Brain & Language 157-158 (2016) 63–71

Contents lists available at ScienceDirect

Brain & Language

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

The neural dynamics of speech perception: Dissociable networks for processing linguistic content and monitoring speaker turn-taking

Dan Foti [a,*], Felicia Roberts [b]

[a] Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907, United States
[b] Brian Lamb School of Communication, Purdue University, 100 N. University Street, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:
Received 30 June 2015
Revised 27 April 2016
Accepted 1 May 2016

Keywords:
EEG
ERP
Speech
Language
Prosody

ABSTRACT

The neural circuitry for speech perception is well-characterized, yet the temporal dynamics therein are largely unknown. This timing information is critical in that spoken language almost always occurs in the context of joint speech (i.e., conversations) where effective communication requires the precise timing of speaker turn-taking—a core aspect of prosody. Here, we used event-related potentials to characterize neural activity elicited by conversation stimuli within a large, unselected adult sample (N = 115). We focused on two stages of speech perception: inter-speaker gaps and speaker responses. We found activation in two known speech perception networks, with functional and neuroanatomical specificity: silence during inter-speaker gaps primarily activated the posterior pathway involving the supramarginal gyrus and premotor cortex, whereas hearing speaker responses primarily activated the anterior pathway involving the superior temporal gyrus. These data provide the first direct evidence that the posterior pathway is uniquely involved in monitoring speaker turn-taking.

1. Introduction

In conversations around the globe, billions of times each day, people shift effortlessly and seamlessly from one speaker to another. This daily exchange of turns at talk, a fundamental structure of human social interaction, remains largely unexplored from a neuroscience perspective. A key entry point for studying these deceptively simple moments is that speaker transition tends to be below the level of awareness, except when something goes awry. The current study leverages that fact to examine neural activity during the processing of delayed and normal turn-taking in conversation. This approach allows for the exploration of the neural pathways involved in processing conversational speech, monitoring the timing of speaker turn-taking, and inferring affect from anomalous speaker transitions.

Over 40 years ago, Sacks, Schegloff, and Jefferson (1974) described several "grossly apparent facts" about conversation and from those observations derived what has become the most widely used, empirically-grounded model of turn-taking for studies of human communication (Roberts & Robinson, 2004). Their model addresses the way in which turns at talk are constructed, ordered, and distributed among speakers. Most important for the current study is their identification of the "transition relevance place," or the moment when one unit of talk comes to an end and a next unit begins (initiated by the same or another speaker). Our goal is to inspect the neural dynamics within this transition space, particularly neural pathways as they relate to the processing of timing cues in turn transitions.

The importance of addressing brain function in more interactional and even true dyadic frameworks, particularly in relation to language and other behavioral processes, has been recently emphasized in neuroscience and psychology (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Scott, McGettigan, & Eisner, 2009; Stephens, Silbert, & Hasson, 2010; Wilson & Wilson, 2005). Indeed, a neuroscience perspective on the motor theory of speech perception suggests that the motor cortex is being recruited primarily to manage the timing of turn-taking in conversation. This is a role that goes beyond deciphering linguistic input, as indicated in some studies (Wilson & Iacoboni, 2006; Wilson, Saygin, Sereno, & Iacoboni, 2004) and implicates the motor cortex in processing social interaction (Scott et al., 2009). Specifically, Scott et al. (2009) propose that speech processing involves two functionally and anatomically distinct neural networks: an anterior pathway that is responsible for decoding word meaning and a posterior pathway that monitors speaker rhythm, rate, and turn-taking. Whereas the anterior pathway encompasses the superior temporal and inferior frontal gyri, it is only the posterior pathway that

* Corresponding author.
E-mail addresses: foti@purdue.edu (D. Foti), froberts@purdue.edu (F. Roberts).

http://dx.doi.org/10.1016/j.bandl.2016.05.001
0093-934X/

encompasses motor areas, along with the supramarginal and inferior frontal gyri. This model is broadly consistent with findings from functional magnetic resonance imaging (fMRI) (Wilson & Iacoboni, 2006; Wilson et al., 2004) and lesion studies (Blank, Bird, Turkheimer, & Wise, 2003; Crinion, Warburton, Lambon-Ralph, Howard, & Wise, 2006), yet direct evidence of functional dissociation between these anterior and posterior pathways is lacking. In the current study we seek to address this empirical gap.

To explore the role of sensorimotor activity in the monitoring of speaker turn-taking, we developed a novel approach for capturing neural dynamics during the processing of naturalistic conversations, using event-related potentials (ERPs). fMRI has excellent spatial resolution but insufficient temporal resolution to precisely map the dynamics of conversation processing. ERPs, meanwhile, have millisecond temporal resolution and are well-suited to characterize how conversations are processed in real time (Bogels, Magyari, & Levinson, 2015; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014). Here, we leveraged the timing information of ERPs to evaluate whether the processing of spoken words and inter-speaker gaps would be linked specifically with the anterior and posterior pathways, respectively—directly testing for the first time the proposed roles of these pathways in speech perception (Scott et al., 2009).

To do this, we examined ERPs elicited by conversation stimuli that were designed with clear turn-taking expectations embedded in their structure: requests and responses. Drawing on the turn-taking model and related descriptive, experimental, and cross-linguistic research, we know that an extended gap where response is expected is a salient social signal of reluctance, uncertainty or other potential cues of "trouble" in the conversation (Brennan & Williams, 1995; Burgoon, Buller, & Guerrero, 1995; Davidson, 1984; Kendrick & Torreira, 2015; Pomerantz, 1984; Roberts & Francis, 2013; Roberts, Francis, & Morgan, 2006; Roberts, Margutti, & Takano, 2011; Swerts & Krahmer, 2005). Building on these findings, we included both normal (200 ms) and extended (700 ms) speaker gaps between mundane requests and affirmative responses, providing a natural window for observing ERPs during two key stages of conversation processing: (a) activity during the silent turn-transition space and (b) activity in response to affirmative responses, following either a normal or extended gap.

