

On the Same Wavelength: Predictable Language Enhances Speaker–Listener Brain-to-Brain Synchrony in Posterior Superior Temporal Gyrus

Suzanne Dikker,^{1,4,5} Lauren J. Silbert,² Uri Hasson,² and Jason D. Zevin^{1,3}

¹Sackler Institute for Developmental Psychobiology, Weill Cornell Medical College, New York, New York 10065, ²Department of Psychology and Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey 08540, ³Haskins Laboratories, New Haven, Connecticut 06511, ⁴New York University, Department of Psychology, New York, New York 10003, and ⁵Utrecht University, Utrecht Institute of Linguistics OTS, 3512 JK Utrecht, The Netherlands

Recent research has shown that the degree to which speakers and listeners exhibit similar brain activity patterns during human linguistic interaction is correlated with communicative success. Here, we used an intersubject correlation approach in fMRI to test the hypothesis that a listener's ability to predict a speaker's utterance increases such neural coupling between speakers and listeners. Nine subjects listened to recordings of a speaker describing visual scenes that varied in the degree to which they permitted specific linguistic predictions. In line with our hypothesis, the temporal profile of listeners' brain activity was significantly more synchronous with the speaker's brain activity for highly predictive contexts in left posterior superior temporal gyrus (pSTG), an area previously associated with predictive auditory language processing. In this region, predictability differentially affected the temporal profiles of brain responses in the speaker and listeners respectively, in turn affecting correlated activity between the two: whereas pSTG activation increased with predictability in the speaker, listeners' pSTG activity instead decreased for more predictable sentences. Listeners additionally showed stronger BOLD responses for predictive images before sentence onset, suggesting that highly predictable contexts lead comprehenders to preactivate predicted words.

Key words: brain-to-brain synchrony; fMRI; language comprehension; language production; lexical-semantics; prediction

Introduction

A growing body of research emphasizes the highly predictive nature of neural processes: our brains are “proactive organs” (Friston, 2003; Bar, 2007) and have even been dubbed prediction machines (Clark, 2013). By anticipating events in our surroundings, we can prepare rapid and targeted behavioral responses (Bar, 2007), improve the isolation and identification of relevant signals in a noisy environment (Obleser and Kotz, 2010), and engage in rapid and efficient language comprehension (Kutas and Federmeier, 2011).

This study asks whether predictions may serve yet another potentially highly useful function, namely that of facilitating neural coupling between interlocutors (Sänger et al., 2011; Hasson et al., 2012; Hari et al., 2013), which has been linked to communicative success: For example, Stephens et al. (2010) found that

story comprehension rates were higher for listeners whose brain activity patterns were more similar to the storyteller's, and suggest that those listeners may have been better able to predict the speaker's intentions.

Does predictability indeed affect neural coupling? If so, which neurocognitive mechanisms might be responsible? Prior research suggests that internal forward models are generated in anticipation of speech acts in both language comprehension and production, leading to relatively more brain activity as predictability increases: Preparing highly predictable speech acts has been proposed to increase the attentional gain for their expected perceptual consequences in language production (Hickok et al., 2011, Tian and Poeppel, 2013), and predictable percepts/words are arguably preactivated before they are seen or heard during language comprehension (Dikker et al., 2010, 2013). When subsequently confronted with unpredicted words, listeners/readers typically show a prediction error response (Kutas and Federmeier, 2011). Crucially, strong prediction error responses are less likely during speech production, because they require speakers to unintentionally violate their own speech plan (e.g., by producing speech errors). Further, whereas listeners may refrain from generating strong predictions in low predictability contexts, speakers will engage in speech planning regardless of context predictability, inducing anticipatory activity in both high and low predictability contexts.

Together, we hypothesize that although predictability affects brain responses in speakers as well as listeners, listener activity is

Received Aug. 29, 2013; revised March 5, 2014; accepted March 31, 2014.

Author contributions: S.D. and J.D.Z. designed research; S.D. and L.J.S. performed research; S.D., L.J.S., and J.D.Z. analyzed data; S.D., U.H., and J.D.Z. wrote the paper.

This work was supported by NIH Grants P01-HD001994 (U.H.) and R01-MH094480, Chinese Academy of Sciences Fellowships for Young International Scientists, Grant No. 2012Y15A0004 (J.D.Z.), and Netherlands Organization for Scientific Research Innovation Scheme Veni Grant 275-89-018 (S.D.).

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. Suzanne Dikker, New York University, Department of Psychology, 6 Washington Place, New York, NY 10003. E-mail: suzanne.dikker@nyu.edu.

