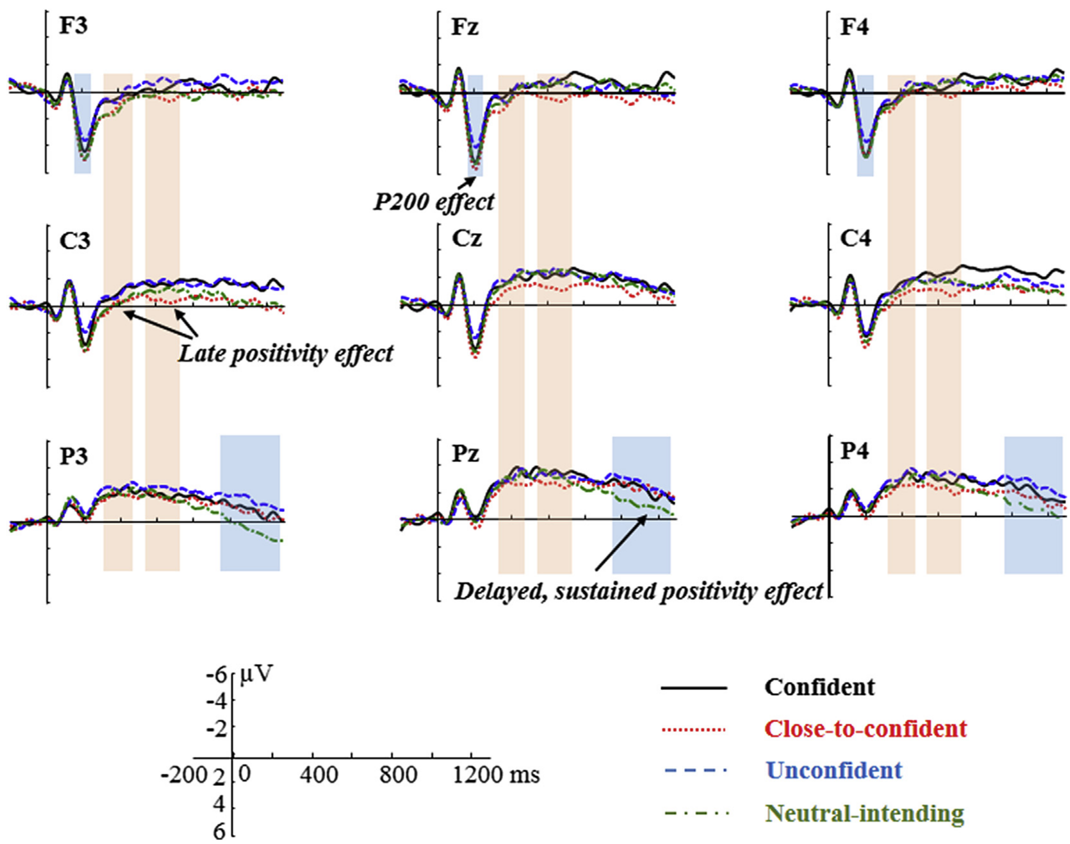

Fig. 2 – Grand average waveforms showing the ERP effect of vocal confidence on 9 representative electrodes, time-locked to the onset of the utterance.

For the N1 and P2 analyses, the optimal LMEM model included level of confidence, hemisphere, region and listener sex as well as by-participant slopes for all fixed main effects. The analysis on N100 did not reveal any significant effect of perceived confidence. The analysis on P200 revealed a significant main effect of perceived confidence level (see Appendix C for the summary of the results). Post-hoc comparisons revealed that unconfident expressions elicited a smaller P200 than confident, neutral-intending, and close-to-confident expressions, whereas no differences were observed between any of the latter three conditions. The optimal model also revealed a significant two-way interaction between perceived confidence and region. Further-going models including only perceived confidence and hemisphere as well as by-participant random effects revealed that the P200 modulation existed in anterior regions (Fig. 3a). Post-hoc comparisons confirmed a reduced P200 for unconfident expressions when compared to confident ($\beta = -.87$, $t = -2.07$, $p = .04$), close-to-confident ($\beta = -1.11$, $t = -2.91$, $p = .0037$) and neutral-intending ($\beta = -1.35$, $t = -3.53$, $p = .0004$) expressions in the anterior region. There was no difference between confident, close-to-confident, and neutral-intending levels, $ts < 1.27$. There was no main effect of sex nor interaction between sex and confidence on P200 amplitude, suggesting that the reduced P200 effect in the unconfident expression was comparable between male and female listeners (see Appendix C).

To examine late ERP effects that may be sensitive to different levels of perceived confidence, the LMEM was built on mean amplitudes in the 330–500 msec, the 550–740 msec, and 980–1270 msec time period, respectively. The optimal model maintained all fixed effects and random slopes for all fixed main effects for all time window analyses. The model conducted in the 330–500 msec time window revealed a significant main effect of confidence (see Appendix D). Post-hoc comparisons revealed that close-to-confident elicited a less negative-going effect (at centro-posterior and more positive-going effect at anterior region) than confident and unconfident expressions in this time interval (Fig. 3b). The neutral-intending utterances also elicited a less negative-going effect than confident and unconfident expressions. No interaction between confidence and topographic factors were found, suggesting that the modulation in this time window was broadly distributed. No effect of sex was observed, suggesting that females and males exhibited comparable effects (see Appendix D).

The model performed on the 550–740 msec window revealed a significant main effect of confidence (see Appendix E). Post-hoc comparisons revealed that close-to-confident expressions elicited a less negative-going effect (at centro-posterior and more positive-going effect at anterior region) than all three other levels of confidence. The interaction between sex and confidence was not significant, although a three-way interaction between sex, confidence and hemisphere reached significance. Further-going models revealed that the positivity effect of close-to-confidence level was broadly-distributed in the male listener whereas it revealed a

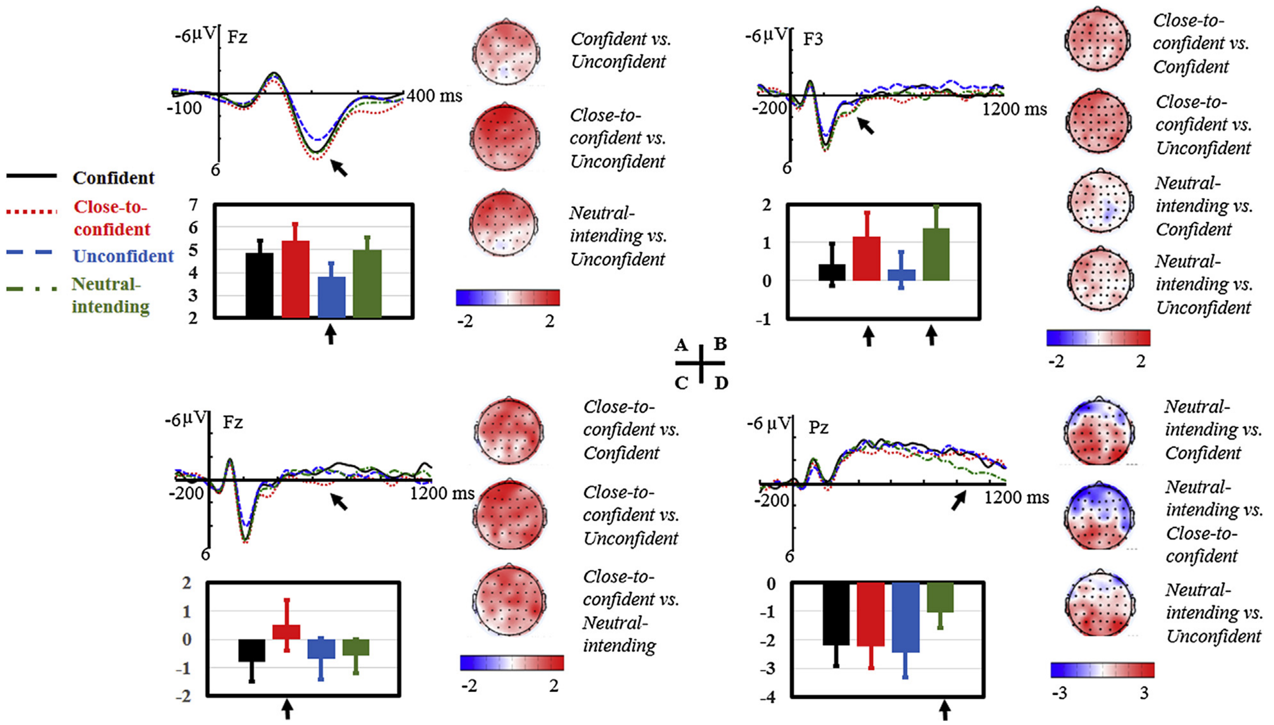

Fig. 3 – Topographic maps, grand average waveforms, and bar graphs of the amplitude of each condition for the significant ERP effects of vocal confidence expressions. For the topographic maps, difference waves are drawn for each ERP effect in reference to a specific baseline. For the bar graphs, both mean and standard deviation of absolute peak or mean amplitudes are shown. Fig. A demonstrates the topographic distribution of reduced P200 peak amplitude for the unconfident than the neutral-intending and all the other levels of confidence in the 183–243 msec window; the waveforms and bar graphs are drawn based on Fz as an exemplar. Unconfident level served as the baseline condition. Fig. B demonstrates the topographic distribution of larger positivity effect for the close-to-confident and the neutral-intending condition in the 330–500 msec; the waveforms and graphs are drawn based on F3 as an exemplar. The confident and unconfident levels each served as baselines for the close-to-confident and the neutral-intending condition. Fig. C demonstrates the topographic distribution of larger positivity effect for the close-to-confident condition than the neutral-intending and the other confidence levels in the 550–740 msec; the waveforms and graphs are drawn based on F3 as an exemplar. The confident, unconfident and neutral-intending levels each served as baselines for the close-to-confident condition. Fig. D demonstrates the topography of late, sustained positivity effect for the neutral-intending utterances than for the other confidence levels in the 980–1270 msec window. Waveforms and graphs are drawn on Pz as an exemplar. The confident, close-to-confident and unconfident levels each served as baselines for the neutral-intending condition.

maximum in the right-hemisphere in the female listener (Appendix E).