We expected that these two stages of conversation processing would elicit unique ERPs: For the inter-speaker gap, the silence between the request and the response, we focused on the auditory stimulus-preceding negativity (SPN), a negative-going slow wave that is involved in stimulus anticipation (Brunia, van Boxtel, & Böcker, 2012) and is maximal at frontal sites (Ohgami, Kotani, Hiraku, Aihara, & Ishii, 2004). For the speaker response, we focused on the P3, which signals the allocation of attention to infrequent or task-relevant stimuli (Polich, 2007). From an affective standpoint, we expected that the affirmative responses following a long gap would be perceived as anomalous and elicit an increased P3 compared to responses following a normal gap. Critically, we also used source estimation to test whether ERPs to the gap and the response would also uniquely map onto the neural pathways as proposed by Scott et al. (2009). We hypothesized that the SPN during the inter-speaker gap would primarily reflect activation within the posterior pathway (i.e., supramarginal gyrus, motor cortex), whereas the P3 to affirmative responses would primarily reflect activation within the anterior pathway (i.e., temporal gyrus).

Given the potential for overlap between activity in the anterior and posterior pathways during conversation processing, we used principal components analysis (PCA) to parse the ERP data. PCA is a highly effective technique for decomposing ERP waveforms into their underlying neural components, separating patterns of activity that overlap in their timing or spatial distributions (Donchin & Heffley, 1979). PCA also improves the accuracy of source localization algorithms compared to localizing the ERP waveform directly (Dien, 2010b); this is important given our aim evaluate the specific activation of the anterior and posterior pathways across stages of processing. To evaluate the likely neural generators at each stage, we used low resolution brain electromagnetic tomography (LORETA) (Pascual-Marqui, 2002, 2007), a localization algorithm that is well-suited to identifying relatively widespread sources of neural activity (i.e., coordinated activity within a network; Pizzagalli, 2007).

2. Material and methods

2.1. Participants

One hundred and thirty-four adults participated in this study. Nineteen were excluded from the current analyses (1 for difficulty hearing, 3 for equipment malfunction, 3 for poor quality EEG data, and 12 for being statistical outliers), leaving 115 in the final sample (Age: $M = 20.23$, $SD = 5.40$; Gender: 67 females, 47 males, 1 declined to answer; Ethnicity: 2 Hispanic, 108 Not Hispanic, 4 declined to answer; Race: 26 Asian, 6 African American, 81 Caucasian, 1 multiracial, 1 declined to answer). Ninety-one participants were current undergraduate students who received course credit for their participation, and the remaining 24 were volunteers from the surrounding community who received monetary compensation for their time. According to a protocol approved by the Institutional Review Board of Purdue University, informed consent was obtained from participants before the experiment.

2.2. Task and materials

2.2.1. Overview of task

A listening task was administered using Presentation software (Neurobehavioral Systems, Inc., Albany, CA) to control the timing of all stimuli. On each trial, participants were presented with an auditory stimulus consisting of a short conversation (10 targets, 4 distractors) that simulated a telephone call between friends. Study participants were thus in the role of a third party overhearing the conversations, as though on a speaker phone. The constructed dialogues concerned mundane themes (e.g., flyers for a school function, going to the gym, going to lunch, homework). After each conversation, study participants provided a judgment, on a six-point scale, about some aspect of the call recipient's affective response to their ostensible friend's request, invitation, or observation about the world. Ratings higher on the scale indicated a perception of more positive affect (e.g., willingness, enthusiasm).

The target dialogues, where the silence manipulations were presented (i.e., 200- or 700-ms inter-speaker gaps), ended with the caller formulating a request (e.g., getting a ride to pick up flyers). The call recipients in these dialogues always answered in the affirmative ("sure") which made compliance with their friend's request lexically specific and clear. Thus, the experimental manipulation (insertion of the silence) modulated the semantics of the "sure" response, coloring them as potentially reluctant or anomalous (Davidson, 1984; Kendrick & Torreira, 2015; Pomerantz, 1984; Roberts & Francis, 2013; Roberts et al., 2006, 2011).

2.2.2. Construction of stimuli

The conversation stimuli were simulated telephone calls based on an actual telephone call between two female friends (Roberts & Robinson, 2004) and were developed using known features of familiarity in telephone interaction (Hopper, 1992; Schegloff, 1979). Each dialogue was approximately 10 s long and began with a friendly greeting sequence followed by the caller reporting some mundane state of affairs (e.g., "I called the copy shop"). After an

acknowledgement of this report by the call recipient ("uh-huh"), the caller would then make the request (e.g., "Can you give me a ride over there?"), to which the recipient answered in the affirmative ("Sure"). The request was thus phrased as a yes/no question. See Appendix A for an example of a target dialogue.

To control for any variation that might be elicited by gender of the caller/responder (Roberts & Norris, 2016), only female voices were used in the targets, and their caller-call recipient roles were held constant. To further control for possible confounding from the acoustic qualities within the actor's slightly different productions of the "sure" responses during recording of the dialogues, we identified a median "sure" response and used that response for all of her target conversations. In other words, the vocal quality of the affirmative response was identical across all target stimuli—the stimulus was the same—and only the silence lengths were manipulated (for additional detail, see Roberts et al., 2006).

2.2.3. Manipulation of inter-turn silence

The stem of each dialogue (ending with the request) and the affirmative "sure" response were presented as separate stimuli; the length of time between these stimuli was manipulated to convey either normal (200 ms) or non-normal (700 ms) speaker turn-taking. These silence lengths were chosen based on descriptive findings for English (Walker & Trimboli, 1984; Wilson & Zimmerman, 1986), recent cross-linguistic evidence (Stivers et al., 2009), and experimental evidence (Roberts & Francis, 2013; Roberts et al., 2006, 2011). The 700 ms gap length was chosen based on evidence that ratings of willingness significantly decrease at that point (Roberts & Francis, 2013) and that, in actual conversation, the likelihood of negative responses increases (Kendrick & Torreira, 2015).

2.2.4. Procedure

Participants first completed two practice trials to familiarize them with the task. The main task consisted of two blocks of 14 trials each (10 targets, 4 distractors). Participants heard each conversation twice, once in the first block of the task and again in the second block; stimulus order within blocks was pseudo-randomized for each participant, such that no more than three targets could occur consecutively. For the first presentation of each target stimulus, the gap was randomly assigned to be either 200 or 700 ms; for the second presentation, the other gap length was used. Distractor stimuli were identical for both presentations. A self-paced break occurred after the first block.