DOI:10.1523/JNEUROSCI.3796-13.2014

Copyright © 2014 the authors 0270-6474/14/346267-06\$15.00/0

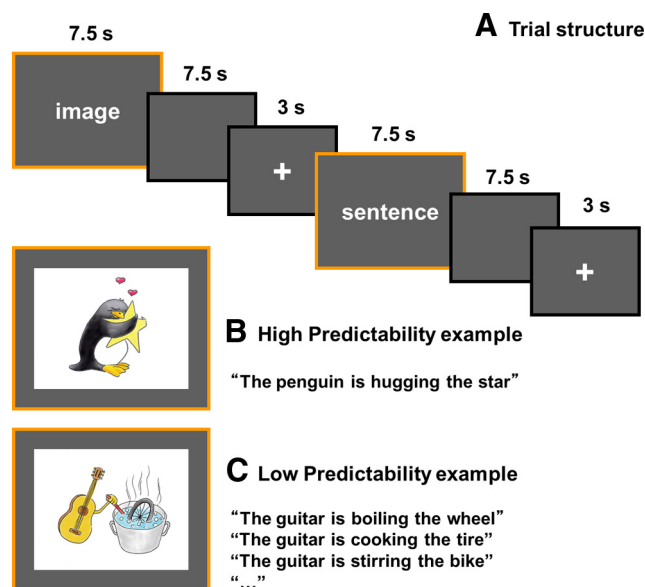


Figure 1. Experimental design and example materials. **A**, Trial structure: 7.5 s image presentation, followed by a 7.5 s blank, then four flashing fixation crosses (375 ms on/off), the sentence (speaker speaks; listeners listen; accompanied by a 7.5 s blank screen) another 3 s of flashing fixation crosses. **B**, **C**, Examples of high predictability (>0.85) and low predictability (<0.35) images and of descriptions provided during the norming study.

more affected by predictability during both the anticipatory stage and the perceptual stage of language processing. As a result, predictability should differentially affect the temporal profiles of brain responses in listeners and speakers respectively, in turn affecting correlated activity between the two.

To investigate whether and how predictability may lead to more synchronous brain activity between speakers and listeners, we used fMRI to record the temporal profiles of BOLD responses in one speaker while describing drawings of improbable/fictitious events, and compared these to brain responses in nine listeners who heard audio recordings of the speaker's descriptions during subsequent fMRI sessions. Crucially, drawings varied in the degree to which they predicted for specific lexical-semantic content of the speaker's utterances. We expected effects of predictability on speaker–listener correlated brain activity to be concentrated in brain regions that have previously been implicated in lexical-semantic prediction during auditory language comprehension, specifically left posterior superior temporal gyrus (for review, see Friederici, 2012).

Materials and Methods

Subjects

One speaker (female; age = 30) and 12 listeners, all right-handed and with normal or corrected-to-normal vision, took part in the fMRI portion of the experiment after providing written informed consent. Two listeners were excluded from analysis for technical reasons during data acquisition and one because he fell asleep during the experiment, leaving nine listeners total to be included for analysis (7 female; mean age = 24.5, $SD = 4.6$). The research protocol was approved by Internal Review Boards for ethics at both the Princeton Neuroscience Institute (PNI; recording site speaker) and Weil-Cornell Medical Center (WCMC; recording site listeners).

Stimuli and experimental design

Materials. Subjects saw 45 hand-drawn color images depicting fictitious scenes in which an animal or object performed an action on another animal or object (e.g., a penguin hugging a star; Fig. 1). Scenes were constructed based on sentences that were created by randomly combin-

ing 45 transitive verbs and 90 nouns denoting common objects, animals, and food items. The speaker was instructed to describe the images using simple declarative sentences in the present tense progressive with a single transitive verb and no adjectives or adverbial phrases (e.g., "The penguin is hugging the star/The dolphin is kissing the tree"). Each image was assigned a predictability score, derived from an offline questionnaire in which 48 volunteers described each of the 45 scenes with the description they deemed most appropriate (23 female; mean age = 31.7, $SD = 7.3$; none participated in the fMRI session of the experiment). For each scene, each participant's entry was assigned a score reflecting the percentage of participants who entered the same response. Predictability was computed as the average across those values. Based on the distribution of predictability across items, items were assigned to one of two "conditions," containing 10 items each: high predictability (>0.85 , $M = 0.9$, $SD = 0.04$) and low predictability (<0.35 , $M = 0.27$, $SD = 0.06$).

