

3-26-2012

# A network model of observation and imitation of speech

Nira Mashal

*University of Chicago; Bar-Ilan University*

Ana Solodkin

*University of Chicago; University of California*

Anthony Steven Dick

*Department of Psychology, Florida International University; University of Chicago, [adick@fiu.edu](mailto:adick@fiu.edu)*

E. Elinor Chen

*University of Chicago; University of California*

Steven L. Small

*University of Chicago; University of California*

Follow this and additional works at: [https://digitalcommons.fiu.edu/psychology\\_fac](https://digitalcommons.fiu.edu/psychology_fac)



Part of the [Psychology Commons](#)

---

## Recommended Citation

Mashal N, Solodkin A, Dick AS, Chen EE and Small SL (2012) A network model of observation and imitation of speech. *Front. Psychology* 3:84. doi: 10.3389/fpsyg.2012.00084

This work is brought to you for free and open access by the College of Arts, Sciences & Education at FIU Digital Commons. It has been accepted for inclusion in Department of Psychology by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).



# A network model of observation and imitation of speech

Nira Mashal<sup>1,2</sup>, Ana Solodkin<sup>1,3,4</sup>, Anthony Steven Dick<sup>1,5</sup>, E. Elinor Chen<sup>1,3</sup> and Steven L. Small<sup>1,4</sup>\*

<sup>1</sup> Department of Neurology, The University of Chicago, Chicago, IL, USA

<sup>2</sup> School of Education, Bar-Ilan University, Ramat Gan, Israel

<sup>3</sup> Department of Anatomy and Neurobiology, University of California, Irvine, Irvine, CA, USA

<sup>4</sup> Department of Neurology, University of California Irvine, Irvine, CA, USA

<sup>5</sup> Department of Psychology, Florida International University, Miami, FL, USA

## Edited by:

Yury Y. Shtyrov, Medical Research Council, UK

## Reviewed by:

Max Garagnani, MRC Cognition and Brain Sciences Unit, UK

Alessandro D'Ausilio, Istituto Italiano di Tecnologia, Italy

## \*Correspondence:

Steven L. Small, Department of Neurology, University of California Irvine, Irvine, CA 92697, USA.

e-mail: small@uci.edu

Much evidence has now accumulated demonstrating and quantifying the extent of shared regional brain activation for observation and execution of speech. However, the nature of the actual networks that implement these functions, i.e., both the brain regions and the connections among them, and the similarities and differences across these networks has not been elucidated. The current study aims to characterize formally a network for observation and imitation of syllables in the healthy adult brain and to compare their structure and effective connectivity. Eleven healthy participants observed or imitated audiovisual syllables spoken by a human actor. We constructed four structural equation models to characterize the networks for observation and imitation in each of the two hemispheres. Our results show that the network models for observation and imitation comprise the same essential structure but differ in important ways from each other (in both hemispheres) based on connectivity. In particular, our results show that the connections from posterior superior temporal gyrus and sulcus to ventral premotor, ventral premotor to dorsal premotor, and dorsal premotor to primary motor cortex in the left hemisphere are stronger during imitation than during observation. The first two connections are implicated in a putative dorsal stream of speech perception, thought to involve translating auditory speech signals into motor representations. Thus, the current results suggest that flow of information during imitation, starting at the posterior superior temporal cortex and ending in the motor cortex, enhances input to the motor cortex in the service of speech execution.

**Keywords:** speech, language, mirror neuron, structural equation modeling, effective connectivity, action observation, ventral premotor cortex, brain imaging

## INTRODUCTION

In everyday communication, auditory speech is accompanied by visual information from the speaker, including movements of the lips, mouth, tongue, and hands. Observing these motor actions improves speech perception, particularly under noisy conditions (MacLeod and Summerfield, 1987) or when the auditory signal is degraded (Sumbly and Pollack, 1954; Ross et al., 2007). One putative neural mechanism postulated to account for this phenomenon is *observation–execution matching*, whereby observed actions (e.g., oral motor actions) are matched by the perceiver to a repertoire of previously executed actions (i.e., previous speech). Support for this matching hypothesis comes from recent studies showing that the brain areas active during action observation and action execution contain many shared components, and that such overlap exists for movements of the finger, hand, and arm (e.g., Tanaka and Inui, 2002; Buccino et al., 2004b; Molnar-Szakacs et al., 2005), as well as those of the mouth and lips during speech (Fadiga et al., 1999; Wilson et al., 2004; Skipper et al., 2005, 2007; D'Ausilio et al., 2011). Although these previous studies demonstrate both commonalities and differences in regional brain activation for observation and execution, they do not characterize the networks that implement these functions in terms of effective connectivity, i.e., the functional influence of one region over those with which it is

anatomically connected. With such network descriptions, as we elucidate here, it is possible to show the quality and degree to which functional brain circuits for observation and execution are intertwined, and thus to test the degree of functional overlap (or lack thereof) related to the interactions established by the activated brain regions.

Studies aiming to characterize the neural mechanisms for observation and imitation of speech have used advanced brain imaging techniques and have shown that some similar brain regions are activated during the two tasks, particularly in motor regions involved in speech [i.e., ventral premotor cortex (vPM) and adjacent pars opercularis of the inferior frontal gyrus]. Although the pars opercularis has been traditionally thought to be critical for speech production (Geschwind, 1970; Ojemann et al., 1989), an increasing number of studies have shown that the underlying implementation of this function may be integrated in a multimodal fashion with visual (MacSweeney et al., 2000; Hasson et al., 2007) and audiovisual speech perception (Skipper et al., 2005, 2007). For example, silent lip-reading increases brain activity bilaterally in the premotor cortex and Broca's area (particularly pars opercularis and its homolog; MacSweeney et al., 2000), and activation in left pars opercularis is associated with individual differences in the integration of visual and auditory speech information

(Hasson et al., 2007). In macaque, related areas appear to be critical for integration of parietal sensory–motor signals with higher-order information originating from multiple frontal areas, with information shared across adjacent areas (Gerbella et al., 2011).

Both passive listening to monosyllables and production of the same syllables leads to overlapping activation in a superior portion of the vPM (Wilson et al., 2004). The time course of activation on a related task – observing and imitating lip forms – successively incorporates the occipital cortex, superior temporal region, inferior parietal lobule, inferior frontal, and ultimately the primary motor cortex, with stronger activation during imitation than observation (Nishitani and Hari, 2002). Using audiovisual stimuli, we previously showed observation/execution overlap in posterior superior temporal cortices, inferior parietal areas, pars opercularis, premotor cortices, primary motor cortex, subcentral gyrus and sulcus, insula, and cerebellum (Skipper et al., 2007). Overall, a number of studies have reported engagement of speech-motor regions in visual (MacSweeney et al., 2000; Nishitani and Hari, 2002), auditory (Fadiga et al., 2002; Wilson et al., 2004; Tettamanti et al., 2005; Mottonen and Watkins, 2009; Sato et al., 2009; D’Ausilio et al., 2011; Tremblay et al., 2011), and audiovisual speech perception (Campbell et al., 2001; Fadiga et al., 2002; Calvert and Campbell, 2003; Paulesu et al., 2003; Watkins et al., 2003; Skipper et al., 2005, 2006, 2007).

The consistent activation of the pars opercularis, inferior parietal lobule, and vPM in studies of speech perception and imitation is predicted by several related accounts of audiovisual speech perception and production, and the relation between them. One set of accounts has emphasized the contribution of motor cortex to speech perception during audiovisual language comprehension (see Schwartz et al., 2012 for review). An influential perspective from this vantage point argues that motor cortex activation in speech perception is the product of “direct matching” of a perceived action with the observer’s previous motor experience with that action (Rizzolatti et al., 2001). This view further hypothesizes that such matching is accomplished, at least in part, by a special class of neurons, called “mirror neurons.” Mirror neurons are sensory–motor neurons, originally characterized from recordings in area F5 of the vPM of the macaque brain, that discharge during both observation and execution of the same goal-oriented actions (Fadiga et al., 1995; Strafella and Paus, 2000; Rizzolatti et al., 2001). Mirror neurons have also been identified in the rostral part of the inferior parietal cortex (areas PF and PFG) in macaque (Fogassi et al., 2005; Fabbri-Destro and Rizzolatti, 2008; Rozzi et al., 2008; for review, see Cattaneo and Rizzolatti, 2009). Mirror neurons have been found in the macaque for both oral actions and manual actions, and human imaging studies have demonstrated task-dependent functional brain activation to observation and execution that fits this pattern and suggests that mirror neurons may also exist in the human (Buccino et al., 2001; Grezes et al., 2003; for review, see Buccino et al., 2004a). Within F5, the mirror neurons are located primarily in the caudal sector in the cortical convexity of F5 (area F5c).