The model on the 980–1270 msec window revealed a significant two-way interaction between confidence and region and a significant three-way interaction between confidence, region and sex. The simpler model on individual region revealed a significant effect of confidence only in the posterior region in the female listener (Appendix F). Post-hoc comparisons revealed that neutral-intending expressions elicited a larger positivity effect than confident ($\beta = 1.90$, $t = 2.14$, $p = .02$), close-to-confident ($\beta = 2.06$, $t = 2.31$, $p = .02$), and unconfident ($\beta = 2.45$, $t = 2.76$, $p = .01$) expressions at this late processing stage (Fig. 3c). There was no difference between confident, close-to-confident, or unconfident expressions for this analysis.

Given the novelty of our investigation, and to provide converging evidence on the temporal data patterns that were obtained in the mean amplitude analysis, we also performed the LMEM on the factor covariance scores at 11 components extracted from PCA analysis (see also Dien, 2006; and Appendix G for detailed results).

The results strongly confirmed the reported findings of confidence-related ERP responses in the four time windows based on the mean amplitude analysis, derived from previous literature. Full details of how the PCA was conducted and complete statistical reporting of our effects using this approach are provided in Appendix G.

4. Discussion

This EEG study sheds light on a largely unexplored topic: how the neurocognitive system responds to vocal cues of expressed confidence during online speech comprehension. Based on participants' *perceived* meaning of different expression types, our data show that unconfident

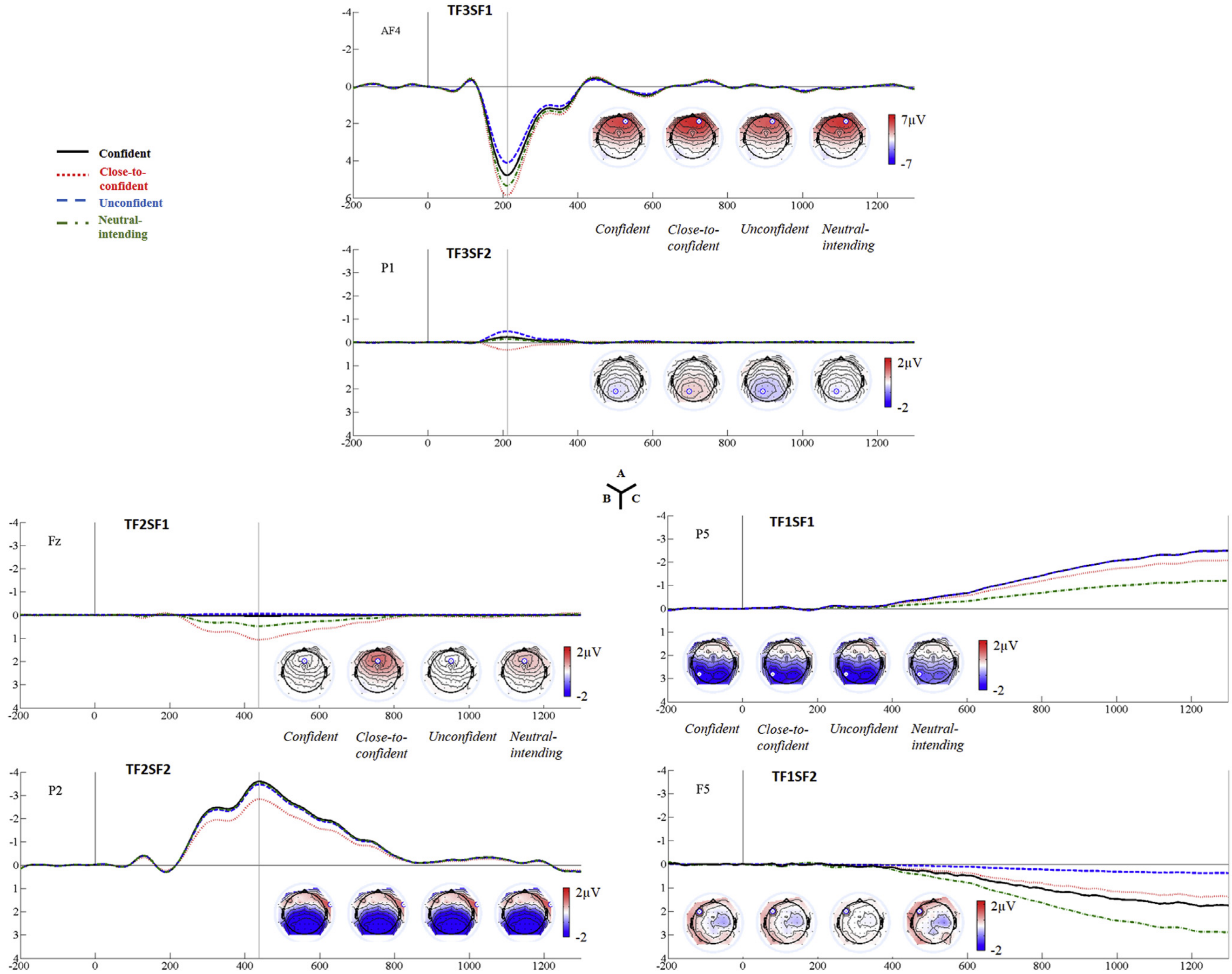

Fig. 4 – Waveforms and topographic distributions at the peak channels for the temporal-spatial factors which explained at least .5% of the total variance in the PCA analysis and corresponded to the ERP components (i.e., P200, LPC/P600, sustained positivity). Fig. A demonstrates the spatial factors of TF3 peaking at 212 msec; Fig. B demonstrates the spatial factors of TF2 peaking at 438 msec; Fig. C demonstrates the spatial factors of TF3 peaking at 1298 msec.

expressions elicited a smaller P200 than confident, close-to-confident, and neutral-intending expressions produced by four distinct speakers. Furthermore, vocal cues that reveal a slight lack of confidence (close-to-confident) or intend neutrality elicited a less negative-going response at centro-posterior region (and a more positive-going response at anterior region) in 330–500 msec, and close-to-confident expressions were treated distinctly from the other expressions in 550–740 msec time range. Finally, neutral-intending expressions elicited a unique posteriorly-distributed, more delayed, sustained positivity (980–1270 msec). These findings establish that listeners rapidly use vocal cues that signal speaker confidence during online speech communication, and provide an important new piece of information that extends multi-stage models of vocal expression processing.

4.1. *Early neural encoding of vocal confidence*

One of our primary goals was to determine how soon the brain responds to vocal cues that encode different levels of confidence in real-time speech processing. We predicted an effect of different confidence expressions on early ERP responses within the N1–P2 complex (Paulmann et al., 2013; Paulmann & Kotz, 2008; Schirmer & Kotz, 2006). The N1 is thought to vary as a function of *acoustic differences* of (emotional) vocal signals, reflecting attentional analysis of features such as frequency and intensity independent of discrete expressive meanings encoded by the stimuli (Schirmer & Kotz, 2006). The emotion literature does not produce a consistent finding of whether or not an N1 response is modulated by the emotional meaning (cf. Liu et al., 2012; Paulmann, Ott, Kotz, 2011) We found no effect of intended confidence meanings on early N1 response, which could be due to the relatively subtle differences in the early acoustic differentiation of different confidence expressions when compared to vocal emotional signals (as shown in Fig. 1). In contrast, we found a notably reduced P200 for unconfident statements than for statements expressed in a more confident voice or in a neutral manner. In studies of vocal emotion expression, the P200 has been associated with the initial deployment of attention to evaluate the significance of vocal signals, in a way that ensures preferential processing of emotional or motivational relevant stimuli ("initial emotional significance detection", see Kotz & Paulmann, 2011 for a review). Different P200 amplitudes are elicited by different types of vocal emotion expressions in speech (Alter et al., 2003; Jessen & Kotz, 2011; Paulmann et al., 2013; Schirmer et al., 2013) and in non-linguistic vocalizations (Charest et al., 2009; Sauter & Eimer, 2010). Although divergent findings still exist, the comparison between emotional vocal signals revealed greater amplitudes of P200 for motivationally *significant* (e.g., arousing) events (Jessen & Kotz, 2011; Paulmann et al., 2013).

Here, the reduced amplitude of P200 in response to only unconfident expressions may be associated with a first-pass evaluation of acoustic signals that are relevant for "tagging" that a speaker is confident, without further differentiation of cues that are relevant for detecting the extent of a speaker's "feeling of not knowing", or lack of confidence, at this early stage. Like angry or disgust vocal expressions that display increased P2 amplitude (Jessen & Kotz, 2011; Paulmann et al., 2013), it can be argued that signals expressing a relatively high degree of confidence—which was true of confident, neutral-intending, and close-to-confident expressions in the experiment—are motivationally salient in many respects; determining that a statement is likely true has critical social significance in person perception (Maslow et al., 1971; London et al., 1971; Pell, 2007), decision-making (London, Meldman, Lanckton, 1970; London et al., 1971), and various learning processes (Sabbagh & Baldwin, 2001; Zimmerman & Ringle, 1981). Thus, it is not surprising that attentional preference might be given to the analysis of vocal features relevant to the detection of speaker confidence, which has major implications on the believability or social credibility of the speaker, affecting future decisions and action tendencies (Pell, 2006; Scherer et al., 1973). It is interesting that this preference extended here to neutral-*intending* stimuli which were rated by listeners as expressing relatively high confidence in our task; in fact, the preferential processing of neutral stimuli at early processing stages has been reported elsewhere as well (Paulmann & Kotz, 2008a).

Acoustic mechanisms could also account for why confident statements received preferential attention processing over unconfident statements as indexed by the P2, i.e., relevance detection may be driven in part by acoustic-perceptual features that grossly differentiate "confident" from "unconfident" signals at the sensory level. Here, acoustic analyses revealed that unconfident statements tended to be higher in pitch and slower than all other expressions, whereas intensity differences did not clearly differentiate unconfident signals from the other three expression types (and mean pitch tended to rise only late in the utterance for unconfident expressions, as shown in Fig. 1). It remains possible that acoustic cues, especially speech rate or rhythmic cues, were rapidly employed by listeners to encode differences between confident and unconfident stimuli as reflected in the P2. However, it should be noted that Paulmann et al. (2013) compared ERP responses to pseudo-utterances conveying six basic emotions (anger, fear, happiness, sadness, disgust, pleasant surprise) and found that the P200 amplitude differed as a function of emotion type but *not* accordingly to specific acoustic features. For example, a disgust voice elicited a larger P2 amplitude than fear although they were similar in f0, loudness and speech rate; conversely, fearful and sad voices did not differ in P200 although the former had a higher f0 mean and range, and the latter had a much slower speech rate.

