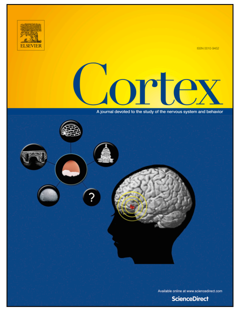
Journal Pre-proof



Multimodal Comprehension in Left Hemisphere Stroke Patients

Gabriella Vigliocco, Anna Krason, Harrison Stoll, Alessandro Monti, Laurel J. Buxbaum

PII: S0010-9452(20)30369-5

DOI: https://doi.org/10.1016/j.cortex.2020.09.025

Reference: CORTEX 3051

To appear in: *Cortex*

Received Date: 22 October 2019

Revised Date: 10 August 2020

Accepted Date: 22 September 2020

Gabriella Vigliocco: conceptualization, supervision, writing first draft, methodology, funding acquisition, revision
Anna Krason: formal analysis, investigation, visualization, revision
Harrison Stoll: formal analysis, investigation, visualization, revision
Alessandro Monti: methodology, software, investigation, revision
Laurel Buxbaum: conceptualization, supervision, methodology, funding acquisition, resources, revision

Multimodal Comprehension in Left Hemisphere Stroke Patients

Gabriella Vigliocco[1,2], Anna Krason[1], Harrison Stoll[2], Alessandro Monti[1] & Laurel J. Buxbaum[2]

1 Experimental Psychology, University College London, UK

2 Cognition and Action Laboratory, Moss Rehabilitation Research Institute, 50 Township Line Rd., Elkins Park, PA, 19027, USA

Corresponding Author: Gabriella Vigliocco, Experimental Psychology, University College London, Gower Street WC16BT, London, UK, g.vigliocco@ucl.ac.uk

Multimodal Comprehension in Left Hemisphere Stroke Patients

Gabriella Vigliocco[1,2], Anna Krason[1], Harrison Stoll[2], Alessandro Monti[1] & Laurel J. Buxbaum[2]

1 Experimental Psychology, University College London, UK

2 Cognition and Action Laboratory, Moss Rehabilitation Research Institute, 50 Township Line Rd., Elkins Park, PA, 19027, USA

Corresponding Author: Gabriella Vigliocco, Experimental Psychology, University College London, Gower Street WC16BT, London, UK, g.vigliocco@ucl.ac.uk

Abstract

Hand gestures, imagistically related to the content of speech, are ubiquitous in face-to-face communication. Here we investigated people with aphasia's (PWA) processing of speech accompanied by gestures using lesion-symptom mapping. Twenty-nine PWA and 15 matched controls were shown a picture of an object/action and then a video-clip of a speaker producing speech and/or gestures in one of the following combinations: speech-only, gesture-only, congruent speech-gesture, and incongruent speech-gesture. Participants' task was to indicate, in different blocks, whether the picture and the word matched (speech task), or whether the picture and the gesture matched (gesture task). Multivariate lesion analysis with Support Vector Regression Lesion-Symptom Mapping (SVR-LSM) showed that benefit for congruent speech-gesture was associated with 1) lesioned voxels in anterior fronto-temporal regions including inferior frontal gyrus (IFG), and sparing of posterior temporal cortex and lateral temporal-occipital regions (pTC/LTO) for the speech task, and 2) conversely, lesions to pTC/LTO and sparing of anterior regions for the gesture task. The two tasks did not share overlapping voxels. Costs from incongruent speech-gesture pairings were associated with lesioned voxels in these same anterior (for the speech task) and posterior (for the gesture task) regions, but crucially, also shared voxels in superior temporal gyri (STG) and middle temporal gyri (MTG), including the anterior temporal lobe. These results suggest that IFG and pTC/LTO contribute to extracting semantic information from speech and gesture, respectively; however, they are not causally involved in integrating information from the two modalities. In contrast, regions in anterior STG/MTG are associated with performance in both tasks and may thus be critical to speech-gesture integration. These conclusions are further supported by associations between performance in the experimental tasks and performance in tests assessing lexical-semantic processing and gesture recognition.

