

Rice, K., Moraczewski, D., & Redcay, E. (2016). Perceived live interaction modulates the developing social brain. *Social Cognitive and Affective Neuroscience*, 11, 1354-1362.

Title: Perceived live interaction modulates the developing social brain

Running title: Social-interactive context in childhood

Authors:

Katherine Rice,^{1,2*} Dustin Moraczewski^{2,3}, & Elizabeth Redcay²

¹Department of Psychology, Texas State University, San Marcos, TX, 78666

²Department of Psychology, University of Maryland, College Park, MD, 20742

³ Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, 20742

* Corresponding author

Correspondence:
Katherine Rice
Department of Psychology
Texas State University
San Marcos, TX 78666
Email: kaw234@umd.edu

Conflict of Interest: None

[Article: 5313 words; Abstract: 165; Figures: 2; Tables: 2; Supplemental materials: 2 figures; 1 table]

Keywords: middle childhood, development, social interaction, theory of mind, fMRI, mentalizing

Funding: This research was partially supported by an NSF graduate fellowship to K.R. and a UMD Research and Scholarship Award to E.R.

Acknowledgements: We thank Vivi Bauman, Seleste Braddock, and Tara Feld for assistance with stimuli creation, Laura Anderson, Kayla Velnoskey, and Briana Viscomi for assistance with data collection and analysis, and the Maryland Neuroimaging Center and staff for project assistance.

ABSTRACT

Although children's social development is embedded in social interaction, most developmental neuroscience studies have examined responses to non-interactive social stimuli (e.g., photographs of faces). The neural mechanisms of real-world social behavior are of special interest during middle childhood (roughly ages 7-13), a time of increased social complexity and competence coinciding with structural and functional social brain development. Evidence from adult neuroscience studies suggests that social interaction may alter neural processing, but no neuroimaging studies in children have directly examined the effects of live social-interactive context on social cognition. In the current study of middle childhood, we compare the processing of two types of speech: speech that children believed was presented over a real-time audio-feed by a social partner and speech that they believed was recorded. Although in reality all speech was prerecorded, perceived live speech resulted in significantly greater neural activation in regions associated with social cognitive processing. These findings underscore the importance of using ecologically-valid and interactive methods to understand the developing social brain.

1. Introduction

Children develop in a world filled with reciprocal social interaction, but social brain development is almost exclusively measured and understood via non-interactive paradigms that examine component pieces of interaction (e.g., looking at photographs of faces). Behavioral evidence, however, from both adults (e.g., Laidlaw et al., 2011; Okita et al., 2007) and children (Goldstein & Schwade, 2008; Kirschner & Tomasello, 2009; Kuhl et al., 2003) suggests that live, interactive context significantly alters response to otherwise matched social stimuli. Adult neuroimaging research has begun to identify the neural bases of social interaction (e.g., Pfeiffer et al., 2014; Pönkänen et al., 2011; Redcay et al., 2010; Rice & Redcay, 2016; Schilbach et al., 2010), but few studies have investigated how the developing brain supports social interaction. Understanding the developmental bases of real-world social behaviors will provide insight into both typical and atypical social development, where disorders such as autism and social anxiety are characterized by interpersonal difficulties (e.g., Heimberg et al., 2010; Klin et al., 2003).

Although social interaction is characterized by a variety of properties (e.g., interaction may be intrinsically rewarding; Mundy & Neal, 2000; Pfeiffer et al., 2014; Schilbach et al., 2010; 2013), one component of successful social interaction is the creation of a shared psychological state between partners (Clark, 1996; Sperber et al., 1996; Tomasello et al., 2005). Consistent with this perspective, recent behavioral (e.g., Teufel et al., 2009) and neural evidence (e.g., Coricelli & Nagel, 2009; Rice & Redcay, 2016) from adults suggests that on-going social interaction involves mental state

inference—or mentalizing—about one’s social partner. Specifically, the mentalizing network (e.g., dorsomedial prefrontal cortex [DMPFC], temporal parietal junction [TPJ]; Frith & Frith, 2006) is consistently activated during social interaction, including when individuals process communicative cues (e.g., Kampe et al., 2003), engage in joint attention (e.g., Redcay et al., 2012; Schilbach et al., 2010), and play games against a human as opposed to a computer (e.g., Coricelli & Nagel, 2009; Gallagher et al., 2002; McCabe et al., 2001). Such studies, however, often involve either explicit mentalizing, as during strategy games, or do not directly compare stimuli that differ solely on interactive context. In a novel paradigm, Rice and Redcay (2016) isolated the potential role of implicit mentalizing in on-going social interaction. Participants listened to audio clips from live versus recorded speakers that contained no explicit mentalizing demands. Live speech resulted in increased activation in regions identified by a separate mentalizing localizer, including DMPFC and TPJ, suggesting social interaction automatically recruits the mentalizing network. Although these lines of converging evidence suggest a role for spontaneous mentalizing in social interaction, little is known about these processes in children.

Middle childhood (roughly ages 7-13) is an important time for considering the role of mentalizing in social interaction. During this age range, children improve on a variety of social cognitive tasks, including measures of mentalizing (e.g., Apperly et al., 2011; Dumontheil et al., 2010; Wang et al., in press). Further, during middle childhood, the brain’s mentalizing network undergoes functional and structural development. For example, regions involved in mentalizing in adults (including precuneus and bilateral

TPJ) become increasingly selective for processing mental states as compared to general social information (Gweon et al., 2012). Further, the degree of mental state specialization in right TPJ correlates with mentalizing ability (Gweon et al., 2012). Both pSTS and TPJ also show protracted structural development (Shaw et al., 2008). These social-cognitive and neural developments coincide with increased complexity of children's social interactions (Farmer et al., 2015; Feiring & Lewis, 1991) as socio-emotional understanding increases (Carr, 2011) and variability in social competence widens (Monahan & Steinberg, 2011). One possibility is that these changes in real-world social behaviors are supported by behavioral and neural changes in the mentalizing system, making middle childhood an important time to study mentalizing during real-time social interaction.

The developmental role of the mentalizing network during social interaction is unknown because the few developmental neuroimaging studies that have employed interactive paradigms have not directly addressed how live context alters social cognition. For example, researchers have investigated how children respond to potential future interaction (e.g., Guyer et al., 2009; 2012), how adolescents make decisions when observed (Chein et al., 2011), and how children respond to social rejection (e.g., Bolling et al., 2011; Will et al., 2016). Such studies, however, do not isolate whether or how an interactive social context alters the neural processing of that interaction's constituent social stimuli.

