Systems/Circuits

Cohesion and Joint Speech: Right Hemisphere Contributions to Synchronized Vocal Production

Kyle M. Jasmin,[1,2] Carolyn McGettigan,[1,3] Zarinah K. Agnew,[4] Nadine Lavan,[3] Oliver Josephs,[5] Fred Cummins,[6] and Sophie K. Scott[1]

[1]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom, [2]Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20492, [3]Department of Psychology, Royal Holloway, University of London, London TW20 0EX, United Kingdom, [4]Department of Otolaryngology, University of California–San Francisco, San Francisco, California 94143, [5]Institute of Neurology, University College London, London WC1N 3BG, United Kingdom, and [6]School of Computer Science, University College Dublin, Dublin 4, Ireland

Synchronized behavior (chanting, singing, praying, dancing) is found in all human cultures and is central to religious, military, and political activities, which require people to act collaboratively and cohesively; however, we know little about the neural underpinnings of many kinds of synchronous behavior (e.g., vocal behavior) or its role in establishing and maintaining group cohesion. In the present study, we measured neural activity using fMRI while participants spoke simultaneously with another person. We manipulated whether the couple spoke the same sentence (allowing synchrony) or different sentences (preventing synchrony), and also whether the voice the participant heard was "live" (allowing rich reciprocal interaction) or prerecorded (with no such mutual influence). Synchronous speech was associated with increased activity in posterior and anterior auditory fields. When, and only when, participants spoke with a partner who was both synchronous and "live," we observed a lack of the suppression of auditory cortex, which is commonly seen as a neural correlate of speech production. Instead, auditory cortex responded as though it were processing another talker's speech. Our results suggest that detecting synchrony leads to a change in the perceptual consequences of one's own actions: they are processed as though they were other-, rather than self-produced. This may contribute to our understanding of synchronized behavior as a group-bonding tool.

Key words: coordinated action; fMRI; joint speech; right hemisphere; social cohesion; speech control

Significance Statement

Synchronized human behavior, such as chanting, dancing, and singing, are cultural universals with functional significance: these activities increase group cohesion and cause participants to like each other and behave more prosocially toward each other. Here we use fMRI brain imaging to investigate the neural basis of one common form of cohesive synchronized behavior: joint speaking (e.g., the synchronous speech seen in chants, prayers, pledges). Results showed that joint speech recruits additional right hemisphere regions outside the classic speech production network. Additionally, we found that a neural marker of self-produced speech, suppression of sensory cortices, did not occur during joint synchronized speech, suggesting that joint synchronized behavior may alter self-other distinctions in sensory processing.

Introduction

In all human cultures, synchronous activities are socially important collective behaviors, which play a central role in establishing and promoting social cohesion (McNeill, 1995). Thus, the U.S. Congress recites the Pledge of Allegiance in unison, protestors shout phrases together; and although armies have not marched into battle in a century, soldiers still train by marching and singing synchronously. Participants are more likely to cooperate on a

Received Nov. 12, 2015; revised Jan. 28, 2016; accepted Feb. 18, 2016.

Author contributions: K.M.J., C.M., Z.K.A., F.C., and S.K.S. designed research; K.M.J., C.M., Z.K.A., N.L., O.J., and F.C. performed research; K.M.J. and F.C. analyzed data; K.M.J. and S.K.S. wrote the paper.

This work was supported by a University College London National Institute of Mental Health Joint Doctoral Training Program in Neuroscience Award to K.M.J., and a Wellcome Trust Senior Research Fellowship in Biomedical (no. WT090961MA) Science to S.K.S.

The authors declare no competing financial interests.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to Dr. Sophie K. Scott, 17 Queen Square, London WC1N 3AR, United Kingdom. E-mail: sophie.scott@ucl.ac.uk.

DOI:10.1523/JNEUROSCI.4075-15.2016

task after playing music together (Anshel and Kipper, 1988; Wiltermuth and Heath, 2009; Kirschner and Tomasello, 2010), and simply tapping a finger synchronously with another person increases feelings of affiliation (Hove and Risen, 2009; Wiltermuth, 2012) while walking in step together boosts confidence (Fessler and Holbrook, 2014).

Synchrony is especially apparent in verbal interactions. For example, in a conversation, participants tend to change their posture and fixate their gaze at around the same time (Shockley et al., 2003; Richardson et al., 2007), and this alignment aids mutual understanding (Shockley et al., 2009). Mutual understanding, or "common ground" (Clark and Brennan, 1991; Clark, 1996), has been shown to correlate with synchrony at the neural level: when listening to a speaker tell a story, the listener's brain activity is synchronized with that of the storyteller's, both simultaneously and at a delay, and the degree of coordination predicts understanding and alignment of mental states (Stephens et al., 2010; Kuhlen et al., 2012).

Our aim was to investigate the neural basis of joint behavior using a joint-speaking task in which two people spoke sentences synchronously with each other. Consistent with its relative frequency in real life (e.g., in joint prayer, protest, and chant), synchronous speaking requires little or no practice and can be performed accurately with a mean latency of ~40 ms (Cummins, 2002, 2003, 2009). In our experiment, participants read sequences of nonmetric speech (i.e., speech with no overt, regular rhythm), as synchronous speech is easy to perform even in the absence of an isochronous (even) rhythm, the speech produced does not need to have a regular metrical structure. Thus, whole groups of people can repeat spontaneous speech in unison, to spread a message through a crowd (King, 2012). To isolate brain activity that was specific to two-way synchrony, with participants mutually adapting their movement to each other, we manipulated whether participants synchronized with the experimenter's live voice in real-time or with a prerecorded version. Importantly, participants were not aware that any recordings would be used in the experiment at all, so any differences in brain activity observed between these conditions must reflect processes outside of conscious awareness.

Under joint speech conditions, we predicted a greater involvement of the posterior "how" pathway connecting auditory and motor cortices, than activation associated with speaking aloud. Auditory processing involves distinctly different anatomical and functional pathways; these include anterior and posterior auditory fields associated, respectively, with recognition/categorization processes (anterior "what" fields) and sensorimotor processes (posterior "how" pathways), which are key to speech production and modulation of speech production (Wise, 2001; Takaso et al., 2010). From an auditory streams perspective, we therefore predicted a greater involvement of the posterior "how" pathway under joint speech conditions relative to speaking aloud. More distributed neural systems have also been identified in studies investigating finger-tapping to computer-generated rhythms, with activation found in sensory and motor regions, such as the cerebellum, basal ganglia, and sensorimotor, premotor. and inferior parietal cortex (Repp, 2005; Repp and Su, 2013). Using EEG, Tognoli et al. (2007) asked participants to produce rhythmic finger movements with another person, under conditions where the other's finger was visible, and when not. When participants could see each other, their movements spontaneously coordinated, which was associated with a change in neural activity in right centroparietal cortex. Oscillations in right centroparietal cortex synchronize across the brains of participants (Dumas et al., 2010), and gamma power in parietal cortex is increased when they freely imitate each other's hand (Dumas et al., 2012).

Materials and Methods

Experiment 1. Eighteen (18) participants (mean age 28 years, SD 8.8 years; 7 females) were tested, who all gave informed consent. Participants were right-handed speakers of British English with no hearing or speech problems (by self-report). The protocol for the experiment was approved by the University College London Psychology Research Ethics Panel, in accordance with the Declaration of Helsinki. The birth date of one participant was not recorded correctly and is omitted from this mean ± SD.

The model texts were five sentences adapted from The Rainbow Passage (Fairbanks, 1960). The sentences were about the same length and could be spoken comfortably during a short presentation window (mean no. of syllables = 20.8 ± 1.3). Sentence 6 (20 syllables) was spoken only by the experimenter, in the Diff-Live condition (see below). Prerecorded utterances were obtained under conditions of synchronous speaking between the experimenter and another British English speaker, to ensure that the recorded sentences were acoustically similar to sentences the experimenter spoke live during the scanning session.

The sentence texts were as follows:

1. When sunlight strikes raindrops in the air, they act as a prism and form a rainbow.
2. There is, according to legend, a boiling pot of gold at one end of a rainbow.
3. Some have accepted the rainbow as a miracle without physical explanation.
4. Aristotle thought that the rainbow was a reflection of the sun's rays by the rain.
5. Throughout the centuries, people have explained the rainbow in various ways.
6. A rainbow is a division of white light into many beautiful colors.

Task. There were three main conditions that required the participant to speak along with the sound of the experimenter's voice:

- Synch-Live—The participant spoke a model sentence synchronously with the experimenter.
- Synch-Rec—The participant spoke a model sentence synchronously with a recording of the experimenter.
- Diff-Live—The participant spoke a model sentence while they heard the experimenter speak Sentence 6.

In three other conditions, the participant did not speak simultaneously with the experimenter:

- Listen-Alone—The participant heard a recording of the experimenter speaking the model sentence.
- Speak-Alone—The participant spoke the model sentence on their own.
- Rest—No stimulus was presented. The participant did nothing and waited for the next instruction.