Participants were further asked to indicate on a 1–5 scale how certain they were that other people entered the exact same sentence. Predictability and certainty were highly correlated (Spearman $r = 0.75$, $p < 0.001$), and sentences that were deemed more likely were also more likely to be produced by the speaker during the fMRI session (Spearman $r = 0.91$, $p < 0.001$; computed as the proportion of participants who filled out the same sentence that was produced by the speaker during the fMRI experiment). Audio files (7.5 s each) containing the sentences were recorded during the speaker's fMRI scanning session, to be played back later to the listeners during their respective scanning sessions. Sentences had an average duration of 2873 milliseconds ($SD = 127.45$). Predictability was uncorrelated with sentence duration ($r = 0.11$, $p = 0.659$), sentence onset (calculated from the start of the audio file; $M = 1162$ milliseconds, $SD = 127.27$; $r = 0.26$, $p = 0.273$), or overall intensity of the audio file ($M = 51.86$ dB, $SD = 1.27$; $r = -0.16$, $p = 0.495$).

Experimental procedure. For both the speaker ($N = 1$) and the listeners ($N = 9$), each image was presented for 7.5 s, followed by a 7.5 s blank and then flashing fixation crosses (375 ms on/off, 3 s total). A visually presented disk cued the speaker to utter the sentence (7.5 s). Listeners instead heard the recorded sentences during this interval, which was followed by another 7.5 s blank screen and 3 s of flashing fixation crosses announcing the beginning of the next trial (Fig. 1A). This design allowed us to keep the trial structure constant across the speaker and listeners, with the only difference that the speaker described the image, and the listeners subsequently listened to her description. Each participant saw a total of 45 trials in random order, distributed over five blocks. Each scanning session lasted ~45 min.

MRI acquisition

The speaker and listeners were scanned at separate times in separate 3T scanners (32-channel head coil; WCMC: Trio, Siemens; PNI: Allegra, Siemens). Acquisition parameters were the same across scanners.

Anatomical (MP-RAGE) images were recorded with 1-mm-thick sagittal slices in a 256×256 matrix, with a 256 mm field-of-view, yielding a resolution of 1 mm^3 . Functional data were recorded using a T2*-weighted echo planar imaging pulse sequence (EPI). Twenty-five functional slices of 3 mm thickness (1 mm gap) were prescribed obliquely, by slightly adjusting an axial prescription so that the middle slice followed the Sylvian fissure in the sagittal plane to ensure maximum coverage of the cerebrum across participants.

Data were collected with a repetition time (TR) of 1500 ms; echo time, 30 ms; field-of-view, 192 mm; matrix size, 64×64 ; in-plane resolution, 3 mm^2 ; flip angle, 90° . Slice acquisition order was interleaved and 257 volumes per run were collected per participant (5 runs).

The speaker's descriptions were recorded with a customized MR-compatible recording system (FOMRI II; Optoacoustics; Stephens et al., 2010), and listeners wore MRI-compatible ear buds (NordicNeuroLab Ear Plugs; 8 Hz to 35 kHz flat frequency response, +30 dB noise attenuation). Stimuli were presented using E-Prime 2.0 software (Psychology Software Tools).

Data analysis

MRI data analysis. Functional data were preprocessed and partially analyzed using AFNI (Cox, 1996). Cortical surface models were created with

FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>), and functional data were projected into anatomical space using SUMA (Saad et al., 2004; Argall et al., 2006; AFNI/SUMA: (<http://afni.nimh.nih.gov/afni>)). Correlations were computed in MATLAB (2010a, The MathWorks).

Preprocessing. All functions referenced below are part of AFNI. For each subject, anatomical and functional data were coregistered with *lpc_align*, which uses positioning information from the scanner to correctly align oblique functional images. Preprocessing of functional datasets included slice timing and head movement corrections (3dTshift and 3dvolreg) extreme values reduction (3dDespike), and linear and quadratic drift detrending from the time series of each run (3dDetrend).

Surface reconstruction and projection of functional data into surface space. First, we converted the T1-weighted MRI structural images into MGH-HMR format. Then, we extracted cortical meshes from structural volumes. Images were inflated to a sphere (Dale et al., 1999) and anatomically registered to a standard sphere in FreeSurfer (Fischl et al., 1999). To allow for group-based analysis, each subject's registered surfaces were imported into SUMA and converted to a standard mesh of an icosahedron, resulting in the same number of corresponding surface nodes for each subject (190,002). This average brain was normalized to Talairach space to provide stereotactic coordinates for the observed activations in AFNI. Data were exported for statistical analysis in MATLAB using customized code based on the intersubject-correlation procedures described by Stephens et al. (2010) and Lerner et al. (2011). Results were then returned to volume space (3dSurf2Vol) for spatial threshold definitions (>1000 voxels for all analyses reported below).

Identifying areas of interest. Regions of interests for the speaker–listener correlation analyses were identified by first establishing which voxels exhibited correlated activity across listeners (at a threshold of $r > 0.6$, $p < 0.0019$, uncorrected).