These findings put into question the extent to which P2 responses we observed between confident and unconfident stimuli reflect sensitivity to only acoustic processing mechanisms. Rather, this finding appears to extend previous studies on vocal emotion processing (e.g., Jessen & Kotz, 2011; Liu et al., 2012; Paulmann et al., 2013; Schirmer et al., 2013) by establishing that not only the relevance of a speaker's emotional state, but also the relevance of their *mental state* (such as the feeling of knowing), is roughly encoded by early neurocognitive processes occurring 200 msec after speech onset, as indexed by the P200 component. Interestingly, it should be noted that the absence of dissociation between confident and close-to-confident expressions in the P2 does not necessarily mean that these expressions were treated as perceptually equal in the vocal channel even at early processing stages; indeed, in our companion study, confident expressions elicited a larger P200 than both close-to-confident and unconfident stimuli when they were preceded by a congruent lexical phrase.

4.2. *Later cognitive elaboration of confidence expressions*

Although early neural responses did not differentiate between confident and close-to-confident vocal expressions, a less-negative going (positive-going) response was elicited by close-to-confident statements at a later processing stage. Upon initial inspection, one may attribute the modulation in the differential responses in the 300–550 msec window to a semantic N400 effect (Kotz & Paulmann, 2007; Strauß, Kotz, Obleser, 2013). Despite previous literature that points to an absence of N400 modulation in the comparison of different vocal expressions (e.g., Paulmann et al., 2013; Rigoulot et al., 2014), visual inspection of the grand average waveform reveals a negative-going wave maximized in the posterior regions following the N1–P2 complex. This negative response, as confirmed by the PCA (see TF2SF2 in Appendix G and Fig. 4b), seems to be driven by a negative temporal-spatial component peaking around 432 msec at the right posterior channel, similar to a typical N400 response (see Kutas & Federmeier, 2000; cf. Dien, Michelson, & Franklin, 2010). A reduction of N400 amplitude has been found on words following a sentence context that strongly constrains these word when compared to those following a weak-constraining context, indicating a reduced effort of semantic access or integration (Kutas & Federmeier, 2000; Strauß, Kotz, Obleser, 2013), or on words with their prosody mismatching the contextual representations (Kotz & Paulmann, 2007; Schirmer et al., 2002; 2005). In the present study, no vocal mismatch was present in the stimuli; nonetheless, the subtle acoustic difference between close-to-confident and confident, or between close-to-confident and unconfident, could hypothetically induce greater difficulty in accessing speaker's meaning in the intermediate condition as compared with other levels, in which case an enhanced N400 response should be observed for the close-to-confident expression. However, here we observed the opposite pattern, with *reduced* N400 response to the close-to-confident expression.

Alternatively, the reduced negative response elicited by the close-to-confident expression may be attributed to an enhanced positivity effect (LPC/P600) that temporally overlapped with the negative response (Molinaro & Carreiras, 2010; Vespignani et al., 2010). This account of our data is likely for two reasons: 1) the reduction of negative response to the close-to-confident expression extended into a later time window at 550–740 msec, which is typically affected by the LPC/P600 (Paulmann & Kotz, 2008b; Paulmann et al., 2013; Schirmer et al., 2002; Schirmer et al., 2006); and, 2) a positivity effect as early as 450 msec has been observed in studies comparing different vocal emotion expressions (Paulmann et al., 2013). The PCA analysis revealed a positive component peaking at Fz at 432 msec, and another positive component peaking at FP2 at 610 msec (Appendix G), which correspond to the 330–500 and 550–740 msec time windows. The temporal proximity and topographic similarity of these two components could reflect similar patterns of a positivity statistically split into two factors. It is clear that the close-to-confident expression elicited a similarly larger positive-going response than other expressions on both of these frontal components.

There could be two interpretations of this positivity effect that treat the positive component differently, as P3a or P3b (Comerchero & Polich, 1999; Picton, 1993). The P3a is an early positivity with a frontal distribution that is sensitive to the processing of novel stimuli regardless of the task relevance, whereas the P3b is a component with a more posterior distribution elicited by the processing of task-relevant stimuli; the P3a occurs as early as the N400 response (Kotz & Paulmann, 2007; Paulmann et al., 2013) while the latency of P3b depends on task demands (Molinaro & Carreiras, 2010; Schirmer et al., 2006; Vespignani et al., 2010).

In the study of vocal emotion, Kotz and Paulmann (2007) reported a larger right-anterior positivity effect (600–950 msec) on utterances violating the contextual expectancy of vocal emotion, irrespective of emotional valence or task relevance. In a related study, a stronger LPC effect (450–700 msec), broadly distributed, was reported for emotionally-inflected pseudo-utterances associated with different emotion types (happiness > pleasant surprise > anger > fear > sadness > disgust), with greater positivity for voices associated with high arousal overall (Paulmann et al., 2013). The modulation of the LPC was present regardless of whether the listener judged the speaker's emotion (relevant task) or reported their subjective feelings (irrelevant task). In the present study, the broadly-distributed enhanced positivity effect seems to be linked to the differential responses on the frontal components (see Appendix G), the topographic distribution of which is similar to P3a. Moreover, the perceived ambiguity of the close-to-confident expression (and of the neutral-intending expression to some extent) could trigger a task-irrelevant monitoring process (Kolk & Chwilla, 2007). However, since the number of trials after reassignment was equal between confident and close-to-confident expressions, we consider it unlikely that the modulation of ERP response between the two expressions is associated with the "novelty" detection of one of the expressions (i.e., P3a effect).

Rather, we tentatively attribute the LPC/P600 effect to an effect of P3b, which assumes the magnitude of the positivity effect was modulated by task relevance (Molinaro & Carreiras, 2010; Vespignani et al., 2010). Under current task conditions (speaker confidence judgment), rating the degree of expressed confidence may be closely related to difficulty in analyzing the vocal stimuli, with those exemplars requiring a more fine-grained, sustained analysis (i.e., close-to-confident) eliciting a larger positive response than the less-demanding levels (i.e., confident, unconfident). If so, the P3b effect should be associated with the level of difficulty/effort that listeners must exert when judging vocal expressions, with the positivity effect being larger for expressions that are most difficult to judge. Indeed, a correlation analysis revealed significant positive correlations between the magnitude of the positivity effect (e.g., the difference in the ERP response between close-to-confident and confident expressions) and differences in response time when participants rated the stimuli in each condition.[8] These findings suggest that the positivity here was task-relevant and may be associated with the P3b effect.

But why would the close-to-confident expressions elicit a larger P3b effect? The late positivity effect has been associated with an in-depth, continuous evaluation of the emotional or contextual significance of an event (Brouwer, Fitz, & Hoeks, 2012; Schirmer & Kotz, 2006). However, the initiation of such processes can be task-dependent (Schact & Sommer, 2009; Schirmer et al., 2006), with deployment of more attentional resources to the task-relevant stimuli. Indeed, there is evidence that the underlying neurocognitive processes of LPC/P600 and P200 effect seem to be inter-dependent; Schirmer et al. (2013) demonstrated a negative correlation between the P200 and the late positivity effect elicited by emotionally-inflected words: the larger the P200, the smaller the late positivity effect.

[8] Although we used a delayed response paradigm (participants responded only after they saw a response cue on the screen), mean response times from onset of the cue until onset of the response revealed differences between the reassigned levels of confidence, which were notably delayed after close-to-confident expressions (1071 msec) than for neutral-intending (1009 msec), confident (980 msec), and unconfident expressions (970 msec). The correlation between the positivity effect and the differential response time between close-to-confident and confident was significantly positive on right centro-parietal electrodes (in the 330–500 msec window: CP2: $r = .46$, $p = .01$; CP4: $r = .40$, $p = .03$; CP6: $r = .48$, $p = .01$; P4: $r = .42$, $p = .02$; POz: $r = .43$, $p = .02$; in the 550–740 msec window: T8: $r = .38$, $p = .04$; CP6: $r = .47$, $p = .01$). A similar correlation was obtained on the positivity effect between close-to-confident and unconfident expressions (in the 550–740 msec window: O1: $r = .38$, $p = .04$; O2: $r = .45$, $p = .01$), between neutral-intending and confident expressions (in the 330–500 msec window: T8: $r = .49$, $p = .01$; P8: $r = .47$, $p = .01$; AF7: $r = .36$, $p = .05$; F1: $r = .41$, $p = .03$; C2: $r = .37$, $p = .05$).