In order to characterize the developmental neural response to real-time social interaction, we extended an fMRI paradigm previously used with adults (Rice & Redcay,

2016) to children aged 7-13. In this paradigm, children listened to two types of content-matched speech: speech that they believed was coming over a live audio feed from a speaker in another room and speech that they believed was recorded. All stimuli were actually prerecorded. On each trial, children heard a short spoken vignette, which they believed to be either live or recorded, that presented two options (e.g., fruit or pancakes). They then heard about someone's preference (e.g., eating healthy), and finally made a choice for that person based on their preference. After each question, children saw positive or negative feedback, in order to match attention and contingency across conditions. Analyses focused on the short vignette, which contained no mentalizing demands or references to people. In adults, the comparison of live versus recorded stimuli resulted in increased activation in each individual's mentalizing network (as defined by a localizer) despite the lack of explicit mentalizing demands. Thus, although regions in the mentalizing network serve a variety of functions, previous findings indicate that this task engages the mentalizing system.

Although this analyzed period of speech did not contain dyadic social interaction, live speech is a cue that often signals the start of social interaction, and, in this paradigm, was always contained within an interactive context (e.g., after listening to the live partner, the participant answered that partner's question and saw the partner give feedback). Thus, unlike studies targeting the intersubject neural synchronization that emerges during interaction (e.g., Dumas et al., 2010; Kawasaki et al., 2013; Koike et al., 2016; Stephens et al., 2010), this paradigm was developed to specifically determine the effects of social-interactive context on speech processing in a well-controlled design.

The current study's comparison between live and recorded speech will help dissociate between several possible patterns of developmental neural selectivity for social interaction. One possibility is that, even in a task without explicit mentalizing demands, children, like adults, show increased activation in regions associated with mentalizing during live speech. Such activation may be due to implicit, ongoing mentalizing about a social partner (Sperber & Wilson, 1996). Another possibility is that children recruit a more diffuse set of regions during live speech, a pattern consistent with functional specialization seen across other domains (reviewed in Johnson, 2011). Finally, children might show no differential activation to live versus recorded speech, suggesting that—at least for well-matched speech stimuli—similar neural mechanisms support the processing of social stimuli regardless of the live context. Consistent with the first possibility, we hypothesized that live interaction would engage regions of the mentalizing network in children. Specifically, given adult findings (Rice & Redcay, 2016), we predicted that sensitivity to live interaction would be strongest in DMPFC and TPJ. Additionally, given evidence that middle childhood corresponds to increased functional specialization within mentalizing regions, we hypothesized that there may be developmental changes in neural sensitivity to live versus recorded speech.

2. Methods

2.1 Participants

Twenty-six typical children aged 7-13 (15 females; average age=10.4 years, SD=1.7) were recruited to participate in the study from a database of local families. All children were full-term, native English speakers, with no history of neurological damage,

psychiatric disorders, head trauma, or psychological medications, no contraindications for MRI scanning, and none had first-degree relatives with autism or schizophrenia, as assessed via parent report. Three of the participants finished one or fewer runs of the experiment, due to general discomfort (1 participant) or discomfort with the headphones (2 participants). Thus, 23 participants (14 females; average age=10.6y, SD= 1.6) completed a sufficient number of runs to examine their behavioral data during the scan (i.e., accuracy and reaction time) and post-test questionnaire ratings.

Of the 23 participants with behavioral data, four participants' neuroimaging data were excluded due to motion (i.e., had more than two runs with over 3.5mm maximum frame displacement or with greater than 10% 1mm outliers). Thus, the final sample with both useable scan and behavioral data included 19 participants aged 7-13 (13 females; average age=10.9y; SD=1.6). All but two of these children were right-handed. Children who provided usable scan data were significantly older than the children who did not (10.9y versus 9.1y; $t(24)=2.57$, $p=.017$).

2.2 Social Interaction Experiment

2.2.1. Creating the live illusion

Although all audio and video stimuli in the experiment were actually prerecorded, a vital component of the design was that children believed that the Live condition was actually live and understood the conceptual difference between live and recorded stimuli. To establish the live illusion before the scan, the main experimenter and the child practiced talking over a truly live video-feed. The main experimenter then explained that the child would hear live and recorded speech during the scan. The child

listened to audio clips of the two recorded speakers: a friendly speaker matched to the Live condition (Social condition) and a more neutral speaker (Standard condition), which was included to ensure that the effect of perceived live speech was not due to differences in likeability or audio characteristics. All three speakers were adult females. All children correctly responded to comprehension questions about each speaker (e.g., “Was the speaker talking to you in real life?”).

2.2.2. Task design

Participants viewed 36 individual trials across 4 runs, 12 from each condition: Live, Social, and Standard (**Fig. 1**). Each trial began with a silent cue screen: either LIVE VOICE (in green text) or RECORDED (in orange text). After two seconds, the story (i.e., two-sentence vignette) began. In addition to the background cue screen, each condition had a different female speaker, to ensure children quickly understood when they were in a live trial. After a 2-4 second jittered fixation cross, participants were presented with a person’s preference (either that of the live speaker or of a third-party character) and then made a choice for that person by selecting one of two options, and, after another 2-4 second jitter, received feedback. Live feedback was a silent video of the live speaker that children believed was presented live video-feed, Social feedback was a standardized picture of a happy or sad female (Tottenham et al., 2009), and Standard feedback was a gold star or red “x”. Trial distribution and timing was determined by OptSeq (<http://surfer.nmr.mgh.harvard.edu/optseq/>), which optimized the estimation of the main effects of each condition. Further, baseline periods (i.e., fixation cross), each lasting 20 seconds, were added to beginning, middle, and end of each run.

2.2.3. Post-test procedure

Participants completed a post-test questionnaire verbally administered by a separate experimenter that assessed their impressions of the experiment and comprehension of the live setup. For each speaker, participants answered questions on a 1 to 5 Likert scale that assessed liveness, (i.e., “How much did it feel like this speaker was talking to you in real life?” and “How much did it feel like this speaker was in the room with you?”), likeability (“i.e., How much did you like this speaker?”), and engagement (i.e., “How much did you want to get the questions that the speaker was asking right?”). The two liveness questions were averaged to create a liveness composite. Further, all participants understood that the live speaker was talking directly to the child and could see the child’s answers, and that the recorded speakers were recorded previously and could not see the child’s answers. No children suspected the live stimuli to be recorded. At the end of the experiment, children and their parents were debriefed.

