

Gestures Orchestrate Brain Networks for Language Understanding

Jeremy I. Skipper,^{1,2,*} Susan Goldin-Meadow,¹
Howard C. Nusbaum,¹ and Steven L. Small^{1,2}

¹Department of Psychology

²Department of Neurology

The University of Chicago

Chicago, IL 60637

USA

Summary

Although the linguistic structure of speech provides valuable communicative information, nonverbal behaviors can offer additional, often disambiguating cues. In particular, being able to see the face and hand movements of a speaker facilitates language comprehension [1]. But how does the brain derive meaningful information from these movements? Mouth movements provide information about phonological aspects of speech [2–3]. In contrast, cospeech gestures display semantic information relevant to the intended message [4–6]. We show that when language comprehension is accompanied by observable face movements, there is strong functional connectivity between areas of cortex involved in motor planning and production and posterior areas thought to mediate phonological aspects of speech perception. In contrast, language comprehension accompanied by cospeech gestures is associated with tuning of and strong functional connectivity between motor planning and production areas and anterior areas thought to mediate semantic aspects of language comprehension. These areas are not tuned to hand and arm movements that are not meaningful. Results suggest that when gestures accompany speech, the motor system works with language comprehension areas to determine the meaning of those gestures. Results also suggest that the cortical networks underlying language comprehension, rather than being fixed, are dynamically organized by the type of contextual information available to listeners during face-to-face communication.

Results and Discussion

What brain mechanisms account for how the brain extracts phonological information from observed mouth movements and semantic information from cospeech gestures? In prior research, we have shown that brain areas involved in the production of speech sounds are active when listeners observe the mouth movements used to produce those speech sounds [7, 8]. The pattern of activity between these areas, involved in the preparation for and production of speech, and posterior superior temporal areas, involved in phonological aspects of speech perception, led us to suggest that when listening to speech, we actively use our knowledge about how to produce speech to extract phonemic information from the face [1, 8]. Here we extrapolate from these findings to cospeech gestures. Specifically, we hypothesize that when

listening to speech accompanied by gestures, we use our knowledge about how to produce hand and arm movements to extract semantic information from the hands. Thus, we hypothesize that, just as motor plans for observed mouth movements have an impact on areas involved in speech perception, motor plans for cospeech gestures should have an impact on areas involved in semantic aspects of language comprehension.

During functional magnetic resonance imaging (fMRI), participants listened to spoken stories without visual input (“No Visual Input” condition) or with a video of the storyteller whose face and arms were visible. In the “Face” condition, the storyteller kept her arms in her lap and produced no hand movements. In the “Gesture” condition, she produced normal communicative deictic, metaphoric, and iconic cospeech gestures known to have a semantic relation to the speech they accompany [5]. These gestures were not codified emblems (e.g., “thumbs-up”), pantomime, or sign language [5, 9]. Finally, in the “Self-Adaptor” condition, the actress produced self-grooming movements (e.g., touching hair, adjusting glasses) with no clear semantic relation to the story. The self-adaptive movements had a similar temporal relation to the stories as the meaningful cospeech gestures and were matched to gestures for overall amount of movement (Movies S1 and S2 available online).

We focused analysis on five regions of interest (ROIs) based on prior research (Figure S1): (1) the superior temporal cortex posterior to primary auditory cortex (STp); (2) the supramarginal gyrus of the inferior parietal lobule (SMG); (3) ventral premotor and primary motor cortex (PMv); (4) dorsal pre- and primary motor cortex (PMd); and (5) superior temporal cortex anterior to primary auditory cortex, extending to the temporal pole (STa) [1, 10]. The first four of these areas have been found to be active not only during action production, but also during action perception [11]. With respect to spoken language, STp and PMv form a “dorsal stream” involved in phonological perception/production and mapping heard and seen mouth movements to articulatory based representations (see above) [1, 7, 8, 12]. Whereas STp is involved in perceiving face movements, SMG is involved in perceiving hand and arm movements [13] and, along with PMv and PMd [11, 14], forms a (another) “dorsal stream” involved in the perception/production of hand and arm movements [15]. In contrast, STa is part of a “ventral stream” involved in mapping sounds to conceptual representations ([1], see [12] for more discussion) during spoken language comprehension; that is, STa is involved in comprehending the meaning of spoken words, sentences, and discourse [16–18]. To confirm that STa was involved in spoken word, sentence, and discourse comprehension in our data, we intersected the activity from all four conditions for each participant. Our rationale was that speech perception and language comprehension are common to all four conditions; activation shared across the conditions should therefore reflect these processes. We found that a large segment of STa and a small segment of STp were bilaterally active for at least 9 of 12 participants (Figure S2).