2.3. Psychophysiological recording, data reduction, and analysis

The continuous EEG was recorded using an actiCAP and the actiCHamp amplifier system (Brain Products, Munich, Germany). The EEG signal was digitized at 24-bit resolution and sampled at 500 Hz. Recordings were taken from 32 scalp electrodes and a ground electrode at Fpz. The electrooculogram was recorded from two auxiliary electrodes placed 1 cm above and below the left eye. Electrode impedances were kept below 30 kΩ.

Brain Vision Analyzer (Brain Products, Munich, Germany) was used for offline analysis. Only target trials were considered here. Data were re-referenced to the mastoid average and band-pass filtered from 0.01 to 30 Hz. For analysis of the inter-speaker gaps, the EEG signal was segmented from −500 to 1500 ms relative to gap onset; for speaker responses, the signal was segmented from −200 to 1000 ms relative to response onset. Ocular correction was performed using a regression method (Gratton, Coles, & Donchin, 1983). Individual channels were rejected for artifacts trial-wise using a semi-automated procedure. ERPs were averaged separately for the long and short gap conditions and baseline corrected (−200 to 0 ms).

Averaged ERP data was then reduced using temporospatial principal components analysis (PCA), an empirical approach for isolating distinct patterns of neural activity within the ERP waveform (for a comparison of PCA versus ICA, see Dien, Khoe, & Mangun, 2007). This analysis was conducted using the ERP PCA Toolkit, version 2.45 (Dien, 2010a) and followed published guidelines for applying PCA to ERP data (Dien, 2010b; Dien & Frishkoff, 2005). In particular, previous simulation studies have found that variance across time points is best captured by temporal PCA with Promax rotation, whereas variance across electrodes is best captured by spatial PCA with Infomax rotation (i.e., an ICA algorithm) (Dien et al., 2007). Here, we used a two-step approach: temporal PCA (Promax) followed by spatial PCA (Infomax). Separate analyses were conducted for ERPs elicited during the inter-speaker gap (long gap condition only) and speaker responses (both long and short gap conditions). In each case, the temporal PCA was applied first, using Promax rotation, and temporal factors were extracted based on Scree plots (Cattell, 1966): 52 for the inter-speaker gap, and 32 for the speaker responses. The spatial distributions of these factors were then analyzed using spatial PCA; a separate spatial PCA was performed on each temporal factor. Infomax rotation was used, and three spatial factors were extracted separately for each factor within the inter-speaker gap and speaker response data. Covariance matrices and Kaiser normalization were used in all cases.

To reduce the number of statistical comparisons, we restricted our analyses to PCA factors that peaked during two windows of interest, based on the pattern of activity in the raw ERPs (Fig. 1): 200–800 ms during the long inter-speaker gap, and 200–500 ms after the speaker response. Analysis of the inter-speaker gap yielded two temporal factors of interest, with Temporal Factor 3 peaking at 353 ms and Temporal Factor 4 peaking at 710 ms. Analysis of the speaker response yielded one temporal factor of interest, with Temporal Factor 2 peaking at 338 ms. For each of these three temporal factors, we analyzed the three associated spatial factors (i.e., nine temporospatial factors total). To control the familywise error rate, we first used MANOVA to test whether the identified temporospatial factors, as a set, were statistically different from zero. This tests the overall effect for the combined temporospatial factors, taking into account the relationships between them. Where the results of the overall MANOVA were significant, we then considered the specific effect of each temporospatial factor separately. Insofar as these analyses were intended to be descriptive, we did not apply correction for multiple comparisons for these follow-up tests.

PCA factors elicited during the inter-speaker gap were analyzed using one-sample *t*-tests. PCA factors elicited by the speaker responses were analyzed using paired-sample *t*-tests contrasting the long and short gap conditions. Statistical analysis was performed within IBM SPSS Statistics (Version 22, IBM Corp.). Factors that were statistically significant were then submitted to source analysis using the LORETA transformation (Pascual-Marqui, 2002, 2007), a distribution source estimation algorithm; the version implemented within Brain Vision Analyzer software is identical to the stand-alone LORETA-KEY software (Pascual-Marqui, 1999; Pascual-Marqui, Michel, & Lehmann, 1994). LORETA accounts for the scalp distribution of ERPs by modeling voxel-wise activation throughout the cortex.

3. Results

3.1. Ratings of affect

Consistent with previous research, conversations with long inter-speaker gaps of 700 ms ($M = 3.19$, $SD = 0.85$) were rated