2.2.4. Stimuli

We piloted 103 Story-Question pairs on a sample of seven typical children (5 males), aged 8-11 (average=9.52, SD=1.6y). After this testing, 30 easy items were selected (on which accuracy was 100%) and six hard items were selected (on which accuracy ranged from 43% to 72%), in order to ensure that participants would see mostly positive feedback after answering questions.

The resulting 36 items were recorded by each of the three speakers (Live, Social, and Standard). These audio stimuli were identical to those used in the previous adult

version of this paradigm (Rice & Redcay, 2016). Thus, although in the adult version participants were listening to similarly-aged speakers (i.e., other adults), for the current study, children were not listening to peers. Each child was assigned one of three stimuli sets, which differed on which 12 short vignette and question pairs were assigned to each condition, and ensured that the total amount of time for each condition was matched. The order of the items was randomized within condition and the order of runs was counterbalanced.

2.3 Control Behavioral Paradigm

We included several differences between the Live and Social conditions to reinforce the live illusion, increase ecological validity, and ensure that children understood which condition they were in. Specifically, the Live but not Social condition included video feedback, first-person language (e.g., “I like”), and briefly meeting the speaker before the experiment. Although our analysis examined the matched audio portion—and not the video feedback or first-person language—we also investigated whether these other factors could produce perceptions of liveness without being told the speaker was live. We conducted a separate control behavioral-only study with $N=19$ typical child participants (7 males, average age=10.4 years) who completed the same task as the fMRI participants including meeting the “live” speaker before the experiment. All control participants, however, were told all stimuli were prerecorded. These control participants also completed the same post-test questionnaire. There were no differences between control participants and the scan participants who provided behavioral data ($N=23$) in age ($t(40)=.44$, $p=.66$) or sex ($\chi^2(1)=.023$, $p=.88$).

2.4 Image acquisition and processing

MRI imaging data were collected using a 12-channel head coil on a single Siemens 3.0-T scanner at the Maryland Neuroimaging Center (MAGNETOM Trio Tim System, Siemens Medical Solutions). The scanning protocol for each participant consisted of four runs of the main experiment (T2-weighted echo-planar gradient-echo; 36 interleaved axial slices; voxel size=3.0 x 3.0 x 3.3 mm; repetition time=2200ms; echo time=24ms; flip angle=90°; pixel matrix=64 x 64) and a single structural scan (three-dimensional T1 magnetization-prepared rapid gradient-echo sequence; 176 contiguous sagittal slices, voxel size=1.0 x 1.0 x 1.0 mm; repetition time=1900ms; echo time=2.52ms; flip angle=9°; pixel matrix= 256 x 256).

fMRI preprocessing was performed using AFNI (Cox, 1996). Data were first slice-time corrected and were aligned to the first volume using a rigid-body transform. The participant's high-resolution anatomical scan was also aligned to the first volume of the first run and then transformed to MNI space using linear and non-linear transforms. The resulting transformation parameters were applied to the functional data. Functional data were spatially smoothed using a 5mm full-width half-maximum (fwhm) Gaussian kernel and then intensity normalized such that each voxel had a mean of 100.

Outliers were defined as volumes in which the difference between two consecutive volumes exceeded 1mm (across translational and rotational movements) and such values were censored in subsequent analyses. Runs were excluded if the number of censored time points exceeded 10% of collected volumes or if total motion exceeded 3.5mm. Participants were included in analyses if they had at least 2 useable

runs. The final sample included one child with two runs, five children with three runs, and 13 children with four runs. Mean frame displacement was not correlated with age in the final sample ($r = -.19$, $p = .12$).

2.5 Data analysis

Response to each condition was analyzed using general linear models in AFNI. Given the long events of the current study and the lack of previous work on developmental response to live interaction, we made no assumptions about the shape of the hemodynamic response. We instead estimated responses for each condition using a cubic spline function beginning at the onset of the cue period and lasting for 24.2 seconds (lasting roughly through when participants answered the question). The spline function allows for a smoother estimation of response than ‘stick’ or finite impulse functions, although the two techniques are conceptually similar. Values were estimated at each TR, resulting in 12 estimated (Beta) values for each condition. Modeled events of no interest included the feedback period (due to differences in stimuli characteristics between the conditions), six motion parameters (x, y, z, roll, pitch, and yaw) and their derivatives, as well as constant, linear, and quadratic polynomial terms to model baseline and drift. To estimate response to speech in the three different conditions, we analyzed the period from Beta 4-6 (Story Window). The Story Window captured 6.8 to 11.2 seconds after story onset, and stories were, on average, around 6 seconds long. Thus, given the hemodynamic response, this window captured the bulk of the story while minimizing any effect of the initial cue screen before the story or preparation for answering the question.

We analyzed two specific contrasts in this Story Window: first, to examine the effect of live interaction we compared Live versus Social speech; second, to isolate the effects of speaker prosody and likeability, we compared Social versus Standard speech. Given that the Standard Story was not well-matched to the Live Story, that comparison was not of interest. Contrast maps were thresholded at a two-tailed $p < .005$, and cluster-corrected for multiple comparisons (overall $\alpha = .05$, $k = 28$) using AFNI's 3dClustSim.

To examine developmental change, we conducted both region of interest (ROI) analyses and whole-brain analyses in order to most fully explore any potential age-related effects. For the whole-brain analyses, we entered age as a covariate in the Live versus Social comparison. For the ROI analyses, we extracted each individual's Live and Social beta values for the Story Window within clusters that showed a significant group-level effect for Live versus Social. We then correlated these individual contrast values with age. Given that whole-brain results for the main effect of live versus recorded speech did not reveal activation in DMPFC, we used a DMPFC ROI defined based on the adult version of this paradigm (Rice & Redcay, 2016), to determine if age-related changes were responsible for the null finding.