Visual action information from STS appears to take two different pathways to the frontal lobe, with distinct projections first to the parietal lobe and then to areas F5c and F5ab of the inferior frontal lobe. One route begins in the upper bank of the STS,

and projects to PF/PFG in the inferior parietal region (Kurata, 1991; Rizzolatti and Fadiga, 1998; Nelissen et al., 2011), which corresponds roughly to the human supramarginal gyrus, and then projects to premotor area F5c. This pathway appears to emphasize information about the agent and the intentions of the agent, and comprises the parieto-frontal mirror circuit involved in visual transformation for grasping (Jeannerod et al., 1995; Rizzolatti and Fadiga, 1998). The other pathway begins on the lower bank of STS, and connects to the frontal region F5ab via the IPS (Luppino et al., 1999; Borra et al., 2008; Nelissen et al., 2011), probably subregion AIP, whereas the second emphasizes information about the object.

We recently observed a related, but topographically different, organization in the human PMv, with a ventral PMv sector containing neurons with mirror properties, and a dorsal PMv sector containing neurons with canonical properties (Tremblay and Small, 2011).

It is not known if motor cortical regions are necessary for speech perception (Sato et al., 2008; D’Ausilio et al., 2009; Tremblay et al., 2011) or are facilitatory, playing a particularly important role in situations of decreased auditory efficiency (e.g., hearing loss, noisy environment; Hickok, 2009; Lotto et al., 2009). In either case, brain networks that include frontal and parietal motor cortical regions are activated during speech perception, and may represent a physiological mechanism by which brain circuits for motor execution aid in the understanding of speech. One way this could occur is by “direct matching” (Rizzolatti et al., 2001; Gallese, 2003), in which an individual recognizes speech by mapping perceptions onto motor representations using a sensory–motor circuit including posterior inferior frontal/ventral premotor, inferior parietal, and posterior superior temporal brain regions (Callan et al., 2004; Guenther, 2006; Guenther et al., 2006; Skipper et al., 2007; Dick et al., 2010).

Although these prior studies demonstrate participation of these visuo-motor regions in speech perception, there does not yet exist a characterization of the organization of these regions into an effectively connected network relating speech production with speech perception. In this paper, we describe such a network organization, and show the relation between the human effective network for observing speech (without a goal of execution) and imitating speech (observing with a goal of execution and then executing). Specifically, we present a formal structural equation (effective connectivity) model of the neural networks used for observation and imitation of audiovisual syllables in the normal state, and compare the structure and effective connectivity of observation and imitation networks in both left and right hemispheres.

In investigating these questions, we have three hypotheses. (i) First, we postulate a gross anatomical similarity between the networks for observation and imitation, i.e., optimal models of the raw imaging data can be described with a core of similar nodes (regions), since there will be overlapping regional activation during both observation and imitation. (ii) Second, we suggest that the effective connections within the network will be of approximately equal strength, particularly those with larger motor biases, such as the connection between the inferior frontal/ventral premotor regions and the inferior parietal regions. (iii) Third, we expect that the networks with the best fit to the data will differ between the left and right hemispheres for both observation and imitation, based

on the postulated left-hemispheric bias for auditory language processing (Hickok and Poeppel, 2007). Further, based on previous findings in speech perception and auditory language understanding (e.g., Mazoyer et al., 1993; Binder et al., 1997) and imitation (e.g., Saur et al., 2008), we expect stronger effective connectivity among relevant regions in the left hemisphere (LH) during imitation compared to observation since the former requires speech output (e.g., see Nishitani and Hari, 2002 for a discussion).

To test these three hypotheses, we focused on six regions that have been shown in previous studies to be involved in speech perception. These regions include (i) vPM and inferior frontal gyrus (combined region); (ii) inferior parietal lobule (including intraparietal sulcus); (iii) primary motor and sensory cortices (M1S1); (iv) dorsal premotor cortex (dPM); (v) posterior superior temporal gyrus and sulcus (combined region); and (vi) anterior superior temporal gyrus and sulcus (combined region).

## MATERIALS AND METHODS

### PARTICIPANTS

Eleven adults (seven females, mean age =  $24 \pm 5$ ) participated. All were right handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), with no history of neurological or psychiatric illness. Participants gave written informed consent and the Institutional Review Board of the Biological Sciences Division of The University of Chicago approved the study.

### STIMULI AND TASK

Participants performed two tasks. In the Observation task, participants passively watched and listened to a female actress (filmed from neck up) articulating four syllables with different articulatory profiles in terms of lip and tongue movements: /pa/, /fa/, /ta/, and /tha/. In the Imitation task, participants were asked to say the syllable out loud immediately after observing the same actress. Each syllable was presented for 1.5 s. The Observation run was 6'30" (6 minutes and 30 seconds) in duration (260 whole-brain images) and the Imitation run lasted 12'30" (12 minutes and 30 seconds) (500 whole-brain images). Each run contained a total of 120 stimuli (30 stimuli for each syllable). In each of these runs, stimuli were presented in a randomized event-related manner with a variable interstimulus interval (ISI; minimum ISI for Observation = 0 s; minimum ISI for Imitation = 1.5 s, maximum ISI = 12 s for both runs). The ISI formed the baseline for computation of the hemodynamic response. Participants viewed the video stimuli through a mirror attached to the head coil that allowed them to see a screen at the end of the scanning bed. The audio track was simultaneously delivered to participants at 85 dB SPL via headphones containing MRI-compatible electromechanical transducers (Resonance Technologies, Inc., Northridge, CA, USA). Before the beginning of the experiment, participants were trained inside the scanner with a set of four stimuli to ensure they understood the tasks and could hear properly the voice of the actress.

### IMAGING AND DATA ANALYSIS

Functional imaging was performed at 3 T (TR = 1.5 s; TE = 25 ms; FA = 77°; 29 axial slices; 5 mm × 3.75 mm × 3.75 mm voxels) on a GE Signa scanner (GE Medical Systems, Milwaukee, WI, USA) using spiral BOLD acquisition (Noll et al., 1995). A volumetric

T1-weighted inversion recovery spoiled grass sequence (120 axial slices, 1.5 mm × 0.938 mm × 0.938 mm resolution) was used to acquire structural images on which anatomical landmarks could be found and functional activation maps could be superimposed.

## DATA ANALYSIS

### Preprocessing and identification of task-related activity

Functional image preprocessing for each participant consisted of three-dimensional motion correction using weighted least-squares alignment of three translational and three rotational parameters, as well as registration to the first non-discarded image of the first functional run, and to the anatomical volumes (Cox and Jesmanowicz, 1999)<sup>1</sup>. The time series were linearly detrended and despiked, the impulse response function was estimated using deconvolution, and analyzed statistically using multiple linear regression. The two principal regressors were for the Observation task and the Imitation task. Nine sources of non-specific variance were removed by regression, including six motion parameters, the signal averaged over the whole-brain, the signal averaged over the lateral ventricles, and the signal averaged over a region centered in the deep cerebral white matter. The regressors were converted to percent signal change values relative to the baseline, and significantly activated voxels were selected after correction for multiple comparisons using false discovery rate (FDR; Benjamini and Hochberg, 1995; Genovese et al., 2002) with a whole-brain alpha of  $p < 0.05$ .

### Whole-brain group analysis of condition differences

A group analysis was conducted on the whole-brain to determine whether there was a significant group level activation relative to a resting baseline, and to compare condition differences at the group level. We conducted one-sample *t* tests to assess activation relative to zero, and dependent paired-sample *t* tests to assess condition differences. These were computed on a voxel-wise basis using the normalized regression coefficients as the dependent variable. To control for multiple comparisons, we used the FDR procedure ( $p < 0.05$ ).