Keywords: Language, Comprehension, Speech-Gesture, Apraxia, Aphasia

Face-to-face communication is multimodal, and representational gestures (i.e., hand movements that iconically evoke properties of objects, events, actions, and spatial relations) are part and parcel with speech. They can imagistically express features of specific concepts (e.g., a speaker making a stirring gesture while saying “mixing”) but they can also express properties that go beyond single words and concepts (properties of complex events, e.g., a speaker making a rolling gesture while describing a circus show with acrobats on trampolines). In production, gestures have been shown to support speakers in retrieving words from memory, and in organizing the semantic/conceptual content of communication (Kita, Alibali & Chu, 2017). In comprehension, listeners process the information provided by gestures (Wu & Coulson, 2007; Kelly, Özyürek, & Maris, 2010; Gunter, Weinbrenner, & Holle, 2015), and if asked later, they are usually unable to tell whether a particular piece of information originated in speech or gesture (Alibali, Flevares, & Goldin-Meadow, 1997). This suggests that, rather than maintaining separate gestural and speech memory traces, listeners combine the semantic information arising from the two modalities into a single coherent semantic representation (Özyürek, 2014).

It has long been known that deficits involving both speech (aphasia) and gestures (limb apraxia) may co-occur after brain damage (Gainotti & Lemmo, 1976; Steinthal, 1871; Finkelnburg, 1870). The German linguist Chaim Steinthal, who introduced the term “apraxia” in 1871 when describing the awkward tool use by an aphasic patient, wrote that “apraxia is an obvious amplification of aphasia”. Similarly, apraxia and aphasia were considered to be two sides of “asymbolia”, namely a disturbance in the expression and comprehension of symbols in any modality by Finkelnburg (1870). The association of apraxia and aphasia is not absolute, however; Weiss, Ubben, Kaesberg, Kalbe, Kessler, Liebig and Fink (2016) reported that 12/50 left hemisphere stroke patients examined exhibited aphasia without apraxia and 2/50 had apraxia without aphasia (see also Kertesz, 1985; Papagno, Della Sala, & Basso, 1993), with co-occurrence of the two disorders associated with inferior frontal gyrus (IFG) damage. Goldenberg and Randerath (2015a, 2015b) found that deficits in pantomime of tool use (a classic test of apraxia) and picture naming were only moderately correlated, in association with damage in the anterior medial temporal lobe. In contrast, deficits in imitation of meaningless movements tended to correlate with language comprehension (as assessed by the Token Test) and to be associated with inferior parietal lesions. Thus, it remains unclear how speech and gesture are orchestrated in production or comprehension, and if so, whether orchestration is achieved by overlapping neural circuits.

Here, we report the first lesion study of people with aphasia (PWA) – accompanied by different degrees of limb apraxia - that investigates their comprehension of speech-gesture pairings. There are two main aims of the study. The first is to establish, using state-of-the-art lesion-symptom mapping methods, the neural regions underscoring benefits of having multimodal congruent speech-gesture pairings (e.g., a speaker moves a fist up and down as if using a hammer while saying "hammer") over unimodal baselines (speech-only or gesture-only), and costs associated to having multimodal incongruent pairings (e.g., a speaker moves a fist up and down as if using a hammer but says "scissors") over unimodal baselines. The second aim is to assess whether lexical-semantic and/or gesture recognition abilities (the latter frequently impaired in limb apraxia) predict benefits from congruent speech-gesture pairings or costs when these pairings are incongruent.