A Pearson product-moment correlation coefficient was computed for each surface node by comparing each listener's Z-scored average BOLD response time course across all trials to the Z-scored time course average for all other listeners combined (Lerner et al. (2011); rationale and further methodological details). The resulting nine r values per surface node were averaged together, yielding one normalized (Fisher's Z) intersubject correlation value per node across all nine listeners.

Speaker–listener intersubject-correlation analysis. We first conducted a correlational analysis comparing the speaker's Z-scored average BOLD response time course to the Z-scored mean time course for all listeners combined ($p < 0.01$, uncorrected), to identify brain regions that exhibited similar activation patterns between the speaker and listeners independent of predictability.

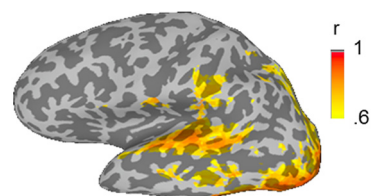
We then proceeded to address our main research question whether Predictability affects speaker–listener intersubject-correlations. First, we created two separate conditions per subject by averaging their Z-scored time courses across the high predictability and low predictability trials respectively (see Materials and Methods). Within each condition, correlation coefficients between each listener and the speaker were then computed on a node-by-node basis. Paired-sample t tests were conducted to compare the resulting normalized nine r values per node for the high predictability versus low predictability items. BOLD time courses from the resulting reliable voxel cluster (>1000, $p < 0.05$, FDR-corrected) were subject to further item analyses.

First, we identified the peak latency of BOLD activity in this region across conditions for the speaker and listeners respectively. At these two time points, we then conducted two 2 (mode: speaker/listeners) \times 2 (predictability: high/low) ANOVAs over the items' BOLD activity, in addition to correlating item BOLD activity and Predictability, sentence duration, audio intensity, and sentence onset.

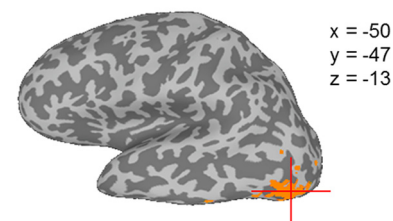
Results

As laid out above, we first conducted an intersubject correlational analysis to identify brain areas where correlated activity between the speaker and listeners was affected by predictability. Then, to explain the observed correlation patterns, we examined how the BOLD signal strength over time in these regions was affected by predictability in the speaker and listeners respectively.

A listener-listener ISC all items



B speaker-listener ISC all items



C speaker-listener ISC high > low predictability

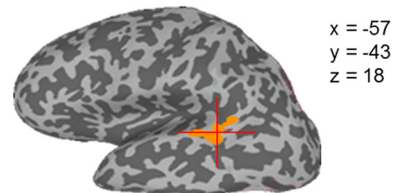


Figure 2. Whole-brain maps of intersubject correlations. **A**, In a whole-brain analysis, high listener–listener intersubject-correlations (ISC) were found in visual and auditory regions associated with image processing and auditory language processing respectively ($N = 9$; only $r > 0.6$, $p < 0.002$, >1000 voxel clusters are shown). **B**, Speaker–listener correlations, time-locked to the image presentation, were concentrated in left fusiform gyrus ($N = 9$; orange cluster: 2400 voxels at $p < 0.01$, uncorrected; cross-hair at center of cluster). **C**, Speaker–listener correlations were significantly higher for high predictability than low predictability items in posterior superior temporal gyrus (orange cluster: 1414 voxels at $p < 0.05$, FDR-corrected; cross-hair at center of cluster).

No reliable speaker–listener correlations were found in the right hemisphere. Thus, only results from the left hemisphere are reported below.

Intersubject correlations

Brain activity was highly correlated across listeners in regions that are typically engaged during image processing (occipital cortex and fusiform gyrus; Grill-Spector et al., 2001), as well as auditory language processing (along superior temporal gyrus, Heschl's gyrus and supramarginal gyrus; for review, see Friederici, 2012; Fig. 2A).

In contrast, when comparing the speaker's time series to the listeners' time series, correlated activity was limited to left fusiform gyrus (Fig. 2B: 2400 voxels; $p < 0.01$, uncorrected; center of cluster: $x = -50$, $y = -47$, $z = -13$).

When comparing the speaker time series to the listeners' time series in high versus low predictability trials, stronger positive speaker–listener correlations were observed for high predictability than low predictability images in posterior superior temporal gyrus (Fig. 2C: 1414 voxels; $p < 0.05$, FDR-corrected; center of cluster: $x = -57$, $y = -43$, $z = 18$). No brain regions exhibited significantly stronger speaker–listener correlations for low predictability items.