Just as research using single-cell electrophysiology in visual cortex examines which neurons prefer or are “tuned” to

*Correspondence: jis2013@med.cornell.edu

(i.e., show a maximal firing rate) particular stimulus properties over others [19], we investigated hemodynamic “tuning” to the meaningfulness of hand and arm movements with respect to the spoken stories by using a peak and valley analysis method ([20, 21], [22] for a similar method). In each ROI for each condition, we averaged the entire time course of the hemodynamic response (henceforth signal) for all voxels that were active in at least one of the four conditions, both within and across participants. We delayed the response by 4 s to align the brain’s hemodynamic response over each entire story to the coded features of that story (i.e., the cospeech gestures in the Gesture condition and the self-adaptor movements in the Self-Adaptor condition). Next, we found peaks in the resulting averaged signal for each condition by using the second derivative of that signal. Gamma functions (with similarity to the hemodynamic response) of variable centers, widths, and amplitudes were placed at each peak and allowed to vary so that the best fit between the actual signal and the summation of the gamma functions was achieved (R^2 s > .97). Half of the full width half maximum (FWHM/2) of a resulting gamma function at a peak determined the search region used to decide whether, for example, an aligned cospeech gesture elicited that peak. Specifically, a particular hand and arm movement (i.e., either a cospeech gesture or self-adaptor movement) was counted as evoking a peak if 2/3rds of that peak contained a hand and arm movement, and was counted as not evoking a peak if less than 1/3rd of that peak contained a hand and arm movement. The distance between the FWHM/2 of two temporally adjacent gamma functions determined which aspects of the stimuli caused a decay or valley in the response. Specifically, a particular hand and arm movement was counted as resulting in a valley if 2/3rds of that valley contained a hand and arm movement, and was counted as not resulting in a valley if less than 1/3rd of that valley contained a hand and arm movement. Regions were considered tuned to cospeech gestures or self-adaptor movements that were represented in peaks but not valleys. Significance was determined by two-way chi-square contingency tables (e.g., gestures versus no-gestures at peaks; gestures versus no-gestures at valleys).

Figure 1 illustrates the peak and valley analysis. Frames from the stories associated with peaks are on the top; frames associated with valleys on the bottom. Note that the speaker’s cospeech gestures are found only on top and thus are associated with peaks, not valleys (Figure 1A); in contrast, her self-adaptor movements are found on top and bottom and thus are associated with both peaks and valleys (Figure 1B). For the Gesture condition, this pattern held in regions PMv, PMd, SMG, and STa where peaks in the response corresponded to gesture movements and valleys corresponded to times when the hands were not moving (Figure 2, gray bars; PMv $\chi^2 = 11.1$, $p < .001$, $\Phi = .51$; PMd $\chi^2 = 8.9$, $p < .003$, $\Phi = .40$; SMG $\chi^2 = 4.46$, $p < .035$, $\Phi = .28$; STa $\chi^2 = 19.5$, $p < .0001$, $\Phi = .62$). Peaks in these regions’ responses were not simply due to movement of the hands—the meaningless hand movements in the Self-Adaptor condition were as likely to result in valleys as in peaks in the PMv, PMd, SMG, and STa response (Figure 2, striped bars; PMv $\chi^2 = 0.06$, $p = .81$, $\Phi = .04$; PMd $\chi^2 = 0.0$, $p = .96$, $\Phi = .01$; SMG $\chi^2 = .42$, $p < .52$, $\Phi = .10$; STa $\chi^2 = 0.56$, $p = .45$, $\Phi = .12$). STp showed no preference for hand movements, either meaningful cospeech gestures or meaningless self-adaptor movements.