In the present study, the larger LPC/P600 was observed when speakers are close-to-confident (i.e., when their vocal cues convey a small source of doubt to the listener). This finding suggests that an enhanced and continuous evaluation was directed to these particular expressions under the current task demand; possibly, the more subtle acoustic determinants of expressing "some doubt" in speech required a more fine-grained analysis by the listener over time. This analysis was insufficient to influence initial attention and early salience decoding. Indeed, the acoustic analysis revealed that close-to-confident and confident expressions were undifferentiated in mean f0 and speech rate, but only in amplitude mean and variability (with close-to-confident sounding less intense and less variable in intensity than confident speech). Given the graded manner in which "full confidence" and "slight doubt" are encoded and decoded in the vocal channel (Jiang & Pell, 2014; Pell, 2007), it seems reasonable that close-to-confident expressions required sustained resources of analysis by the neurocognitive system in order to recognize the precise extent of the speaker's feeling of knowing. The same principle holds true for the neutral-intending expression. In the 330–500 msec window, the neutral-intending expression also elicited a larger positive response than the confident and unconfident expression. Neutral-intending statements share some acoustic proximity with the confident expression (e.g., mean pitch and mean intensity), which may also require some additional resources to analyze. The continuous evaluation process was corroborated by a correlation analysis between the magnitude of the P200 response and the magnitude of the late positive responses elicited by the close-to-confident expression (and neutral-intending expression) across the electrodes.[9]

In addition to this LPC/P600 effect, a qualitatively different positivity was observed for neutral-intending statements when compared to all three expressions of intended confidence: this positivity was more delayed, sustained, and associated with more than one spatial component. Neutral-intending statements, which were not intended to encode confidence, were judged less frequently as conveying confidence by listeners and showed marked acoustic distinctions from confidence expressions, with a low pitch, high intensity, and rapid speech rate (see Table 2 and Fig. 1). At the same time, when explicit attention is directed to vocal confidence, neutral-intending utterances tend to be perceived as relatively high in confidence, greater than both unconfident and close-to-confident expressions (Jiang & Pell, 2014). Based on past knowledge, a delayed sustained positivity like the one we observed often represents the listener's attempt to infer the goals or intentions of a speaker or third party, for example, when social partners speak or act in an expected way (Jiang et al., 2013; Van der Cruyssen et al., 2009; Van Overwalle, Van den Eede, Baetens, Vandekerckhove, 2009). A late, sustained positivity was found on the honorific form of the second-person pronoun which was used in an utterance spoken by a higher-status person (e.g., a professor) to a lower-status addressee (e.g., a student), promoting an ironic interpretation (Jiang et al., 2013). A long-sustaining positivity was also witnessed for words that violated implied characteristics of a protagonist described in a context, suggesting that this response reflects a demanding inferential process of deriving contextually appropriate interpretations (Van der Cruyssen et al., 2009; Van Overwalle et al., 2009). Similar positivity effects were also found on words mismatching the conversational context associated with ironic, sarcastic or sincere interpretations (Regel, Gunter, Coulson, 2010; Regel et al., 2011; Regel et al., 2014; Rigoulot et al., 2014; Spotorno et al., 2013). Based on these results, we speculate that the positivity evoked by neutral-intending utterances is associated with ongoing inferential demands associated with the "mismatch" between task goals (*rate the speaker's confidence*) and the evolving perception that distinct acoustic characteristics of the neutral-intending stimuli (e.g., lower mean pitch, faster speech rate) were not *meant* to communicate confidence to the listener.

4.3. *Implications for neurocognitive models of vocal processing*

Our report highlighting that expressed confidence is processed at multiple neural processing stages provides new fodder for broader proposals on the nature of speech comprehension, and how the neurocognitive system establishes a mental representation from a continuous and changing vocal signal that encodes multiple simultaneous meanings (Kotz & Schwartze, 2010; Schwartze & Kotz, 2013). According to three-stage models of vocal emotion processing (e.g., Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), a first stage involves sensory analysis of the vocal signal and extraction of emotion-relevant acoustic cues, which is quickly followed by a second stage where the emotional significance is derived and evaluated from emerging acoustic patterns with different values. In the last stage, the emotional significance is applied to high-order cognition in various ways (Kotz & Paulmann, 2011; Mitchell, 2013; Mitchell & Ross, 2013; Schirmer & Kotz, 2006).

Electrophysiological studies have demonstrated that the first and the second stages of processing a vocal signal can be clearly identified in the early (N1, Liu et al., 2012) and late ERP components (e.g. P200, LPC/P600, late negativity, Paulmann & Kotz, 2008a; Paulmann, Ott, & Kotz, 2011), although stages two and three were suggested to be differentiated in the late ERP components (Kotz & Paulmann, 2011). Despite the fact that both vocal expressions of confidence and vocal emotions are encoded by an evolving constellation of acoustic parameters over the course of the speech, and the fact that they may share underlying neural systems (Monetta et al., 2008), an important distinction is that expressed confidence encodes *relational* meanings relevant to the speaker – hearer in the interpersonal context in which they appear, marking the reliability of

[9] A moderate but significant negative correlation was shown between the P200 magnitude on the frontal electrode and the late positive response on the left posterior electrode (for coupling between P200 and 330–500 msec positivity: FCz – PO9: Pearson $r = -.39$, $p = .03$; F7 – PO9: $r = -.37$, $p = .05$; for coupling between P200 and 550–740 msec positivity: FCz – FT9: $r = -.36$, $p = .05$; FCz – T7: $r = -.37$, $p = .04$). Similar negative correlations were also found between the P200 and the late positivity elicited by neutral-intending expressions (for coupling between P200 and 330–500 msec positivity: F6 – PO9: $r = -.36$, $p = .05$; for coupling between P200 and 550–740 msec positivity: Fz – TP9: $r = -.40$, $p = .03$; F3 – PO9: $r = -.36$, $p = .05$; Fz – PO9: $r = -.37$, $p = .05$; FC2 – PO9: $r = -.37$, $p = .05$; F1 – PO9: $r = -.39$, $p = .04$; F2 – PO9: $r = -.39$, $p = .04$; F6 – PO9: $r = -.38$, $p = .04$).

the speaker or correctness of the assertive speech (Caffi & Janney, 1994).

By demonstrating the temporal neural dynamics of processing different shades of speaker confidence, we show that multi-stage models of vocal expression processing can also characterize how listeners uncover a speaker's *mental state* (e.g., confidence) during utterance comprehension. In the first stages, vocal cues relevant for detecting confidence are presumably identified and extracted by the listener, and these are used to dissociate and "tag" cues indicative of (high) confidence as early as 200 msec into speech (i.e., in the stage of P200 effect). Stimulus relevance at this early stage may be inherently due to acoustic properties, or at times biased by the listener's expectation of a speaker's feeling of knowing when other cues, such as verbal information, are simultaneously available (Jiang & Pell, 2014). In the late stages (stage three of most multi-stage models), different cognitive processes are applied to elaborate and "fine-tune" inferences about speaker confidence in the social context. This includes a fine-grained (re)-evaluation of ongoing acoustic features that provide increasingly sensitive indications of the speaker's feeling of knowing (i.e., the speaker is "close-to-confident" but has some doubt about what is being said), which could be influenced to a large extent by perceived loudness according to our data. This process seems to occur within the 330–740 msec time window, as reflected by an LPC/P600 effect, at least when vocal signals encode the speaker's mental state of knowing or believing in relation to the content of the utterance (see also Rigoulot et al., 2014).

At later stages, an initial representation of the speaker's confidence based on their vocal expression would also be mapped onto higher-level details about the speaker to infer the nature of speech acts, speaker intentions, and underlying motivations as reflected by a tertiary, sustained positivity effect. Based on our neutral-intending condition, it appears that this tertiary positivity effect is increased in cases where vocal signals elicit resource-demanding inferences to satisfy task requirements and/or in the context of conflicting cues. To test this latter suggestion, new studies could explore whether the sustained positivity disappears for neutral-intending statements in a similar task where listeners do not explicitly focus on speaker confidence.

4.4. *Sex differences*

While not the main focus of our paper, it is noteworthy that sex differences characterized both behavioral judgments of vocal confidence and ERP responses to these stimuli in late time windows. Behaviorally, female listeners were better at judging subtle differences between confident, neutral-intending and close-to-confident expressions; this advantage is consistent with the claim that socio-emotional cues in vocal expressions are assigned greater significance by females versus males (Schirmer & Kotz, 2006; Schirmer, Simpson, & Escoffier, 2007) and that females demonstrate stronger interpersonal communication skills (Allison, Baron-Cohen, Wheelwright, Stone, & Muncer, 2011). Electrophysiologically, we noted slight differences in the laterality of the late positivity effect elicited by close-to-confident expressions between 550 and 740 msec for female and male listeners; moreover, only females exhibited a delayed sustained positivity in the 980–1270 msec time window, which uniquely differentiated neutral-intending and confidence-intending expressions.

Sex-related variation in the lateralization of ERP effects has been previously reported in the perception of utterances with incongruent versus congruent emotional prosody (Kotz & Paulmann, 2007), with males exhibiting a bilateral positivity and females exhibiting a right-lateralized positivity to the prosodic incongruence. Moreover, ERP evidence implies that females use socially-relevant information from the vocal channel in an earlier and more automatic fashion than males; notably, females showed an N400 on a target word when its semantics mismatched the emotional valence of vocal cues 200 msec prior to the target, regardless of whether they were asked to guide their attention to the semantics or not, whereas males only showed a task-relevant mismatch effect (Schirmer et al., 2002; 2005). In another study, a mismatch between the utterance content and the voice-implied speaker's identity (e.g. An adult saying, *I cannot sleep without my teddy bear in my arms*) elicited a larger N400 response in female but not male listeners when their attention was not tuned to the speaker (Van den Brink et al., 2012). Here, it can be argued that the sustained positivity we observed was triggered by a task-irrelevant pragmatic inference based on the perceived acoustic distinctions of the neutral-intending expressions, an effect that appears to be female-specific (Allison et al., 2011). Our data strongly encourage further studies of the neural mechanisms that appear to confer a female advantage in vocal communication and in the derivation of certain social inferences about a speaker's interpersonal stance and/or mental state.

4.5. *Limitations and conclusion*

Our results establish that a listener's brain is rapidly attuned to vocal cues that signal one's feeling of knowing and can differentiate subtle vocal variations as a function of the perceived *degree* of knowing. Like other ERP studies on vocal perception, the vocal expressions presented in our study were produced by actors (i.e., simulated expressions) and elicited in simple utterances to carefully control the psycholinguistic properties of the stimuli, somehow limiting the ecological validity of the data. Brain responses recorded during interactive speech, for example, a dialog in which vocal cues such as pauses and hesitations between turns are measured as natural indices of certain confidence levels, would be highly innovative and informative as the field moves forward. Examining how non-verbal (facial expressions, e.g., *eye gaze*) and verbal cues (linguistic phrase, e.g., *I think*) interact with the vocal signal (Swerts & Krahmer, 2005; Walker, 1977) to communicate confidence, and how these multimodal cues affect underlying brain responses associated with a feeling of another's knowing, would also be highly informative.