Figure 3A shows the averaged time series per condition for the speaker (red) and listeners (blue) across this left posterior supe-

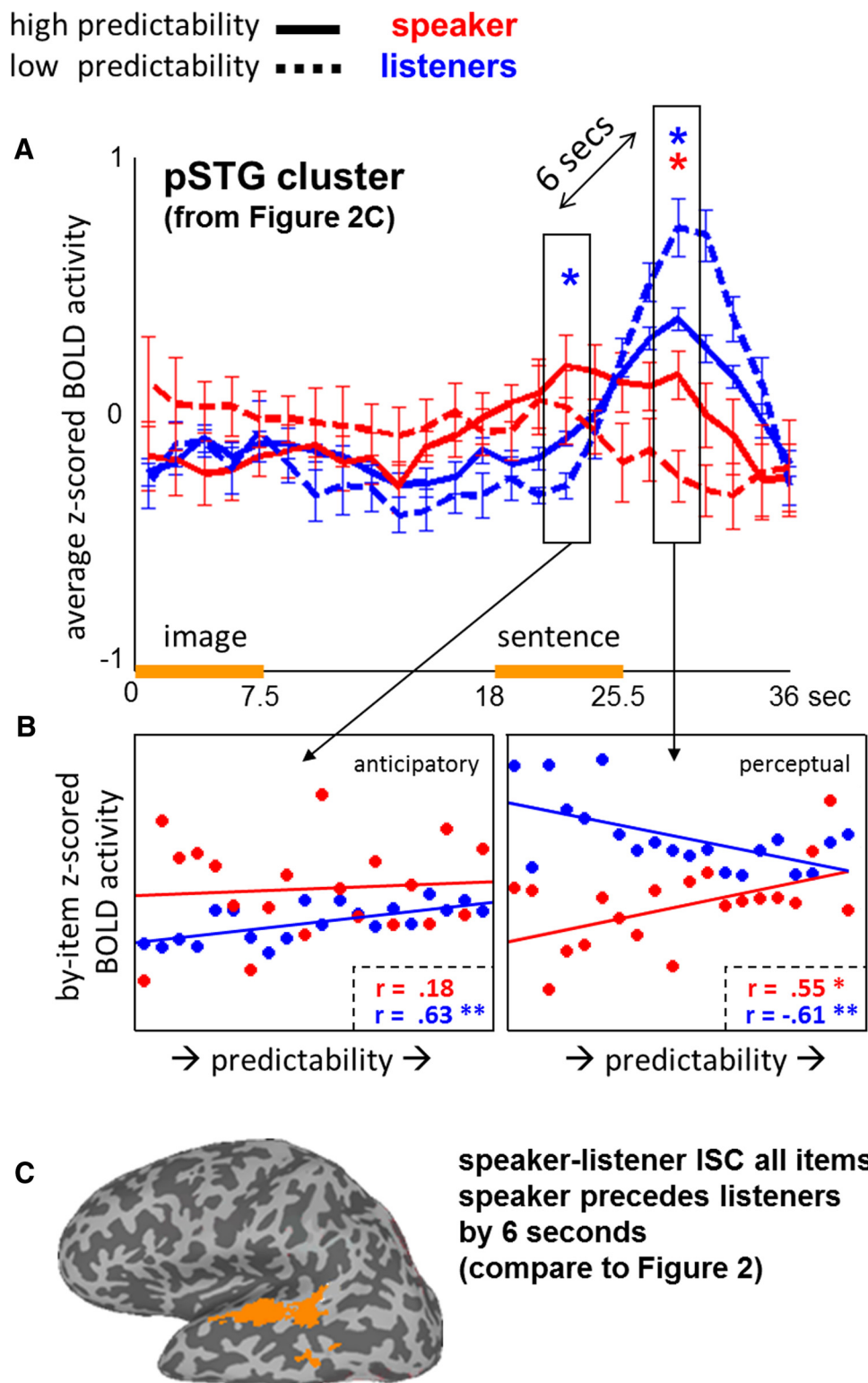


Figure 3. Effects of predictability in pSTG. **A**, Z-scored BOLD time course activation for the speaker (red) and listeners (blue) in high predictability (solid) and low predictability (dashed) items respectively over voxels extracted from the pSTG cluster displayed in Fig. 2C. Error bars reflect by-item SEs for each TR (see timeline on the bottom). Time points for the average speaker (anticipatory) peak and listeners' (perceptual) peak are marked (significant at $*p < .01$). **B**, Scatter plots of Z-scored BOLD activity and item predictability for the anticipatory peak (left) and perceptual peak (right), respectively. In the speaker, predictability and perceptual peak activity over individual items were positively correlated. In the listeners, perceptual peak activity and predictability were negatively correlated, preceded by a positive correlation at the anticipatory peak. **C**, When time-shifting BOLD time series for all items combined, speaker–listener correlations (ISC) were reliable across left superior temporal gyrus (compare with the stimulus-locked analysis in Fig. 2B).

rrior temporal gyrus (pSTG) cluster shown in Figure 2C. Speaker–listener correlations were positive in the high predictability condition ($r = 0.53$, $p = .0077$) but negative in the low predictability condition ($r = -0.69$, $p = .0002$).