To understand how PMv, PMd, SMG, and STa work together in language comprehension, we analyzed the network

interactions among all five regions. The fMRI signals corresponding to the brain’s response to each condition from each of the five regions were used in structural equation modeling (SEM) to find the functional connectivity between regions. The output of these models reflects between-area connection weights that represent the statistically significant influence of one area on another, controlling for the influence of other areas in the network [10, 23]. SEMs are subject to uncertainty because a large number of models could potentially fit the data well, and there is no a priori reason to choose one model over another. To address this concern, we solved all of the possible models for each condition [24] and averaged the best-fitting models (i.e., models with a nonsignificant chi-square) by using a Bayesian model averaging approach [10, 25, 26]. Bayesian model averaging has been shown to produce more reliable and stable results and provide better predictive ability than choosing any single model [27].

Bayesian averaging of SEMs resulted in one model for each condition with nine physiologically plausible connections. Results show that the brain dynamically organizes activity patterns in response to the demands and available information of the immediate language task (Figure S3). Specifically, during the Gesture condition, connection weights were stronger between SMG and PMd (bidirectionally), from SMG to STp, from STa to PMv, and from PMd to STa than in any other condition. In contrast, in the Face condition, connection weights were stronger between STp and PMv (bidirectionally) and from PMv to STa. In the Self-Adaptor condition, connection weights were stronger only from STa to PMd, and, in the No Visual Input condition, no connection weights were stronger than any others. Although this is a complex set of results, a pattern emerges that can be more easily understood by considering the mean of the connection weights between STa and pre- and primary motor cortex (i.e., PMv and PMd) and between STp and pre- and primary motor cortex (Figure 3). This representation of the results shows that the statistically strongest connection weights associated with STa and pre- and primary motor cortex correspond to the Gesture condition. In contrast, the statistically strongest connection weights associated with STp and pre- and primary motor cortex correspond to the Face condition. This result is consistent with the peak and valley analysis showing that STp does not respond to hand movements (see Figure 2). Thus, when cospeech gestures are visible, PMd, STa, and SMG work together more strongly to interpret the meanings conveyed by the hand movements, which could enhance interpretation of the speech accompanying the gestures. In contrast, when only facial movements are visible, PMv and STp work together more strongly to simulate face movements relevant to speech production, a simulation that could aid phonological interpretation; this finding is consistent with previous work in our lab showing that PMv and STp are sensitive to the correlation between observable mouth movements and speech sounds [7, 8].

Results suggest that cospeech manual gestures provide information germane to the semantic goal of communication, whereas oral gestures support phonemic disambiguation. This interpretation is supported by two additional pieces of evidence. First, percent of items recalled correctly was 100%, 94%, 88%, and 84% for Gesture, Self-Adaptor, Face, and No Visual Input conditions, respectively, as probed by three true/false questions 20 min after scanning. Participants were significantly more accurate at recalling information in the Gesture condition, which displayed face movements and