3. Results

3.1 Behavioral results

Overall accuracy was high (mean=89.4%, SD=8.5%, Range=67-97%) and average reaction time was well within the 5s response window (mean=1.79s, SD=.51s, Range: 1.04-3.10). Although there were no significant differences in accuracy across conditions, comparison of reaction times indicated that children were significantly slower

at answering Live items as compared to Social items ($t(22)=2.11$, $p=.046$; **Table 1A**). There was no difference in reaction time between Social and Standard items ($t(22)=-1.39$, $p=.18$). Children became faster at answering questions with age, but this effect did not interact with condition type (main effect of age on overall RT: $F(1,21)=7.74$, $p=.011$; Condition x Age interaction on RT: $F<1$). In contrast, for participants in the control behavioral study—who were told that the Live condition’s stimuli were recorded—there were no between-condition differences in RT (**Supplementary Table 1**).

On the post-test questionnaire, children in the scan study perceived the Live condition as significantly more live than the Social condition (**Table 1B**), which in turn was perceived as more live than the Standard condition. There were no relations between any of the post-test rating measures and reaction time or between post-test ratings and age ($ps>.05$). Among the behavior-only control participants, who were told that the live stimuli were prerecorded, there were no differences in perceived likeability, engagement, or liveness between the Live and Social speaker, although both conditions were rated as more live than the Standard speaker. A repeated-measures ANOVA indicated a significant interaction between whether participants were told the Live speaker was actually live (i.e., whether a child was a control versus scan participant) and perceived liveness ($F(2,80)=5.48$, $p=.006$). This interaction was not significant for likeability or engagement.

3.2 Neuroimaging results

3.2.1. Main Effect of Live versus Recorded Social Stimuli

For the Story Window (Betas 4-6), whole-brain analyses revealed significantly

greater activation for Live than Social speech in regions often associated with mentalizing (Frith & Frith, 2006), including left TPJ and precuneus, although no differences in DMPFC activation were observed (**Table 2, Figure 2A**). Additionally, at a more liberal voxel-wise correction threshold of $p < .01$ (cluster-corrected $p < .05$), a significant cluster for Live versus Social speech also emerged in right pSTS (MNI coordinates=[58 66 -46], $k=58$, $t=4.09$). In contrast to the comparison of live versus matched recorded speech, comparison of the two recorded conditions, which also differed on subjective engagement and likeability, revealed no activation differences. Using the more liberal threshold of $p < .01$ (cluster corrected $p < .05$), two clusters were identified as more active for Social than Standard speech: one in lingual gyrus (MNI coordinates=[-4 -94 -18], $k=44$, $t=3.38$) and one in superior temporal gyrus (MNI coordinates=[68 -24 0], $k=47$, $t=3.43$), a region associated with pitch processing (e.g., Hyde et al., 2008; Scott et al., 2000).

For some of the significant Story Window clusters, we noted that differential response to the Live condition began before the analysis window, and thus followed a different time course than response to speech characteristics (i.e., the response when comparing the two recorded conditions; **Supplementary Figure 1**). This earlier response may capture cue-related differences between the live and recorded conditions. Thus, we conducted a post-hoc analysis for Betas 2-4 (Cue Window). The Cue Window corresponded from 4.4 seconds after the beginning of the cue (the two-second screen reading “Live Voice” or “Recording” before the start of the story) through 6.8 seconds after the start of the story. Similar to the Story Window, this analysis

revealed significantly increased activation in regions associated with mentalizing, including TPJ and precuneus (**Figure 2B**). Given that the cue screen was identical for both Social and Standard speech, we did not compare those conditions.

Although, as in previous adult work (Rice & Redcay, 2016), the TPJ was sensitive to live versus recorded speech, the current study did not employ a mentalizing localizer to assess whether the region of the TPJ recruited was involved selectively in mentalizing tasks. The TPJ has been implicated in domain-general processes beyond mentalizing, including attention (Decety & Lamm, 2007; Mitchell, 2008); however, previous studies have indicated that the region's roles in attention and mentalizing are spatially separable (Carter & Huettel, 2013; Scholz et al., 2009). Thus, we used the meta-analytic database Neurosynth (Yarkoni et al., 2011; www.neurosynth.org) to examine the peak TPJ coordinates from both the Story and Cue Window. Both clusters had a strong association with meta-analytic maps of *mentalizing* (Story Window: $z=4.54$, posterior probability=0.82; Cue Window: $z=4.76$, posterior probability=0.83), but not with maps for *attention*, *selective attention*, or *attentional control* ($z=0$ for all terms for both clusters).

3.2.3. Age-related Differences in Response to Live versus Recorded Social Stimuli

Using regions identified as more sensitive to Live than Social speech during the Story Window, there were no significant relations between age and activation to Live versus Social conditions, nor to Live or Social speech versus baseline. Whole-brain analyses of the Story Window indicated no significant effects of age on processing Live versus Social speech, although a more liberal voxel-wise correction of $p<.01$ (cluster

corrected $p < .05$) did reveal a significant cluster in left superior frontal gyrus (MNI coordinates= $[-24\ 48\ 18]$, $k=42$, $t=-4.15$; **Supplementary Figure 2**). Post-hoc analyses examining each condition versus baseline within this cluster indicated that response to recorded speech increased with age (Social: $r=.49$, $p=.03$; Standard: $r=.60$, $p=.007$), but response to Live speech was unchanged ($r=-.17$, $p=.50$). For the Cue Window, whole-brain and ROI analyses for the Live versus Social contrast showed no relation between activity and age.

Given the unexpected whole-brain finding that DMPFC was not more active for Live than Social speech in the Story window, we conducted additional analyses to determine if age-related changes obscured differences in DMPFC activation. Specifically, we used the peak coordinates of right and left DMPFC activation from adults in this same paradigm (Rice & Redcay, 2016) in order to create spherical ROIs with 6mm radii. Then, within both DMPFC ROIs, we extracted each child's response to each condition's speech during the Story Window. Left DMPFC activation for *recorded* speech increased with age (Social: $r=.48$, $p=.04$; Standard: $r=.41$, $p=.081$), but sensitivity to live speech did not change ($r=.02$, $p=.94$). Right DMPFC activation, however, was not related to age.