Item analysis

Consistent with previous findings (Stephens et al., 2010), average peak activity for the speaker preceded the listeners' peak activity by 6 s (Fig. 3A, boxes). Given the timing of each peak (2 s after

sentence onset for the speaker vs 8 s for the listeners), we henceforth refer to these time points as the “anticipatory” (planning) peak, and the “perceptual” peak respectively. In addition to main effects of mode at both peaks (anticipatory: $F_{(1,18)} = 8.98$, $p = .0049$; perceptual: $F_{(1,18)} = 43$, $p < .0001$), there was a mode \times predictability interaction at the perceptual peak ($F_{(2,36)} = 17$, $p = .0002$): although for the listeners, high predictability items triggered lower BOLD amplitude than low predictability items ($t_{(18)} = 2.97$, $p = 0.008$; HP: $M = 0.42$, $SD = 0.13$; LP: $M = 0.75$, $SD = 0.33$), this pattern was reversed for the speaker, with significantly more activity for the high predictability items than low predictability items ($t_{(18)} = 2.89$, $p = 0.0098$; HP: $M = 0.21$, $SD = 0.28$; LP: $M = -0.16$, $SD = 0.29$). At the anticipatory peak, in contrast, listeners showed higher amplitudes for high predictability as opposed to low predictability items ($t_{(18)} = 3.12$, $p < 0.0005$; HP: $M = -0.02$, $SD = 0.09$; LP: $M = -0.19$, $SD = 0.15$), whereas there was no reliable difference between conditions for the speaker ($t_{(18)} = 0.12$, $p = 0.4028$; HP: $M = 0.25$, $SD = 0.36$; LP: $M = 0.1$, $SD = 0.42$).

BOLD amplitudes by item are plotted by predictability in Figure 3B: BOLD peak activity was positively correlated with predictability for the listeners during the anticipatory peak (left: $r = 0.62$, $p = 0.0029$) but negatively correlated during the perceptual peak (right: $r = -0.61$, $p = 0.0039$). For the speaker, instead, there was a positive correlation between BOLD activity and predictability during the perceptual peak (right: $r = 0.55$, $p = 0.0111$), and no reliable correlation for the anticipatory peak (left: $r = 0.19$, $p = 0.4146$). BOLD activity was not correlated with audio intensity, sentence duration, or sentence onset time.

Time-shifted speaker–listener correlations

In further support of Stephens et al.’s (2010) findings, when time-shifting the BOLD time series for all voxels to align the speaker and listeners’ peaks (i.e., the same whole-brain intersubject-correlation analysis as above, but shifting back the listeners’ time series by 3 TR/6 s), speaker–listener correlations were reliable across left-superior temporal gyrus for all items combined (Fig. 3C). Recall that in the same whole-brain analysis over BOLD time series for all participants time-locked to the stimulus, areas responsible for visual processing were reliably correlated instead (Fig. 2B). Together, Figures 2B and 3C suggest that although the visual world is processed concurrently independent of mode, language-related processes associated with mapping this visual world onto linguistic descriptions typically recruit superior-temporal gyrus in language production before they do so in comprehension. This is in line with the assumed temporal dissociation between speech planning and speech comprehension respectively.

In sum: (1) regions associated with auditory language comprehension across left temporal cortex as well as image processing (occipital cortex and fusiform gyrus) were consistently activated across listeners, (2) activity in fusiform gyrus was correlated between the speaker and listeners regardless of predictability, and (3) significantly stronger correlations were found between the speaker and listeners for high predictability than low predictability items in left pSTG. This effect was driven by an interaction between mode and predictability in pSTG: whereas activity for individual items increased with predictability in the speaker during sentence perception, the listeners’ activity increased as predictability decreased. In addition, listeners’ activity was positively correlated with predictability during sentence anticipation.