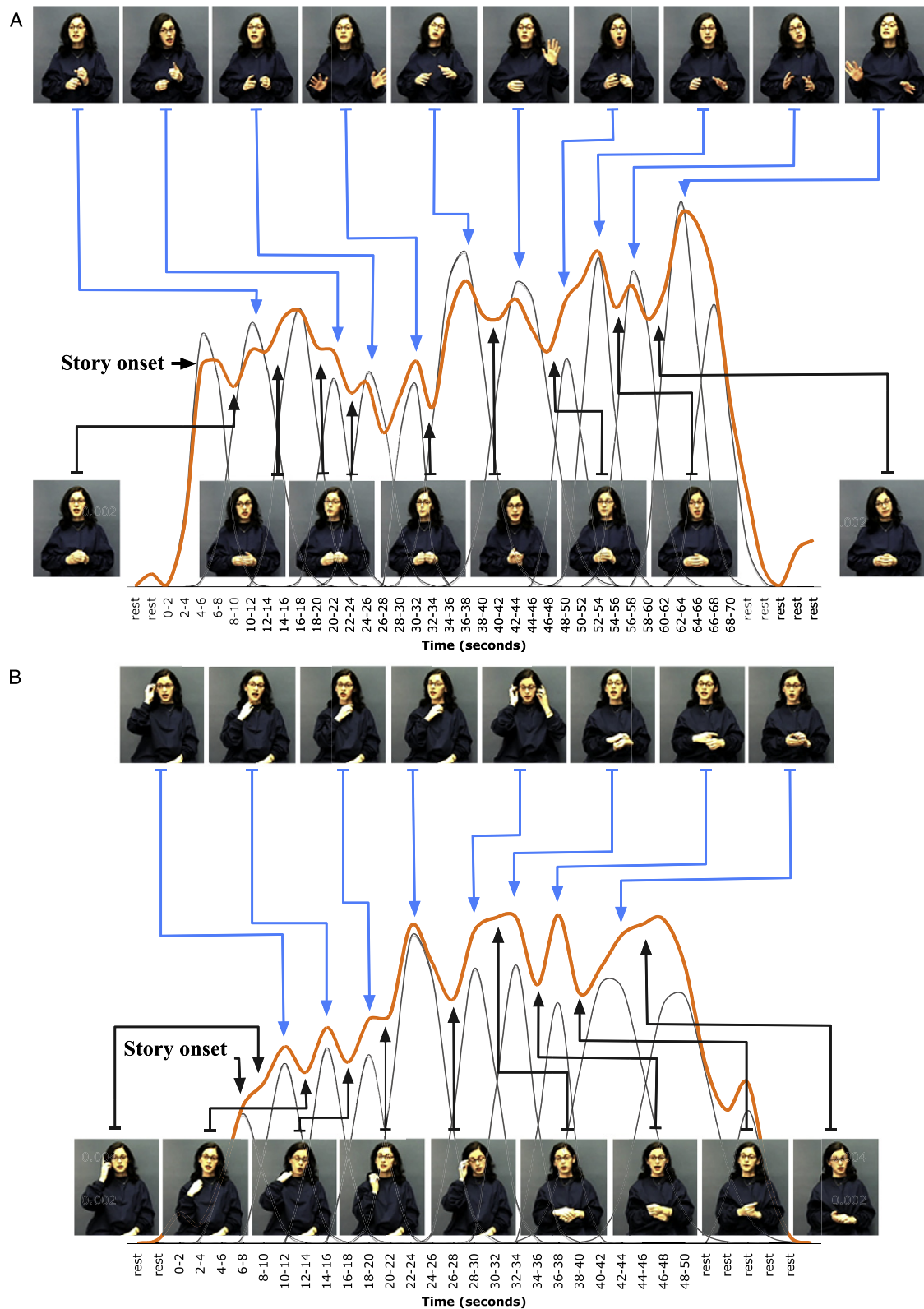


Figure 1. Peak and Valley Analysis of the Gesture or Self-Adaptor Conditions

Peaks and valleys associated with the brain's response to the (A) Gesture and (B) Self-Adaptor conditions in ventral pre- and primary motor cortex (PMv; see Figure S2). Frames from the movies associated with peaks are on the top row, and frames associated with valleys are on the bottom. The orange line is the brain's response. Gray lines are the gamma functions fit to each peak in the response that were used to determine which aspects of the stimulus resulted in peaks and valleys (see text). Note that in the Gesture condition (A), cospeech gestures are associated with peaks and hands-at-rest are associated with valleys. In contrast, in the Self-Adaptor condition (B), self-adaptor movements are associated with both peaks and valleys.