The Neural Substrate of Processing Gestures Accompanying Speech

Shared processing, or integration, between speech and gestures has been argued to involve left inferior frontal gyrus (IFG) and, to different extents, left (or bilateral) posterior temporal cortices (pTC). Some previous imaging studies reported overlap between the processing of speech and gestures in left IFG and bilateral posterior middle temporal gyrus (pMTG) (Straube, Green, Weis and Kircher, 2012; Xu, Gannon, Emmorey, Smith and Braun, 2009). However, these results do not tell us whether the two channels are integrated or merely complementary. This is an important and often-overlooked distinction. In order to address whether/which regions are engaged in integrating information from speech and gestures, looking at overlap in activation is not sufficient. It is necessary to use paradigms where multimodal stimuli (e.g., simultaneous presentation of congruent speech and gestures) are compared to unimodal ones (e.g., speech-only), or in which gestures coupled with degraded speech are compared to those with clear speech, or where the semantic relations between the two channels is manipulated (i.e., congruent speech-gesture pairing; incongruent speech-gesture pairing; or pairings in which speech and gesture supplement/complement each other).

Studies that have focused on the integration of speech and gesture suggest left IFG, left posterior superior temporal sulcus (pSTS) and posterior middle temporal gyrus (pMTG) as key nodes contributing to this integration process (Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008; Holle, Obleser, Rüschemeyer, & Gunter 2010; Willems, Özyürek & Hagoort, 2007, 2009). There is a further suggestion that while left IFG would contribute to integration of speech and iconic co-speech gestures (i.e., idiosyncratic gesticulation that is time locked with speech and is not meaningful if considered in isolation, Kendon, 2004), pSTS would be engaged in processing speech-gesture pairings, where the gestures were clear pantomimes (Willems et al., 2009). Note however that Dick et al. (2014) using more naturalistic stimuli did not find activity in pSTS as related to processing of co-speech gestures, in the context of clear effects in left IFG and pMTG.

A general issue with those studies that have identified left IFG and left pMTG (in addition to pSTS) as potential convergence sites for speech and gesture relates to the interconnectivity between IFG and pMTG (Friederici, 2009). It is possible to observe correlated patterns of activation between these regions such that activations in one or the other may simply be a consequence of strong connectivity, rather than having a causal role in the integration of semantic information from speech and gestures (Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). Zhao, Riggs, Schindler and Holle (2018) disrupted left IFG and left pMTG with transcranial magnetic stimulation (TMS) using congruent and incongruent word-gesture pairings (all referring to actions with objects/tools). Participants were presented with speech-gesture pairs and asked to indicate the gender of the voice producing the word. Trials included congruency/incongruency in gender between the body gesturing and the voice as well as semantic congruency/incongruency between the speech and the gesture. They found that TMS applied at both sites reduced the difference between semantically incongruent and congruent pairings (reduced the reaction time cost for the incongruent trials), therefore suggesting that disrupting both of these nodes affects the ability to integrate information from each modality compatibly with an integration account.

<u>Patient Studies of Speech-Gesture Processing</u>

Those studies that have looked at aphasics' performance in tasks combining speech and gesture indicate that PWA show congruence and incongruence effects when presented with speech-gesture pairings. For example, Eggenberger, Preisig, Schumacher, Hopfner, Vanbellingen et al.

(2016) asked PWA and control participants to judge if a spoken word and a co-speech gesture matched. Stimuli were either congruent (same meaning), incongruent (different meaning) or baseline (words produced in the context of a meaningless gesture). PWA showed both an accuracy advantage in processing congruent pairings as well as a disadvantage in processing incongruent pairings compared to baseline (see Perniss, Vinson, & Vigliocco, 2020 for related results in neurotypical individuals). No correlation with measures of limb apraxia was found in this study. However, apraxia was assessed by means of gesture production, and it is unclear whether the patients had deficits in gesture comprehension, as would be relevant to the word-gesture matching task. In addition, the task used by Eggenberger et al. (2016) did not test the integration of semantic information from the speech and gestures, rather PWA's ability to compare the two channels given that the task merely asked them whether the speech and the gesture referred to the same object. Cocks, Byrne, Pritchard, Morgan, & Dipper (2018; see also Cocks, Sautin, Kita, Morgan and Zlotowitz, 2009) asked whether PWA would show a multimodal gain by integrating different information from speech and gestures (e.g., hearing "paying" along with a writing gesture to indicate "paying with a check"). They found that PWA, in contrast to neurotypical controls, did not benefit from the multimodal presentations, indicating a deficit in semantic integration of the two channels. However, the type of integration required in the task may be more similar to the inferential processes engaged in understanding a complex event rather than a manifestation of a mandatory and automatic integration across the two modalities. More broadly, these studies do not relate performance in these tasks and patient's aphasia type, apraxia scores or lesion loci; thus, they are not revealing with respect to the neural substrate nor do they allow us to identify PWAs who could benefit from co-speech gestures.