### Network analysis using structural equation modeling

The primary analysis was a network analysis using SEM (McIntosh, 2004), which was performed using AMOS software (Arbuckle, 1989), which can be used to model fMRI data from both block and event-related designs (Gates et al., 2011). We first specified a theoretical anatomical model, which consisted of the regions comprising the nodes of the network, and the directional connections (i.e., paths) among them. Our hypotheses focused on six anatomical regions, identified on each individual participant. The regions of the model, which are specified further in Table 1, included M1S1, dPM, vPM including *pars opercularis* of the inferior frontal gyrus and the inferior portion of the precentral sulcus and gyrus, inferior parietal lobule (IP) including the intraparietal sulcus, posterior superior temporal gyrus and sulcus (pST), and anterior superior temporal gyrus and sulcus (aST). Connections were specified with reference to known macaque anatomical connectivity (e.g.,

<sup>1</sup><http://afni.nimh.nih.gov/afni/>

**Table 1 | Anatomical description of the cortical regions of interest.**

ROI	Anatomical structure	Brodmann's area	Delimiting landmarks
IFGOp/PMV	<i>Pars opercularis</i> of the inferior frontal gyrus, inferior precentral sulcus, and inferior precentral gyrus	6, 44	A = anterior vertical ramus of the sylvian fissure P = central sulcus S = inferior frontal sulcus, extending a horizontal plane posteriorly across the precentral gyrus I = anterior horizontal ramus of the sylvian fissure to the border with insular cortex
PMd	<i>Pars opercularis</i> of the inferior frontal gyrus, inferior precentral sulcus, and inferior precentral gyrus	6	A = vertical plane through the anterior commissure P = central sulcus S = medial surface of the hemisphere I = inferior frontal sulcus, extending a horizontal plane posteriorly across the precentral gyrus
IP	Supramarginal gyrus; angular gyrus; intraparietal sulcus	39, 40	A = postcentral sulcus P = sulcus intermedius secundus S = superior parietal gyrus I = horizontal posterior segment of the superior temporal sulcus
STa	Anterior portion of the superior temporal gyrus, superior temporal sulcus, and planum polare	22	A = inferior circular sulcus of insula P = a vertical plane drawn from the anterior extent of the transverse temporal gyrus S = anterior horizontal ramus of the sylvian fissure I = middle temporal gyrus
STp	Posterior portion of the superior temporal gyrus, superior temporal sulcus, and planum temporale	22, 42	A = a vertical plane drawn from the anterior extent of the transverse temporal gyrus P = angular gyrus S = supramarginal gyrus I = middle temporal gyrus
	Central sulcus; postcentral gyrus	1, 2, 3, 4	A = precentral gyrus P = postcentral sulcus S = medial surface of the hemisphere I = parietal operculum

A, anterior; P, posterior; S, superior; I, inferior.

Petrides and Pandya, 1984; Matelli et al., 1986; Seltzer and Pandya, 1994; Rizzolatti and Matelli, 2003; Schmahmann et al., 2007)

Definition of these regions on each individual participant was obtained using the automated parcellation procedure in *Freesurfer*<sup>2</sup>. Cortical surfaces were inflated (Fischl et al., 1999a) and registered to a template of average curvature (Fischl et al., 1999b). The surface representations of each hemisphere of each participant were then automatically parcellated into regions (Fischl et al., 2004). Small modifications to this parcellation were made manually (see **Table 1** for anatomical definition).

For SEM, we first re-sampled the (rapid event-related) time series to enable assessment of variability and thus quantification of goodness of fit. We first obtained time series from the peak voxel in each ROI (voxel associated with the highest *t* value from all active voxels; corrected FDR  $p < 0.05$ ). The peak voxel approach was chosen because it has been shown empirically in comparison with other approaches to result in robust models across individual participants contributing to a group model (Walsh et al., 2008). We

then re-sampled these time series (260 and 500 time points for the Observation and Imitation conditions, respectively) down to 78 in the LH and 77 points in the RH using a locally weighted scatterplot smoothing (LOESS) method. In this method, each re-sampled data point is estimated with a weighted least-squares function, giving greater weight to actual time points near the point being estimated, and less weight to points farther away (Cleveland and Devlin, 1988). Non-significant Box's *M* tests indicated no differences in the variance-covariance structure of the re-sampled and original data. The SEM analysis was conducted on these re-sampled time series.

To specify a theoretical model constrained by known anatomy, and to determine whether it was able to reproduce the observed data, we used maximum likelihood estimation. We first estimated the path coefficients based on examination of the interregional correlations, which were used as starting values to facilitate maximum likelihood estimation (McIntosh and Gonzalez-Lima, 1994). We assessed the difference between the predicted and the observed solution using the stacked model (multiple group) approach (Gonzalez-Lima and McIntosh, 1994; McIntosh and Gonzalez-Lima, 1994; McIntosh et al., 1994). If the  $\chi^2$  statistic characterizing

<sup>2</sup><http://surfer.nmr.mgh.harvard.edu>

the difference between the models is not significant, then the null hypotheses (i.e., that there is no difference between the predicted and the observed data) should be retained, and the model represents a good fit. Note that in cases where two models have different degrees of freedom, missing nodes are included with random-constant time series and its connections are added to the less specified model with connection strength of zero to permit comparison (Solodkin et al., 2004).

## RESULTS

### WHOLE-BRAIN ANALYSIS: ACTIVATION COMPARED TO RESTING BASELINE AND ACROSS CONDITIONS

Patterns of activation in Imitation and Observation conditions were quite similar, with activation in the occipital cortex, anterior and posterior superior temporal regions, inferior frontal gyrus, and primary sensory-motor cortex bilaterally. All activations were of higher volume and intensity during Imitation compared to Observation (see **Table 2** for the quantitative data). Activation during Imitation but not Observation extended to anterior parts of the IFG (i.e., pars orbitalis) bilaterally. The activation profile from a representative participant is shown in **Figure 1**.

### STRUCTURAL EQUATION MODELS: MODELS OF OBSERVATION AND IMITATION IN THE LEFT HEMISPHERE

The predicted model for the LH fit the data for both Observation and Imitation (for Observation:  $\chi^2 = 1.8$ ,  $df = 1$ ,  $p = 0.18$ ; for Imitation  $\chi^2 = 0.01$ ,  $df = 1$ ,  $p = 0.92$ ; see **Figure 2**). The strongest effective connections ( $EF > 0.4$ ) for both Observation and Imitation models included those from pST to IP (0.60, 0.72, in Observation and Imitation, respectively), from IP to vPM (0.63, 0.47, respectively), from vPM to M1S1 (0.81, 0.73), and from pST to aST (0.54, 0.50, respectively).

There were also important differences based on comparison with the stacked model approach (Gonzalez-Lima and McIntosh, 1994; McIntosh and Gonzalez-Lima, 1994; McIntosh et al., 1994). Overall, the models for Observation and Imitation differed ( $\chi^2 = 55.2$ ,  $df = 18$ ,  $p < 0.0001$ ), suggesting differences in the magnitude of some of the path coefficients of

the models. Although many of the coefficients did not differ (**Figure 3**; including IP  $\rightarrow$  vPM, pST  $\rightarrow$  aST, pST  $\rightarrow$  IP, aST  $\rightarrow$  IP, IP  $\rightarrow$  dPM, IP  $\rightarrow$  M1S1, vPM  $\rightarrow$  M1S1), connections from pST to vPM ( $\chi^2 = 11.7$ ,  $df = 1$ ,  $p < 0.001$ ), vPM to dPM ( $\chi^2 = 5.2$ ,  $df = 1$ ,  $p < 0.05$ ), and dPM to M1S1 ( $\chi^2 = 14.0$ ,  $df = 1$ ,  $p < 0.001$ ) were stronger during Imitation than during Observation (**Figure 4**).