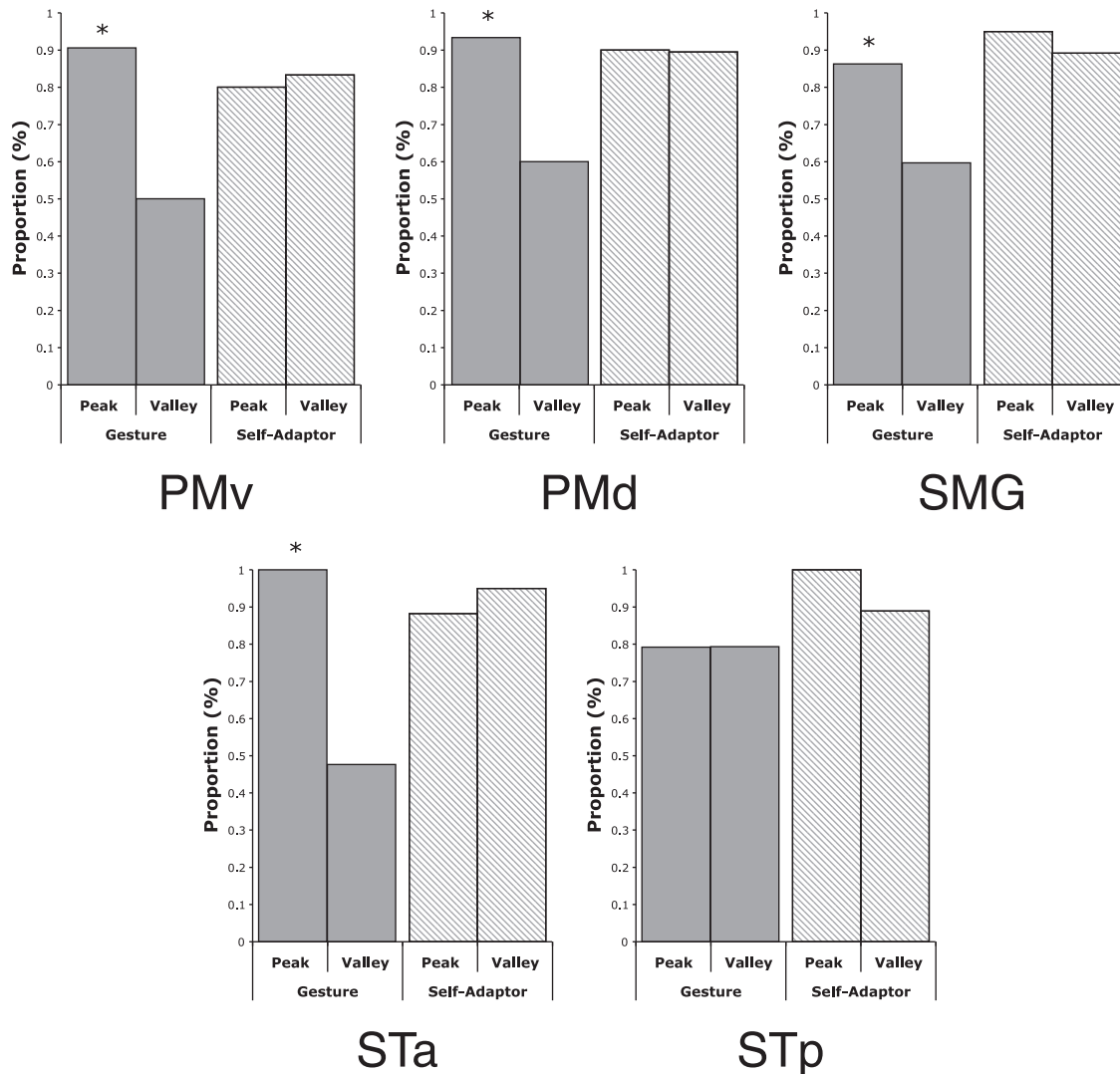


Figure 2. Percentage of All Peaks or Valleys in Regions of Interest in which a Hand Movement Occurred or Did Not Occur in the Gesture or Self-Adaptor Conditions

Asterisks indicate a significant difference between the actual occurrences of hand movements or no hand movements at peaks or valleys and the expected value ($p < .05$). Abbreviations: superior temporal cortex anterior (STa) and posterior (STp) to Heschl's gyrus; supramarginal gyrus (SMG); ventral (PMv) and dorsal (PMd) premotor and primary motor cortex. See text and Figure S2 for further information.

meaningful hand movements, than in the other three conditions, which displayed at most face movements or hand movements unrelated to the speech ($t = 3.7$; $df = 11$; $p < .004$; Figure S4). Second, in a prior analysis of the present data [10], we found that the weakest connection weights between Broca's area and other cortical areas were associated with the Gesture condition, whereas the strongest weights were associated with the Self-Adaptor and Face conditions. This pattern is consistent with research showing that Broca's area plays a role in semantic selection or retrieval [28–31]. In other words, selection and retrieval demands are reduced because cospeech gestures provide semantic information (see [32] for supporting data).

Our findings provide evidence that brain areas involved in preparing and producing actions are also involved in processing cospeech gestures for language comprehension. The routes through which cospeech gestures and face movements affect comprehension differ. Mouth movements are used by

listeners during language comprehension to aid speech perception—by covertly simulating the speaker's mouth movements (through activation of motor plans involved in producing speech), listeners can make the acoustic properties of speech less ambiguous and thereby facilitate comprehension [8]. In contrast, for hand movements to play a role in language comprehension, listeners must not only simulate the speaker's hand and arm movements, they must also use those simulations to get to the meanings conveyed by the movements. Thus, in some real sense, we have shown that the hands become the embodiment of the meaning they communicate and, in this way, improve communication. Furthermore, we show that the cortical networks supporting language comprehension are not a static set of areas. What constitutes language in the brain dynamically changes when listeners have different visual information available to them. When cospeech gestures are observed, the networks underlying language comprehension are differently tuned and