<u>Left IFG and pMTG involvement in verbal and action semantics</u>

Left IFG and pMTG have long been considered as key in semantic processing from language (words and sentences) and action (gesture recognition), respectively. Many imaging studies have shown left IFG involvement in a large variety of tasks requiring processing semantic information from verbal (spoken, written or signed) material, both in production as well as in comprehension (Binder, Desai, Graves, & Conant, 2009; Hickok & Poeppel, 2007; Hickok, 2012). In particular, IFG activation has been associated with tasks in which participants choose among semantic competitors (e.g., Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997) and when semantic ambiguity must be resolved (Bedny, McGill, & Thompson-Schill, 2008; Rodd, Davis, & Johnsrude, 2005; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007). Lesions to IFG have also been associated with semantic control deficits (Metzler, 2001; Noonan, Jefferies, Corbett, & Lambon Ralph, 2010; Robinson, Shallice, Bozzali, & Cipolotti, 2010). In a large lesion study of 99 PWA, IFG, and more precisely the white matter track "bottleneck" underlying IFG (convergence of inferior frontal-occipital, uncinate fasciculi and anterior thalamic radiations) were found to load on a "semantic recognition" factor (Mirman, Chen, Zhang, Wang, Faseyitan, Coslett, & Schwartz, 2015).

Left pMTG (more broadly pTC and lateral temporal-occipital area LTO), in addition to frontoparietal regions, has been shown to support action comprehension (e.g., Kalenine, Buxbaum, & Coslett, 2010a, 2010b; Kalenine & Buxbaum, 2016; Hoeren, Kümmerer, Bormann, Beume, Ludwig, Vry et al., 2014; Kilner, 2011; Lingnau & Petris, 2013; Spunt & Lieberman, 2012; see Watson, Cardillo, Ianni, & Chatterjee, 2013, for a meta-analysis). In particular, pMTG has been argued to act as a semantic "hub" for tools and tool actions (Martin, Kyle, Simmons, Beauchamp, & Gotts, 2014; van Elk, van Schie, & Bekkering, 2014a, 2014b). For example, Kalenine et al. (2010a, 2010b) reported that PWA with left pMTG lesions were impaired in their gesture recognition ability. On the basis of a large lesion study (131 left hemisphere patients), Tarhan, Watson and Buxbaum (2015) suggested an “anterior shift” within pTC with lesions in LTO (overlapping with motion-sensitive region hMT+) giving rise to disproportionate problems in action recognition and more anterior lesions including pMTG responsible for both action recognition and object-related action production (pantomime to show object use) problems.

Thus, both left IFG and pTC/LTO contribute to processing speech-gesture pairings in terms of their role in processing semantics from verbal (IFG) and gestural (pTC/LTO) inputs. For patients with lesions in one or the other nodes, the presence of information in the other channel, when available, may be sufficient for successful comprehension in everyday communication. It is, therefore, important to establish how integrity of the two nodes correlates with greater benefit in processing.