Prediction mechanisms in production and comprehension

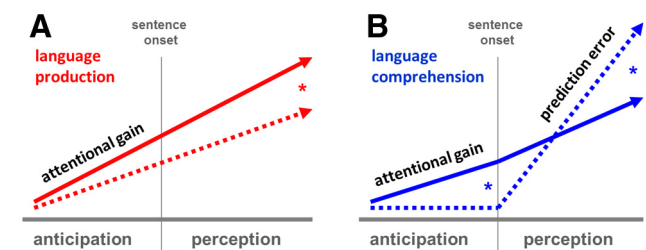


Figure 4. Prediction mechanisms in language production and comprehension. **A**, Increased attentional gain for highly predictable auditory percepts (solid) during production (red) is initiated during the anticipatory stage before sentence onset, persisting into the perceptual stage, yielding significantly stronger pSTG activity for highly predictable sentences. **B**, Increased attentional gain for highly predictable auditory percepts (solid) similarly lead to increased pSTG activation in comprehension (blue). Because predictions are less strong in low predictability contexts (dashed), BOLD signal amplitudes are higher for the high predictability items during the anticipatory stage (left). Prediction error responses reverse this pattern during the perceptual stage, inducing more pSTG activation for unpredictable as opposed to predictable sentences (right).

Discussion

This study used intersubject correlation analyses of fMRI data to ask whether the ability to predict a speaker’s intentions might result in synchronous brain activity between a speaker and a group of listeners. We report a significant increase in brain-to-brain synchrony for highly predictive contexts compared with nonpredictive contexts in pSTG, a region that has been implicated in lexical-semantic processing as well as prediction (for review, see Friederici, 2012). The temporal profiles of pSTG activation suggest that predictability has different effects on brain responses in speakers and listeners respectively, in turn affecting the extent to which neural response patterns are synchronous between language production and comprehension.

Our findings can be explained within existing models whereby language processing is comprised of an anticipatory stage and a perceptual stage: both speakers and listeners take advantage of predictability by “preprocessing” predictable representations during the anticipatory stage, which subsequently affects how those representations are processed during perception. We propose that the neurocognitive mechanisms that govern these processes are similar across production and comprehension, at least at the level of granularity at which they may affect increases and decreases in the BOLD signal.

A number of terms have been used to describe how prediction, top-down processing, and attention may affect action preparation and perception: predictive coding (Friston, 2003), biased competition (Desimone and Duncan, 1995), efference copy (Wolpert and Ghahramani, 2000), preactivation (Dikker and Pykkänen, 2013), spreading activation and priming (Kutas and Federmeier, 2011), etc. We here adopt the term “attentional gain” (borrowed from Tian and Poeppel, 2013) to describe how generating internal forward model/prediction may increase the excitability of neuronal populations associated with predicted representations in language production as well as comprehension (Wlotko and Federmeier, 2007). During speech planning, it has been argued that speakers internally simulate articulatory commands, and that highly predictable speech acts increase the attentional gain for their expected perceptual consequences (Hickok et al., 2011), the neural effects of which persist into the perceptual stage (Fig. 4A; Tian and Poeppel, 2013). Predictability more strongly affects attentional gain in comprehension, not only during anticipa-

tion (Dikker and Pykkänen, 2013), but also during perception: listeners show prediction error responses to unpredicted words (Fig. 4B), whereas lexical-semantic prediction error appears to play no role in the speaker (Fig. 4A): the speaker likely produced each sentence exactly as planned/anticipated (see Introduction).

Thus, as summarized in Figure 4, our results suggest that both speakers and listeners take predictability into account when generating estimates of upcoming linguistic stimuli. These changes in activation resulting from predictive processing, in turn, impact the extent to which brain activity is correlated between speakers and listeners.

For the listeners, our data and the explanation provided above are compatible with a large body of research on lexical-semantic comprehension (Lau et al., 2008; Kutas and Federmeier, 2011) and with recent models suggesting that prediction induces the preactivation of modality-specific representations associated with predicted words (Dikker and Pykkänen, 2013).

In contrast, very little is known about the mechanisms underlying predictive language production, and the research that does exist mainly focuses on low-level properties of the speech signal (Tourville et al., 2008) or repetition priming (Bergerbest et al., 2004; Menenti et al., 2012; for review, see Indefrey and Levelt, 2004; Hickok, 2012), leaving much to be investigated about prediction in language production. In addition, future studies will have to replicate our findings on a larger sample size.

It is further important to emphasize that while fMRI allowed us to localize effects of prediction on correlated brain activity between speakers and listeners, it is not the most suitable tool to dissociate predictive and perceptive processes, by virtue of its low temporal resolution. Because anticipation may precede perception by as little as 200 milliseconds (Dikker and Pykkänen, 2013), measures with a high temporal resolution such as electroencephalography will have to be used to further disentangle neural responses associated with predictive as opposed to perceptual mechanisms. Such methods will also enable investigations into whether the effects observed here may be modulated by direct, face-to-face interaction between speakers and listeners (Jiang et al., 2012). For example, prior research has demonstrated that visual information derived from the speaker's face as she is talking can affect auditory language prediction and comprehension, even in the subsequent absence of concurrent visual information (von Kriegstein et al., 2008). In the present study, listeners received no (visual) information about the speaker, so they could not benefit from such cues. Future studies will further have to explore the relationship between prediction and communicative success, and how the neurobiology of predictive processing may support synchronized linguistic behavior (Richardson et al., 2008) and conversational convergence (Garrod and Pickering, 2004).