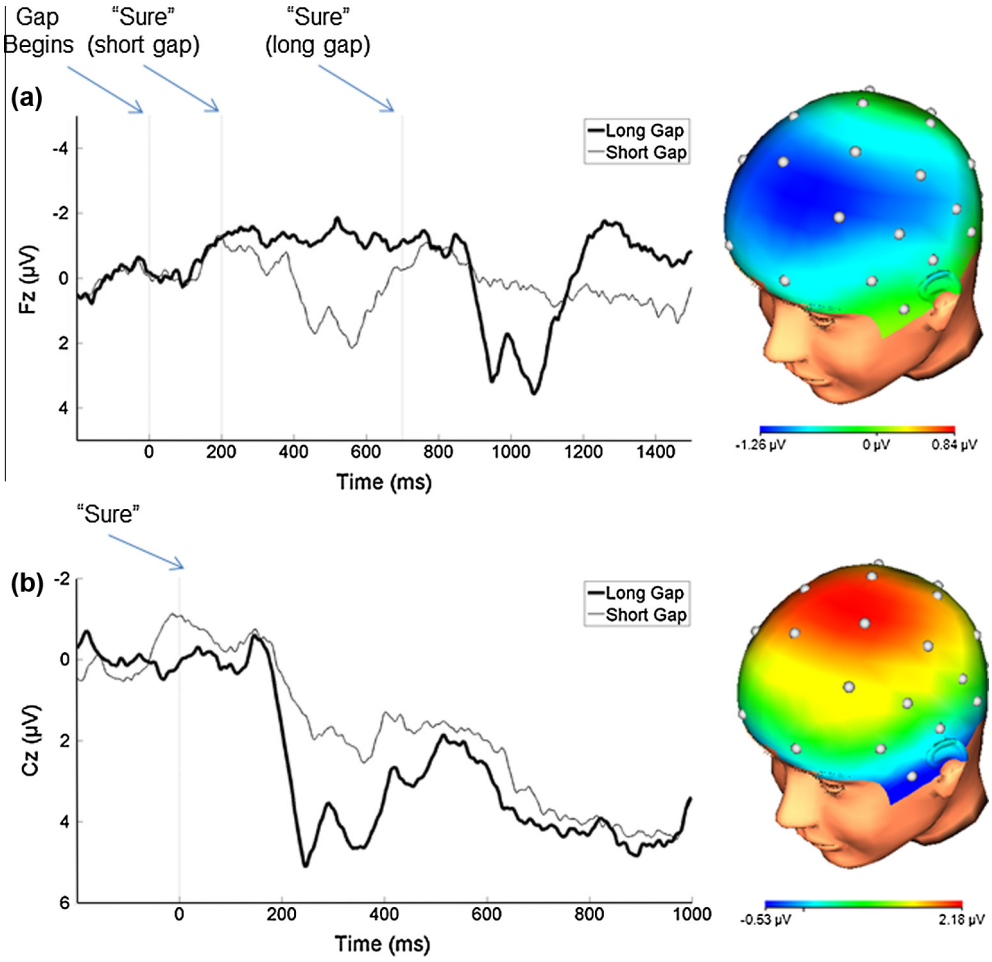

Fig. 1. Event-related potentials, prior to principal components analysis. (a) Activity during inter-speaker gaps at electrode Fz. The gap began at 0 ms and proceeded for either 200 (short gap) or 700 ms (long gap). The headmap depicts activity during the long gap (200–700 ms). (b) Activity to the speaker response at electrode Cz. Response onset was at 0 ms. The headmap depicts the difference between the long and short gap conditions from 200 to 500 ms.

consistently more negative than those with inter-speaker gaps of 200 ms ($M = 3.60$, $SD = 0.79$; $t(112) = 7.88$, $p < 0.001$, Cohen's $d = 0.75$). This indicates that the long gap condition was sufficient to impact the perceived affective tone of the conversations, despite the fact that the affirmative response stimulus was acoustically identical in each case.

3.2. Event-related potentials

Raw ERPs (i.e., prior to PCA) elicited during conversation processing are presented in Fig. 1. During the inter-speaker gap (Fig. 1a), the SPN was apparent at frontal sites with an onset of approximately 100 ms, for both the long and short gap conditions. For the long gap, the SPN was sustained throughout the inter-speaker gap (i.e., until the speaker response at 700 ms). Speaker responses elicited a P2/N2/P3 complex that was maximal at central electrodes and was increased for the long versus short gap (Fig. 1b).

3.2.1. Inter-speaker gap

The PCA on ERP activity during the long inter-speaker gap decomposed the SPN into two temporal factors of interest (Fig. 2): Temporal Factor 3 had an onset of approximately 80 ms and peaked at 352 ms, at the midpoint of the extended gap. Temporal Factor 4, meanwhile, had an onset of approximately 400 ms and peaked at the end of the gap (710 ms). When the six total temporospatial factors (i.e., three spatial factors each for TF3 and TF4) were entered together as dependent variables within a MANOVA, the intercept was significant ($F(6,109) = 3.85$, $p < 0.01$, $\eta_p^2 = 0.18$), indicating that the mean for this set of factors was significantly different from zero.

Next, we considered the six temporospatial factors separately. For Temporal Factor 3, Spatial Factor 2 (TF3SF2) had a mean that was significantly different from zero ($t(114) = 3.49$, $p < 0.001$, Cohen's $d = 0.32$); TF3SF1 and TF3SF3 were non-significant (p's > 0.10). TF3SF2 was characterized as a frontal negativity, and source analysis indicated a likely neural generator in the posterior temporal gyrus (Brodmann Areas 19, 37, 39; MNI peak: −52, −65, 1).

For Temporal Factor 4, all three associated spatial factors had means that were significantly different from zero (TF4SF1: $t(114) = 1.98$, $p < 0.05$, $d = 0.18$; TF4SF2: $t(114) = 2.20$, $p < 0.05$, $d = 0.20$; TF4SF3: $t(114) = 2.04$, $p < 0.05$, $d = 0.19$). As shown in Fig. 2b, TF4SF1 was characterized as a parietal negativity, and source analysis indicated a likely neural generator in the mPFC (Brodmann Areas 10, 11, 32; MNI peak: 4, 52, 1). TF4SF2 was characterized as a frontal negativity, with a likely neural generator in the supramarginal/angular gyri (Brodmann Areas 13, 39, 40; MNI peak: 53, −50, 29). TF4SF3 was characterized as a frontotemporal negativity (Fig. 2, right), and source analysis indicated coactivation between the mPFC (MNI peak: −3, 52, 1) and premotor cortex (Brodmann Area 6; MNI peak: −3, −8, 61).