Participants were briefed on the task: they were told that they would see sentences presented on the screen, and, depending on the instruction prompt, should (1) speak along with the experimenter while trying to synchronize as closely as possible, (2) speak alone, (3) listen to the experimenter, (4) speak while the experimenter spoke a different sentence, (5) or rest. They also had a chance to briefly practice speaking synchronously with the experimenter. Finally, when it was clear the subject understood their instructions, they were equipped with an MRI-safe microphone and headset and placed in the MRI scanner. Before each trial, participants saw an instruction, which told them what to do with the sentence they would see next: (1) speak it aloud synchronously with the experimenter's voice ("SYNCHRONIZE" displayed on screen); (2) speak it aloud while the experimenter spoke a different sentence ("DO NOT SYNCH" displayed on screen); (3) speak it alone ("SPEAK"); (4) listen to the experimenter speak the sentence ("LISTEN"); or (5) do nothing ("REST"). After 3 s, the instruction disappeared, a fixation cross appeared for 1 s, then the

short sentence appeared in its place and stayed on the screen for 6 s while the participant executed the instruction. On half of the trials where the participant synchronized, the experimenter was speaking in unison at the same time ('Synch-Live'); thus, as the participant attempted to synchronize with the experimenter, both parties could reciprocally influence each other. In a covert manipulation (of which the participant was unaware), the other half of the trials required the participant to synchronize with a nonreciprocating voice, a prerecorded presentation of the same experimenter speaking synchronously with a different partner on a different day (Synch-Rec). This crucial control isolated the neural profile of mutual synchronization while keeping the auditory and motor components of the task as similar as possible across conditions. The prompts for the Synch-Live and Synch-Rec conditions were identical to ensure the participants were naive to our manipulation throughout the experiment. In debriefing, it was confirmed that participants were indeed unaware that they had heard any recordings during the course of the experiment.

In the design, the 5 main model sentences were fully crossed with each of the 5 conditions (i.e., excepting Rest). Each combination of model sentence and condition appeared twice in an experimental run. The order of conditions was pseudo-randomized such that every 5 trials included one trial in each condition and one with each of the 5 model sentences. Each functional run consisted of 55 trials, including 10 from each of the 5 main conditions, and 5 Rest trials. Each participant participated in 3 functional runs.

Apparatus. The subject was equipped with a FOMRI-III noise cancelling microphone from Optoacoustics and insert earphones by Sensimetrics. The experimenter sat at a table inside the control room in front of a microphone and with over-ear headphones. Visual prompts and audio recordings were presented with a Macbook Pro running MATLAB and PsychToolbox (Brainard, 1997). An Alto ZMX-52 Mixer was used to direct audio signals between the subject's and experimenter's microphones and earphones, and also to present prerecorded audio from the stimulus presentation computer. The subject and experimenter heard each other in both ears. All audio output from the Alto ZMX-52 was saved to the experimental laptop, with the experimenter's voice (live or prerecorded) in channel 1 and the subject's microphone in channel 2 of a WAV file. Separate recordings were generated for each trial.

Only the experimenters were aware that some "Synchronize" trials would involve synchronizing with a recording. To alert the experimenter that an upcoming trial was "live" and she would have to speak, a color code was used such that the prompt appearing on the laptop (and scanner room projector) was yellow in the Synch-Live condition, and blue in the Synch-Rec condition. To disguise this code to the subject, the color of the prompt (yellow or blue) for prompts in all other conditions was also varied such that the prompt was yellow or blue half the time, with the order varied randomly.

MRI data collection. Subjects were scanned with a Siemens Avanto 1.5 tesla MRI scanner with a 32-channel head coil. The functional data were collected using a dual-echo, sparse-clustered EPI pulse sequence (40 slices, slice time 92.9 ms, 25 degree tilt transverse to coronal, ascending sequential acquisition; $3 \times 3 \times 3$ mm voxel size) with a 14 s intertrial interval that included a 6.5 s silent period to allow speech production and the presentation of auditory stimuli in quiet. This quiet period was followed by 2 volume acquisitions, each with an acquisition time (TA) of 3.7 s. Gradient echo images were collected at 24 and 58 ms. To acquire better signal in anterior temporal regions, a *Z*-shim of 2 millitesla/m was applied to the first echo time. After collected functional runs, a high-resolution T1-weighted structural scan was collected (160 slices, sagittal acquisition, 1 mm isotropic voxels).

MRI data processing. Preprocessing was performed with SPM8 (www.fil.ion.ucl.ac.uk/spm). Each participant's fMRI time series was spatially realigned to the first volume of the run, coregistered to their anatomical T1-weighted image, normalized into MNI space using parameters obtained from unified segmentation of the anatomical image, resampled at $3 \times 3 \times 3$ mm, and smoothed with an 8 mm Gaussian kernel. Only the first volume from each cluster was analyzed because much of the hemodynamic response had dissipated by the end of the first volume's TA. Although images at two echo times (TE) were acquired, we chose to simplify the analysis by only analyzing the images from the second TE, which, at 58 ms, is close to the standard used in most EPI sequences.

Statistical analysis. At the single-subject level, SPM8 was used to model event-related responses for each event type with a canonical hemodynamic response function. This was done by setting a repetition time of 14 s, which was divided into 32 time bins of 438 ms each. The micro-time onset was set to occur with the middle slice of the first volume acquisition (at bin number 18, i.e., 7884 ms after the beginning of the stimulus presentation window). The event duration was set as 6 s (which was the length of time participants spoke and heard speech in a trial). Motion parameters for the 6 degrees of freedom were included as nuisance regressors. For each subject, contrast images were calculated for the following comparisons: 1 * Synch-Live + 1 * Synch-Rec > 2 * Speak, 1 * Synch-Live + 1 * Synch-Rec > 2 * Listen, Synch-Live > Speak, Synch-Live > Listen, Synch-Rec > Speak, Synch-Rec > Listen, Synch-Live > Synch-Rec, Synch-Live > DiffLive.

Group-level statistical contrast images were generated with SPM8 by performing one-sample *t* tests on the individual subject-level contrast images. The statistical significance in the group-level contrasts was assessed by first setting a voxel height threshold of $p < 0.005$ and then determining the minimum cluster size necessary to reach whole-brain FEW correction to $p < 0.05$. To do this, the smoothness of the data was estimated from the square root of the residuals images for the second-level contrasts generated by SPM. Then a Monte Carlo simulation with 10,000 iterations was run using AlphaSim (AFNI). Minimum cluster sizes necessary for rejection of the null hypothesis at $p < 0.05$ ranged from 62 to 71 voxels.

Where conjunction analyses are reported, they were calculated using the Conjunction Null Hypothesis method (Nichols et al., 2005), which identifies voxels that are deemed statistically significant (surviving both the voxel height and cluster correction thresholds) in all component contrasts.

In the ROI analysis, parameter estimates were extracted using SPM's MarsBaR toolbox (Brett et al., 2002). *Post hoc t* tests on the resulting β weights from the models were tested against a Bonferroni-corrected significance level of 0.005.

Functional connectivity was assessed with psycho-physiological interactions (PPI) (Friston et al., 1997). In this procedure, clusters were converted to binary masks and the signal within the mask was extracted and deconvolved with the HRF to get a time series of the underlying neural signal. New subject-level GLMs of the data were then constructed, which included the ROI signal time series, the experimental design vector, and an interaction term formed from the signal time series and a contrast of Synch-Live > Synch-Rec.

Experiment 2. To further ensure that the difference between live and recorded speech was not readily detectable, we conducted an additional behavioral experiment with a different set of participants under ideal listening conditions, in which participants had full awareness that they would hear some trials with recordings and others with a live voice.

Participants. We tested 14 participants from the University College Dublin campus who gave informed consent. Participants were right-handed speakers of Hiberno-English with no hearing or speech problems (mean age 36 years, SD 8.0 years; 6 females). A single experimenter was used (F.C.). The protocol was approved by the University College Dublin Life Sciences Research Ethics Committee, in accordance with the Declaration of Helsinki.