The present study

We use a case series approach to characterize the behavioral and anatomical profile of PWA’s comprehension of speech and of gestures when presented in combination and in isolation. We compare multimodal speech-gesture pairings to unimodal baselines (speech-only or gesture-only) to establish benefits (difference between congruent speech-gesture pairings in which both the speech and the gesture refer to the same meaning and unimodal baseline) and costs (difference between incongruent pairings -- speech and gesture refer to different meanings -- and unimodal baseline), and to relate these behavioral results to lesion patterns and performance on gesture recognition and lexical-semantic tasks.

Participants carried out a picture-word (speech task) and a picture-gesture matching task (gesture task) in which they were first presented with a picture of an object or an action, and then a video of a speaker. In the speech task, we assessed the effects of representational gestures on speech comprehension accuracy by asking participants to judge the match between the picture and the speech. In the gesture task, we assessed the effects of speech on gesture comprehension accuracy by asking participants to judge the match between the picture and the gesture. Support Vector Regression Lesion-Symptom Mapping (SVR-LSM, Zhang, Kimberg, Coslett, Schwartz, & Wang, 2014), a state-of-the-art machine learning-based method for multivariate lesion-symptom mapping, was used to examine the relationships between patients' lesions and the magnitude of benefits (in congruent speech-gesture pairings) and costs (in incongruent pairings) in multimodal comprehension. Assessing benefits and costs for both the speech and the gesture tasks allowed us to identify nodes in the network that are specifically engaged in semantic processing of one or the other modality as well as nodes genuinely involved in integration across modalities by investigating overlap between lesioned voxels associated to benefits (or costs) in the speech task and those in the gesture task.

Finally, we further assessed the relationship between performance in the experimental task and scores in tests assessing lexical-semantic processing and gesture recognition in order to more broadly clarify the contribution of these critical aspects of aphasia and apraxia to the patients' ability to comprehend speech-gesture pairings.

Most previous studies have used stimuli that lack ecological validity as the face of the model was obscured, covered or cropped (see Kelly et al. 2010; Habets, Kita, Shao, Özyürek, & Hagoort, 2011; Holle & Gunter, 2007; Holle et al. 2010; Kelly, Hirata, Manansala, & Huang, 2014; Obermeier, Dolk, & Gunter 2012; Özyürek & Kelly, 2007; Willems et al. 2007, 2009; Wu & Coulson, 2014). Crucially, this raises the question of whether the integration effects that have been found for speech and gesture in spoken language comprehension are due to the absence of important visual information from the face (mouth movements). Here, as in previous work from our group with neurotypical individuals (Perniss, Vinson, & Vigliocco, 2020), we edited our videos to combine the face from one video and the body from another in order to create more natural incongruent stimuli (see Figure 1).

We contrast predictions from two accounts: integration of speech and gesture in IFG and pTC/LTO and use of complementary processing (speech and gesture) in IFG (extracting semantic information from speech) and pTC/LTO (extracting semantic information from gesture). Note that this latter account does not preclude integration of information from the two channels, it does however, argue that such integration would not occur in left IFG and pTC/LTO but rather in general semantic hubs involved in combining conceptual information from different sources, such as anterior temporal lobe (ATL; e.g., Holland & Lambon Ralph, 2010).

The integration account predicts:

(a) *For Congruent Speech-Gesture Pairings*, performance should be more accurate (i.e., there should be a benefit) for multimodal than unimodal stimuli in both the speech and the gesture tasks because of the integration of the two channels. The integrity of IFG and pTC/LTO should be associated with the amount of benefit from multimodal presentations in both tasks.

(b) *For Incongruent Speech-Gesture Pairings*, performance should be more disrupted (i.e., there should be a cost) for multimodal than unimodal stimuli in both the speech and the gesture tasks. The integrity of IFG and pTC/LTO should be associated with the amount of cost from multimodal presentations in both tasks.

(c) If PWA are integrating information from the two channels, benefits and costs associated with processing multimodal pairings should not be related to their scores on tasks separately assessing the ability to derive meaning from words or from gestures.