### STRUCTURAL EQUATION MODELS: MODELS OF OBSERVATION AND IMITATION IN THE RIGHT HEMISPHERE

Connectivity models with similar nodes characterized both the Observation and Imitation conditions (for Observation:  $\chi^2 = 2.5$ ,  $df = 1$ ,  $p = 0.11$ ; for Imitation:  $\chi^2 = 3.1$ ,  $df = 2$ ,  $p = 0.21$ ; **Figure 2**). As in the case of the LH, there were differences in the magnitude of the path coefficients across conditions ( $\chi^2 = 174.3$ ,  $df = 17$ ,  $p < 0.001$ ). As can be seen from **Figure 3**, some of the connections are different during Imitation from Observation: IP  $\rightarrow$  vPM ( $\chi^2 = 15.4$ ,  $df = 1$ ,  $p < 0.01$ ), pST  $\rightarrow$  vPM ( $\chi^2 = 7.4$ ,  $df = 1$ ,  $p < 0.01$ ), pST  $\rightarrow$  M1S1 ( $\chi^2 = 7.2$ ,  $df = 1$ ,  $p < 0.01$ ), vPM  $\rightarrow$  dPM ( $\chi^2 = 11.5$ ,  $df = 1$ ,  $p < 0.001$ ), and vPM  $\rightarrow$  M1S1 ( $\chi^2 = 28.5$ ,  $df = 1$ ,  $p < 0.001$ ). In contrast with the LH, some of the connections were different during Observation from Imitation: pST  $\rightarrow$  IP ( $\chi^2 = 14.4$ ,  $df = 1$ ,  $p < 0.001$ ), pST  $\rightarrow$  dPM ( $\chi^2 = 4.5$ ,  $df = 1$ ,  $p < 0.05$ ), and IP  $\rightarrow$  M1S1 ( $\chi^2 = 12.2$ ,  $df = 1$ ,  $p < 0.001$ ).

### STRUCTURAL EQUATION MODELS: MODELS OF IMITATION IN LH VS RH

The models for Imitation differed across hemispheres ( $\chi^2 = 84.2$ ,  $df = 18$ ,  $p < 0.0001$ ). Specifically, three connections were different in the LH compared to the RH: pST  $\rightarrow$  IP ( $\chi^2 = 16.8$ ,  $df = 1$ ,  $p < 0.001$ ), vPM  $\rightarrow$  M1S1 ( $\chi^2 = 17.7$ ,  $df = 1$ ,  $p < 0.001$ ), aST  $\rightarrow$  dPM ( $\chi^2 = 7$ ,  $df = 1$ ,  $p < 0.05$ ). No connections were different in the RH from the LH.

### STRUCTURAL EQUATION MODELS: MODELS OF OBSERVATION IN THE LH VS RH

The models for Observation differed across hemispheres ( $\chi^2 = 50.8$ ,  $df = 16$ ,  $p < 0.001$ ). Specifically, three connections were different in the LH than the RH: IP  $\rightarrow$  vPM ( $\chi^2 = 13.8$ ,  $df = 1$ ,  $p < 0.001$ ), IP  $\rightarrow$  dPM ( $\chi^2 = 4.6$ ,  $df = 1$ ,  $p < 0.05$ ), and pST  $\rightarrow$  M1S1 ( $\chi^2 = 6.7$ ,  $df = 1$ ,  $p < 0.01$ ).

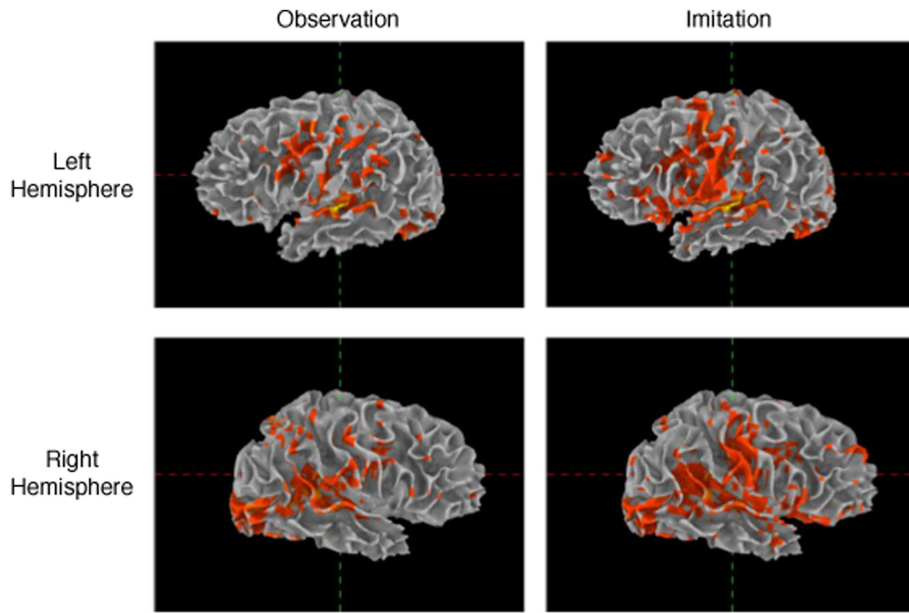
## DISCUSSION

The present study examined three hypotheses regarding effective connectivity among brain regions important for observation and imitation of audiovisual syllables in the healthy adult. In our first hypothesis, we predicted structural similarity (i.e., similar active regions) across conditions, and our findings support this: The networks for Observation and Imitation incorporate the same nodes (brain regions). In our second hypothesis, we predicted similarity in regional interconnectivity across Observation and Imitation, and found partial support for this: while we did find considerable similarity across Imitation and Observation (e.g., see **Figure 4**), we also found several differences in connectivity in both hemispheres. Interestingly, the effective connectivity differences are not restricted to connections between historically identified “motor” areas (e.g., the connection between ventral premotor to dorsal premotor), as would be expected when motor execution is necessary for Imitation but not for Observation. While we did find this,

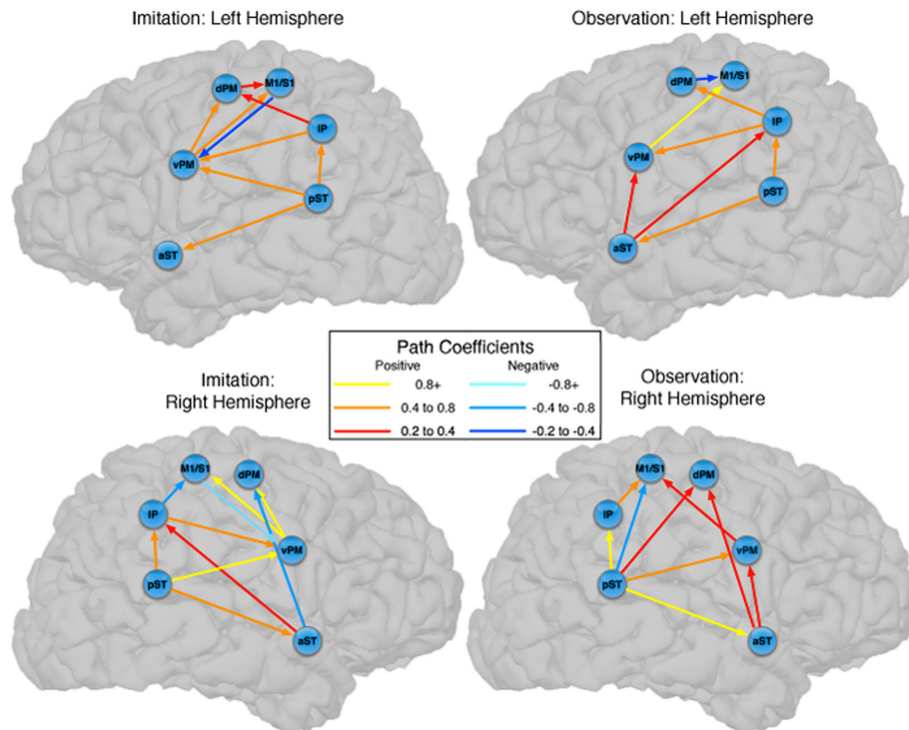
**Table 2 | Average number of active voxels in each of the regions of interest (FDR corrected  $p < 0.05$ ).**

Region	Observation		Imitation	
	LH	RH	LH	RH
IP	32	36	79	79
M1S1	5	2	52	47
pST	22	53	38	67
aST	16	20	36	37
vPM	25	25	95	67
dPM	17	15	78	60

LH, left hemisphere; RH, right hemisphere; IP, inferior parietal lobule; M1S1, primary motor/somatosensory cortex; pST, posterior superior temporal gyrus and sulcus; aST, anterior superior temporal gyrus and sulcus; vPM, ventral premotor cortex; dPM, dorsal premotor cortex. Anatomical definition of the regions is provided in **Table 1**.