Stimuli and task. The task used the same text of 5 sentences as in the imaging study. Participants sat in a quiet room wearing full-cup Beyerdynamic DT 100 headphones, with their back to an experimenter. In an initial calibration, a single recording and repeated live utterances were played until participants judged them to be of equal loudness. On each trial, participants were prompted to either produce the same sentence with the experimenter, merely to listen, or listen while imagining they were speaking synchronously with the experimenter, depending on the condition. Then participants heard three introductory taps made by the experimenter, who proceeded to either (1) speak with the participant via the headphones or (2) play a recording of his own speech over the headphones as the participant executed their instruction. After each trial, participants judged whether what they heard was a recording or a live

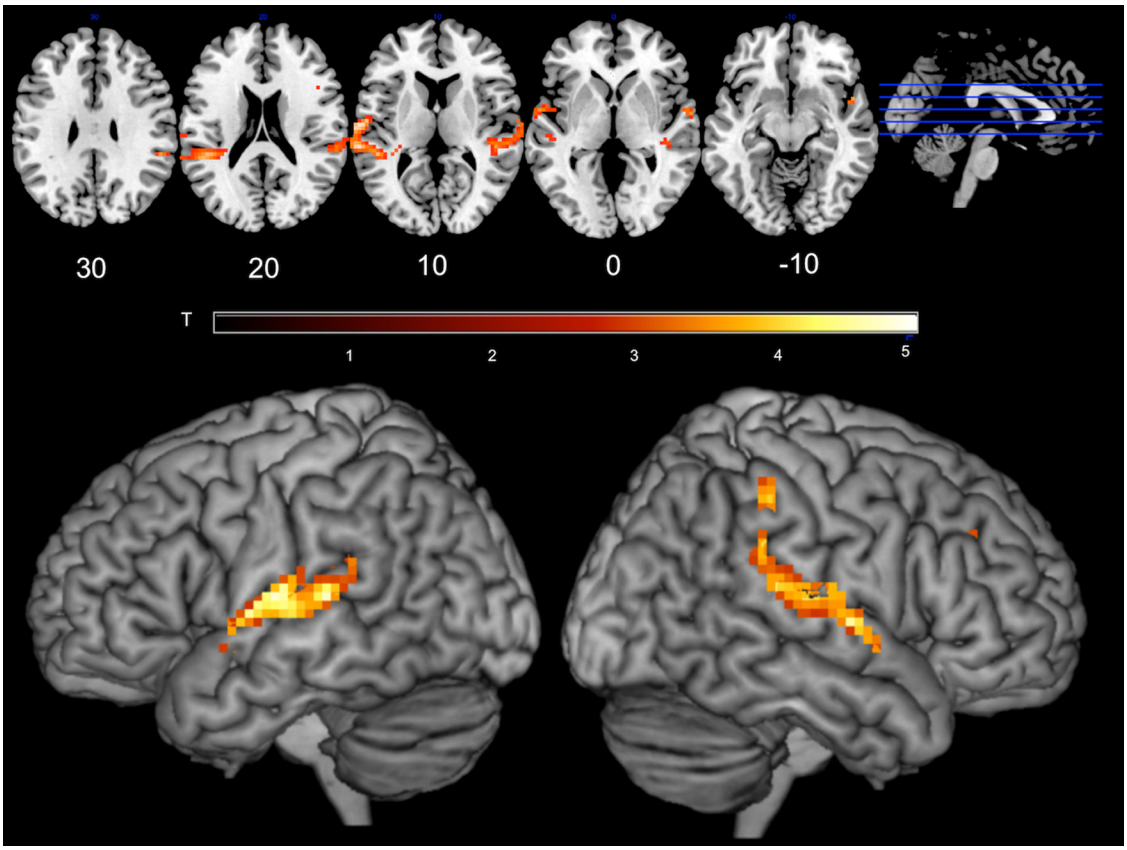

Figure 1. Cortical regions more active during synchronous speaking than solo speaking and listening. This conjunction analysis shows voxels more active during synchronous speaking (with both live and recorded voice) than either speaking alone or passive listening (Synch-Live + Synch-Rec > Speak ∩ Synch-Live + Synch-Rec > Listen. Activated regions include the anterior and posterior auditory processing streams and medial regions on the superior temporal plane. Colors represent voxel *T* values.

Table 1. Conjunction of Synch-Live + Synch-Rec > Speak ∩ Synch-Live + Synch-Rec > Listen[a]

Voxels	Region	*x*	*y*	*z*	*T* (peak voxel)
323	STG	−30	−34	13	5.03
228	STG	42	−28	−5	4.70
1	IFG	44	15	23	2.95

[a]Coordinates in MNI space.

Table 2. Conjunction of Synch-Rec > Speak ∩ Synch-Rec > Listen[a]

Voxels	Region	*x*	*y*	*z*	*T* (peak voxel)
317	STG	−63	−31	10	5.23
169	STG	57	−31	16	4.36

[a]Coordinates in MNI space.

Table 3. Conjunction of Synch-Live > Speak ∩ Synch-Live > Listen[a]

Voxels	Region	*x*	*y*	*z*	*T* (peak voxel)
229	STG	42	−28	−5	5.5
268	STG	−63	−4	4	4.48
31	IFG (opercularis)	48	17	7	3.64

[a]Coordinates in MNI space.

voice, and rated their confidence on a 5 point scale, with 1 representing maximal uncertainty. Participants never heard the same recording twice. Each block contained 5 sentences/trials in a single condition. Six blocks were run in each condition, with the order of conditions counterbalanced across participants.

Results

Synchronous speech versus sensorimotor baselines

The first set of analyses examined brain activity during synchronous speech compared with sensory and motor baselines. In a conjunction analysis, activity in the Synch-Live and Synch-Rec conditions was contrasted against activity associated with both (1) speaking alone and (2) with passively listening to the experimenter speak [(Synch-Live + Synch-Rec > Speak-Alone) ∩ (Synch-Live + Synch-Rec > Listen-Alone)].

The results showed bilateral activity in the temporal plane and the superior temporal gyrus (STG), which extended into inferior parietal cortex in the right hemisphere, as well as a single voxel in the right homolog of Broca's area in the right inferior frontal gyrus (RIFG; Fig. 1; Table 1).

Next, the same conjunction was performed for the Synch-Live and Synch-Rec conditions separately. Synchrony with a recorded voice (Synch-Rec) activated bilateral STG (Synch-Rec > Speak-Alone) ∩ (Synch-Rec > Listen-Alone) (Table 2). Synch-Live also showed largely overlapping activity in bilateral STG, although the right hemisphere cluster was larger, extending further to the anterior and posterior (Table 3; Fig. 2). Additionally, a cluster of 31 activated voxels was detected in RIFG for the Live-Synch conjunction (3), which was not significantly active during the Synch-Rec condition (Table 2).

Parametric effect of alignment between speakers

For each trial in which the subject or experimenter spoke, a WAV file was recorded for each talker. This allowed us to analyze the speech output of both talkers with respect to each other and relate

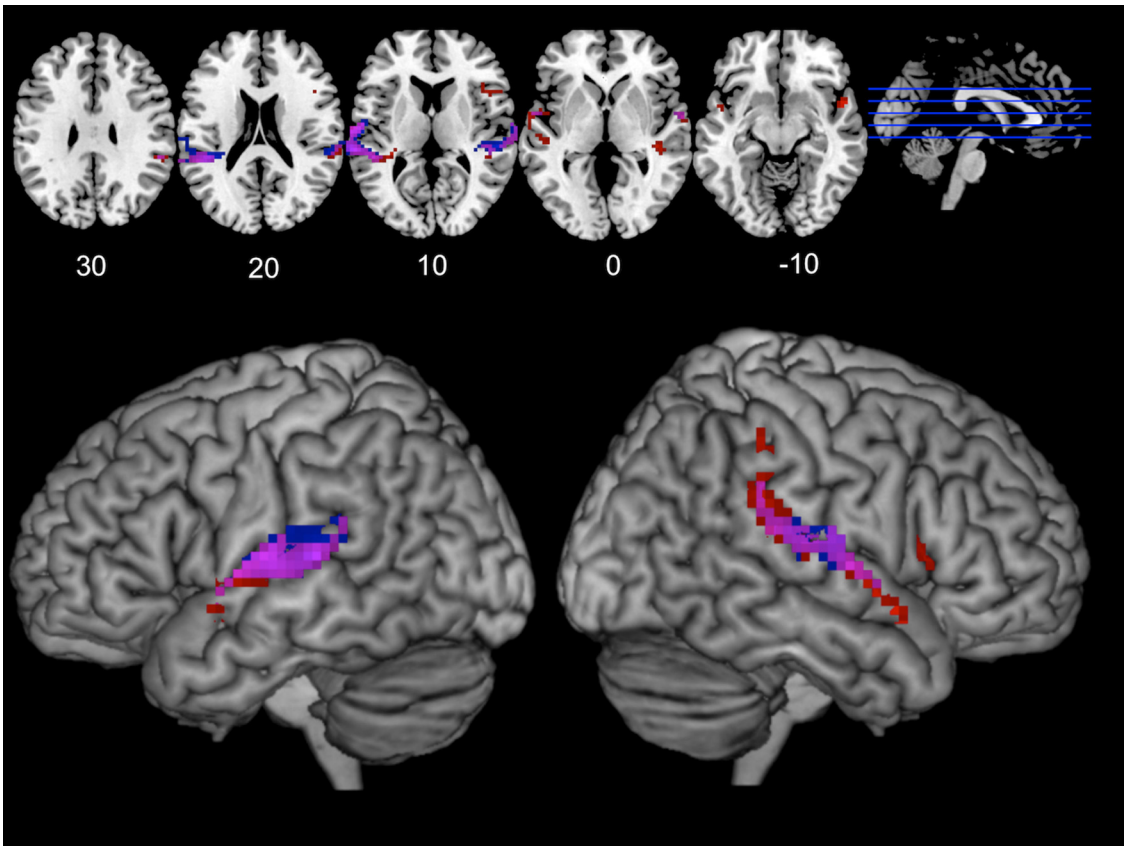

Figure 2. Cortical regions more active during synchronous speaking than solo speaking and listening, shown for Synch-Live and Synch-Rec separately. This conjunction analysis shows voxels more active during synchronous speaking, displaying activation for Synch-Live and Synch-Rec separately. Voxels in red represent significant voxels ($p < 0.005$, cluster corrected) for Synch-Live. Blue voxels represent significant voxel for Synch-Rec. Purple voxels represent significance in both conjunctions. Activated regions were largely overlapping, although activity in right inferior frontal gyrus was only detected for the Synch-Live condition. A statistical test for differences between Synch-Live and Synch-Rec is reported in the next section.

behavioral performance to the fMRI data. A Dynamic Time Warping algorithm was used to compute a synchrony score, which indicated how closely synchronized the subject and experimenter were on each trial (Cummins, 2009). This was only computable for the Synch-Live and Synch-Rec conditions, where the subject and experimenter simultaneously spoke the same text. In the Dynamic Time Warping procedure, the speech in the subject's and experimenter's audio recordings on each trial were converted to sequences of Mel frequency-scaled cepstral coefficients. A similarity matrix was then created, and a least-cost "warp path" through the matrix was computed, which reflected how one speaker's utterance was warped in time relative to the other. A larger area under the warp path, relative to the diagonal, indicated greater asynchrony on the trial.