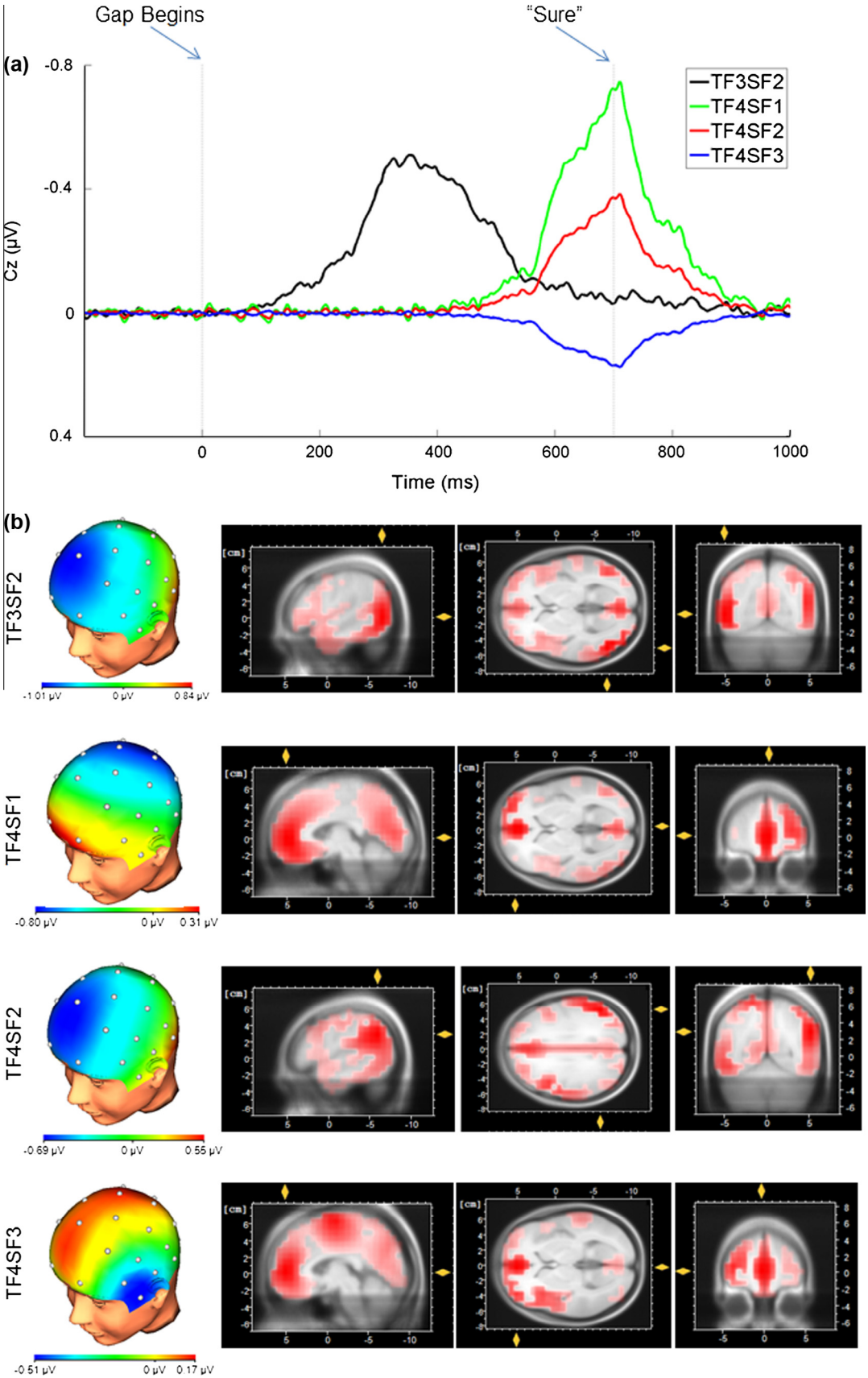

Fig. 2. PCA factors during the long inter-speaker gap. The gap began at 0 ms, and the speaker response was at 700 ms. (a) PCA waveforms at electrode Cz representing the portion of ERP activity associated with each PCA factor. (b) Headmaps depicting the spatial distribution of each factor at its temporal peak, and source estimates derived from LORETA.

3.2.2. Speaker response

The PCA on ERP activity to the speaker response yielded one temporal factor of interest. Temporal Factor 2 peaked at 338 ms, which is consistent with the P2/N2/P3 complex apparent in the raw ERP data. To test for the overall effect in this time range, we entered the differences scores (long minus short gap) for all three difference factors into a MANOVA. The intercept was significant ($F(3,112) = 16.28$, $p < 0.001$, $\eta_p^2 = 0.31$), indicating that the mean for this set of factors was significantly different from zero.

Next, we considered the three spatial factors separately. Two showed a significant effect of inter-speaker gap, with increased activation for the long versus short gaps (TF2SF1: $t(114) = 5.08$, $p < 0.001$, $d = 0.47$; TF2SF2: $t(114) = 3.98$, $p < 0.001$, $d = 0.35$; TF2SF3: $t(114) = 1.86$, $p = 0.06$, $d = 0.18$). TF2SF1 was characterized as a parietal positivity (Fig. 3a), with a likely neural generator in the mPFC (Brodmann Areas 10, 11, 32; MNI peak: −3, 52, 1). TF2SF2 was characterized as a frontal positivity (Fig. 3b), with a likely neural generator spanning the superior temporal and angular gyri (Brodmann Areas 22, 39; MNI peak: 53, −60, 22).

4. Discussion

The current study provides new insight into the neural dynamics of speech perception by considering neural activity elicited by naturalistic conversation stimuli. Inter-speaker gaps (i.e., silence) following a simple request elicited an SPN that was sustained throughout gap duration, reflecting the anticipation of speaker response. In the environment of extended gaps that deviated from normative conversational rhythms, affirmative speaker responses elicited an increased P2/N2/P3 complex, reflecting the increased allocation of attention to these anomalous responses. Critically, source estimation revealed that these ERPs also mapped onto distinct neural networks: the SPN during the extended inter-speaker gap was accounted for by initial activity in the posterior temporal gyrus, followed by activity in the supramarginal gyrus, premotor cortex and mPFC (i.e., posterior pathway); the P2/N2/P3 complex to the speaker response, meanwhile, was accounted for by more widespread activation along the superior temporal gyrus, as well as activity in the mPFC (i.e., anterior pathway)—but not the motor cortex. These two networks strongly converge with a proposed neuroanatomical model of speech perception (Scott et al., 2009). The pattern of activation across speech perception stages supports their proposed functional dissociation.

The current study thus provides direct evidence for the first time that the posterior pathway plays a primary role in monitoring speaker turn-taking, whereas the anterior pathway is primarily involved in decoding speech content (Scott et al., 2009). Moreover, we found that the timing of activity during the gap was consistent with the neuroanatomy of the posterior pathway, with activity first in the posterior temporal gyrus (352 ms), followed by activity in the supramarginal gyrus, premotor cortex, and mPFC (710 ms). Crucially, activation of the posterior pathway was observed during silence and not acoustic stimulation—silence conveying socially meaningful deviations from normative speaker turn-taking. On the other hand, hearing speaker responses did not activate this posterior pathway, instead eliciting widespread activation along the superior temporal gyrus that is consistent with the anterior speech perception pathway. These data strongly support the notion that the sensorimotor system plays a critical role in coordinating joint speech by monitoring conversational rhythms. While there is almost certainly a complex inter-relationship between monitoring conversational rhythms and decoding word meaning, particularly in the context of anticipating utterance completion (Levinson & Torreira, 2015), the relatively simple structure of our stimuli allowed us to isolate specific patterns of neural activity and thereby reveal a direct link between the posterior pathway and the monitoring of speaker turn-taking rather than speech content.