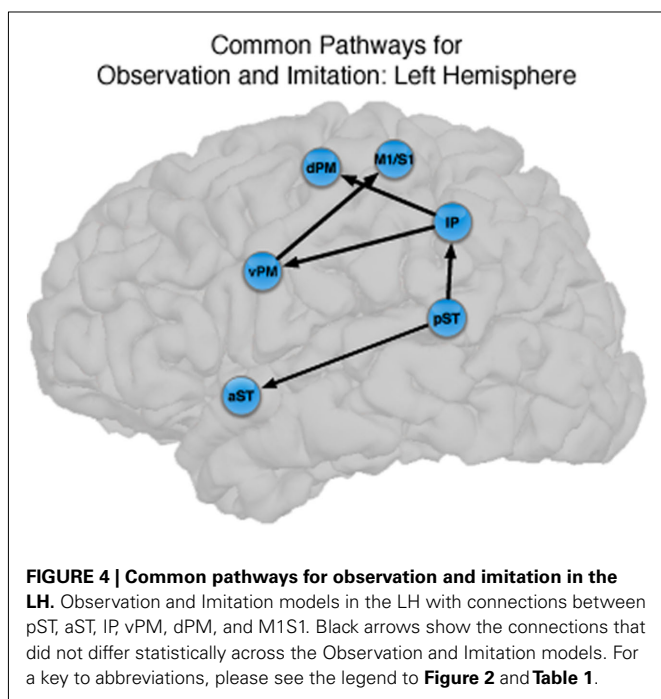
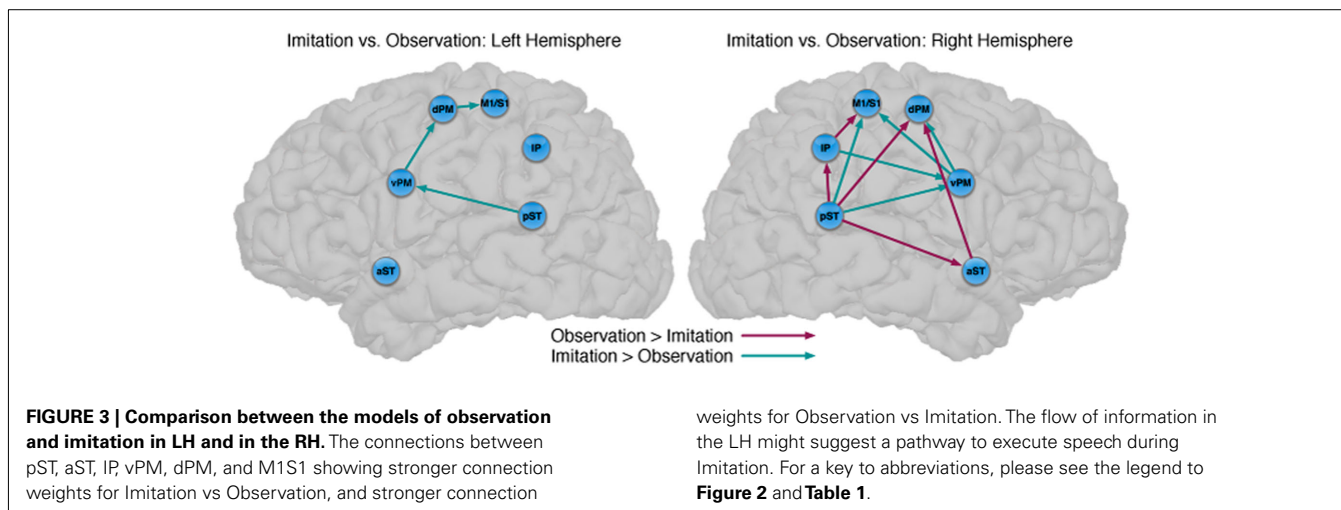


**FIGURE 1 | Activation during observation and imitation.** Voxels were selected using general linear model after adjusting for false positives using false discovery rate ( $p < 0.05$ ). The figure shows the data obtained from a representative single subject.



**FIGURE 2 | Observation and Imitation models in both the LH and RH with connections between pST, aST, IP, vPM, dPM, and M1S1.** IP, inferior parietal lobule. M1S1, primary motor/somatosensory cortex; pST, posterior superior temporal

gyrus and sulcus; aST, anterior superior temporal gyrus and sulcus; vPM, ventral premotor cortex; dPM, dorsal premotor cortex; M1/S1, primary motor/somatosensory cortex. Anatomical definitions of the regions are provided in **Table 1**.



we also found differences in sensory–motor interactions (e.g., the connection between posterior superior temporal region and vPM). In our third hypothesis, we predicted stronger effective connectivity during Imitation compared to Observation, particularly in the LH, since the former requires speech output and the latter does not. For the “dorsal stream” pathway connecting pST → vPM → dPM and M1/S1, we found with stronger connectivity for Imitation compared to Observation in both hemispheres. Additional differences in connectivity were found across the two conditions in the right hemisphere (RH).

It is important to note that these models reflect effective connectivity and not anatomical connectivity. Thus, whereas we show the presence of overlapping networks for Observation and Imitation, characterized by similar anatomical regions and similar statistical covariation among activity in these regions, we cannot

make conclusions about brain anatomy, i.e., the white matter connections among these regions. SEM does not assess anatomical pathways directly, but rather statistical covariance in the BOLD response. Nevertheless, these networks present strong evidence on effective connectivity, which incorporates an *a priori* anatomical model (based largely on what is known about connectivity in the non-human primate), but still represents statistical covariation and not explicit anatomical evidence, supporting a human system for observation–imitation matching in speech perception.

**OBSERVATION AND IMITATION IN THE LH**

The results we report with respect to BOLD signal amplitude replicate previous studies in audiovisual speech perception that show brain activation in regions involved in planning and execution of speech (Calvert et al., 2000; Callan et al., 2003, 2004; Calvert and Campbell, 2003; Jones and Callan, 2003; Sekiyama et al., 2003; Wright et al., 2003; Miller and D’Esposito, 2005; Ojanen et al., 2005; Skipper et al., 2005, 2007; Pekkola et al., 2006; Pulvermuller et al., 2006; Bernstein et al., 2008). Specifically, we showed that several regions were active during both speech production and speech perception (**Table 2**; cf. (Pulvermuller et al., 2006; Skipper et al., 2007)). These regions were also activated in an event-related MEG study (Nishitani and Hari, 2002), which showed temporal progression of activity for both observation and imitation (of static lip forms) from the occipital cortex to the pST, the IP, IFG, to the sensory–motor cortex (M1/S1). In our current work, we elaborate on these activation studies by elucidating the functional relationships between the relevant regions, i.e., showing the basic organization of the network in terms of effective connections and relative strengths across conditions and hemispheres.

The novel contribution of the present work is a characterization of the networks for observation and imitation of dynamic speech stimuli, and we found that the functional interactions among brain regions that were active during Observation and Imitation share both similarities and differences. **Figure 4** illustrates the connections with similar strength during both conditions in the LH. The current models are consistent with the time course demonstrated by prior MEG results (Nishitani and Hari, 2002) and with previous models of effective connectivity during the perception of



intelligible speech (between superior temporal and inferior frontal regions; Leff et al., 2008) and speech production (between inferior frontal/ventral premotor regions and primary motor cortex during production; Eickhoff et al., 2009). We consider these similarities below.

The models presented here include integral connections from pST  $\rightarrow$  IP, from IP  $\rightarrow$  IFG/vPM, and from IFG/vPM  $\rightarrow$  M1S1. Both pST and IP have been implicated in speech perception, and both are activated during acoustic and phonological analyses of speech (e.g., Binder et al., 2000; Burton et al., 2000; Wise et al., 2001). pST is activated by observation of biologically relevant movements including mouth, hands, and limb movements (Allison et al., 2000). This model represents a hierarchical network from the sensory temporal and parietal lobules, to inferior frontal and ventral premotor regions, to execution by primary motor cortex. In fact, it is quite similar, in many respects, to the results presented by Nishitani and Hari (2002) in their MEG study of observation and imitation of static speech stimuli. These authors identified a flow of information from posterior superior temporal sulcus, to inferior parietal lobule, to inferior frontal cortex, to primary motor cortex.