In a comparison of the scores for the Synch-Live and Synch-Rec conditions, participants' voices were more closely synchronized to the experimenter's during live trials than prerecorded ones ($df = 17$, $T = 3.9$, $p = 0.001$, two-tailed). This is unsurprising given that, in two-way mutual synchronization, both partners are able to respond dynamically to minimize the lag (Cummins, 2002); in contrast, during the Synch-Rec condition, the timing of the experimenter's voice cannot be adjusted contingently, and the minimization of the lag must be done wholly by the subject.

To identify brain regions that were associated with greater accuracy, synchrony scores were included as parametric modulators (one for the Synch-Live condition and one for Synch-Rec) in a new first-level model of the data. Second-level one-sample t tests in SPM then explored the correlates of these modulators. Although no clusters survived the minimum cluster size threshold, for the sake of transparency, here we report clusters surviving an uncorrected threshold of voxelwise $p < 0.005$ and a minimum cluster size of 10 voxels (Lieberman and Cunningham, 2009). Closer synchrony was linked with increased activity in pericentral regions, such as bilateral postcentral gyrus, right precentral gyrus, as well as other regions (Fig. 3; Table 4).

Contrast of Synch-Live > Synch-Rec

The previous analyses showed the loci of increased during synchronous speech compared with solo speech and listening. Next, we directly examined results of the covert manipulation of whether participants spoke with a live or a prerecorded voice using a contrast of Synch-Live > Synch-Rec. Several regions were more active during synchrony with a live voice: right IFG, right supramarginal gyrus and angular gyrus, right temporal pole, and bilateral parahippocampal gyri (Table 5; Fig. 4). Because subjects were unaware that any recordings were used at all during the experiment (as confirmed in debriefing), we posit that these activations reflect processing differences operating outside of awareness. Indeed, in Experiment 2, we provide additional behavioral evidence that people cannot reliably distinguish between a live and a prerecorded partner during synchronous speaking, even when they are aware that that recordings are being used.

Contrast of Synch-Live > Diff-Live

Synch-Live and Synch-Rec differed in that the experimenter's voice could respond contingently in the Live but not the Rec

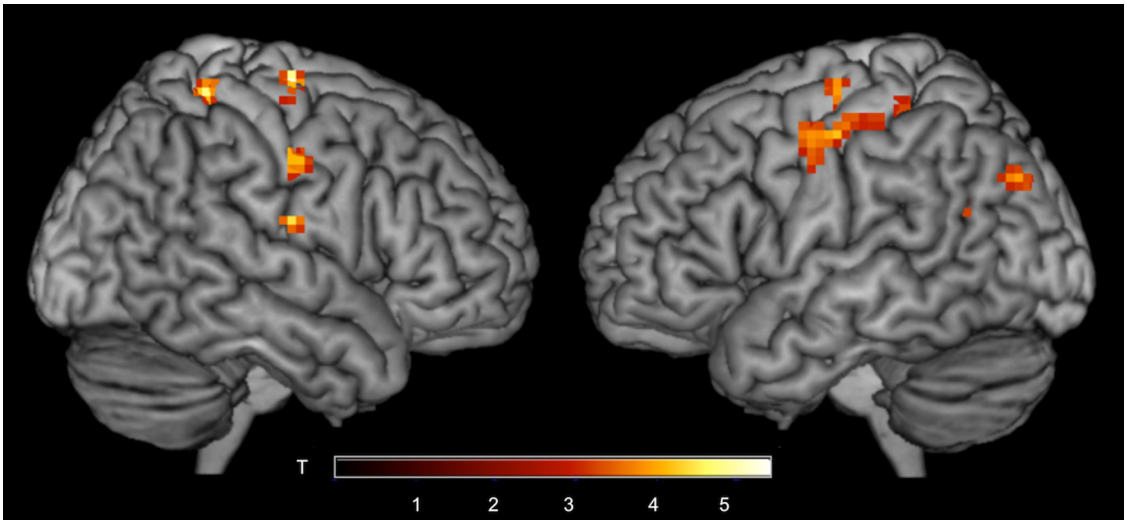

Figure 3. Parametric effect of closer synchrony. Pericentral regions showed increased activity on trials where synchrony between talkers was closer ($p < 0.005$, uncorrected).

Table 4. Parametric correlation of closer synchrony[a]

Voxels	Region	*x*	*y*	*z*	*T*
33	Superior frontal gyrus	24	−13	67	5.39
23	Postcentral gyrus	24	−43	64	4.73
36	Posterior cingulate cortex	12	−52	28	4.6
12	Rolandic operculum	57	−13	19	4.41
36	Postcentral gyrus	66	−10	40	4.33
22	Insula	36	−1	19	4.24
63	Postcentral gyrus	−54	−16	49	4.15
22	Precentral gyrus	−27	−16	61	4.1
10	Rolandic operculum	36	−28	25	3.93
15	Middle occipital gyrus	−42	−79	34	3.87
14	Rolandic operculum	−39	−10	16	3.81
15	Precuneus	18	−70	28	3.77
10	Middle temporal gyrus	−39	−58	22	3.54
14	Postcentral gyrus	−33	−37	55	3.28
10	Putamen	30	−7	1	3.24

[a]Coordinates in MNI space.

Table 5. Contrast of Synch-Live > Synch-Rec[a]

Voxels	Region	*x*	*y*	*z*	*T*	*Z*
76	IFG	48	23	13	5.37	4.05
106	Temporal pole	36	20	−29	5.1	3.92
84	Parahippocampal/lingual gyrus	−18	−43	−5	5.03	3.88
78	Parahippocampal gyrus	18	−34	−5	4.46	3.58
73	Supramarginal/angular gyrus	66	−28	31	3.79	3.18

[a]Coordinates in MNI space.

condition. However, it is possible that the increased activity observed in the contrast could be a general response to contingent interaction, and might have nothing to do with synchrony. That is to say, contingent, live interaction of any kind might be sufficient to produce these differences. As an additional control, the Synch-Live condition was contrasted against Diff-Live. In the Diff-Live condition, the subject and experimenter could hear each other (and therefore could interact contingently), but because the texts they were given to read in these trials were different, synchronization of speech and vocal movements was impossible. The results of the Synch-Live > Diff-Live contrast showed increased activity in the right temporal pole, as well as a cluster appearing in the corpus callosum and posterior cingulate (Table 6).

Response to contingent and congruent interaction in temporal pole

In the Synch-Live and Synch-Rec conditions, the talkers' speech was congruent, but only Synch-Live featured contingent interaction. Conversely, Synch-Live and Diff-Live both featured contingent interaction but only the Synch-Live condition featured congruent speech. A combination of contrasts between these conditions was used to isolate increases in brain activity specific to joint speaking that is both contingent and congruent. To do this, a conjunction analysis of the two contrasts was performed (Synch-Live > Synch-Rec ∩ Synch-Live > Diff-Live). The only region to be significantly active for both contrasts was the right temporal pole, a region associated with higher-order auditory and social processing (Table 7; Fig. 5) (Kling and Steklis, 1976; Olson et al., 2007).

Speech-responsive temporal cortex is typically suppressed during speech production. This has been demonstrated during both speaking aloud (Wise et al., 1999; Houde et al., 2002; Flinker et al., 2010; Agnew et al., 2013) and extends to other other sensory modalities, such as self-directed touch (Blakemore et al., 1998). The interpretation of this effect varies, with some theories associating the suppression with varieties of error-detection mechanisms in the superior temporal gyrus (Guenther, 2006; Hickok, 2012) and others suggesting that the suppression is an index of the distinction between self-produced sensations and sensations caused by external agents (Blakemore et al., 1998). When an action is produced by the self, the perceptual consequences of the action are suppressed; when an action is produced by another person, the perceptual consequences of the action are processed "normally," without suppression. Suppression of activity has been shown to extend into the temporal pole during speech production (Chang et al., 2013). It was therefore hypothesized that the activity in right temporal pole observed during live, mutual synchronous speaking might reflect a release of this suppression: that although the participant was speaking, this region was responding as though the participant were simply listening, and not speaking. This prediction was testable given the design of the experiment. To argue that the activity in temporal pole during synchronous speech reflects a release of suppressed activity during speech production, it is first necessary to demonstrate that suppression occurs in this region.