As the surface distribution of ERP effects reflects the summation of multiple underlying brain sources, it will also be necessary to conduct studies using functional MRI to pinpoint the neural networks underlying each stage of vocal confidence processing. Based on recent frameworks (Schwartze & Kotz, 2013), these studies may show that distinct neural pathways underlie the decoding of different (formal and temporal) aspects of vocal confidence expressions. In addition to a "formal" pathway that targets temporal cortex and aims to establish sensory-memory representations of vocal expressions, a “temporal” pathway representation targeting frontal cortex via rapid cerebellar transmission may prepare specific regions for perceptual integration from memory representation, while engaging basal ganglia and striato-thalamo-cortical circuits to represent temporal relations (see also Monetta et al., 2008). This dual-pathway architecture would allow the neuro-cognitive system to detect regularity in vocal expressions, including expressions of confidence, and generate temporally specific predictions with increasing sensitivity during ongoing speech processing and person perception.

Acknowledgment

We are grateful for generous funding from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery grant, RGPIN/203708-2011) and a James McGill Professor award to M.D. Pell. We thank Mr. François Anderson-de Serres and Mr. Shawn Wen for their precious time in running the tests and Mr. Caleb Harrison for his patience in editing auditory materials. We thank Miss Sonia Kroll, Miss Lorraine Chuen, and Miss Nabitha Kanagaratnam for their efforts in recording and validating the vocal stimuli. We thank Mr. Yi'nan Cao and Mr. David Sears for discussions on linear mixed effects models. We also thank Mary Giffen for her precious help in manuscript preparation.

Appendix A

Table A1 – The spoken sentence exemplars (with the communicative type listed in brackets) selected from the perceptual study for one male speaker (SK) as critical statements in the experiment that were each produced in the confident, close-to-confident, the unconfident and the neutral-intending voice.

1. She has access to the building. (fact)
2. He turned left at the lights. (fact)
3. She left the key by the fridge. (fact)
4. There's diabetes in his family. (fact)
5. He misjudged the cost of the project. (fact)
6. They have four children. (fact)
7. They don't drink alcohol. (fact)
8. The store will go bankrupt. (fact)
9. I'll finish the essay tonight. (intention)
10. We'll arrive during the carnival. (intention)
11. I'll be picked for the team. (intention)
12. I'll convince the judge. (intention)
13. We'll complete our assignment. (intention)
14. She'll pick you up after work. (intention)
15. They'll visit Japan during the tour. (intention)
16. She'll sell her car soon. (intention)
17. You skate better than your brother. (judgment)
18. You won't miss the connecting flight. (judgment)
19. He's the right person to do this. (judgment)
20. He likes this author's writing. (judgment)
21. She'll do a good job. (judgment)
22. He has a good sense of humor. (judgment)
23. This medicine will help you. (judgment)
24. The table is worth the money. (judgment)

Table A2 – Means (Standard Deviations) of the Behavioral Responses to the elicited items for each level of confidence in the perceptual study

Intended level of confidence	Percentage of “Yes” response in judging whether the speaker conveyed some level of confidence (%)[b]	Speaker confidence rating (1–5)[c]						
		Speaker 1 male (CM)	Speaker 2 female (CS)[a]	Speaker 3 female (MW)	Speaker 4 male (NH)[a]	Speaker 5 male (SK)[a]	Speaker 6 female (ST)[a]	Total
Confident	87.7 (13.1)	4.24 (.29)	4.21 (.35)	4.51 (.22)	4.26 (.30)	4.18 (.32)	4.36 (.26)	4.30 (.31)
Close-to-confident	80.1 (15.1)	4.04 (.33)	3.22 (.56)	3.95 (.53)	4.06 (.39)	3.97 (.36)	3.39 (.77)	3.76 (.61)
Unconfident	81.0 (11.3)	2.83 (.59)	1.80 (.37)	2.59 (.50)	2.99 (.69)	1.61 (.42)	2.33 (.58)	2.35 (.73)
Neutral	59.8 (10.6)	4.16 (.30)	3.53 (.48)	4.04 (.33)	4.06 (.34)	3.99 (.30)	3.94 (.37)	3.97 (.41)

[a] The speaker that was selected for the EEG study.
[b] The mean percentage was calculated for each item as the number of “Yes” response divided by total number of response ($n = 10$); the standard deviation reveals inter-item variation.
[c] The mean speaker confidence rating was calculated as the mean rating of each item that has been judged as conveying some level of confidence ($n = 10$); the standard deviation reveals inter-item variation.

Perceptual validation of vocal expressions in the elicitation study

Table A2 displays the results of the perceptual ratings on 3624 vocal expressions based on 60 native English listeners. Two LMEMs were built separately on task 1 and task 2 measures. The first LMEM was built on the mean percentage of "yes" responses (calculated for each item), including level of confidence as the fixed effect, and by-item intercepts as random effects. The LMEM revealed a significant effect of level of confidence, $F(3, 2714) = 336.84$, $p < .0001$. The post-hoc comparisons revealed that the neutral-intending stimuli were less judged as conveying some level of confidence than statements intending some level of confidence (neutral *vs* confident: $\beta = -.18$, $t = -32.95$, $p < .0001$; neutral *vs* close-to-confident: $\beta = -.11$, $t = 20.65$, $p < .0001$; neutral *vs* unconfident: $\beta = -.11$, $t = -20.6$, $p < .0001$), while no difference was observed between the three confidence expressions, $ps > .1643$. The LMEM was also built on the confidence rating, including level of confidence as the fixed effect, and by-rater intercept and by-rater and by-item slopes for level of confidence as random effects. The result revealed a significant effect of level of confidence, $F(3, 33433) = 596.04$, $p < .0001$. The confidence rating was highest in the confident expression (confident *vs* neutral: $\beta = .12$, $t = 2.07$, $p = .0347$; confident *vs* close-to-confident: $\beta = .21$, $t = 3.69$, $p = .0002$ confident *vs* unconfident: $\beta = 1.41$, $t = 24.56$, $p < .0001$), followed by the neutral-intending expression (neutral *vs* close-to-confident: $\beta = .12$, $t = 2.10$, $p = .0361$; neutral *vs* unconfident: $\beta = 1.22$, $t = 21.76$, $p < .0001$), the close-to-confident expression (close-to-confident *vs* unconfident $\beta = 1.20$, $t = 20.59$, $p < .0001$), and the unconfident expression. The LMEM also revealed a significant interaction between speaker and level of confidence, $F(15, 33433) = 161.09$, $p < .0001$, suggesting that the perceived confidence between high and low intended level of confidence varied as a function of the speaker (see also Table B1). The perceptual rating revealed that the neutral-intending statement can be distinguished from the statements intending confidence meanings; different intended levels of confidence by the actor can be distinguished by an independent group of listener, with some actors better distinguished than the other.

Appendix B

Table B1 – ANOVA Table for the Online Behavioral Judgments, for LMEMs, by Sex.

Coefficient	*df*	*SS*	*MS*	*F*	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
				Male		
Confidence	3	106.84	35.61	50.17	< .0001	< .0001
Comparisons between level of confidence	Estimate		*t*		*p* (Lower Bound)	*p* (Upper Bound)
Close-to-confident vs. confident	-.40		-2.53		.01	.01
Unconfident vs. confident	-1.72		-10.98		<.0001	< .0001
Neutral-intending vs. confident	-.16		-1.01		.31	.31
Unconfident vs. close-to-confident	-1.33		-8.45		<.0001	<.0001
Neutral-intending vs. close-to-confident	.24		1.52		.13	.13
Neutral-intending vs. unconfident	1.56		9.97		<.0001	<.0001
				Female		
Confidence	3	221.61	73.87	134.84	< .0001	< .0001
Close-to-confident vs. confident	-.53		-4.04		.0001	.0001
Unconfident vs. confident	-2.37		-18.07		<.0001	<.0001
Neutral-intending vs. confident	-.25		-1.92		.05	.05
Unconfident vs. close-to-confident	-1.84		-14.03		<.0001	<.0001
Neutral-intending vs. close-to-confident	.28		2.12		.03	.03
Neutral-intending vs. unconfident	2.12		16.15		<.0001	<.0001

Note: confidence as the fixed effect, and by-item and by-participant intercepts, and by-item and by-participant slopes for confidence as random effects.
[a]Anterior: Denominator: $df = 2550$.[b]Denominator: $df = 2864$.

Appendix C

Table C1 – ANOVA Table for the P200, for the LMEM.

Coefficient	*df*	*SS*	*MS*	*F*	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
Confidence	3	34.71	11.57	3.45	.02	.02
Hemisphere	2	12.66	6.33	1.89	.15	.15
Region	2	748.33	374.17	111.73	<.0001	<.0001
Sex	1	.23	.23	.07	.79	.79
Confidence: Hemisphere	6	20.79	3.47	1.03	.40	.40
Confidence: Region	6	126.21	21.04	6.28	<.0001	<.0001
Hemisphere: Region	4	39.66	9.91	2.96	.02	.02
Confidence: Sex	3	11.43	3.81	1.53	.12	.11
Hemisphere: Sex	2	24.07	12.04	3.59	.03	.03
Region: Sex	2	153.84	76.92	22.97	<.0001	<.0001
Confidence: Hemisphere: Region	12	16.06	1.34	.40	.96	.96
Confidence: Hemisphere: Sex	6	25.91	4.32	1.29	.26	.26
Confidence: Region: Sex	6	22.89	3.81	1.14	.34	.34
Hemisphere: Region: Sex	4	33.69	8.42	2.52	.04	.04
Confidence: Hemisphere: Region: Sex	12	17.34	1.45	.43	.95	.95

Comparisons between level of confidence	Estimate	*t*	*p* (Lower Bound)[b]	*p* (Upper Bound)[c]
Close-to-confident *vs* confident	.43	1.15	.25	.25
Unconfident *vs* confident	-.97	−2.10	.01	.01
Neutral-intending *vs* confident	.47	1.58	.17	.17
Unconfident *vs* close-to-confident	−1.11	−2.96	.003	.003
Neutral-intending *vs* close-to-confident	.24	.63	.53	.53
Neutral-intending *vs* unconfident	−1.35	−3.59	.0003	.0003

Note: confidence, hemisphere, region and listener sex as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Denominator: $df = 6708$.
[b] Denominator: $df = 6888$.

Table C2 – ANOVA Table for the P200, for LMEMs, by Region.