The current findings complement those of two recent EEG/ERP studies that also assessed the monitoring of speaker turn-taking. In one study, participants listened to spoken sentences and were asked to press a button when the speaker finished their turn, and of interest was EEG activation as participants anticipated the speaker turn end (Magyari et al., 2014). For stimuli that followed a predictable sentence structure, participants were highly accurate and exhibited EEG desynchronization in the beta frequency band—a physiological response generally associated with motor preparation—beginning 1500 s before the turn end. This beta activity partially corresponded to activation of the posterior network, including the inferior parietal lobule and posterior temporal gyrus. Notably, the beta response was not localized to the motor cortex per se, likely due to the task design: participants were required to make button presses while listening to the stimuli, thereby confounding the monitoring of turn-taking with motor movement. In a related study, participants were asked to respond to spoken trivia questions as quickly as possible, and of interest was ERP/EEG activity elicited as participants anticipated the end of each question (Bogels et al., 2015). Neural activity (positive slow wave and alpha-band desynchronization) was apparent more than 2 s before question end, signaling response planning that occurred concurrent with auditory processing. This activation was localized to activity in both the anterior and posterior pathways, including the middle and inferior frontal gyri, superior temporal gyrus, and motor cortex. Both of these studies are generally consistent with the proposed model of speech processing (Scott et al., 2009), demonstrating activation of the posterior pathway in anticipation of speaker turn ends, but due to the design of each study they were unable to disentangle the processing of turn-taking from word meaning. Here, we show similar activation of the posterior pathway specifically during inter-speaker gaps, supporting the functional specificity of this neural pathway in the monitoring of speaker turn-taking.

In addition to observing activation of the anterior and posterior pathways, we also observed prominent mPFC activity during both inter-speaker gaps and speaker responses. In fact, the current source analyses indicate significant mPFC/premotor coactivation during the inter-speaker gap as indicated by distributed sources in these regions for the relevant PCA factor. The mPFC activation observed here is likely a direct result of the task design: participants were required to make explicit judgments about the speaker's intent or affect within each conversation. Other research has shown similar mPFC activation on tasks requiring the attribution of mental states (Gallagher et al., 2000) and emotional perspective-taking (Hynes, Baird, & Grafton, 2006). Insofar as listening to conversation stimuli could arguably never be truly passive (i.e., to the extent that conversation stimuli naturally capture attention), it would be difficult to design a task that reduces perspective-taking to nil. Rather, a further test of the relative activation of the mPFC and speech perception pathways might be achieved by using stimuli that are potentially more emotionally salient to participants (e.g., using similar requests spoken by actual friends of study participants). Such a condition might yield relatively increased activation of the mPFC compared with listening to conversations between strangers. This approach could provide a contrast that might help disentangle the impact of speech processing from other effects of perspective-taking, empathy, or emotional salience.

We also note that motor cortex activation during speech processing is thought to reflect a broader role of tracking stimulus rhythm (i.e., both verbal and non-verbal stimuli) in order to facilitate coordinated action, rather than an effect that is specific

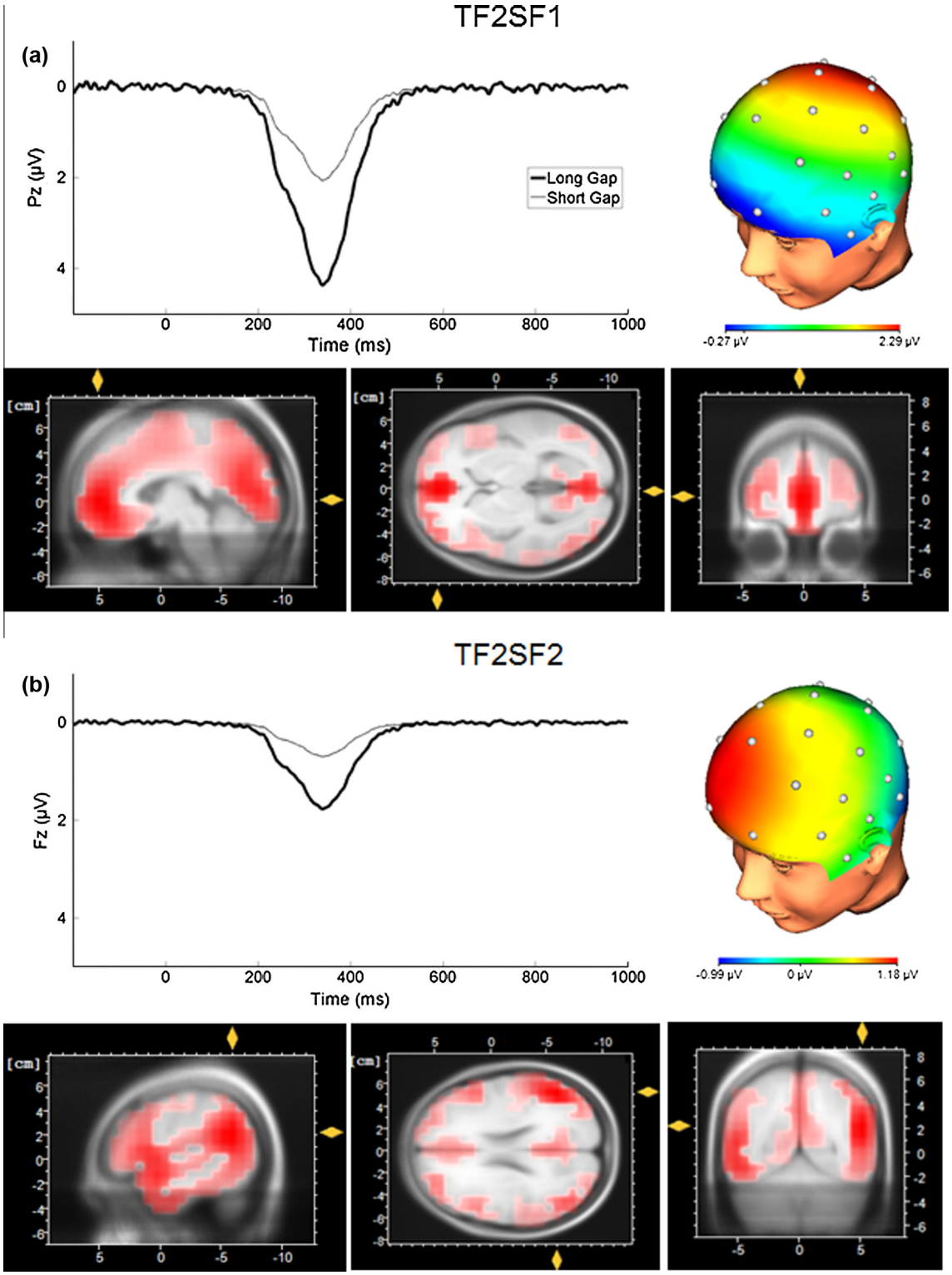

Fig. 3. PCA factors elicited by the speaker response: (a) Temporal Factor 2/Spatial Factor 1. (b) Temporal Factor 2/Spatial Factor 2. PCA waveforms are presented separately for the short and long gap conditions. Headmaps depict the spatial distribution of each factor at its temporal peak (long minus short). Source estimates for each factor were derived from LORETA.