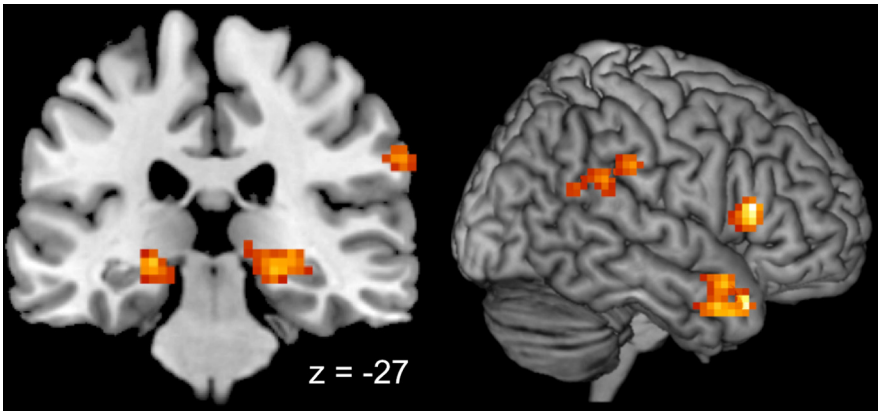

Figure 4. Regions that showed greater activity during synchrony with a live voice (Synch-Live) compared with a recorded voice (Synch-Rec). Several regions showed increased activity during live synchrony compared with recorded synchrony, including the right temporal pole, inferior frontal gyrus, supramarginal gyrus, and bilateral parahippocampal gyrus, extending into hippocampus and lingual gyrus ($p < 0.005$, cluster corrected).

Table 6. Synch-Live > Diff-Live[a]

Voxels	Region	*x*	*y*	*z*	*T*	*Z*
109	Posterior cingulate/corpus callosum	−3	−31	10	5	3.87
100	Temporal pole	48	17	−23	4.86	3.8

[a]Coordinates in MNI space.

Table 7. Conjunction of (Synch-Live > Synch-Rec) ∩ (Synch-Live > Diff-Live)[a]

Voxels	Region	*x*	*y*	*z*	*T*
59	Temporal pole	39	20	−25	4.24

[a]Coordinates in MNI space.

To test this, the right temporal pole cluster defined by the second-level group conjunction described above was used as an ROI. The signal within this cluster was extracted to compare the degree of activation across all conditions. *Post hoc t* tests between each possible pair of conditions were conducted at a Bonferroni-corrected Type I error rate of .005 (correcting for 10 tests; Fig. 5).

As in previous studies in which people speak normally, a speaking-induced suppression effect was apparent: when participants spoke alone, activity in the temporal pole was significantly attenuated compared with when they listened to the experimenter speak ($t_{(17)} = 4.0, p = 0.0009$, two-tailed). Suppression of activity relative to Listen was also found for Synch-Rec ($t_{(17)} = 3.9, p = 0.001$), and for Diff-Live ($t_{(17)} = 4.1, p = 0.0008$).

In contrast, suppression was not seen during Synch-Live, which did not differ from the passive listening condition ($t_{(17)} = 1.0, p = 0.33$). This result suggests that the suppression that routinely and automatically occurs during speech production does not occur when speech production occurs in the context of two people speaking synchronously together (i.e., when the speech is congruent and contingent). The conditions that were used to define the ROI (Synch-Live, Synch-Rec, Diff-Live) were independent from those used to test for the basic sensory suppression effect within the ROI (Speak-Alone, Listen-Alone), thereby avoiding circularity.

Functional connectivity of posterior temporal regions and RIFG

The conjunction contrast that controlled for sensorimotor baselines (Synch-Live > Speak ∩ Synch-Live > Listen) found several regions that showed increased activity during synchronous speaking relative to control conditions: left STG, right STG with IPL, right IFG, and right temporal pole. PPI functional connectivity analyses were used to explore how each of these regions varied in its pattern of functional connectivity with the rest of the brain during Synch-Live compared with Synch-Rec. When subjects spoke with a live voice, the right superior temporal gyrus (RSTG) showed greater functional connectivity with a large part of right parietal cortex (309 voxels; Table 8). The results are shown at a voxel height threshold of $p < 0.005$, cluster corrected to $p < 0.05$ (minimum cluster size threshold of 64 voxels). The peak voxel was in postcentral gyrus in somatosensory cortex.

RIFG was strongly correlated with somatosensory cortex in dorsal postcentral gyrus bilaterally in the Synch-Live condition, compared with the Synch-Rec condition (Table 9). The LSTG showed increased functional connectivity with one cluster that just barely passed significance, exceeding the minimum cluster size by only one voxel (73 voxels, with a minimum cluster size of 72). This region was located in left hemisphere frontal white matter (Table 10).

Functional connectivity of right anterior temporal lobe

A final PPI was constructed to assess the functional connectivity profile of the right temporal pole where the release of speech-induced suppression was observed. As with the other PPIs, functional connectivity during Synch-Live was compared with functional connectivity during Synch-Rec. No regions survived cluster correction at a minimum cluster size of 65 voxels and height threshold of $p < 0.005$. Here, we therefore report exploratory results with a minimum cluster size of 20 voxels. It has been argued that, at a height threshold of $p < 0.005$, even an arbitrary cluster size of 10 voxels is an acceptable trade-off between Type I and Type II error in exploratory analyses (Lieberman and Cunningham, 2009). Activity in the right temporal pole was significantly more correlated with clusters in the left middle frontal gyrus, right inferior occipital gyrus, insula, and the right IFG (opercularis) (Table 11). Interestingly, voxels in RIFG overlapped with those detected in the conjunction analysis of (Synch-Live > Speak) ∩ (Synch-Live > Listen).

For illustration purposes, the results of the above analyses are plotted together in Figure 6. As is apparent, the results of the PPI from RIFG and RSTG, as well as the parametric effect of closer synchrony, all converged in right somatosensory cortex. The functional connectivity relationship between the results regions, as well as the effect of closer synchrony, are plotted schematically in Figure 7.

Experiment 2

Mean d′ accuracy scores of "liveness" judgments across participants was 0.42 (56%) across trials in which the participant and experimenter synchronized and 0.51 (57%) for trials in which the participant listened passively. The difference in d′ scores was not significant ($t_{(13)} = 0.24, p = 0.8$). Higher confidence ratings were associated with correct judgments ($t_{(13)} = 3.5, p = 0.01$), although confidence scores did not differ by condition ($t_{(13)} = 1.4, p = 0.2$). Even in ideal conditions, participants could not reliably distinguish live from recorded speech, regardless of whether they