The complementary processing account predicts:

(a) *For Congruent Speech-Gesture Pairings* performance will also be more accurate for multimodal than unimodal stimuli (i.e., there will be benefits) because of reliance on the other channel - i.e., better performance in the speech task because of reliance on gesture and vice versa for the gesture task. In this case we should observe dissociations between the speech and gesture tasks: the amount of benefit from having the additional channel should be inversely associated with the severity of lesions affecting specifically the processing of semantics from words (IFG lesions) or gestures (pTC/LTO lesions).

(b) *For Incongruent Speech-Gesture Pairings,* costs for incongruent pairings can come about as a consequence of damage to the regions engaged in extracting semantic information from speech (IFG) in the speech task; or engaged in extracting semantic information from gesture (pTC/LTO) in the gesture task. In this latter scenario, no integration needs to be assumed as costs would reflect reliance on the unimpaired modality.

(c) If PWA are relying on a relatively intact channel, we should observe significant relationships between performance in the speech-gesture experimental tasks and scores on tasks assessing the ability to derive meaning from words and gestures. On this account, deficits in one channel (speech or gesture) will increase sensitivity to incongruence in the other channel. Specifically, if patients are relying upon gesture recognition in the face of impairments in lexical-semantics, we should see that the degree of impairment in lexical-semantic processing should be associated with increased costs of incongruent gestures in the speech task. Similarly, if patients are relying on gesture recognition in the face of deficits in lexical-semantic processing, impairments in gesture recognition should be associated with increased costs of incongruent speech in the gesture task.

Methods

Participants

Forty-five right-handed native American English speakers participated in the study: 30 chronic aphasic/apraxic patients[1] and 15 healthy controls who were equivalent in age ($t(42) = -1.59$, $p = .12$) and education ($t(42) = -1.59$, $p = .12$).

All subjects were recruited from the Moss Rehabilitation Research Institute (MRRI) Research Registry (Schwartz, Brecher, Whyte, & Klein, 2005) and tested in the MRRI laboratories (Elkins Park, Pennsylvania, USA). Healthy controls were included in the study provided that they had a minimum score of 27 on the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975). All patients were right-handed, between the ages of 21 and 80, had suffered left-hemisphere stroke at least six months before the experiment, and had an auditory comprehension score above 4 (out of 10; suggesting moderate impairment) in the revised Western Aphasia Battery (WAB; Kertesz, Kertesz, Raven, & PsychCorp, 2007). Given the novelty of the experimental design, we were unable to derive prior sample size estimates. We sampled the largest possible number of participants we could test in the allocated time frame for the study.

[1] One patient did not complete the experiment, leaving 29 patients for the analyses reported here.

In compliance with the guidelines of the Institutional Review Board (IRB) of Einstein Healthcare Network, all participants gave informed consent and were compensated for travel expenses and participation. The informed consents obtained did not include permission to make data publicly available; as such, the conditions of our IRB approval do not permit anonymized study data to be publicly archived. To obtain access to the data, individuals should contact the corresponding author. Requests for data are assessed and approved by the IRB of Einstein Healthcare Network.

Image acquisition. Research-quality structural MRI (n=20) or CT (n = 8) scans were acquired for all but one patient. Research MRI scans included whole-brain T1-weighted MR images collected on a 3T (Siemens Trio, Erlangen, Germany; repetition time = 1620 msec, echo time = 3.87 msec, field of view = 192 × 256 mm, 1 × 1 × 1 mm voxels) or 1.5T (Siemens Sonata, repetition time = 3,000 msec, echo time = 3.54 msec, field of view = 24 cm, 1.25 × 1.25 × 1.25 mm voxels) scanner, using a Siemens eight-channel head coil. Participants for whom MRI scanning was contraindicated underwent whole-brain research CT scans without contrast (60 axial slices, 3–5 mm slice thickness) on a 64-slice Siemens SOMATOM Sensation scanner.