Coefficient	*df*	*SS*	*MS*	*F*	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
					Anterior	
Confidence	3	55.49	18.50	6.04	.0004	.0004
Hemisphere	2	6.09	3.05	.99	.37	.37
Sex	1	.59	.59	.19	.66	.66
Confidence: Hemisphere	6	10.08	1.68	.55	.77	.7
Confidence: Sex	3	53.59	17.86	5.83	.16	.16
Hemisphere: Sex	2	5.41	2.70	.88	.41	.41
Confidence: Hemisphere: Sex	6	3.14	.52	.17	.99	.99
					Central	
Confidence	3	19.83	6.61	1.92	.12	.12
Hemisphere	2	15.12	7.56	2.20	.11	.11
Sex	1	1.40	1.40	.41	.52	.52
Confidence: Hemisphere	6	18.94	3.16	.92	.48	.48
Confidence: Sex	3	40.79	13.60	3.95	.29	.29
Hemisphere: Sex	2	23.70	11.85	3.45	.03	.03
Confidence: Hemisphere: Sex	6	20.56	3.43	1.00	.42	.42
					Posterior	
Confidence	3	10.81	3.60	1.42	.24	.24
Hemisphere	2	9.94	4.97	1.95	.14	.14
Sex	1	2.59	2.59	1.02	.31	.31
Confidence: Hemisphere	6	6.30	1.05	.41	.87	.87
Confidence: Sex	3	60.12	20.04	7.88	.10	.10
Hemisphere: Sex	2	31.63	15.81	6.22	.002	.002
Confidence: Hemisphere: Sex	6	19.36	3.23	1.27	.27	.27

Note: confidence, hemisphere and listener sex as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Anterior Denominator: $df = 2481$; Central/Posterior Denominator: $df = 2001$.
[b] Anterior Denominator: $df = 2616$; Central/Posterior Denominator: $df = 2136$.

Appendix D

Table D1 – ANOVA Table for the Effect in 330–500 msec Window, for the LMEM.

Coefficient	*df*	*SS*	*MS*	*F*	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
Confidence	3	41.07	13.69	3.61	.01	.01
Hemisphere	2	471.79	235.89	45.13	<.0001	<.0001
Region	2	446.02	223.01	42.66	<.0001	<.0001
Sex	1	7.76	7.76	1.48	.22	.22
Confidence: Hemisphere	6	27.15	4.52	.87	.52	.52
Confidence: Region	6	16.70	2.78	.53	.78	.78
Hemisphere: Region	4	253.26	63.32	12.11	<.0001	<.0001
Confidence: Sex	3	21.72	7.24	1.22	.41	.41
Hemisphere: Sex	2	14.59	7.30	1.40	.25	.25
Region: Sex	2	392.51	196.26	37.55	<.0001	<.0001
Confidence: Hemisphere: Region	12	20.84	1.74	.33	.98	.98
Confidence: Hemisphere: Sex	6	30.24	5.04	.96	.45	.45
Confidence: Region: Sex	6	42.84	7.14	1.37	.22	.22
Hemisphere: Region: Sex	4	60.54	15.14	2.90	.02	.02
Confidence: Hemisphere: Region: Sex	12	18.58	1.55	.30	.99	.99

Comparisons between level of confidence	Estimate	*t*	*p* (Lower Bound)[b]	*p* (Upper Bound)[c]
Close-to-confident *vs* confident	.72	2.07	.04	.04
Unconfident *vs* confident	−.01	−.01	.98	.98
Neutral-intending *vs* confident	.70	1.99	.05	.05
Unconfident *vs* close-to-confident	−.72	−2.08	.04	.04
Neutral-intending *vs* close-to-confident	−.03	−.09	.93	.93
Neutral-intending *vs* unconfident	.70	1.99	.05	.05

Note: confidence, hemisphere and region as fixed effects, and by-participant intercept and slopes for all fixed effects as random effects.
[a] Denominator: $df = 6708$.
[b] Denominator: $df = 6888$.

Appendix E

Table E1 – ANOVA Table for the Late Positivity in 550–740 msec Window, for the LMEM.

Coefficient	*df*	*SS*	*MS*	*F*	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
Confidence	3	737.55	245.85	31.82	<.0001	<.0001
Hemisphere	2	563.61	281.81	36.48	<.0001	<.0001
Region	2	344.67	172.34	22.31	<.0001	<.0001
Sex	1	6.09	6.09	.79	.37	.37
Confidence: Hemisphere	6	42.57	7.09	.92	.48	.48
Confidence: Region	6	59.39	9.90	1.28	.26	.26
Hemisphere: Region	4	109.26	27.32	3.54	.01	.01
Confidence: Sex	3	39.02	13.01	1.68	.17	.17
Hemisphere: Sex	2	117.28	58.64	7.59	.0005	.0005
Region: Sex	2	74.72	37.36	4.84	.01	.01
Confidence: Hemisphere: Region	12	35.52	2.96	.38	.97	.97
Confidence: Hemisphere: Sex	6	99.00	16.50	2.14	.05	.05
Confidence: Region: Sex	6	29.62	4.94	.64	.70	.70
Hemisphere: Region: Sex	4	67.40	16.85	2.18	.07	.07
Confidence: Hemisphere: Region: Sex	12	18.13	1.51	.20	.99	.99

Comparisons between level of confidence	Estimate	*t*	*p* (Lower Bound)[b]	*p* (Upper Bound)[c]
Close-to-confident *vs* confident	.77	2.14	.03	.03
Unconfident *vs* confident	−.04	−.11	.91	.91
Neutral-intending *vs* confident	.22	.78	.44	.44
Unconfident *vs* close-to-confident	−.81	−2.25	.02	.02
Neutral-intending *vs* close-to-confident	−.74	−1.96	.05	.05
Neutral-intending *vs* unconfident	.27	.90	.39	.39

Note: confidence, hemisphere, region and listener sex as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Denominator: $df = 6708$.
[b] Denominator: $df = 6888$.

Table E2 – ANOVA Table for the Late Positivity in the 500–740 msec Window, for LMEMs, by Sex and Hemisphere.

Coefficient	*df*	SS	MS	F	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
				Male, Left		
Confidence	3	182.49	60.83	7.79	<.0001	<.0001
Region	2	113.29	56.65	7.26	.0007	.0007
Confidence: Region	6	19.35	3.22	.41	.87	.87
				Male, Medial		
Confidence	3	176.32	58.77	6.69	.0002	.0002
Region	2	55.36	27.68	3.15	.04	.04
Confidence: Region	6	5.35	.89	.10	.99	.99
				Male, Right		
Confidence	3	158.83	52.94	6.95	.0001	.0001
Region	2	117.74	58.87	7.73	.0005	.0005
Confidence: Region	6	32.07	5.35	.70	.65	.65
				Female, Left		
Confidence	3	105.33	35.11	5.33	.001	.001
Region	2	112.42	56.21	8.53	.0002	.0002
Confidence: Region	6	24.73	4.12	.63	.71	.71
				Female, Medial		
Confidence	3	123.35	41.12	6.47	.0002	.0002
Region	2	111.19	55.59	8.74	.0002	.0002
Confidence: Region	6	36.15	6.02	.95	.46	.46
				Female, Right		
Confidence	3	171.80	57.27	8.87	<.0001	<.0001
Region	2	119.45	59.73	9.25	.0001	.0001
Confidence: Region	6	25.02	4.17	.65	.69	.69

Note: confidence and region as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Denominator: $df = 1083$.
[b] Denominator: $df = 1188$.

Appendix F

Table F1 – ANOVA Table for the More Delayed, Sustained Positivity in 980–1270 msec Window, for the LMEM.

Coefficient	*df*	SS	MS	F	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
Confidence	3	21.52	7.17	.73	.53	.53
Hemisphere	2	501.54	250.77	25.52	<.0001	<.0001
Region	2	112.95	56.47	5.75	.0032	.0032
Sex	1	19.96	19.96	2.03	.15	.15
Confidence: Hemisphere	6	46.43	7.74	.79	.58	.58
Confidence: Region	6	594.19	99.03	10.08	<.0001	<.0001
Hemisphere: Region	4	55.83	13.96	1.42	.22	.22
Confidence: Sex	3	235.44	78.48	7.99	<.0001	<.0001
Hemisphere: Sex	2	178.18	89.09	9.06	.0001	.0001
Region: Sex	2	188.10	94.05	9.57	.0001	.0001
Confidence: Hemisphere: Region	12	65.79	5.48	.56	.88	.88
Confidence: Hemisphere: Sex	6	100.25	16.71	1.70	.12	.12
Confidence: Region: Sex	6	133.31	22.22	2.26	.04	.04
Hemisphere: Region: Sex	4	45.63	11.41	1.16	.33	.33
Confidence: Hemisphere: Region: Sex	12	43.74	3.65	.37	.97	.97

Note: confidence, hemisphere, region and listener sex as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Denominator: $df = 6708$.
[b] Denominator: $df = 6888$.

Table F2 – ANOVA Table for the More Delayed, Sustained Positivity in 980–1270 msec Window, for the LMEM, by Region.

Coefficient	*df*	SS	Mean Sq	F	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
				Anterior		
Confidence	3	16.85	5.62	.60	.61	.61
Hemisphere	2	204.48	102.24	10.98	<.0001	<.0001
Sex	1	11.98	11.98	1.29	.26	.26
Confidence: Hemisphere	6	29.97	5.00	.54	.78	.78

(continued on next page)

Table F2 – (continued)

Coefficient	df	SS	Mean Sq	F	p (Lower Bound)[a]	p (Upper Bound)[b]
Confidence: Sex	3	102.68	34.23	3.68	.01	.01
Hemisphere: Sex	2	25.74	12.87	1.38	.25	.25
Confidence: Hemisphere: Sex	6	75.76	12.63	1.36	.23	.23
				Central		
Confidence	3	17.47	5.82	.59	.62	.62
Hemisphere	2	374.58	187.29	19.03	<.0001	<.0001
Sex	1	31.53	31.53	3.20	.07	.07
Confidence: Hemisphere	6	41.60	6.93	.70	.65	.65
Confidence: Sex	3	68.41	22.80	2.32	.07	.07
Hemisphere: Sex	2	99.27	49.63	5.04	.01	.01
Confidence: Hemisphere: Sex	6	57.95	9.66	.98	.44	.44
				Posterior		
Confidence	3	62.51	20.84	2.66	.05	.05
Hemisphere	2	361.01	180.50	23.08	<.0001	<.0001
Sex	1	.79	.79	.10	.75	.75
Confidence: Hemisphere	6	34.95	5.82	.74	.61	.61
Confidence: Sex	3	164.92	54.97	7.03	.0001	.0001
Hemisphere: Sex	2	102.45	51.23	6.55	.002	.002
Confidence: Hemisphere: Sex	6	43.03	7.17	.92	.48	.48

Note: confidence, hemisphere and listener sex as fixed effects, and by-participant intercept and slopes for all fixed effects as random effects.
[a] Anterior Denominator: $df = 2481$; Central/Posterior Denominator: $df = 2001$.
[b] Anterior Denominator: $df = 2616$; Central/Posterior Denominator: $df = 2136$.