4. Discussion

This study investigated the neural mechanisms supporting social interaction in middle childhood. Specifically, we used a well-controlled fMRI paradigm—a paradigm that engages the mentalizing network in adults—in order to compare the brain's response to two types of matched speech: speech that children believed was coming

from a live social partner (Live) and speech that children believed was recorded (Social). Behavioral results indicate that children understood the distinction between live and recorded speech and perceived the live speaker to be significantly more live (e.g., felt like she was in the same room). Consistent with previous research examining the neural correlates of social interaction (e.g., Hampton et al., 2008, Kampe et al., 2003; Redcay et al., 2010; Rice & Redcay, 2016), simply believing that speech was live resulted in increased activation in social cognitive regions frequently associated with mentalizing, including precuneus and TPJ. Additional control analyses and experiments suggested that this difference in activation was unlikely to be attributable to differences in low-level audio characteristics, speaker identity, or speaker likeability. These results indicate that neural sensitivity to interactive contexts is present by middle childhood and that mentalizing systems may support on-going social interaction.

Post-hoc examination of neural response to the “cue” screen (which informed participants whether they were about to hear live or recorded speech), also suggested that social cognitive brain regions, including TPJ, were differentially activated by potential live interaction. These findings suggest a possible preparatory response in mentalizing regions, perhaps in anticipation of needing to consider the mental states of a social partner. The current study’s design, however, did not allow for dissociation between response to the cue and the beginning of speech. Future studies should dissociate these mechanisms, to determine if preparatory mentalizing employs different neural substrates than mentalizing during an on-going interaction.

Findings for both the Cue and Story Window provide mixed evidence for

developmental continuity in the neural mechanisms supporting social interaction. Like adults, children showed increased activity in regions associated with mentalizing, specifically TPJ, when processing live versus recorded speech—speech with no explicit mentalizing demands—suggesting a role for automatic mentalizing during interaction. Also similar to adults, mentalizing, but not reward or attentional regions, were sensitive to live speech. Unlike adults, however, children showed no significant differential response to live versus recorded speech in DMPFC. This null finding is surprising given that DMPFC is consistently implicated as sensitive to social context across both interactive and ‘offline’ paradigms (reviewed in Van Overwalle, 2011).

One potential explanation for the lack of DMPFC activation is not a lack of response to live stimuli, but rather changes in response to recorded stimuli. In this middle childhood sample, DMPFC response to recorded social stimuli increased with age, whereas response to live stimuli remained constant. The finding of increased activation to recorded stimuli is consistent with past studies of ‘offline’ social cognition, which have found higher DMPFC response to non-interactive social stimuli in adolescence than adulthood (e.g., reviewed in Blakemore, 2008). Thus, one speculative possibility is that DMPFC response to live speech is early-emerging and relatively invariant across age, whereas early adolescence represents a time of peak sensitivity to communicative cues regardless of the interactive context. That is, perhaps the end of middle childhood corresponds to a general increase in social sensitivity that extends broadly to all social stimuli, including recorded human speech. Future work should examine larger samples and compare response to live versus recorded stimuli across a

variety of modalities and ages.

In contrast to DMPFC, the current study did not find age-related changes in selectivity for live interaction in TPJ or precuneus—regions associated with the mentalizing network that were more active for live versus recorded speech. This null result is in contrast to research finding increased middle childhood specialization for explicit mentalizing in similar regions (Gweon et al., 2012). One possible explanation for this discrepancy is although similar brain regions are implicated in explicit and implicit mentalizing (Kovacs et al., 2014; Schneider et al., 2014), specialization for the more implicit mentalizing required by on-going interaction—the type displayed in interactive contexts even by very young children—happens before explicit specialization. Perhaps regions implicated in explicit mentalizing have an ontogenetically-prior role in supporting social interaction more broadly (e.g., Grossmann & Johnson, 2010) and, later in development, become loci of explicit mentalizing (e.g., false belief tasks) due to children’s cumulative experiences employing mental state reasoning during social interaction. The current data, however, cannot speak directly to these possibilities.

The behavioral data also indicated that children were sensitive to the distinction between live and recorded stimuli and were slower to respond to questions from the live speaker. Although the exact mechanism spurring slower responses to a live partner is unknown, one possibility—consistent with the brain data—is that children engaged in more mentalizing or different mentalizing about the live social partner. Consistent with this explanation, no difference in reaction time emerged when participants in a control study were told that the live stimuli were recorded. Future research involving

interference tasks could help determine if increased mentalizing is the predominant cognitive mechanism driving behavioral differences in responding to live versus recorded partners (e.g., Kiesel et al., 2010).

Interpretation of the current results is complicated by the fact that regions in the mentalizing network are involved in processes beyond mentalizing, spanning both social (e.g., animacy detection, Shultz & McCarthy, 2014; narrative processing; Mar, 2011) and non-social (e.g., attention; Decety & Lamm, 2007; Mitchell, 2008) domains. Although comparison of this study's pattern of results to a meta-analytic database show strong overlap with other studies of mentalizing, future developmental research could adopt a localizer approach in order to isolate, for each individual, regions involved in mentalizing or language processing (see Rice & Redcay, 2016). Additionally, although the live and recorded speech was content-matched, the current paradigm did have several differences between conditions, designed to heighten the saliency of the live social partner. Although both the comparison between the two recorded conditions and the results from the control behavioral study suggest that differences in audio characteristics or speaker likeability are not responsible for the observed results, future research should develop even more well-controlled paradigms. Ultimately, however, it is possible that isolating the effect of social interaction will be difficult, as attentional or motivational processes may be inextricably linked to the emergent properties of real-world social interaction (e.g., see Koike et al., 2016 for evidence that social partners' eyeblinks become synchronized). The current study was not well-suited to examine such emergent properties, as the analyzed portion of live speech did not involve a

temporally-unfolding interaction between actors. Future research should continue to make social neuroscience more interactive by examining two or more social partners simultaneously (e.g., Dumas et al., 2010; Kawasaki et al., 2013).

Overall, the current study provides some of the first developmental evidence that neural response to otherwise matched social stimuli is modulated by social-interactive context. The brain regions sensitive to a live social partner in children are similar, but not identical, to those identified in adult studies of social interaction (e.g., Kampe et al., 2003; Redcay et al., 2010; Rice & Redcay, 2016), suggesting that mentalizing network activity accompanies social interaction across development. Specialization for live interaction did not increase with age in this sample, but paradigms with more complex dyadic interaction may yet reveal specialization. Although future research should more finely parse the components of live interaction that may be driving the current findings, the current study is a novel first step in embedding developmental social neuroscience in the social world.

