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