to language per se (Scott et al., 2009). For example, the motor cortex is also activated when listening to music, particularly when the motor actions necessary to perform the music are known (Lahav, Saltzman, & Schlaug, 2007). Interestingly, recent ERP studies have found links between musicianship and language processing, including the perceptual discrimination (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011) and learning of verbal stimuli (Kuhnis, Elmer, Meyer, & Jancke, 2013), supporting the notion that music and language processing utilize overlapping neural networks. It would be interesting for future studies to test how sensitivity to speaker turn-taking during conversations may similarly be modulated by musical expertise.

From a clinical perspective, by providing a baseline observation of neural activity to speaker turn-taking among an unselected adult sample, this study opens the possibility of comparative work with impaired populations. Indeed, prosody perception has been advanced as a core clinical construct within the Research Domain Criteria project (Reception of Non-Facial Communication, as part of the Social Processes domain; Cuthbert, 2014), and such research is of relevance to multiple diagnostic groups: impaired prosody detection is a prominent cognitive deficit in schizophrenia (Hoekert, Kahn, Pijnenborg, & Aleman, 2007) and has been suggested as a potential mechanism of illness risk (Kee, Horan, Mintz, & Green, 2004). Other research has shown deficient prosody detection in autism (Wang, Lee, Sigman, & Dapretto, 2006). The current paradigm may be useful for isolating subtle forms of impairment within specific stages of conversation processing, helping to clarify which aspects of speech perception are impaired for a given population. If, as our data indicate, there is a separate neural pathway for processing timing in turn-taking, than this particular social cue could be a viable area for developing approaches to early detection and diagnosis.

Because word meaning is never fully dissociable from affective stance or emotional content, and since accumulating evidence indicates that integration of word meaning and emotional prosody is what may be impaired in populations with social processing deficits (Brazo, Beaucousin, Lecardeur, Razafimandimby, & Dollfus, 2014), then the cue we have chosen, silence prior to affirmative response, is something that can serve as a non-linguistic, non-intonation cue that still has social relevance. It is the delay of response that colors the valence of the responder's attitude as reluctant or uncertain, not the intonation of the response itself. While timing is a key component of conversational prosody, previous studies of impaired prosody detection have largely focused on other characteristics, particularly intonation. In our study, it is the affirmative "sure" appearing after the long gap that generates the discrepancy in expectation for a negative response. Delays reaching 700 ms are often followed by refusals, not acceptances (Kendrick & Torreira, 2015), and thus affirmation after a long gap elicits increased attention. It could be that in impaired populations this discrepancy is not noted or is more tolerated (i.e., perhaps there is less aversion to long gaps), which could potentially be exhibited by reduced activation of the posterior pathway. Based on the literature concerning social deficit processing (Brazo et al., 2014) and other recent studies of conversation processing (Bogels et al., 2015; Magyari et al., 2014), this is a question that warrants further investigation.

Strengths of the current study include the examination of a relatively large sample, as well as the use of temporospatial PCA to parse the ERP data. For example, the PCA applied to the inter-speaker gap clarified that the SPN in this context may be better understood as a set of overlapping neural responses with distinct neural generators, rather than a single, sustained response. An additional strength is the use of naturalistic, well-controlled stimuli. By maintaining a plausible conversational context and varying only the length of the silent gap between requests and responses, we isolated the effect of the delay without interference from speaker characteristics. A limitation of this study is the relatively small number of trials used in the task; however, because these were naturalistic conversations, to overly repeat the same conditions could have alerted participants to the nature of the study. We opted, therefore, to utilize fewer trials but with a wider range of participants. It would be potentially useful for follow-up studies to incorporate more trials as well as a wider range of gap lengths which, by being more variable, may elicit greater activation within the posterior pathway as more precise monitoring of speaker turn-taking is required.

While fMRI research has identified the relevant neural network for speech perception, the temporal dynamics of speech processing are not well understood, particularly with regard to naturalistic conversation stimuli. Here, we used ERPs to characterize the functional specificity of neural activation across two key stages of speech processing: inter-speaker gaps and speaker responses. The former primarily activated the posterior pathway, including sensorimotor regions, whereas the latter primarily activated the anterior pathway. These data provide direct evidence for the first time that the posterior pathway is uniquely involved in monitoring the timing of speaker turn-taking, a critical aspect of joint speech and conversational behavior.

Disclosures

The authors report no conflicts of interest, financial or otherwise.

Funding sources

This research was funded by Purdue University. The funding source had no involvement in study design, implementation, manuscript preparation, or publication.

Appendix A

Example of a target dialogue. All target dialogues contained the same yes/no request formulation "Can you give me a ride over there?" (line 9) with slightly varied mundane frames for the request (e.g., flyers, new desk, going to the gym.) The same "sure" affirmative response was edited into all target dialogues (line 10). The gap conditions were manipulated between presentation of line 9 and line 10.

1	((Telephone rings))
2	A: Hello?
3	B: Rachel?
4	A: Yeah,
5	B: Hey it's me.
6	A: Hey. How's it goin.
7	B: Good. I just called the copy shop,
8	A: uh huh
9	B: And the flyers are ready. Can you give me a ride over there?
10	A: Sure.

References

Blank, S. C., Bird, H., Turkheimer, F., & Wise, R. J. (2003). Speech production after stroke: The role of the right pars opercularis. *Annals of Neurology, 54*(3), 310–320. http://dx.doi.org/10.1002/ana.10656.

Bogels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*, 12881. http://dx.doi.org/10.1038/srep12881.