Table F3 – ANOVA Table for the Effect in the 980–1270 msec Window, for LMEMs, by Sex and Region.

Coefficient	df	SS	MS	F	p (Lower Bound)[a]	p (Upper Bound)[b]
				Male, Anterior		
Confidence	3	10.25	3.42	.53	.66	.66
Hemisphere	2	74.40	37.20	5.82	.003	.003
Confidence: Hemisphere	6	86.11	14.35	2.24	.04	.04
				Male, Central		
Confidence	3	3.41	1.14	.21	.89	.89
Hemisphere	2	44.29	22.15	4.07	.02	.02
Confidence: Hemisphere	6	20.40	3.40	.63	.71	.71
				Male, Posterior		
Confidence	3	7.73	2.58	.43	.73	.73
Hemisphere	2	33.99	17.00	2.82	.06	.06
Confidence: Hemisphere	6	26.09	4.35	.72	.63	.63
				Female, Anterior		
Confidence	3	.81	.27	.04	.99	.99
Hemisphere	2	77.49	38.74	5.42	.005	.005
Confidence: Hemisphere	6	19.63	3.27	.46	.84	.84
				Female, Central		
Confidence	3	10.67	3.56	.49	.69	.69
Hemisphere	2	190.29	95.15	13.10	<.0001	<.0001
Confidence: Hemisphere	6	79.15	13.19	1.82	.09	.09
				Female, Posterior		
Confidence	3	33.66	11.22	2.54	.05	.05
Hemisphere	2	143.71	71.86	16.29	<.0001	<.0001
Confidence: Hemisphere	6	51.88	8.65	1.96	.07	.07

Note: confidence and region as fixed effects, and by-participant slopes for all fixed effects as random effects.
[a] Anterior: Denominator: $df = 1203$; Central/Posterior: Denominator: $df = 963$.
[b] Denominator: $df = 1308$; Central/Posterior: Denominator: $df = 1068$.

Appendix G

Principal Component Analysis

Method

In order to provide converging evidence on the temporal data patterns that were obtained in the mean amplitude analysis, a two-step principal component analysis (temporal-spatial PCA) was performed to decompose the time/space-locked correlational structure of the ERP amplitudes (ERP PCA Toolkit, Version 2.33, Dien, 2010; 2012). The PCA provides a data-driven approach that is free of the theoretical preconceptions that are implied by standard ERP component labeling (e.g., P200, N400, LPC/P600). The PCA is aimed at confirming what independent spatial component the experimental factor reveals its effect within a certain temporal window and if the component response is positive or negative. The temporal and spatial variances entered the PCA by including the sampling points of the entire ERP segment (200 msec before the onset and 1300 msec after the onset of the vocal stimulus) on all EEG channels. Individual variance was taken account by including averaged ERPs for each subject into PCA. The expression types entered the PCA as a condition variance. A temporal PCA was first performed on the item-averaged data. A Parallel Scree Test was conducted to determine the retained factors by comparing the size of the unrotated factors of the experimental data against that of a same-sized random dataset (Horn, 1965). A sequential spatial PCA was performed on each temporal factor which passed the Parallel Test. Oblique rotations were performed to achieve the largest representation of a factor as one ERP component in the temporal (Promax Rotation) and the spatial PCA (Infomax Rotation; Dien, 2012). Factor loadings were rescaled to microvolts by converting them into covariance loadings (Dien, 2006). The covariance loadings on the peak channel at peak time point for a temporal-spatial factor which explained more than the minimum threshold of .5% total variance were modeled using LMEM (Dien et al., 2010), treating level of confidence as a fixed effect and participant as a random effect.

Result

The temporal PCA resulted in twenty-three factors that passed the Scree test, which accounted for 94.9% of the total variance. The sequential spatial PCA revealed four factors on each temporal component, which accounted for 65.3% of the variance. Ten temporal-spatial components accounted for at least .5% of the total EEG variance (Dien et al., 2010). All of them fell in 6 temporal factors that contributed mostly to the mean ERP amplitudes analyzed above. In particular, TF6 (peaking at 118 msec) and TF3 (peaking at 212 msec) can be regarded as contributing to the N100 and P200 window respectively, TF2 (peaking at 438 msec) and TF4 (peaking at 610 msec) as contributing to the N400/LPC/P600 window, and the remaining factor TF5 (peaking at 848 msec) and TF1 (peaking at 1298 msec) as contributing to the more delayed window after 800 msec post-onset of the stimuli.

The LMEM was performed on the factor covariance scores at all 11 components (see also Dien, 2006) and five components revealed significant differences between levels of confidence. The TF3SF2 (with a positive peak at AF4 in 212 msec) revealed a reduced positive response for unconfident than for confident ($\beta = -.58$, $t = -1.91$, $p = .05$), neutral-intending ($\beta = -.74$, $t = -2.40$, $p = .02$) and close-to-confident expressions ($\beta = -.75$, $t = -2.45$, $p = .02$). The TF2SF1 (with a positive peak at Fz in 438 msec) revealed a larger positive response for close-to-confident than for neutral-intending ($\beta = .54$, $t = 1.82$, $p = .06$), confident ($\beta = 1.34$, $t = 2.63$, $p = .01$), and for unconfident expressions ($\beta = 1.35$, $t = 2.76$, $p = .01$). The TF4SF1 (with a positive peak on FP2 in 610 msec) revealed a larger response for close-to-confident than for neutral ($\beta = 1.19$, $t = 1.92$, $p = .05$), confident ($\beta = 1.26$, $t = 2.06$, $p = .05$), and for unconfident expressions ($\beta = 1.72$, $t = 2.18$, $p = .02$). The TF1SF1 (with the negative peak on P5 at 1298 msec) revealed a less negative response for the neutral-intending level, as compared with confident ($\beta = 1.69$, $t = 2.38$, $p = .01$), close-to-confident ($\beta = 1.28$, $t = 1.94$, $p = .05$), unconfident ($\beta = 1.69$, $t = 2.38$, $p = .01$), the three of which did not differ. The TF1SF2 (with a positive peak on F5 at 1298 msec) revealed a similar pattern: the positive response was strongest in the neutral-intending level, followed by the confident ($\beta = 2.90$, $t = 1.92$, $p = .05$), the close-to-confident ($\beta = 2.79$, $t = 2.00$, $p = .05$), and the unconfident level ($\beta = 3.36$, $t = 2.88$, $p = .01$).[10] In sum, the PCA analysis generally confirmed the findings of confidence-related ERP responses in the mean amplitude analysis, which was conducted on four time windows based on previous literature.

Table G1 – The list of the principal components which explained at least .5% of the total variance of the averaged ERP data.

Principal component	Peak latency[a] (msec)	Peak channel[b]	Peak polarity[c]	Total variance explained (%)
TF1SF1	1298	P5	−	11.6
TF1SF2	1298	F5	+	2.4
TF2SF1	438	Fz	+	7.2
TF2SF2	438	P2	−	5.3
TF3SF1	212	AF4	+	3.4
TF3SF2	212	P1	−	1.5
TF4SF1	610	FP2	+	1.6
TF5SF1	848	Pz	−	1.4
TF5SF2	848	FP2	+	1.0
TF6SF1	118	AF3	−	0.9
TF6SF2	118	Pz	−	0.6

Note: the PCA was based on the ERP average of each of the 750 sampling point of the entire epoch of the vocal stimuli for each of the four types of expressions on 64 channels and 30 participants.
[a] Peak Latency: the time point with the greatest absolute voltage (of all the conditions after computing the grand average).
[b] Peak Channel: the channel with the greatest absolute voltage (out of all conditions after computing the grand average).
[c] Peak Polarity: whether the voltage of the peak latency is positive or negative.

[10] The TF1/SF1 accounts for the increased ERP variances which is maximal at the end of the epoch, which may or may not be necessarily meaningful (Kayser & Tenke, 2003). However, we chose to retain this factor because 1) it differentiated between levels of confidence, indicating that it contains systematic variance relevant to the effect of interest (see also Foti, Hajcak, & Dien, 2009); 2) the effect of confidence level emerged at around 500 msec and slowly accumulated at the later peak.

Table G2 – ANOVA Table for each principal component which explained at least .5% of the total variance of the averaged ERP data, for LMEMs.

Principal component	*df*	SS	MS	F	*p* (Lower Bound)[a]	*p* (Upper Bound)[b]
TF01SF1	3	33.48	11.16	2.47	.05	.05
TF01SF2	3	87.55	29.18	2.87	.05	.05
TF02SF1	3	23.53	7.84	3.31	.03	.03
TF02SF2	3	11.14	3.71	1.84	.14	.14
TF03SF1	3	51.21	17.07	4.19	.01	.01
TF03SF2	3	9.98	3.33	1.01	.39	.39
TF04SF1	3	18.24	6.08	2.44	.05	.05
TF05SF1	3	2.70	.90	.58	.63	.63
TF05SF2	3	3.57	1.19	.95	.42	.42
TF06SF1	3	4.68	1.56	1.28	.29	.28
TF06SF2	3	5.38	1.79	.34	.80	.80

Note: confidence as a fixed effect, and by-participant intercept as a random effect.
[a] Denominator: $df = 86$.
[b] Denominator: $df = 116$.