Our results are also consistent with those of Leff et al. (2008), who investigated word-level language comprehension, and exhaustively constructed all models of effective connectivity across the posterior superior temporal, anterior superior temporal, and inferior frontal gyrus. They found that the optimal model exhibited a “forward” architecture originating in the posterior superior temporal sulcus, with a directional projection to the anterior superior temporal sulcus, and a subsequent termination in the anterior inferior frontal gyrus. Notably, unlike the model proposed by Nishitani and Hari (2002), Leff et al.’s (2008) model of temporal–inferior frontal connectivity did not pass through the inferior parietal lobule. We found a similar pathway, in addition to the “forward” architecture from pST  $\rightarrow$  IP  $\rightarrow$  vPM pathway, in which there was significant directional connectivity from pST  $\rightarrow$  aST. We showed that this directional pST  $\rightarrow$  aST connection is present during both Observation and Imitation. This finding provides support for the notion that shared network interactions during production and perception allow for the development of and maintenance of speech representations, in particular between anterior and posterior superior temporal regions typically emphasized during speech perception and comprehension. However, the fact that we did not find strong connectivity between the aST and IFG/vPM during Imitation suggests the core interactions shared by Observation and Imitation proceed through the “dorsal” pST  $\rightarrow$  IP  $\rightarrow$  IFG/vPM  $\rightarrow$  M1/S1 pathway identified by Nishitani and Hari (2002).

Of particular interest is that the connection strengths from IP to IFG/vPM, and from IFG/vPM to M1S1 were not significantly different across conditions. Interactions between IP and IFG/vPM have been shown to be important for speech production, as electrical stimulation of both of these structures and the fiber pathways connecting them impairs speech production (Duffau et al., 2003). Further, both of these regions are sensitive to the incongruence between visual and auditory speech information during audiovisual speech perception (Hasson et al., 2007; Bernstein et al., 2008). In conjunction with the primary sensory–motor cortex, the vPM, and posterior inferior frontal gyrus are also necessary for speech

production (Ojemann et al., 1989; Duffau et al., 2003), and there is evidence that even perception of audiovisual and auditory-only speech elicits activity in both premotor and primary motor cortices (Pulvermuller et al., 2006; Skipper et al., 2007). Thus, in addition to the overlapping regional activation for observation and imitation of audiovisual speech, we show similar connectivity from IP to IFG/vPM and from IFG/vPM to M1S1 during these tasks, suggesting similar interactivity among these regions during perception and production of speech. This finding suggests that the flow of information during speech perception involves a motor execution circuit, and this motor circuit (IP – IFG/vPM – M1S1) supporting speech production relies on the relevant sensory experience.

Although there are similarities, the networks implementing perception and production of speech are dissociated by stronger effective connectivity in the LH for Imitation compared to Observation (Figure 3). Connections that differed included those from pST to vPM, vPM to dPM, and dPM to M1S1, all of which were stronger during Imitation than during Observation. The first two connections (pST  $\rightarrow$  vPM, vPM  $\rightarrow$  dPM) are implicated in Hickok and Poeppel’s (2007) “dorsal stream” of speech perception. By their account, the dorsal stream helps translate auditory speech signals into motor representations in the frontal lobe, which is essential for speech development and normal speech production (Hickok and Poeppel, 2007). Our results are consistent with this view by pointing to a partially overlapping network for Observation and Imitation as part of a larger auditory–motor integration circuit. The presence of a connection from dPM to M1S1 that is stronger during Imitation than Observation represents a novel finding of potential relevance, suggesting a flow of information during Imitation from pST to vPM, vPM to dPM, and dPM to M1S1, which provides stronger input to M1S1 in triggering speech execution.

The LH models also differed across tasks by the inclusion of a negative influence from dPM to M1S1 during Observation that was positive during Imitation. Such negative influence in the motor system has been previously shown in a motor imagery task, compared to overt motor execution (Solodkin et al., 2004). In that study participants were asked to execute finger–thumb opposition movement or to imagine it kinetically (with no overt motor output). In the model that describes the execution of movement, dPM had positive influence on M1S1 whereas during kinetic imagery M1S1 received strong negative influence. This is consistent with the recent argument that action observation involves some sort of covert simulation (Lamm et al., 2007) that has similarities with kinetic motor imagery (Fadiga et al., 1999; Solodkin et al., 2004). Although the precise nature of such a mechanism remains elusive, and appears not to make use of identical circuits (Tremblay and Small, 2011), the present network models, with their shared but distinctive features, suggest a more formal notion of what such “simulation” might mean in terms of network dynamics.

Our data also support the notion that imitation of speech in the human brain involves a hierarchical flow of information from pST to IP to vPM through the “dorsal stream.” As noted, Nishitani and Hari (2002) found evidence for this pathway during observation and imitation of static speech, and Iacoboni (2005) called this small circuit the “minimal neural architecture for imitation.”

By this account, the STS sends a visual description of the observed action to be imitated to posterior parietal mirror neurons, then augments it with additional somatosensory input before sending to inferior frontal mirror neurons, which code for the associated goal of the action. Efferent copies of the motor commands providing the predicted sensory output are then sent to sensory cortices and comparisons are made between real and predicted sensory consequences, and corrections are made prior to execution. We have previously developed a model of speech perception based on an analogous mechanism (Skipper et al., 2006).

Thus, our data demonstrate that the core circuit underlying imitation of speech in the LH overlaps with that for observation, and that this circuit is embedded in larger networks that differ statistically. This supports the view that the core circuitry of imitation (pST, IP, vPM) in a context-dependent manner (McIntosh, 2000) depending on the nature of the actions to be imitated (Iaconi et al., 2005).

### OBSERVATION AND IMITATION IN THE RH

We have established similarities and differences for speech imitation and observation in the LH, but how is the observation of speech processed similarly or differently from the imitation of speech in the RH? We first focus on the similarities across hemispheres. For the pST → vPM → dPM component of a “dorsal” pathway, both hemispheres showed stronger connectivity during Imitation compared to Observation (turquoise in **Figure 3**). However, unlike in the LH, in the RH during Imitation the M1/S1 region was more influenced by activity in the vPM rather than the dPM. Inferior parietal → vPM connectivity was also stronger during Imitation in the RH. These results suggest strong similarities in the pathways for speech production across hemispheres, although there are differences, primarily in the interactions between vPM and inferior parietal and primary motor/somatosensory regions. Further, with the exception of some differences in premotor-motor interactions, a dorsal stream implemented through pST-vPM interactions appears to be a prominent component for both Imitation and Observation in both hemispheres.

We are not making the claim that the two hemispheres are involved in speech production in an identical manner. It is well known that LH damage leads to more severe impairments in speech production and articulation (Dronkers, 1996; Borovsky et al., 2007) and there is evidence for increasing LH involvement in speech production with development (Holland et al., 2001). However, we do emphasize that the predominant focus of the prior literature on LH involvement minimizes the involvement of the RH, it ignores the fact that different regions show different patterns of lateralization, and it does not provide a sufficient characterization of how different regions in the speech production network interact. For example, there are regional differences in the developmental trajectory of lateralization for speech production. While the left inferior frontal/vPM shows increasing lateralization with age during speech production tasks (Holland et al., 2001; Brown et al., 2005; Szaflarski et al., 2006), this pattern does not hold for posterior superior temporal and inferior parietal regions, which show a more bilateral pattern of activation (Szaflarski et al., 2006). With respect to connectivity, despite evidence for the participation of both hemispheres in speech production (Abel et al.,

2011; Elmer et al., 2011; see Indefrey, 2011 for review), other effective connectivity models of speech production (e.g., Eickhoff et al., 2009), have failed to model the connectivity of RH regions. We have done so here, and have revealed interesting differences in the interactions of sensory and motor regions during speech perception and production across both hemispheres.