Brazo, P., Beaucousin, V., Lecardeur, L., Razafimandimby, A., & Dollfus, S. (2014). Social cognition in schizophrenic patients: The effect of semantic content and emotional prosody in the comprehension of emotional discourse. *Frontiers in Psychiatry, 5*, 120. http://dx.doi.org/10.3389/fpsyt.2014.00120.

Brennan, S. E., & 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 Memory and Language, 34*(3), 383–398.

Brunia, C. H., van Boxtel, G. M., & Böcker, K. B. (2012). Negative slow waves as indices of anticipation: The Bereitschafts potential, the contingent negative variation, and the stimulus-preceding negativity. In *The Oxford handbook of event-related potential components* (pp. 189–207). Oxford: Oxford University Press.

Burgoon, J. K., Buller, D. B., & Guerrero, L. K. (1995). Interpersonal deception IX. Effects of social skill and nonverbal communication on deception success and detection accuracy. *Journal of Language and Social Psychology, 14*(3), 289–311.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.
Crinion, J. T., Warburton, E. A., Lambon-Ralph, M. A., Howard, D., & Wise, R. J. (2006). Listening to narrative speech after aphasic stroke: The role of the left anterior temporal lobe. *Cerebral Cortex, 16*(8), 1116–1125. http://dx.doi.org/10.1093/cercor/bhj053.
Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry, 13*(1), 28–35. http://dx.doi.org/10.1002/wps.20087.
Davidson, J. (1984). Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 102–128). Cambridge: Cambridge University Press.
Dien, J. (2010a). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138–145. http://dx.doi.org/10.1016/j.jneumeth.2009.12.009.
Dien, J. (2010b). Evaluating two-step PCA of ERP data with geomin, infomax, oblimin, promax, and varimax rotations. *Psychophysiology, 47*(1), 170–183. http://dx.doi.org/10.1111/j.1469-8986.2009.00885.x.
Dien, J., & Frishkoff, G. (2005). Principal components analysis of event-related potential datasets. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 189–208). Cambridge, MA: The MIT Press.
Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: Promax vs. Infomax rotations. *Human Brain Mapping, 28*(8), 742–763. http://dx.doi.org/10.1002/hbm.20304.
Donchin, E., & Heffley, E. (1979). Multivariate analysis of event-related potential data: A tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research*. Washington, DC: U.S. Government Printing Office.
Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11–21.
Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484. http://dx.doi.org/10.1016/0013-4694(83)90135-9.
Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences, 16*(2), 114–121.
Hoekert, M., Kahn, R. S., Pijnenborg, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research, 96*(1), 135–145.
Hopper, R. (1992). *Telephone conversation*. Bloomington: Indiana University Press.
Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia, 44*(3), 374–383.
Kee, K. S., Horan, W. P., Mintz, J., & Green, M. F. (2004). Do the siblings of schizophrenia patients demonstrate affect perception deficits? *Schizophrenia Research, 67*(1), 87–94.
Kendrick, K. H., & Torreira, F. (2015). The timing and construction of preference: A quantitative study. *Discourse Processes, 52*(4), 255–289.
Kuhnis, J., Elmer, S., Meyer, M., & Jancke, L. (2013). Musicianship boosts perceptual learning of pseudoword-chimeras: An electrophysiological approach. *Brain Topography, 26*(1), 110–125. http://dx.doi.org/10.1007/s10548-012-0237-y.
Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience, 27*(2), 308–314. http://dx.doi.org/10.1523/JNEUROSCI.4822-06.2007.
Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731. http://dx.doi.org/10.3389/fpsyg.2015.00731.
Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*(11), 2530–2539. http://dx.doi.org/10.1162/jocn_a_00673.
Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23*(10), 2701–2715. http://dx.doi.org/10.1162/jocn.2010.21585.
Ohgami, Y., Kotani, Y., Hiraku, S., Aihara, Y., & Ishii, M. (2004). Effects of reward and stimulus modality on stimulus-preceding negativity. *Psychophysiology, 41*(5), 729–738. http://dx.doi.org/10.1111/j.1469-8986.2004.00203.x.
Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *International Journal of Bioelectromagnetism, 1*(1), 75–86.
Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology, 24*(Suppl. D), 5–12.
Pascual-Marqui, R. D. (2007). Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: Exact, zero error localization. arXiv preprint arXiv:0710.3341.
Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology, 18*(1), 49–65.
Pizzagalli, D. A. (2007). Electroencephalography and high-density electrophysiological source localization. In J. T. Cacioppo, L. G. Tassinaru, & G. G. Bernston (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 56–84). Cambridge, England: Cambridge University Press.
Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. http://dx.doi.org/10.1016/j.clinph.2007.04.019.
Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 57–101). Cambridge: Cambridge University Press.
Roberts, F., & Francis, A. L. (2013). Identifying a temporal threshold of tolerance for silent gaps after requests. *The Journal of the Acoustical Society of America, 133*(6), EL471–EL477.
Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of "trouble" in conversation. *Speech Communication, 48*(9), 1079–1093.
Roberts, F., Margutti, P., & Takano, S. (2011). Judgments concerning the valence of inter-turn silence across speakers of American English, Italian, and Japanese. *Discourse Processes, 48*(5), 331–354.
Roberts, F., & Norris, A. M. (2016). Gendered expectations for "agreeableness" in response to requests and opinions. *Communication Research Reports, 33*(1), 16–23.
Roberts, F., & Robinson, J. D. (2004). Interobserver agreement on first-stage conversation analytic transcription. *Human Communication Research, 30*(3), 376–410.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 696–735.
Schegloff, E. A. (1979). Identification and recognition in telephone conversation openings. In G. Psathas (Ed.), *Everyday language: studies in ethnomethodology* (pp. 23–78). New York: Irvington.
Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10*(4), 295–302.
Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America, 107*(32), 14425–14430.
Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10587–10592.
Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language, 53*(1), 81–94.
Walker, M. B., & Trimboli, C. (1984). The role of nonverbal signals in co-ordinating speaking turns. *Journal of Language and Social Psychology, 3*(4), 257–272.
Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain, 129*(4), 932–943.
Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage, 33*(1), 316–325.
Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*(7), 701–702.
Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review, 12*(6), 957–968.
Wilson, T. P., & Zimmerman, D. H. (1986). The structure of silence between turns in two-party conversation. *Discourse Processes, 9*(4), 375–390.