References

- Apperly, I. A., Warren, F., Andrews, B. J., Grant, J., & Todd, S. (2011). Developmental continuity in theory of mind: Speed and accuracy of belief–desire reasoning in children and adults. *Child Development, 82*(5), 1691-1703.
- Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience, 9*(4), 267-277.
- Bolling, D. Z., Pitskel, N. B., Deen, B., Crowley, M. J., Mayes, L. C., & Pelphrey, K. A. (2011). Development of neural systems for processing social exclusion from childhood to adolescence. *Developmental Science, 14*(6), 1431-1444.
- Carr, A. (2011). Social and emotional development in middle childhood. In D. Skuse, H. Bruce, L. Dowdney, & D. Mrazek (Eds.), *Child Psychology and Psychiatry: Frameworks for Practice* (2nd ed., pp. 56-61). Hoboken, NJ: Wiley-Blackwell.
- Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in cognitive sciences, 17*(7), 328-336.
- Chein, J., Albert, D., O'Brien, L., Uckert, K., & Steinberg, L. (2011). Peers increase adolescent risk taking by enhancing activity in the brain's reward circuitry. *Developmental Science, 14*(2), F1-F10.
- Clark, H.H. (1996). *Using Language*. Cambridge, Cambridge University.
- Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences, 106*(23), 9163-9168.

- Cox, R.W. 1996. AFNI: software for analysis and visualization of functional magnetic resonance NeuroImages. *Computer Biomed. Res.* 29, 162-173.
- Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist*, 13, 580-593.
- Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-brain synchronization during social interaction. *PloS One*, 5(8), e12166.
- Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 13(2), 331-338.
- Farmer, T. W., Irvin, M. J., Motoca, L. M., Leung, M., Hutchins, B. C., Brooks, D. S., & Hall, C. M. (2015). Externalizing and internalizing behavior problems, peer affiliations, and bullying involvement across the transition to middle school. *Journal of Emotional and Behavioral Disorders*, 23, 3-16.
- Feiring, C., & Lewis, M. (1991). The transition from middle childhood to early adolescence: Sex differences in the social network and perceived self-competence. *Sex Roles*, 24, 489-509.
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531-534.
- Gallagher, H.L., Jack, A.I., Roepstorff, A., & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, 163, 814-821.
- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19, 515-523.

- Grossmann, T., & Johnson, M. H. (2010). Selective prefrontal cortex responses to joint attention in early infancy. *Biology Letters*, 6(4), 540-543.
- Guyer, A. E., Choate, V. R., Pine, D. S., & Nelson, E. E. (2012). Neural circuitry underlying affective response to peer feedback in adolescence. *Social Cognitive and Affective Neuroscience*, 7(1), 81-92.
- Guyer, A. E., McClure-Tone, E. B., Shiffrin, N. D., Pine, D. S., & Nelson, E. E. (2009). Probing the neural correlates of anticipated peer evaluation in adolescence. *Child Development*, 80(4), 1000-1015.
- Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, 83(6), 1853-1868.
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, 105(18), 6741-6746.
- Heimberg, R. G., Brozovich, F. A., & Rapee, R. M. (2010). A cognitive-behavioral model of social anxiety disorder: Update and extension. *Social anxiety: Clinical, developmental, and social perspectives*, 2, 395-422.
- Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2), 632-639.
- Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development?. *Developmental Cognitive Neuroscience*, 1(1), 7-21.

- Kampe, K.K., Frith, C.D., & Frith, U. (2003). “Hey John”: signals conveying communicative intention toward the self activate brain regions associated with “mentalizing,” regardless of modality. *Journal of Neuroscience*, 23, 5258-5263.
- Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., & Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific Reports*, 3, 1-8.
- Kirschner, S., & Tomasello, M. (2009). Joint drumming: Social context facilitates synchronization in preschool children. *Journal of Experimental Child Psychology*, 102(3), 299-314.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, 136(5), 849.
- Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358, 345-360.
- Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., ... & Sadato, N. (2016). Neural substrates of shared attention as social memory: A hyperscanning functional magnetic resonance imaging study. *NeuroImage*, 125, 401-412.
- Kovács, A. M., Kühn, S., Gergely, G., Csibra, G., Brass, M. (2014). Are All Beliefs Equal? Implicit Belief Attributions Recruiting Core Brain Regions of Theory of Mind. *PLOS ONE*, 9(9), e106558.

- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, 100(15), 9096-9101.
- Laidlaw, K. E., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, 108(14), 5548-5553.
- Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62, 103-134.
- McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, 98(20), 11832-11835.
- Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, 18, 262-271.
- Monahan, K. C., & Steinberg, L. (2011). Accentuation of individual differences in social competence during the transition to adolescence. *Journal of Research on Adolescence*, 21(3), 576-585.
- Mundy, P., & Neal, A. R. (2000). Neural plasticity, joint attention, and a transactional social-orienting model of autism. *International review of research in mental retardation*, 23, 139-168.
- Okita, S.Y., Bailenson, J., Schwartz, D.L. (2007). The mere belief of social interaction improves learning. In: McNamara, D.S., Trafton, J.G. (Eds.), *The Proceedings of the 29th Meeting of the Cognitive Science Society*, pp. 1355–1360.

- Pfeiffer, U.J., et al. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, 101, 124-137.
- Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, 6(4), 486-494.
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., & Saxe, R. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage*, 50(4), 1639-1647.
- Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: the neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience*, 6, 169.
- Rice, K., & Redcay, E. (2016). Interaction matters: a perceived social partner alters the neural processing of human speech. *NeuroImage*, 129, 480-488. doi: 10.1016/j.neuroimage.2015.11.041
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah, N. J., Fink, G. R., & Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, 22, 2702-2715.
- Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36, 393-414.

- Schneider, D., Slaughter, V. P., Becker, S. I., & Dux, P. E. (2014). Implicit false-belief processing in the human brain. *NeuroImage*, 101, 268-275.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PloS one*, 4(3), e4869.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400-2406.
- Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., ... & Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience*, 28, 3586-3594.
- Shultz, S., & McCarthy, G. (2014). Perceived animacy influences the processing of human-like surface features in the fusiform gyrus. *Neuropsychologia*, 60, 115-120.
- Sperber, D., & Wilson, D. (1996). *Relevance: Communication and Cognition*, second ed. Blackwell, Oxford.
- Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32), 14425-14430.
- Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S., & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology*, 19(15), 1274-1277.

- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(05), 675-691.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., ... & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, 168(3), 242-249.
- Will, G. J., van Lier, P. A., Crone, E. A., & Güroğlu, B. (2016). Chronic childhood peer rejection is associated with heightened neural responses to social exclusion during adolescence. *Journal of Abnormal Child Psychology*, 44, 43-55.
- Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage*, 54(2), 1589-1599.
- Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665-670.

Table 1. Behavioral performance and post-test questionnaire ratings.

	Live	Social	Standard	$F_{(2,44)}$	Pairwise Comparisons
A. BEHAVIORAL PERFORMANCE					
Accuracy (%)	89.61 (12.74)	90.46 (8.8)	88.04 (15.45)	.263	Live=Soc=Std
RT (ms)	1847 (580)	1718 (474)	1811 (601)	1.63	Live>Soc=Std
B. POST-TEST QUESTIONNAIRE RATINGS					
Liveness	4.16 (.70)	2.75 (1.29)	2.39 (1.34)	29.37***	Live>Soc>Std
Likeability	4.46 (.69)	4.01 (.96)	3.44 (1.07)	12.53***	Live>Soc>Std
Engagement	4.74 (.60)	4.38 (.83)	4.17 (.94)	6.42**	Live>Soc=Std

Note: Values are mean (standard deviation). All post-test questionnaire ratings are composites of items scored on a 1 to 5 scale. Post-hoc pairwise comparisons were made using a Tukey's test with an alpha of .05. **, $p < .01$; ***, $p < .001$. Soc=Social; Std=Standard.

Table 2. *Regions Sensitive to Live Interaction.*

Region	Side	Peak	Cluster	MNI Coordinates		
		<i>t</i>	<i>k</i>	<i>x</i>	<i>y</i>	<i>z</i>
A. STORY WINDOW						
1. Effect of live speech (matched content and prosody)						
Live>Social						
Precuneus	L/R	3.38	194	0	-72	36
Occipital gyrus	R	7.32	97	32	-90	20
TPJ	L	4.28	35	-48	-66	44
Social>Live						
None						
2. Effect of friendly speech (both recorded)						
Social>Standard						
None						
Standard>Social						
None						
B. CUE WINDOW						
Effect of live speech (matched content and prosody)						
Live>Social						
Precuneus/PCC	R	3.47	316	2	-76	44
TPJ	R	3.52	90	60	-54	36
Lingual gyrus	L/R	3.35	89	0	-78	0
Angular gyrus	L	3.36	29	-34	-70	50
Social>Live						
None						

Note. TPJ=temporal parietal junction; PCC=posterior cingulate cortex. Coordinates in Montreal Neurologic Institute (MNI) space. First corrected at $p<.005$ and cluster corrected at $p<.05$ ($k=28$).

Figures

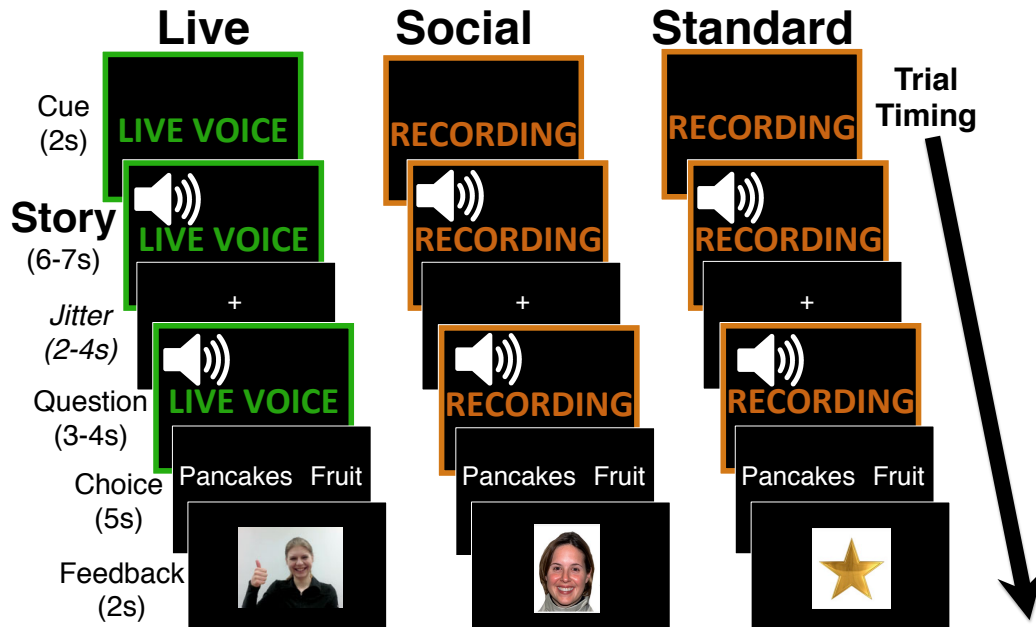


Figure 1. Experimental trial structure. Each of the three conditions (Live, Social, Standard) is represented in a column. Children believed that the Live condition was presented via a real-time audio-feed by an experimenter who could see their answers and that the other two conditions were recorded. In each trial, after the Cue screen, children heard a two-sentence Story that presented two options with no mention of social information (e.g., “There are two things on the breakfast menu. One is pancakes and one is a bowl of fruit.”). After this Story, children heard a question, either about the Live speaker or, for the Social and Standard conditions, about a third-party character (“I/Megan am/is trying to eat healthy. Which food should I/she eat?”). After answering the question, children saw feedback dependent on their answer. Analyses focused on the matched Story portion. s=seconds.