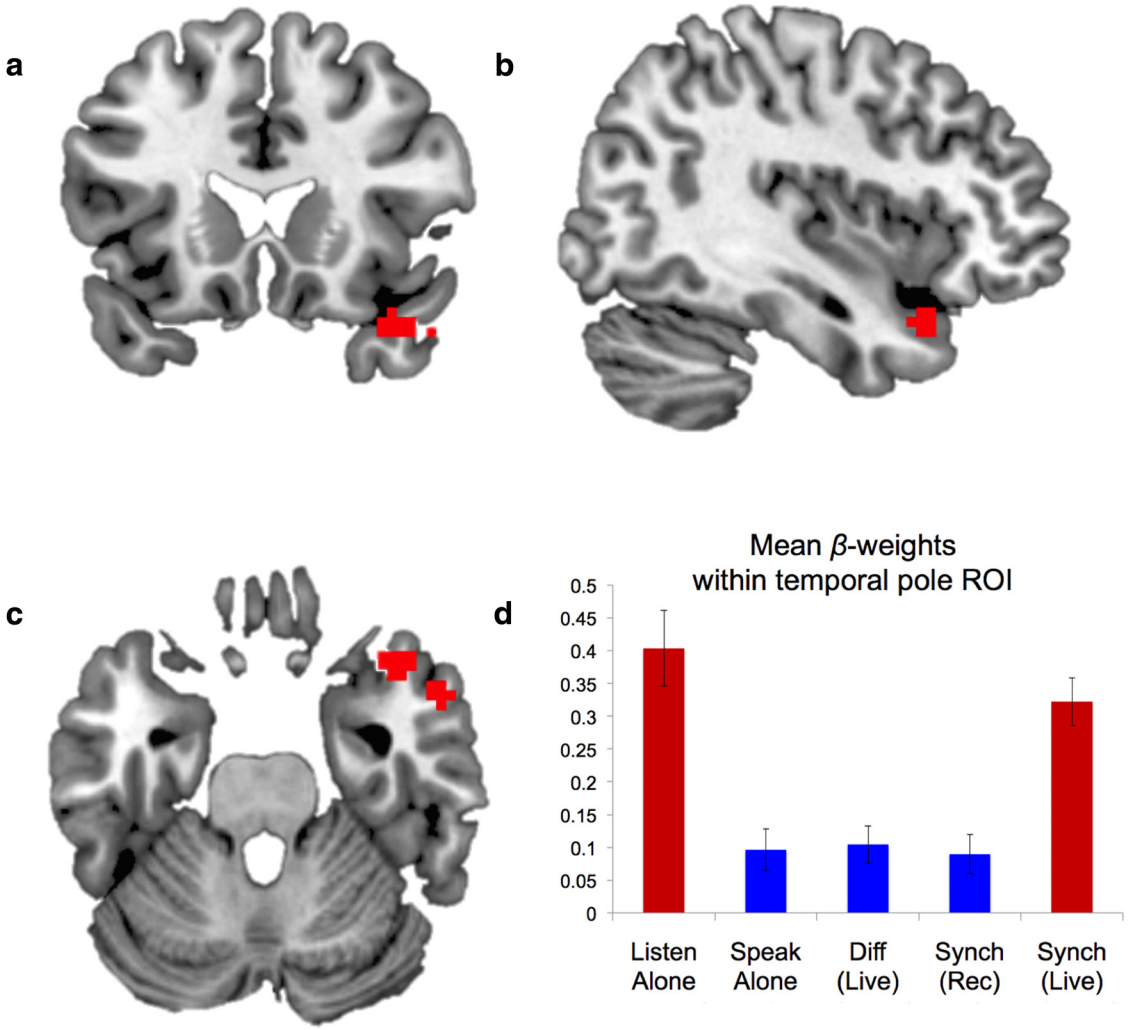

Figure 5. Release of auditory suppression during live synchronous speaking. Red voxels represent increased activity during speech that is produced synchronously with a live person (***a–c***) as defined by the conjunction of Synch-Live > Synch-Rec ∩ Synch-Live > Diff-Live. The profile of activity in this region (***d***) indicates that this activity reflects an absence of suppression that typically occurs during speech production. The conditions are colored to indicate significant differences: red bars are significantly greater than blue bars; bars of the same color did not differ significantly. Significance was set at a Bonferroni-corrected level of $\alpha = 0.005$ for all 10 pairwise tests. Error bars indicate SEM.

Table 8. Clusters with increased functional connectivity with rSTG during Synch-Live versus Synch-Rec[a]

Voxels	Anatomy	*x*	*y*	*z*	*T*
309	Postcentral gyrus	33	−28	43	6.68

[a]Coordinates in MNI space.

Table 9. Clusters with increased functional connectivity with rIFG during Synch-Live versus Synch-Rec[a]

Voxels	Region	*x*	*y*	*z*	*T*
563	Postcentral gyrus	21	−34	55	7.35
151	Postcentral gyrus	−39	−31	58	8.75

[a]Coordinates in MNI space.

were speaking in synchrony or merely listening. Although the imagined synchrony condition was not analogous with any condition in the fMRI and irrelevant to arguments made in this paper, we additionally report it here: d′ accuracy was 0.48 and did not differ from significantly from the synchronous speaking condition ($t_{(13)} = 0.27$, $p = 0.8$) or the passive listening condition ($t_{(13)} = 0.09$, $p = 0.9$).

Table 10. Clusters with increased functional connectivity with LSTG during Synch-Live versus Synch-Rec[a]

Voxels	Anatomy	*x*	*y*	*z*	*T*
72	White matter	−24	17	37	6.27

[a]Coordinates in MNI space.

Table 11. PPI of right temporal pole: Synch-Live > Synch-Rec[a]

Voxels	Region	*x*	*y*	*z*	*T* (peak voxel)
44	Frontal white matter	21	29	25	4.58
34	IFG (opercularis)	45	17	13	3.92
27	Middle frontal gyrus	−33	20	52	4.10
24	Inferior occipital gyrus	47	−76	−5	4.02
23	Insula	48	−7	4	4.00

[a]Coordinates in MNI space. These regions showed increased functional connectivity with the right temporal pole during Synch-Live trials, compared with Synch-Rec trials. $p < 0.005$, uncorrected (20 voxel extent threshold).

Discussion

Summary of main findings

This study showed that synchronous speaking, compared with sensorimotor baselines, was associated with increased activity in bilateral posterior superior temporal lobes and the right inferior

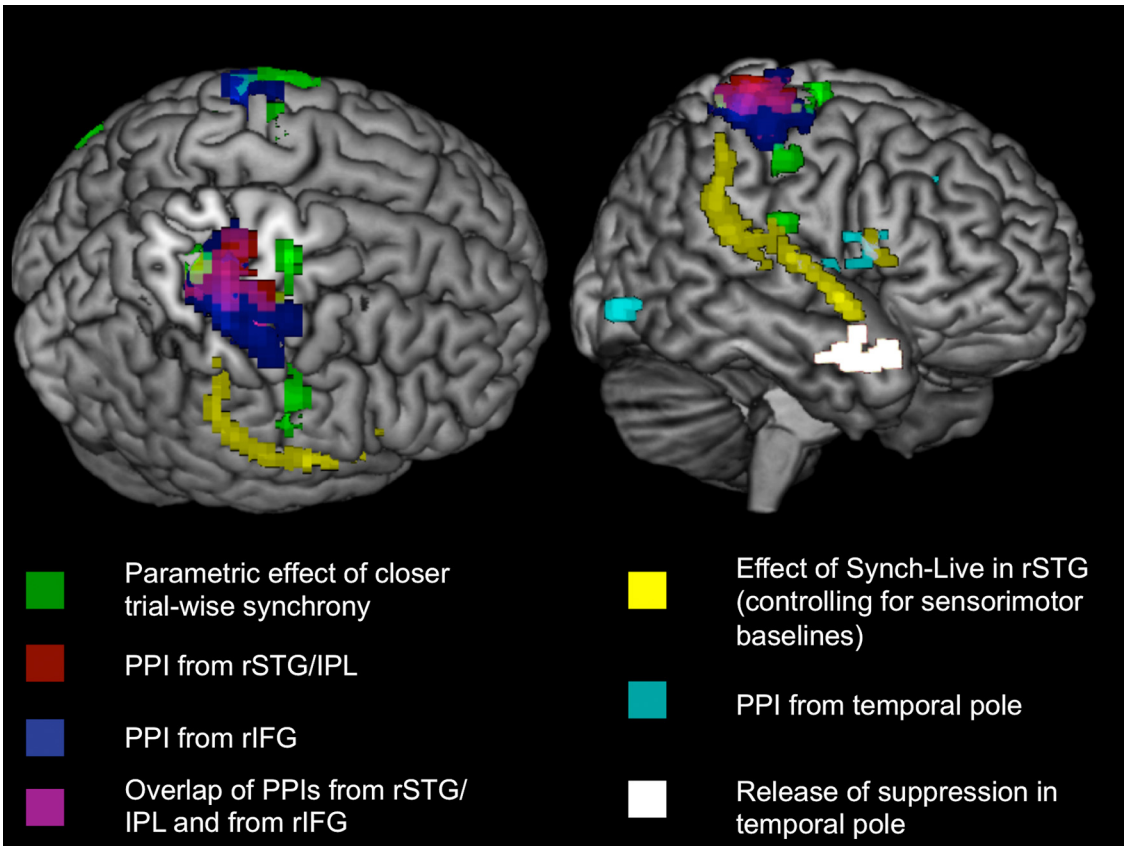

Figure 6. Effects related to synchronous speaking. Functional connectivity (measured with PPI) from RSTG/IPL and RIFG, as well as the parametric effect of closer synchrony, all converged in right somatosensory cortex. Both rendered images show the same activity, from two different angles. Left image is sectioned to show overlap of results in somatosensory cortex.

frontal gyrus. Furthermore, performing joint speech with a live partner was associated with increased activity in the right temporal pole. A parametric analysis with the accuracy of synchrony revealed positively correlated activity in postcentral and precentral gyri: connectivity analyses revealed that activity in the right postcentral gyrus was also correlated with activity in the rIFG and right dorsal auditory fields.

Release of suppression in the ventral stream

The selective activity in the right anterior temporal lobe to joint speech with a live talker was a result of a release from the suppression effects associated with speech production (Wise et al., 1999; Chang et al., 2013). This release from speech-induced suppression was only observed in the live condition: possibly because, in live synchrony, performance can converge on a highly accurate alignment of timing and prosody across the participants. Here, the self-other distinction, which sensory suppression has been argued to support (in other sensory modalities), may have been overridden to some degree (Blakemore et al., 1998). In somatosensation, synchronous (but not asynchronous) touch of one's own face, whereas watching another person's face being touched, leads to participants self-identifying photos of themselves that are partially morphed with the face of the other (Tsakiris, 2008).

These cortical activations may correlate with a change in the locus of control of the interaction: when people coordinate their actions, they minimize movement degrees of freedom by becoming part of a self-organizing system with the locus of control "outside" either individual (Riley et al., 2011). Konvalinka et al. (2010) showed, using simple dyadic finger-tapping tasks, that participants function as one integrated dynamic system, mutually adapting their timing to the other. A more formal characterization of such emergent dynamics as "strongly anticipatory," in contrast to the weak anticipation of interacting systems with internal models, is provided in Stepp and Turvey (2010). The network seen in the Synch-Live condition (where mutual influence was possible) could be associated with this shift in the control of timing to a higher order between both talkers, rather than within each talker individually.