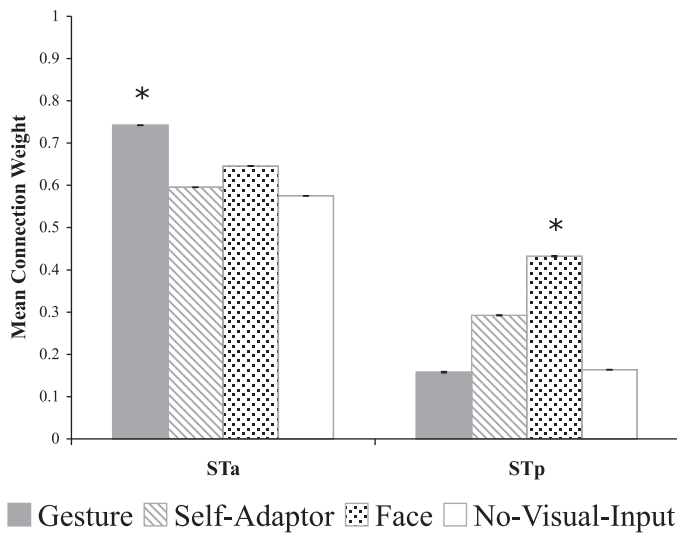


Figure 3. Structural Equation Model Connection Weights

Mean of connection weights from Bayesian averaged structural equation models between STa or STp and premotor and primary motor cortex (i.e., PMv and PMd) regions of interest. Asterisks indicate that a condition resulted in significantly stronger connection weights compared to all of the other conditions ($p < .05$). See Figure 2 for abbreviations.

connected than when meaningless hand movements or face movements alone are observed. We suggest that in real-world settings, the brain supports communication through multiple cooperating and competing [33] networks that are organized and weighted by the type of contextual information available to listeners and the behavioral relevance of that information to the goal of achieving understanding during communication.

Experimental Procedures

Participants

Twelve right-handed [34] native English speakers (mean age = 21 ± 5 years; 6 females) with normal hearing and vision, no history of neurological or psychiatric illness, and no early exposure to a second language participated. The study was approved by the Institutional Review Board of the Biological Science Division of The University of Chicago and participants provided written informed consent.

Stimuli and Task

Stories were adaptations of Aesop's Fables told by a female speaker. Each story lasted 40–50 s. Participants were asked to listen attentively to the stories. Each story type (i.e., Gesture, Self-Adaptor, Face, and No Visual Input) was presented once in a randomized manner in each of the two 5 min runs. Note that, because of the dependence of gestures on speech, a "Gesture Only" control was not included in this study because of the likelihood of introducing unnatural processing strategies (see [35] for an elaboration). See [10] for further details and Movies S1 and S2 for example video clips. Participants did not hear the same story twice. Stories were matched so that one group heard, for example, a story in the Gesture condition while another group heard the same story in the Self-Adaptor condition. Conditions were separated by a baseline period of 12–14 s. During baseline and No Visual Input, participants saw only a fixation cross but were not explicitly asked to fixate. Audio was delivered at a sound pressure level of 85 dB-SPL through headphones containing MRI-compatible electromechanical transducers (Resonance Technologies, Inc., Northridge, CA). Participants viewed video stimuli through a mirror attached to a head coil that allowed them to clearly see the hands and face of the actress on a large screen at the end of the scanning bed. Participants were monitored with a video camera.

Imaging Parameters

Functional imaging was performed at 3 Tesla (GE Medical Systems, Milwaukee, WI) with a gradient echo T2* spiral acquisition sequence sensitive to BOLD contrast [36]. A volumetric T1-weighted inversion recovery spoiled grass sequence was also used to acquire anatomical images.