Lesion segmentation and warping to template. For high-resolution MRI scans, lesions were manually segmented on the patients' T1-weighted structural images. Lesioned voxels, consisting of both grey and white matter, were assigned a value of 1 and preserved voxels were assigned a value of 0. Binarized lesion masks were then registered to a standard template (Montreal Neurological Institute "Colin27") using a symmetric diffeomorphic registration algorithm (Avants, Epstein, Grossman, & Gee, 2008, www.picsl.upenn.edu/ANTS). Volumes were first registered to an intermediate template comprising healthy brain images acquired on the same scanner. Then, volumes were mapped onto the "Colin27" template to complete the transformation into standardized space. To ensure that no errors occurred during the transformation process, lesion maps were subsequently inspected by a neurologist (H.B. Coslett), who was naïve to the behavioral data. Research CT scans were drawn directly onto the "Colin27" template by the same neurologist using MRIcron (http://www.mccauslandcenter.sc.edu/mricro/mricron/index.html). For increased accuracy, the pitch of the template was rotated to approximate the slice plane of each patient's scan. Inter-rater reliability (Cohen's Kappa) between the neurologist and other trained segmenters in Buxbaum's lab was at least 85% (Schnur et al., 2009). Specifically, mean percentage volume difference = 23 ± 11; mean percentage discrepant voxels = 6 ± 5, where discrepant is defined as > 2 voxels from the other manually drawn lesion volume.

Each patient in our study was also assessed with a well-studied measure of lexical-semantic processing, the Synonymy Triplets task (Martin, Schwartz, & Kohen, 2006; Mirman et al., 2015), in which participants were asked to decide which two of three written words are most similar in meaning. Half of the trials involve nouns (e.g., violin, fiddle, clarinet), the other half verbs (e.g., to repair, to design, to fix). Performance is measured by number correct of 30 trials. Finally, patients performed a well-studied Gesture Recognition Test that has been associated with posterior temporal functions (e.g., Buxbaum, Kyle, & Menon, 2005; Kalenine et al., 2010a, 2010b). Participants heard an action verb (e.g., sawing) read aloud, and simultaneously viewed the written verb presented on a computer screen. After a 2-s pause, they saw a videotaped examiner performing a gesture "A", and after an additional 2-s pause, a second gesture "B". One gesture was the correct match to the verb (e.g., sawing), and the other was incorrect by virtue of a postural, spatial, or temporal error (e.g., sawing with an open hand posture). The order of the correct and incorrect gesture videos was randomized. Patients had to select which gesture "A" or

"B" (by verbalizing or pointing) correctly matched the action verb. They were allowed to respond at any point while the videos were being shown. There were 24 trials. Patients also completed a control task to ensure they understood the verbs used in the gesture recognition tasks. The control task involved forced-choice matching between action verbs and picture stimuli. Participants heard the same action verbs as in the Gesture Recognition Task and had to choose an object picture from an array of three objects to match with the action name (e.g., matching a saw to the verb "sawing"). Any verb failed on the control task was excluded from scoring for the Gesture Recognition Task. Few trials were excluded for this reason (Md. 2 trials excluded).

Table 1 provides demographic information and semantics tasks scores for patient participants.

Table 1: Patient Demographic and Task Performance

Subject	Age	Gender	Education (in years)	Months Post Stroke	WAB Comprehension	Lesion Volume (in mm^3)	Gesture Recognition (% Accuracy)	Lexical-Semantics (% Accuracy)
S01	64	M	15	112	8.95	179606	78	73
S02	73	M	14	79	4.6	64793	85	73
S03	53	M	12	49	9	92744	92	83
S04	66	M	19	54	9.55	51399	83	87
S05	57	F	14	96	9.4	37091	96	93
S06	53	M	12	87	8.5	52416	81	70
S07	46	M	18	13	9.3	64375	100	93
S08	36	M	13	57	5.65	88046	92	37
S09	53	F	13	140	9.2	80020	91	80
S10	51	F	12	134	6.25	94536	77	43
S11	65	M	12	137	8.5	69778	91	9
S12	37	F	16	14	9	93628	100	8
S13	73	M	19	112	8.7	76301	100	87
S14	62	M	12	81	9.8	23141	97	77
S15	60	M	12	106	8.9	47442	75	26
S16	80	M	12	131	8.05	144857	86	53
S17	79	F	12	96	8.6	16547	86	83
S18	54	F	16	129	9.4	78357	84	56