Posterior auditory stream

The role of posterior auditory and inferior parietal fields in the synchronous production of speech (with both a live and recorded talker) is consistent with a role for dorsal auditory pathways in coordinating the tight auditory-motor coupling of one's own voice with that of a partner. Activity in these pathways has been well described during speech production (Wise et al., 1999; Blank et al., 2002; Rauschecker and Scott, 2009; Rauschecker, 2011), and also sensory modulation of speech production (Tourville et al., 2008; Takaso et al., 2010). These posterior auditory-motor "how" pathways therefore play a critical role in the coordination of the sound production (vocal and other) with concurrent auditory information.

RIFG and joint speech

Speech production is classically associated with activity in the left inferior frontal gyrus (Broca's area). We found that the right hemisphere homolog (RIFG) was recruited during synchronous speaking. Notably, whereas rIFG has been reported to be suppressed during speech production (Blank et al., 2003), we found

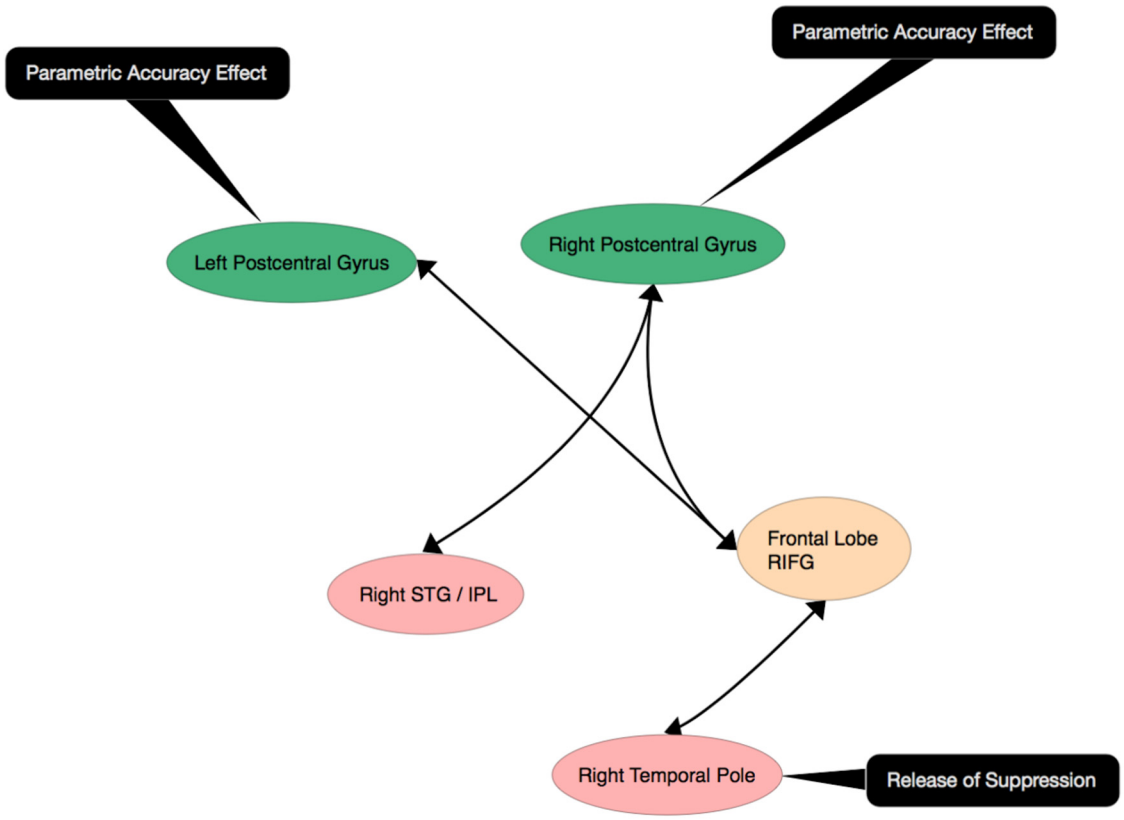

Figure 7. Schematic diagram of brain areas recruited by synchronized speaking. Arrows indicate pairs of regions that showed increased correlated activity during two-way (Synch-Live) synchronous speaking compared with one-way (Synch-Rec) synchronous speaking. Increased accuracy was parametrically related to activity in postcentral gyrus bilaterally during live synchronous speaking. The right temporal pole showed release of speech-induced suppression during live synchronous speaking.

that rIFG activation was enhanced during synchronous speech production, especially with the live talker. This could reflect an increase in the perceptual control of prosody: previous studies have shown that RIFG may be important for processing both pitch and rhythm of speech to detect sarcasm or irony (Wang et al., 2006). Alternatively, this could relate to the timing of joint actions, as rIFG is consistently activated in MRI studies that investigate the perception and motoric entrainment of temporally structured behaviors (Wiener et al., 2010; Repp and Su, 2013).

The rIFG is also sensitive to the perceived ownership of a stimulus, responding more to the static image of one's own face, and the sound of one's own voice, than to another's (Kaplan et al., 2008). Here, the rIFG correlated in activity with the right temporal lobe cluster (at a reduced threshold), suggesting that this area could be associated with the release of suppression in the joint live talker condition. It is possible, for instance, that the self-other distinction is processed in rIFG, which in turn modulates suppression in anterior temporal fields. However, our data do not allow us to make a strong claim about either the locus of the self versus other processing (because we did not manipulate this explicitly) or a causal effect of activity in rIFG on the right temporal pole.

Unconscious detection of a "live" speaker

Experiment 2 confirmed that participants are poor at detecting the differences between live and recorded joint speech. The commonality of the activity for live and recorded synchrony conditions in posterior auditory fields is consistent with a reduced awareness of such sensorimotor processes. For example, using a tapping task in which subjects tapped to a beat that is subtly perturbed, it has been shown that subjects implicitly accommodate to the perturbations, even when they are unaware they are doing so (Repp and Keller, 2004).

Right hemispheric networks for interactive speech

Most studies of speech production have used tasks in which participants speak on their own, and are not required to engage with another talker. Here, our data suggest that the lateralization of speaking may vary with the interactive context in which the speech is spoken. Synchronous speech, especially with a live talker, recruits additional right hemisphere regions in frontal cortex (rIFG), temporal cortex (STG/IPL, motor cortex) and parietal cortex (somatosensory cortex).

PPI analyses from rSTG and rIFG showed that these two regions shared connectivity with a third region in the right parietal cortex. The parametric analysis of trial-by-trial synchrony scores showed that closer synchrony between talkers was associated with bilateral clusters of activity in superior parietal cortex, which were adjacent with the regions functionally connected with both rIFG and rSTG. The PPI analysis of the right temporal pole area where suppression was released showed that during synchronous speech the region showed greater functional connectivity with rIFG.

These results suggest that processing in right parietal cortex may be particularly important for synchronized behavior. Functional connectivity was observed between right somatosensory cortex and rIFG and rSTG, and the region also showed increased activity on trials where subjects had closer accuracy (Fig. 6). This may suggest that the sensory-motor coupling seen during syn-

chronous behavior could be mediated by somatosensory processing in parietal cortex.

Changes in parietal cortex activity were implicated in previous studies of visual guidance of synchronized hand and/or finger movements (Tognoli et al., 2007; Dumas et al., 2010, 2012). Although our participants aligned their speech articulators using auditory information, we still found right partietal cortex activation, suggesting that this region may play an important role in synchrony during social interaction regardless of the particular effectors or sensory modality involved.

Future work

Our study investigated a form of vocal synchrony that naturally occurs in verbal behavior. It remains to be seen whether the patterns of neural behavior we found in this study might apply to conversational speech. Turn-taking in conversation is known to occur on a rapid timescale across cultures (Stivers et al., 2009). Is the complex coordination that occurs during conversation mediated by the same types of neural systems as the simpler test case of synchronous speech, which we examined here? We might hypothesize, as Scott et al. (2009) have, that turn-taking relies on the dorsal auditory pathway and links to motor fields. Following the present study, we might also predict that turn-taking in an interactive context may depend on right lateralized networks. Many of the alignments that occur in synchronized speech (matching of timing and speech rate, pitch, and breathing) are similar to the kinds of matching that occur in natural conversation (Menenti et al., 2012). Alignment and coordination of behavior in free conversation may therefore recruit the same neural processes that support synchronized speaking. Testing these ideas would involve imaging participants while they have a conversation, a difficult but not impossible task so long as motion artifacts are properly dealt with (Xu et al., 2014).

In conclusion, we studied joint speech, which is found across human cultures where unified, cohesive performance is sought. Speaking in strict unison is a task people are frequently surprised to find they can do with ease. We demonstrated a network of right hemisphere regions that are not typically addressed in models of speech production, and a key role for bilateral auditory dorsal pathways, in joint speech. We also showed that the auditory suppression that commonly accompanies speech production is absent during synchronous speech with a live person, and hypothesize that this may alter the subjective experience of self versus other when participating in unison speech. It remains to be seen whether this is affected during other joint behaviors, such as the alignment of posture and eye gaze (Shockley et al., 2009), and conversational speech (Pickering and Garrod, 2004; Fusaroli et al., 2014).