Data Analysis

Functional images were spatially registered in three-dimensional space by Fourier transformation of each of the time points and corrected for head

movement, with AFNI software [37]. Signal intensities greater than 4.5 standard deviations from the mean signal intensity were considered artifact and removed from the time series (TS), which was linearly and quadratically detrended. The TS data were analyzed by multiple linear regression with separate regressors for each condition. These regressors were waveforms with similarity to the hemodynamic response generated by convolving a gamma-variant function with onset time and duration of the target blocks. The model also included a regressor for the mean signal and six motion parameters for each of the two runs. The resulting t-statistic associated with each condition was corrected for multiple comparisons by Monte Carlo simulation [38]. The TS was mean-corrected by the mean signal from the regression. The logical conjunction analysis (Figure S2) was performed for each participant individually by first thresholding each of the four conditions at a corrected p value of $p < .05$ by Monte Carlo simulation [38] and taking the intersection of active voxels for all conditions.

Anatomical volumes were inflated and registered to a template of average curvature with Freesurfer software [39–41]. Surface representations of each hemisphere for each participant were then automatically parcellated into ROIs [42], which were manually subdivided into further ROIs (Figure S1). The next step involved interpolating the corrected t-statistic associated with each regression coefficient and the TS data from the volume domain to the surface representation of each participant's anatomical volume via SUMA software [43]. A relational database was then created, with individual tables for each hemisphere of each participant's corrected t-statistic, TS, and ROI representations [21].

By using the R statistical package [44], we queried the database to extract from each ROI the TS of only those surface nodes that were active in at least one of the four conditions for each hemisphere of each participant. A node was determined to be active if any of the conditions was active above a corrected threshold of $p < .05$. The TSs corresponding to each of the active nodes in each of the ROIs for each hemisphere of each participant were averaged. The resulting TS for each participant was then split into the eight TSs corresponding to the TS for each condition from each run. The TS for each condition from each run was then concatenated by condition. In each of the four resulting TSs, time points with signal change values greater than 10% were replaced with the median signal change. Finally, the resulting four TSs were averaged over participants, thus establishing for each ROI one representative TS in each of the four conditions.

After this step, the second derivative of the TS was calculated for each condition. This second derivative was used to perform peak and valley analysis (as described in the text). Statistics were based on two-by-two contingency tables of the form peaks corresponding to and not corresponding to a stimulus feature, and valleys corresponding to and not corresponding to a stimulus feature. For contingency tables associated with Figure 2, a peak or valley was counted as containing or not containing a hand movement if that feature occupied at least 2/3 of or less than 1/3 of the total peak or valley time, respectively. See [20] and [21] for a more detailed description of the peak and valley analysis and [22] for a similar method.

The second derivative was also used to perform SEM because, as shown by the peak analysis (see e.g., Figure 1), it encodes peaks in the TS at second derivative minima that reflect events in the stories. We tested all possible SEMs by using the resulting second derivative of the TSs from the 5 regions for each condition for physically plausible connections, utilizing up to 150 processors on a grid computing system (see acknowledgments and [21]). To make computation manageable, models had only forward or backward connections between two regions. For each model, the SEM package was provided the correlation matrix derived from the second derivative of the TS between all regions within that model. Only models indicating a good fit, i.e., models whose χ^2 was not significant ($p > .05$), were retained.

Resulting connection weights from each model were averaged. Averaging was weighted by the Bayesian information criterion for each model. The Bayesian information criterion adjusts the chi-square of each model for

the number of parameters in the model, the number of observed variables, and the sample size. Individual connection weights were compared for the different models by using *t* tests correcting for heterogeneity of variance and unequal sample sizes by the Games-Howell method [45]. Degrees of freedom were calculated with Welch's method [45] (though the number of models for each condition were not significantly different). See [10] for a detailed description of the Bayesian averaging SEM approach.

Supplemental Data

Supplemental Data include four figures and two movies and can be found with this article online at [http://www.current-biology.com/supplemental/S0960-9822\(09\)00804-5](http://www.current-biology.com/supplemental/S0960-9822(09)00804-5).

Acknowledgments

Support to the senior author (S.L.S.) from NIH grants DC003378 and HD040605 is greatly appreciated. Thanks also to Bennett Bertenthal and Rob Gardner for TeraPort access and to Bernadette Brogan, E. Chen, Peter Huttenlocher, Goulven Josse, Uri Hasson, Susan Levine, Robert Lyons, Xander Meadow, Philippa Lauben, Arika Okrent, Fey Parrill, Linda Suriyakhham, Ana Solodkin, and Alison Wiener. In memory of Leo Stengel (6/20/31–11/27/08) from the first author (J.I.S.).