References

- Argall BD, Saad ZS, Beauchamp MS (2006) Simplified intersubject averaging on the cortical surface using SUMA. *Hum Brain Mapp* 27:14–27. [CrossRef Medline](#)
- Bar M (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289. [CrossRef Medline](#)
- Bergerbest D, Ghahremani DG, Gabrieli JD (2004) Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *J Cogn Neurosci* 16:966–977. [CrossRef Medline](#)
- Clark A (2013) Whatever next? Predictive brains situated agents and the future of cognitive science. *Behav Brain Sci* 36:181–204. [CrossRef Medline](#)
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173. [CrossRef Medline](#)
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis I segmentation and surface reconstruction. *Neuroimage* 9:179–194. [CrossRef Medline](#)
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. [CrossRef Medline](#)
- Dikker S, Pykkänen L (2013) Predicting language: MEG evidence for lexical preactivation. *Brain Lang* 127:55–64. [CrossRef Medline](#)
- Dikker S, Rabagliati H, Farmer TA, Pykkänen L (2010) Early occipital sensitivity to syntactic category is based on form typicality. *Psychol Sci* 21:629–634. [CrossRef Medline](#)
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284. [CrossRef Medline](#)
- Friederici AD (2012) The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci* 16:262–268. [CrossRef Medline](#)
- Friston K (2003) Learning and inference in the brain. *Neural Netw* 16:1325–1352. [CrossRef Medline](#)
- Garrod S, Pickering MJ (2004) Why is conversation so easy? *Trends Cogn Sci* 8:8–11. [CrossRef Medline](#)
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vis Res* 41:1409–1422. [CrossRef Medline](#)
- Hari R, Himberg T, Nummenmaa L, Hämäläinen M, Parkkonen L (2013) Synchrony of brains and bodies during implicit interpersonal interaction. *Trends Cogn Sci* 17:105–106. [CrossRef Medline](#)
- Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci* 16:114–121. [CrossRef Medline](#)
- Hickok G (2012) Computational neuroanatomy of speech production. *Nat Rev Neurosci* 13:135–145. [CrossRef Medline](#)
- Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422. [CrossRef Medline](#)
- Indefrey P, Levelt WJ (2004) The spatial and temporal signatures of word production components. *Cognition* 92:101–144. [CrossRef Medline](#)
- Jiang J, Dai B, Peng D, Zhu C, Liu L, Lu C (2012) Neural synchronization during face-to-face communication. *J Neurosci* 32:16064–16069. [CrossRef Medline](#)
- Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62:621–647. [CrossRef Medline](#)
- Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de-)constructing the N400. *Nat Rev Neurosci* 9:920–933. [CrossRef Medline](#)
- Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 31:2906–2915. [CrossRef Medline](#)
- Menenti L, Segaert K, Hagoort P (2012) The neuronal infrastructure of speaking. *Brain Lang* 122:71–80. [CrossRef Medline](#)
- Oblers J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb Cortex* 20:633–640. [CrossRef Medline](#)
- Richardson DC, Dale R, Shockley K (2008) Synchrony and swing in conversation: coordination, temporal dynamics, and communication. In: *Embodied communication in humans and machines* (Wachsmuth I, Lenzen M, Knoblich G, eds.), pp 75–93. Oxford: Oxford UP.
- Saad Z, Reynolds R, Argall B, Japee S, Cox RW (2004) SUMA: an interface for surface-based intra- and inter-subject analysis with AFNI. In: *Proceedings of the 2004 IEEE international symposium on biomedical imaging*, pp 1510–1513.
- Sänger J, Lindenberger U, Müller V (2011) Interactive brains, social minds. *Commun Integr Biol* 4:655–663. [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](#)
- Tian X, Poeppel D (2013) The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *J Cogn Neurosci* 25:1020–1036. [CrossRef Medline](#)
- Tourville JA, Reilly KJ, Guenther FH (2008) Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39:1429–1443. [CrossRef Medline](#)
- von Kriegstein K, Dogan Ö, Grüter M, Giraud AL, Kell CA, Grüter T, Kleinschmidt A, Kiebel SJ (2008) Simulation of talking faces in the human brain improves auditory speech recognition. *Proc Natl Acad Sci U S A* 105:6747–6752. [CrossRef Medline](#)
- Wlotko EW, Federmeier KD (2007) Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia* 45:3001–3014. [CrossRef Medline](#)
- Wolpert DM, Ghahramani Z (2000) Computational principles of movement neuroscience. *Nat Neurosci* 3:1212–1217. [CrossRef Medline](#)