References

Agnew ZK, McGettigan C, Banks B, Scott SK (2013) Articulatory movements modulate auditory responses to speech. Neuroimage 73:191–199. CrossRef Medline

Anshel A, Kipper DA (1988) The influence of group singing on trust and cooperation. J Music Ther 25:145–155. CrossRef

Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. Nat Neurosci 1:635–640. CrossRef Medline

Blank SC, Scott S, Wise R (2001) Neural systems involved in propositional and non-propositional speech. Neuroimage 13:509. CrossRef

Blank SC, Scott SK, Murphy K, Warburton E, Wise RJ (2002) Speech production: Wernicke, Broca and beyond. Brain 125:1829–1838. CrossRef Medline

Blank SC, Bird H, Turkheimer F, Wise RJ (2003) Speech production after stroke: the role of the right pars opercularis. Ann Neurol 54:310–320. CrossRef Medline

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436. CrossRef Medline

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage 16:S497.

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A, 110:2653–2658. CrossRef Medline

Clark HH (1996) Using language. Cambridge, MA: Cambridge UP.

Clark HH, Brennan SE (1991) Grounding in communication. Perspect Socially Shared Cogn 13:127–149.

Cummins F (2002) On synchronous speech. Acoust Res Lett Online 3:7–11. CrossRef

Cummins F (2003) Practice and performance in speech produced synchronously. J Phonetics 31:139–148. CrossRef

Cummins F (2009) Rhythm as entrainment: the case of synchronous speech. J Phonetics 37:16–28. CrossRef

Dumas G, Nadel J, Soussignan R, Martinerie J, Garnero L (2010) Inter-brain synchronization during social interaction. PLoS One 5:e12166. CrossRef Medline

Dumas G, Martinerie J, Soussignan R, Nadel J (2012) Does the brain know who is at the origin of what in an imitative interaction? Front Hum Neurosci 6:128. CrossRef Medline

Fairbanks G. Voice and articulation drillbook. New York: Harper, 1960.

Fessler DM, Holbrook C (2014) Marching into battle: synchronized walking diminishes the conceptualized formidability of an antagonist in men. Biol Lett 10:20140592. CrossRef Medline

Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. J Neurosci 30:16643–16650. CrossRef Medline

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 6:218–229. CrossRef Medline

Fusaroli R, Rączaszek-Leonardi J, Tylén K (2014) Dialog as interpersonal synergy. New Ideas Psychol 32:147–157. CrossRef

Guenther FH (2006) Cortical interactions underlying the production of speech sounds. J Commun Disord 39:350–365. CrossRef Medline

Guionnet S, Nadel J, Bertasi E, Sperduti M, Delaveau P, Fossati P (2012) Reciprocal imitation: toward a neural basis of social interaction. Cereb Cortex 22:971–978. CrossRef Medline

Hickok G (2012) Computational neuroanatomy of speech production. Nat Rev Neurosci 13:135–145. CrossRef Medline

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. J Cogn Neurosci 14:1125–1138. CrossRef Medline

Hove MJ, Risen JL (2009) It's all in the timing: interpersonal synchrony increases affiliation. Soc Cogn 27:949–960. CrossRef

Kaplan JT, Aziz-Zadeh L, Uddin LQ, Iacoboni M (2008) The self across the senses: an fMRI study of self-face and self-voice recognition. Soc Cogn Affect Neurosci 3:218–223. CrossRef Medline

King H (2012) Antiphon: notes on the people's microphone. J Popular Music Studies 24:238–246. CrossRef

Kirschner S, Tomasello M (2010) Joint music making promotes prosocial behavior in 4-year-old children. Evol Hum Behav 31:354–364. CrossRef

Kling A, Steklis HD (1976) A neural substrate for affiliative behavior in nonhuman primates. Brain Behav Evol 13:216–238. Medline

Konvalinka I, Vuust P, Roepstorff A, Frith CD (2010) Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. Q J Exp Psychol (Hove) 63:2220–2230. CrossRef Medline

Kuhlen AK, Allefeld C, Haynes JD (2012) Content-specific coordination of listeners' to speakers' EEG during communication. Front Hum Neurosci 6:266. CrossRef Medline

Lieberman MD, Cunningham WA (2009) Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc Cogn Affect Neurosci 4:423–428. CrossRef Medline

McNeill WH (1995) Keeping together in time. Cambridge, MA: Harvard UP.

Menenti L, Pickering MJ, Garrod SC (2012) Toward a neural basis of interactive alignment in conversation. Front Hum Neurosci 6:185. CrossRef Medline

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005) Valid conjunction inference with the minimum statistic. Neuroimage 25:653–660. CrossRef Medline

Olson IR, Plotzker A, Ezzyat Y (2007) The Enigmatic temporal pole: a re-

view of findings on social and emotional processing. Brain 130:1718–1731. CrossRef Medline

Pickering MJ, Garrod S (2004) Toward a mechanistic psychology of dialogue. Behav Brain Sci 27:169–190; discussion 190–226. Medline

Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271:16–25. CrossRef Medline

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: how work in non-human primates has contributed to our understanding of human speech processing. Nat Neurol 12:718–724. CrossRef Medline

Repp BH (2005) Sensorimotor synchronization: a review of the tapping literature. Psychon Bull Rev 12:969–992. CrossRef Medline

Repp BH, Keller PE (2004) Adaptation to tempo changes in sensorimotor synchronization: effects of intention, attention, and awareness. Q J Exp Psychol 57:499–521. CrossRef Medline

Repp BH, Su YH (2013) Sensorimotor synchronization: a review of recent research (2006–2012). Psychon Bull Rev 20:403–452. CrossRef Medline

Richardson MJ, Marsh KL, Isenhower RW, Goodman JR, Schmidt RC (2007) Rocking together: dynamics of intentional and unintentional interpersonal coordination. Hum Movement Sci 26:867–891. CrossRef Medline

Riley MA, Richardson MJ, Shockley K, Ramenzoni VC (2011) Interpersonal synergies. Front Psychol 2:38. CrossRef Medline

Scott SK, McGettigan C, Eisner F (2009) A little more conversation, a little less action: candidate roles for the motor cortex in speech perception. Nat Rev Neurosci 10:295–302. CrossRef Medline

Sforza A, Bufalari I, Haggard P, Aglioti SM (2010) My face in yours: visuo-tactile facial stimulation influences sense of identity. Soc Neurosci 5:148–162. CrossRef Medline

Shockley K, Santana MV, Fowler CA (2003) Mutual interpersonal postural constraints are involved in cooperative conversation. J Exp Psychol 29:326–332. CrossRef Medline

Shockley K, Richardson DC, Dale R (2009) Conversation and coordinative structures. Top Cogn Sci 1:305–319. CrossRef Medline

Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. Proc Natl Acad Sci U S A 107:14425–14430. CrossRef Medline

Stepp N, Turvey MT (2010) On strong anticipation. Cogn Syst Res 11:148–164. CrossRef Medline

Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, Hoymann G, Rossano F, de Ruiter JP, Yoon KE, Levinson SC (2009) Universals and cultural variation in turn-taking in conversation. Proc Natl Acad Sci U S A 106:10587–10592. CrossRef Medline

Takaso H, Eisner F, Wise RJ, Scott SK (2010) The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study. J Speech Lang Hear Res 53:226–236. CrossRef Medline

Tognoli E, Lagarde J, DeGuzman GC, Kelso JA (2007) The phi complex as a neuromarker of human social coordination. Proc Natl Acad Sci U S A 104:8190–8195. CrossRef Medline

Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. Neuroimage 39:1429–1443. CrossRef Medline

Tsakiris M (2008) Looking for myself: current multisensory input alters self-face recognition. PLoS One 3:e4040. CrossRef Medline

Wang AT, Lee SS, Sigman M, Dapretto M (2006) Neural basis of irony comprehension in children with autism: the role of prosody and context. Brain 129:932–943. CrossRef Medline

Wiener M, Turkeltaub P, Coslett HB (2010) The image of time: a voxel-wise meta-analysis. Neuroimage 49:1728–1740. CrossRef Medline

Wiltermuth S (2012) Synchrony and destructive obedience. Soc Influence 7:78–89. CrossRef

Wiltermuth SS, Heath C (2009) Synchrony and cooperation. Psychol Sci 20:1–5. CrossRef Medline

Wise RJ, Greene J, Büchel C, Scott SK (1999) Brain regions involved in articulation. Lancet 353:1057–1061. CrossRef Medline

Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within Wernicke's area. Brain 124:83–95. CrossRef Medline

Xu Y, Tong Y, Liu S, Chow HM, AbdulSabur NY, Mattay GS, Braun AR (2014) Denoising the speaking brain: toward a robust technique for correcting artifact-contaminated fMRI data under severe motion. Neuroimage 103:33–47. CrossRef Medline