REFERENCES

Alain, C., & Winkler, I. (2012). Recording event-related potentials: application to study auditory perception. In D. Poeppel, T. Overath, A. Popper, & R. Fay (Eds.), *The human auditory cortex* (pp. 69–96). New York: Springer.

Allison, C., Baron-Cohen, S., Wheelwright, S., Stone, M., & Muncer, S. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences, 51*, 829–835.

Alter, K., Rank, E., Kotz, S., Toepel, U., Besson, M., Schirmer, A., et al. (2003). Affective encoding in the speech signal and in event-related brain potentials. *Speech Communication, 40*, 61–70.

Apple, W., Streeter, L., & Krauss, R. (1979). Effects of pitch and speech rate on personal attributes. *Journal of Personality and Social Psychology, 37*, 715–727.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin & Review, 10*, 462–467.

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129–135.

Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Phonology: Vol. 2. Universals of human language* (pp. 471–524). Stanford, CA: Stanford University Press.

Brennan, S., & Williams, M. (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Language and Memory, 34*, 383–398.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446*, 127–143.

Brouwer, H., & Hoeks, J. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience, 7*, 758.

Caffi, C., & Janney, R. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics, 22*, 325–373.

Charest, I., Pernet, C., Rousselet, G., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*, 127.

Comerchero, C., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*, 24–30.

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*, 9–21.

DePaulo, B., Stone, J., & Lassiter, D. (1985). Telling ingratiating lies: effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology, 48*, 1191–1203.

Dien, J. (2006). Progressing towards a consensus on PCA of ERPs. *Clinical Neurophysiology, 117*, 699–702.

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology, 47*, 170–183.

Dien, J. (2012). Applying principal components analysis to event-related potentials: a tutorial. *Developmental Neuropsychology, 37*, 497–517.

Dien, J., Michelson, C., & Franklin, M. (2010). Separating the visual sentence N400 effect from the P400 sequential expectancy effect: cognitive and neuroanatomical implications. *Brain Research, 1355*, 126–140.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology, 46*, 521–530.

Hemsley, G., & Doob, A. (1978). The effect of looking behavior on perceptions of a communicator's credibility. *Journal of Applied Social Psychology, 8*, 135–144.

Jessen, S., & Kotz, S. (2011). The temporal dynamics of processing emotions from vocal, facial and bodily expressions. *NeuroImage, 58*, 665–674.

Jiang, X., Li, Y., & Zhou, X. (2013). Is it over-respectful or disrespectful? Differential patterns of brain activity in perceiving pragmatic violation of social status information during utterance comprehension. *Neuropsychologia, 51*, 2210–2223.

Jiang, X., & Pell, D. M. (2014). Encoding and decoding confidence information in speech. In *Proceedings of the 7th International Conference in Speech Prosody (Social and Linguistic Speech Prosody)* (pp. 573–576). ISSN: 2333-2042 http://fastnet.netsoc.ie/sp7/sp7book.pdf. Retrieved 04.12.14.

Kayser, J., & Tenke, C. (2003). Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. *Clinical Neurophysiology, 114*, 2307–2325.

Kimble, C., & Seidel, S. (1991). Vocal signs of confidence. *Journal of Nonverbal Behavior, 15*, 99–105.
Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language, 100*, 257–261.
Kotz, S., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151*, 107–118.
Kotz, S., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass, 5*, 108–125.
Kotz, S., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences, 14*, 392–399.
Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
Liu, P., & Pell, D. M. (2012). Recognizing vocal emotions in Mandarin Chinese: a validated database of Chinese vocal emotional stimuli. *Behavioral Research Method, 44*, 1042–1051.
Liu, T., Pinheiro, A., Deng, G., Nestor, P., McCarley, R., & Niznikiewicz, M. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport, 23*, 108–112.
London, H., McSeveney, D., & Tropper, R. (1971). Confidence, overconfidence and persuasion. *Human Relations, 24*, 359–369.
London, H., Meldman, P., & Lanckton, A. (1970a). The jury method: how the persuader persuades. *Public Opinion Quarterly, 34*, 171–183.
Lopez-Calderon, J., & Luck, S. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213.
Luck, S. J. (2010). *Is it legitimate to compare conditions with different numbers of trials?*. http://erpinfo.org/Members/sjluck/Mean_Peak_Noise.pdf/view. Retrieved 04.12.14.
Makeig, S., Bell, A., Jung, T., & Sejnowski, T. (1996). Independent component analysis of electroencephalographic data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8* (pp. 145–151). Cambridge: MIT Press.
Maslow, C., Yoselson, K., & London, H. (1971). Persuasiveness of confidence expressed via language and body language. *British Journal of Clinical Psychology, 10*, 234–240.
Mitchell, R. L. C. (2013). Further characterisation of the functional neuroanatomy associated with prosodic emotion decoding. *Cortex, 49*, 1722–1732.
Mitchell, R. L. C., & Ross, E. (2013). Attitudinal prosody: what we know and directions for future study. *Neuroscience and Biobehavioral Reviews, 37*, 471–479.
Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology, 83*, 176–190.
Monetta, L., Cheang, H., & Pell, M. D. (2008). Understanding speaker attitudes from prosody by adults with Parkinson's disease. *Journal of Neuropsychology, 2*, 415–430.
Näätänen, R., & Picton, T.. (1987). The N1 wave of the human electric and magnetic response to sound. *A review and analysis of the component structure, 24*, 375–425.
Newman, A., Tremblay, A., Nichols, E., Neville, H., & Ullman, M. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience, 24*, 1205–1223.
Pantev, C., Elbert, T., Ross, B., Eulitz, C., & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research, 100*, 164–170.
Paulmann, S., Bleichner, M., & Kotz, S. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology, 4*, 345.
Paulmann, S., & Kotz, S. (2008a). Early emotional prosody perception based on different speaker voices. *NeuroReport, 19*, 209–213.
Paulmann, S., & Kotz, S. (2008b). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language, 105*, 59–69.
Paulmann, S., Ott, D., & Kotz, S. (2011). Emotional speech perception unfolding in time: the role of basal ganglia. *PLoS ONE, 6*, e17694.
Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain and Language, 96*, 221–234.
Pell, M. D. (2007). Reduced sensitivity to prosodic attitudes in adults with focal right hemisphere brain damage. *Brain and Language, 101*, 64–79.
Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 275–288.
Picton, T. (1993). The P300 wave of the human event-related potentials. *Journal of Clinical Neurophysiology, 9*, 456–479.
Pon-Barry, H., & Shieber, S. (2010). Assessing self-awareness and transparency when classifying a speaker's level of certainty. *Speech Prosody*.
Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *NeuroReport, 11*, 1329–1333.
Regel, S., Coulson, S., & Gunter, T. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research, 1311*, 121–135.
Regel, S., Gunter, T., & Friederici, A. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience, 23*, 277–293.
Regel, S., Meyer, L., & Gunter, T. (2014). Distinguishing neurocognitive processes reflected by P600 effects: evidence from ERPs and neural oscillations. *PloS ONE, 9*, e96840.
Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. *Brain Research, 1565*, 48–62.
Sabbagh, M., & Baldwin, D. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development, 72*, 1054–1070.
Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience, 22*, 474–481.
Schact, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, & Behavioral Neuroscience, 9*, 28–43.
Scherer, K., London, H., & Wolf, J. (1973). The voice of confidence: paralinguistic cues and audience evaluation. *Journal of Research in Personality, 7*, 31–44.
Schirmer, A., Chen, C., Ching, A., Tan, L., & Hong, R. (2013). Vocal emotions influence verbal memory: neural correlates and interindividual differences. *Cognitive, Affective and Behavioral Neuroscience, 13*, 80–93.
Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*, 25–30.
Schirmer, A., Kotz, S., & Friederici, A. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14*, 228–233.
Schirmer, A., Kotz, S., & Friederici, A. (2005). On the role of attention for the processing of emotions in speech: sex differences revisited. *Cognitive Brain Research, 24*, 442–452.

Schirmer, A., Lui, M., Maess, B., Escoffier, N., Chan, M., & Penney, T. (2006). Task and sex modulate the brain response to emotional incongruity in Asian listeners. *Emotion, 6*, 406–417.

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research, 1176*, 103–112.

Schwartze, M., & Kotz, S. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience and Biobehavioral Reviews, 37*, 2587–2596.

Smith, V., & Clark, H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25–38.

Spotorno, N., Cheylus, A., Van Der Henst, J., & Noveck, I. (2013). What's behind a P600? Integration operations during irony processing. *PloS ONE, 8*, 1–10.

Spreckelmeyer, K., Kutas, M., Urbach, T., Altenmüller, E., & Münte, T. (2009). Neural processing of vocal emotion and identity. *Brain and Cognition, 69*, 121–126.

Strauß, A., Kotz, S., & Obleser, J. (2013). Narrowed expectancies under degraded speech: revisiting the N400. *Journal of Cognitive Neuroscience, 25*, 1384–1395.

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*, 81–94.

Tremblay, A. (2014). *LMERConvenienceFunctions: A suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions*. R-version 2.5 http://cran.r-project.org/web/packages/LMER ConvenienceFunctions/index.html.

Van Overwalle, F., Van den Eede, S., Baetens, K., & Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. *Social, Cognitive and Affective Neuroscience, 4*, 177–190.

Van den Brink, D., Van Berkum, J., Bastiaansen, M., Tesink, C., Kos, M., Buitelaar, J., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social, Cognitive & Affective Neuroscience, 7*, 173–183.

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., & Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience, 4*, 165–184.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22*, 1682–1700.

Walker, M. (1977). The relative importance of verbal and nonverbal cues in the expression of confidence. *Australian Journal of Psychology, 29*, 45–57.

Wang, L., Zhu, Z., Bastiaansen, M., Hagoort, P., & Yang, Y. (2013). Recognizing the emotional valence of names: an ERP study. *Brain and Language, 125*, 118–127.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody. *International Journal of Speech-Language Pathology, 11*, 277–281.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics, 38*, 1559–1579.

Zimmerman, B., & Ringle, J. (1981). Effects of model persistence and statements of confidence on children's self-efficacy and problem solving. *Journal of Educational Psychology, 73*, 485–493.