It is notable that the only connections in which Observation was stronger than Imitation were found in the RH, and this provides further support for the notion that speech perception also relies on the participation of the RH (McGlone, 1984; Boatman, 2004; Hickok and Poeppel, 2007). For the RH, our model exhibits a similar “forward” pST → aST architecture described by Leff et al. (2008) in their network study of speech comprehension, such that in our study for Observation compared to Imitation activation in the pST modulated activation in M1/S1 via IP, and in dPM both directly and via aST. This latter connection also mirrors the pST-aST influence from Leff and colleagues, but our results further suggest that these interactions continue to influence nodes of the network typically associated with motor output.

Our findings are also consistent with evidence from electrocortical mapping suggesting that information transfer during speech perception proceeds from the posterior superior temporal cortex in both an anterior direction (to the anterior superior temporal cortex; Leff et al., 2008) and in a posterior direction through the inferior parietal lobe (see Boatman, 2004 for review). We suggest that the modulation of premotor and motor regions via these functional paths (pST → IP → M1/S1; and pST → aST → dPM; pST → dPM) during Observation could reflect the modification of speech-motor representations through perception. That is, while to this point we have focused on action/motor influences on speech perception, there is also evidence that speech perception shapes articulatory/motor representations. This notion is more evident over the course of development, where speech perception and speech production emerge in concert over an extended period (Doupe and Kuhl, 1999; Werker and Tees, 1999), but such influences remain a part of models of adult speech perception (e.g., Guenther, 2006; Guenther et al., 2006; Schwartz et al., 2012) and have some support from functional imaging studies of the role of the pST in the acquisition and maintenance of fluent speech (Dhanjal et al., 2008).

This explanation requires further empirical investigation, and it also raises the question of why such differences during Observation and Imitation were not revealed in the LH. Functional interactions among these regions (e.g., aST → dPM) were either not significant in the LH, or did not differ significantly across conditions (see **Figure 4**). Null findings are difficult to interpret, and we cannot rule out the possibility that intermediate nodes that were not modeled have an influence on the connectivity profile of these networks. For example, although Eickhoff et al. (2009) found significant cortico-cerebellar and cortico-striatal interactions in their dynamic causal model of speech production, we did not model these interactions, and this could account for the difference in the connectivity profile across hemispheres. Alternatively, it may also reflect that the core circuit underlying imitation and observation of speech in the LH overlaps considerably not only in the structure of the functional connections, but also in the strength of the interactions.

In summary, similar, if not identical, brain networks mediate the observation and imitation of audiovisual syllables, suggesting strong overlap in the neural implementation of speech production and perception. The network for Imitation in particular appears to be mediated by the two cerebral hemispheres in similar ways. In both hemispheres during both Observation and Imitation, there is significant directional connectivity between pST → aST. However, the primary flow of audiovisual speech information involves a “dorsal” pathway proceeding from pST → IP → vPM → M1/S1, with additional modulation of M1/S1 through dPM. The regions that appear to have mirror

properties in humans, IP and vPM, are functionally integrated with temporal regions involved in speech perception and motor and somatosensory regions involved in speech production, and comprise the core of this network.

## ACKNOWLEDGMENTS

This study was supported by the National Institutes of Health (NIH) under grants NIH R01 DC007488 and DC03378 (to Steven L. Small), and NS-54942 (to Ana Solodkin), and by the James S. McDonnell Foundation under a grant to the Brain Network Recovery Group.

## REFERENCES

- Abel, S., Huber, W., Weiller, C., Amunts, K., Eickhoff, S., and Heim, S. (2011). The influence of handedness on hemispheric interaction during word production: insights from effective connectivity analysis. *Brain Connect.*
- Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci. (Regul. Ed.)* 4, 267–278.
- Arbuckle, J. L. (1989). AMOS: analysis of moment structures. *Am. Stat.* 43, 66–67.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Method.* 289–300.
- Bernstein, L. E., Lu, Z. L., and Jiang, J. (2008). Quantified acoustic-optical speech signal incongruity identifies cortical sites of audiovisual speech processing. *Brain Res.* 1242, 172–184.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., and Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J. Neurosci.* 17, 353–362.
- Boatman, D. (2004). Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92, 47–65.
- Borovsky, A., Saygin, A. P., Bates, E., and Dronkers, N. (2007). Lesion correlates of conversational speech production deficits. *Neuropsychologia* 45, 2525–2533.
- Borra, E., Belmalih, A., Calzavara, R., Gerbella, M., Murata, A., Rozzi, S., and Luppino, G. (2008). Cortical connections of the macaque anterior intraparietal (AIP) area. *Cereb. Cortex* 18, 1094–1111.
- Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., and Schlaggar, B. L. (2005). Developmental changes in human cerebral functional organization for word generation. *Cereb. Cortex* 15, 275–290.
- 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.
- Buccino, G., Binkofski, F., and Riggio, L. (2004a). The mirror neuron system and action recognition. *Brain Lang.* 89, 370–376.
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., and Rizzolatti, G. (2004b). Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron* 42, 323–334.
- Burton, M. W., Small, S. L., and Blumstein, S. E. (2000). The Role of Segmentation in phonological processing: an fMRI investigation. *J. Cogn. Neurosci.* 12, 679–690.
- Callan, D. E., Jones, J. A., Callan, A. M., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22, 1182–1194.
- Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., and Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14, 2213–2218.
- Calvert, G. A., and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cogn. Neurosci.* 15, 57–70.
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.
- Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G., McGuire, P., Suckling, J., Brammer, M. J., and David, A. S. (2001). Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Brain Res. Cogn. Brain Res.* 12, 233–243.
- Cattaneo, L., and Rizzolatti, G. (2009). The mirror neuron system. *Arch. Neurol.* 66, 557–560.
- Cleveland, W. S., and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* 83, 596–610.
- Cox, R. W., and Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magn. Reson. Med.* 42, 1014–1018.
- D’Ausilio, A., Bufalari, I., Salmas, P., and Fadiga, L. (2011). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*. PMID: 21676385. [Epub ahead of print].
- D’Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385.
- Dhanjal, N. S., Handunnetthi, L., Patel, M. C., and Wise, R. J. (2008). Perceptual systems controlling speech production. *J. Neurosci.* 28, 9969–9975.
- Dick, A. S., Solodkin, A., and Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain Lang.* 114, 101–114.
- Doupe, A. J., and Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* 22, 567–631.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature* 384, 159–161.
- Duffau, H., Capelle, L., Denvil, D., Gatignol, P., Sichez, N., Lopes, M., Sichez, J. P., and Van Effenterre, R. (2003). The role of dominant premotor cortex in language: a study using intraoperative functional mapping in awake patients. *Neuroimage* 20, 1903–1914.
- Eickhoff, S. B., Heim, S., Zilles, K., and Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philos. Transact. A Math Phys. Eng. Sci.* 367, 2399–2421.
- Elmer, S., Hanggi, J., Meyer, M., and Jancke, L. (2011). Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. *Hum. Brain Mapp.* 32, 2064–2074.
- Fabbri-Destro, M., and Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology (Bethesda)* 23, 171–179.
- Fadiga, L., Buccino, G., Craighero, L., Fogassi, L., Gallese, V., and Pavesi, G. (1999). Corticospinal excitability is specifically modulated by motor imagery: a magnetic stimulation study. *Neuropsychologia* 37, 147–158.
- Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402.
- Fadiga, L., Fogassi, L., Pavesi, G., and Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *J. Neurophysiol.* 73, 2608–2611.
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999a). Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.

- Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284.
- Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667.
- Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 358, 517–528.
- Gates, K. M., Molenaar, P. C., Hillary, F. G., and Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *Neuroimage* 54, 1151–1158.
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878.
- Gerbella, M., Belmalih, A., Borra, E., Rozzi, S., and Luppino, G. (2011). Cortical connections of the anterior (F5a) subdivision of the macaque ventral premotor area F5. *Brain Struct. Funct.* 216, 43–65.
- Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.
- Gonzalez-Lima, F., and McIntosh, A. R. (1994). Neural network interactions related to auditory learning analyzed with structural equation modelling. *Hum. Brain Mapp.* 2, 23–44.
- Grezes, J., Armony, J. L., Rowe, J., and Passingham, R. E. (2003). Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study. *Neuroimage* 18, 928–937.
- Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J. Commun. Disord.* 39, 350–365.
- Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang.* 96, 280–301.
- Hasson, U., Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2007). Abstract coding of audiovisual speech: beyond sensory representation. *Neuron* 56, 1116–1126.
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J. Cogn. Neurosci.* 21, 1229–1243.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Holland, S. K., Plante, E., Weber Byars, A., Strawsburg, R. H., Schmithorst, V. J., and Ball, W. S. Jr. (2001). Normal fMRI brain activation patterns in children performing a verb generation task. *Neuroimage* 14, 837–843.
- Iacoboni, M. (2005). Neural mechanism of imitation. *Curr. Opin. Neurobiol.* 15, 632–637.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., and Rizzolatti, G. (2005). Grasping the intentions of others with one’s own mirror neuron system. *PLoS Biol.* 3, e79. doi:10.1371/journal.pbio.0030079
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2:255. doi:10.3389/fpsyg.2011.00255
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., and Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci.* 18, 314–320.
- Jones, J. A., and Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14, 1129–1133.
- Kurata, K. (1991). Corticocortical inputs to the dorsal and ventral aspects of the premotor cortex of macaque monkeys. *Neurosci. Res.* 12, 263–280.
- Lamm, C., Fischer, M. H., and Decety, J. (2007). Predicting the actions of others taps into one’s own somatosensory representations – a functional MRI study. *Neuropsychologia* 45, 2480–2491.
- Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., and Price, C. J. (2008). The cortical dynamics of intelligible speech. *J. Neurosci.* 28, 13209–13215.
- Lotto, A. J., Hickok, G. S., and Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 110–114.
- Luppino, G., Murata, A., Govoni, P., and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). *Exp. Brain Res.* 128, 181–187.
- MacLeod, A., and Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *Br. J. Audiol.* 21, 131–141.
- MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B., and Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport* 11, 1729–1733.
- Matelli, M., Camarda, R., Glickstein, M., and Rizzolatti, G. (1986). Afferent and efferent projections of the inferior area 6 in the macaque monkey. *J. Comp. Neurol.* 251, 281–298.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrrier, O., Salamon, G., Dehaene, S., Cohen, L., and Mehler, J. (1993). The cortical representation of speech. *J. Cogn. Neurosci.* 5, 467–479.
- McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain Lang.* 22, 150–157.
- McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural Netw.* 13, 861–870.
- McIntosh, A. R. (2004). Contexts and catalysts: a resolution of the localization and integration of function in the brain. *Neuroinformatics* 2, 175–182.
- 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.
- McIntosh, A. R., Grady, C. L., Ungerleider, L. G., Haxby, J. V., Rapoport, S. I., and Horwitz, B. (1994). Network analysis of cortical visual pathways mapped with PET. *J. Neurosci.* 14, 655–666.
- Miller, L. M., and D’Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893.
- Molnar-Szakacs, I., Iacoboni, M., Koski, L., and Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cereb. Cortex* 15, 986–994.
- Mottronen, R., and Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *J. Neurosci.* 29, 9819–9825.
- Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G. A. (2011). Action observation circuits in the macaque monkey cortex. *J. Neurosci.* 31, 3743–3756.
- Nishitani, N., and Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron* 36, 1211–1220.
- Noll, D. C., Cohen, J. D., Meyer, C. H., and Schneider, W. (1995). Spiral K-space MRI of cortical activation. *J. Magn. Reson. Imaging* 5, 49–56.
- Ojanen, V., Mottronen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T., and Sams, M. (2005). Processing of audiovisual speech in Broca’s area. *Neuroimage* 25, 333–338.
- Ojemann, G., Ojemann, J., Lettich, E., and Berger, M. (1989). Cortical language localization in left, dominant hemisphere: an electrical stimulation mapping investigation in 117 patients. *J. Neurosurg.* 71, 316–326.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Paulesu, E., Perani, D., Blasi, V., Silani, G., Borghese, N. A., De Giovanni, U., Sensolo, S., and Fazio, F. (2003). A functional-anatomical model for lipreading. *J. Neurophysiol.* 90, 2005–2013.
- Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottronen, R., and Sams, M. (2006). Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Hum. Brain Mapp.* 27, 471–477.
- Petrides, M., and Pandya, D. N. (1984). Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. *J. Comp. Neurol.* 228, 105–116.
- Pulvermuller, F., Huss, M., Kherif, F., Moscoso Del Prado Martin, F., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870.
- Rizzolatti, G., and Fadiga, L. (1998). Grasping objects and grasping action meanings: the dual role of monkey rostroventral premotor cortex (area F5). *Novartis Found. Symp.* 218, 81–95; discussion 95–103.
- Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.
- Rizzolatti, G., and Matelli, M. (2003). Two different streams form the dorsal visual system: anatomy and functions. *Exp. Brain Res.* 153, 146–157.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153.

- Rozzi, S., Ferrari, P. F., Bonini, L., Rizzolatti, G., and Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.* 28, 1569–1588.
- Sato, M., Mengarelli, M., Riggio, L., Gallese, V., and Buccino, G. (2008). Task related modulation of the motor system during language processing. *Brain Lang.* 105, 83–90.
- Sato, M., Tremblay, P., and Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang.* 111, 1–7.
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.
- Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'arceuil, H. E., De Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.
- Schwartz, J. L., Basirat, A., Mènard, L., and Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): a perceptuo-motor theory of speech perception. *J. Neurolinguistics*. (in press).
- Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci. Res.* 47, 277–287.
- Seltzer, B., and Pandya, D. N. (1994). Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study. *J. Comp. Neurol.* 343, 445–463.
- 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.
- Skipper, J. I., Nusbaum, H. C., and Small, S. L. (2006). “Lending a helping hand to hearing: a motor theory of speech perception,” in *Action To Language via the Mirror Neuron System*, ed. M. A. Arbib (Cambridge: Cambridge University Press), 250–286.
- 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.
- Solodkin, A., Hlustik, P., Chen, E. E., and Small, S. L. (2004). Fine modulation in network activation during motor execution and motor imagery. *Cereb. Cortex* 14, 1246–1255.
- Strafella, A. P., and Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport* 11, 2289–2292.
- Sumbly, W. H., and Pollack, I. (1954). Visual contribution of speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215.
- Szaflarski, J. P., Schmithorst, V. J., Altaye, M., Byars, A. W., Ret, J., Plante, E., and Holland, S. K. (2006). A longitudinal functional magnetic resonance imaging study of language development in children 5 to 11 years old. *Ann. Neurol.* 59, 796–807.
- Tanaka, S., and Inui, T. (2002). Cortical involvement for action imitation of hand/arm postures versus finger configurations: an fMRI study. *Neuroreport* 13, 1599–1602.
- Tettamanti, M., Buccino, G., Saccumano, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., and Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J. Cogn. Neurosci.* 17, 273–281.
- Tremblay, P., Sato, M., and Small, S. L. (2011). TMS-induced modulation of action sentence priming in the ventral premotor cortex. *Neuropsychologia* 50, 319–326.
- Tremblay, P., and Small, S. L. (2011). From language comprehension to action understanding and back again. *Cereb. Cortex* 21, 1166–1177.
- Walsh, R. R., Small, S. L., Chen, E. E., and Solodkin, A. (2008). Network activation during bimanual movements in humans. *Neuroimage* 43, 540–553.
- Watkins, K. E., Strafella, A. P., and Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41, 989–994.
- Werker, J. F., and Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annu. Rev. Psychol.* 50, 509–535.
- Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.
- Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., and Warburton, E. A. (2001). Separate neural subsystems within ‘Wernicke’s area’. *Brain* 124, 83–95.
- Wright, T. M., Pelphrey, K. A., Allison, T., Mckeown, M. J., and McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb. Cortex* 13, 1034–1043.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 October 2011; accepted: 05 March 2012; published online: 26 March 2012.

Citation: Mashal N, Solodkin A, Dick AS, Chen EE and Small SL (2012) A network model of observation and imitation of speech. *Front. Psychology* 3:84. doi: 10.3389/fpsyg.2012.00084

This article was submitted to *Frontiers in Cognition*, a specialty of *Frontiers in Psychology*.

Copyright © 2012 Mashal, Solodkin, Dick, Chen and Small. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.