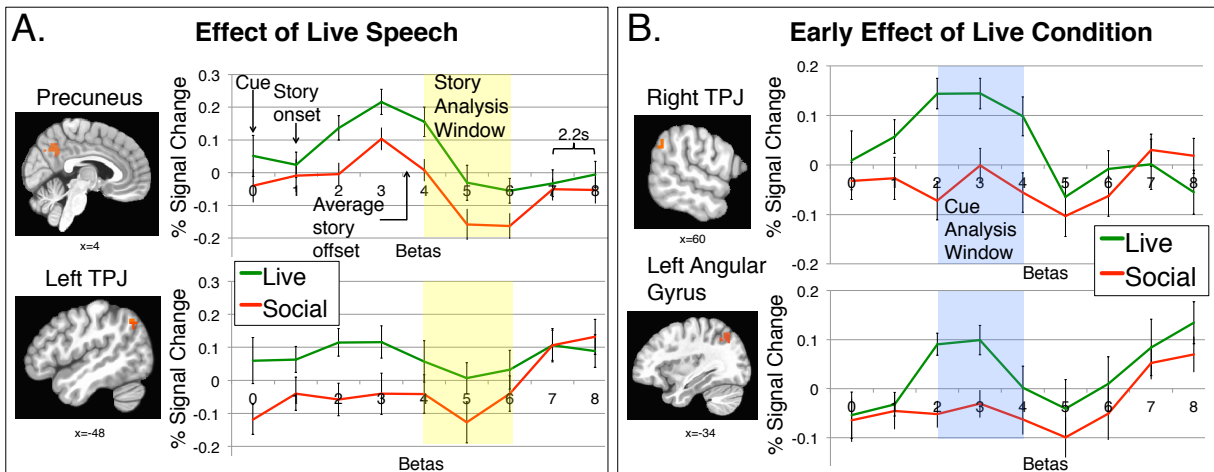


Figure 2. Time series plots for selected clusters defined from group-level comparisons ($p < .05$ corrected). Individual spline estimates versus baseline for a subset of regions defined by the (A) Live > Social comparison in the Story Analysis Window (Betas 4-6) and (B) Live > Social comparison in the Cue Analysis window (Betas 2-4). Error bars represent standard error of the mean. s=seconds; TPJ=temporal parietal junction (See also Table 2).

SUPPLEMENTARY MATERIAL

Supplementary Table 1

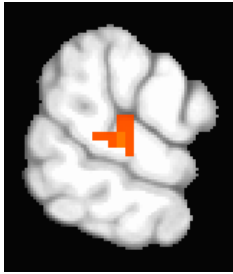
Behavioral Performance and Post-test Questionnaire Ratings for the Control Behavioral Experiment

	Live	Social	Standard	$F_{(2,36)}$	Pairwise Comparisons
BEHAVIORAL PERFORMANCE					
Accuracy (%)	84.72 (11.16)	82.87 (19.3)	86.57 (11.12)	.442	Live=Soc=Std
RT (ms)	1384 (244)	1421 (360)	1438 (312)	.348	Live=Soc=Std
POST-TEST QUESTIONNAIRE RATINGS					
Liveness	3.42 (.95)	3.13 (.85)	2.39 (.88)	9.89***	Live=Soc>Std
Likeability	3.84 (1.07)	4.00 (1.20)	3.39 (1.12)	2.02	Live=Soc=Std
Engagement	4.58 (.61)	4.47 (.77)	4.16 (.83)	2.11	Live=Soc=Std

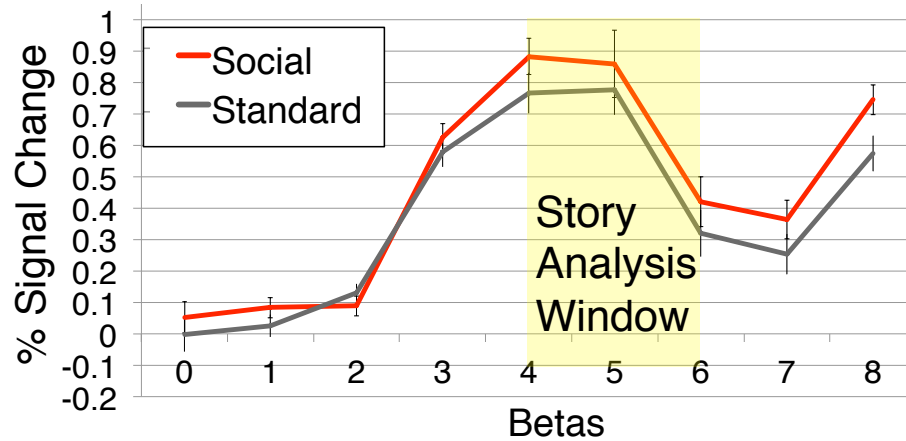
Note: Values are mean (standard deviation). Due to a technical error, task data were not recorded for one participant, resulting in N=18 for accuracy and RT measures and N=19 for post-test ratings. All post-test questionnaire ratings are based on items scored on a 1 to 5 scale. The control study had a 3s response window due to timing constraints at the behavioral session. Post-hoc pairwise comparisons were made using a Tukey's test with an alpha of .05. ***, $p < .001$. Soc=Social; Std=Standard.

Control Comparison

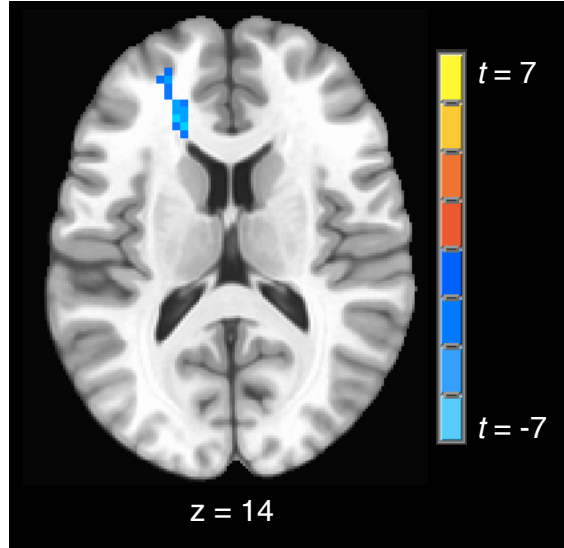
Right STG



x=65



Supplementary Figure 1. Time series plot for a selected cluster defined from group-level comparison of the two recorded speech conditions. Individual spline estimates are depicted for a region active for the Social > Standard comparison in the Story Window (Betas 4-6; voxel-wise correction of $p < .01$ and cluster correction of $p < .05$). Error bars represent standard error of the mean. s=seconds; STG=superior temporal gyrus.



Supplementary Figure 2. Whole-brain analysis examining the relation between age and the magnitude of the Live > Social contrast. Using a voxel-wise correction of $p < .01$ and cluster correction of $p < .05$, a single significant cluster was identified. This cluster demonstrated a negative relation between age and sensitivity to live speech.