Received: July 23, 2008

Revised: February 13, 2009

Accepted: February 13, 2009

Published online: March 26, 2009

References

1. Skipper, J.I., Nusbaum, H.C., and Small, S.L. (2006). Lending a helping hand to hearing: Another motor theory of speech perception. In *Action to Language Via the Mirror Neuron System*, M.A. Arbib, ed. (Cambridge, MA: Cambridge University Press).
2. Sumbly, W.H., and Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.
3. McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.
4. Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think* (Cambridge, MA: Harvard University Press).
5. McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought* (Chicago, IL: The University of Chicago Press).
6. Cassell, J., McNeill, D., and McCullough, K.E. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics & Cognition* 7, 1–33.
7. Skipper, J.I., Nusbaum, H.C., and Small, S.L. (2005). Listening to talking faces: motor cortical activation during speech perception. *Neuroimage* 25, 76–89.
8. Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., and Small, S.L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399.
9. Ekman, P., and Friesen, W. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–58.
10. Skipper, J.I., Goldin-Meadow, S., Nusbaum, H.C., and Small, S.L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang.* 101, 260–277.
11. Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., and Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404.
12. Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
13. Wheaton, K.J., Thompson, J.C., Syngieniotis, A., Abbott, D.F., and Puce, A. (2004). Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage* 22, 277–288.
14. Rizzolatti, G., Luppino, G., and Matelli, M. (1998). The organization of the cortical motor system: new concepts. *Electroencephalogr. Clin. Neurophysiol.* 106, 283–296.
15. Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157.
16. Crinion, J., and Price, C.J. (2005). Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain* 128, 2858–2871.
17. Humphries, C., Willard, K., Buchsbaum, B., and Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport* 12, 1749–1752.
18. Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432.
19. Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.
20. Chen, E.E., Skipper, J.I., and Small, S.L. (2007). Revealing tuning functions with brain imaging data using a peak fit analysis. *Neuroimage* 36, S1–S125.
21. Hasson, U., Skipper, J.I., Wilde, M.J., Nusbaum, H.C., and Small, S.L. (2008). Improving the analysis, storage and sharing of neuroimaging data using relational databases and distributed computing. *Neuroimage* 39, 693–706.
22. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Inter-subject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640.
23. McIntosh, A.R., and Gonzalez-Lima, F. (1994). Structural equation modelling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2, 2–22.
24. Hanson, S.J., Matsuka, T., Hanson, C., Rebbecki, D., Halchenko, Y., and Zaimi, A. (2004). Structural equation modeling of neuroimaging data: exhaustive search and markov chain Monte Carlo. 10th Annual Meeting of the Organization for Human Brain Mapping, Budapest, Hungary.
25. Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian model averaging: a tutorial. *Stat. Sci.* 14, 382–417.
26. Kass, R.E., and Raftery, A.E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795.
27. Madigan, D., and Raftery, A.E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Am. Stat. Assoc.* 89, 1535–1546.
28. Gough, P.M., Nobre, A.C., and Devlin, J.T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J. Neurosci.* 25, 8010–8016.
29. Moss, H.E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., and Tyler, L.K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cereb. Cortex* 15, 1723–1735.
30. Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., and Farah, M.J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. USA* 94, 14792–14797.
31. Wagner, A.D., Pare-Blagoev, E.J., Clark, J., and Poldrack, R.A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338.
32. Willems, R.M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333.
33. Arbib, M.A., Erdi, P., and Szentágothai, J. (1998). *Neural Organization: Structure, Function and Dynamics* (Cambridge, MA: The MIT Press).
34. Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
35. Holle, H., Gunter, T.C., Ruschemeyer, S.A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024.
36. Noll, D.C., Cohen, J.D., Meyer, C.H., and Schneider, W. (1995). Spiral k-space MRI imaging of cortical activation. *J. Magn. Reson. Imaging* 5, 49–56.
37. Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.
38. Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., and Noll, D.C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.
39. Fischl, B., Sereno, M.I., Tootell, R.B., and Dale, A.M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284.

40. Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.
41. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
42. Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
43. Argall, B.D., Saad, Z.S., and Beauchamp, M.S. (2005). Simplified inter-subject averaging on the cortical surface using SUMA. *Hum. Brain Mapp.* 27, 14–27.
44. Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Statist.* 5, 299–314.
45. Kirk, R.E. (1995). *Experimental Design: Procedures for the Behavioral Sciences* (Pacific Grove, CA: Brooks-Cole Publishing Company).