S19	71	M	18	192	7.65	258736	86	66
S20	45	M	18	52	7.9	71905	80	93
S21	66	F	12	60	6.9	117809	67	63
S22	63	F	12	60	5.6	26714	76	43
S23	44	F	18	60	9.1	60457	72	83
S24	51	F	11	27	9.35	55118	75	60
S25	63	M	16	44	8.7	N/A	83	67
S26	52	F	12	113	9.2	131776	65	77
S27	52	M	11	50	9.5	16977	86	83
S28	64	M	13	218	8.55	99980	83	80
S29	46	F	16	65	8.75	68764	100	83
Average (*SD*)	57.89 (*11.49*)	12 Fs	14.13 (*2.64*)	88.55 (*48.85*)	8.36 (*1.32*)	80832.61 (*51470.34*)	85.06 (*9.78*)	66.51 (*23.66*)
Range	36-80	n/a	11-19	13-218	4.6-9.8	16547-258736	65-100	8-93

Materials

Stimuli for the experimental task were 47 pictures[2] and short video clips (mean clip length: 3s). Half of the pictures were of objects, the other half of actions (see https://osf.io/pvube for the full list of stimuli). Pictures were taken from various sources. The video clips showed an actress producing a word and/or a gesture. For video clips containing both speech and gestures, we recorded the actress producing words denoting objects and actions accompanied by a representational gesture iconically evoking features of that object or action (no specific instructions about the form of the gesture were given). For objects, the gestures either depicted an action associated with the object (e.g., a loosely closed hand twisting back and forth to represent “screwdriver”) or outlined the object’s shape (e.g., the hands tracing a circle to represent a “ball”). For actions, gestures depicted the manual manipulation of the object involved (e.g., holding an iron and moving it back and forth to represent “ironing”) or represented the

[2] The original task included 48 items; however, initial testing indicated that most individuals did not know one of the items (vault) which was therefore excluded.

bodily movement involved in the action (e.g., moving open hands away from the body to represent “pushing”). For speech-only video clips, the actress produced the word keeping her hands still in her lap. For gesture-only video clips, she produced the gesture while remaining silent. In all videos containing gestures, the actress’ hands were in her lap at video onset and returned to her lap after production of each item (i.e., we did not trim the video to include only the stroke portion of the gesture, as in Zhao et al., 2018).

Speech and gesture were congruent (they expressed the same meaning, e.g., *pushing* in speech accompanied by a pushing gesture) in half of the videos and incongruent (they expressed different meanings, e.g., *pushing* in speech accompanied by a tearing gesture) in the other half. We constructed congruent and incongruent speech-gesture pairings using the video editing software Final Cut Pro 6.0 (www.apple.com/finalcutpro/). We created them by overlaying the face from one video onto the body from another video (see Figure 1). We retained only the audio from the face video (top), deleting the audio track from the body video (bottom). In this way, we could mismatch speech and gesture while maintaining congruence between the heard word and the visible movements of the face/mouth. In overlaying the two videos, we took care that the timing of speech and gesture onset looked natural, by aligning speech onset for both clips. As a result, gesture onsets slightly preceded speech onset as it occurs in natural communication (Morrel-Samuels & Krauss, 1992; Schegloff, 1984). We used the same editing procedure for both congruent and incongruent stimuli so that the videos did not differ in this respect.

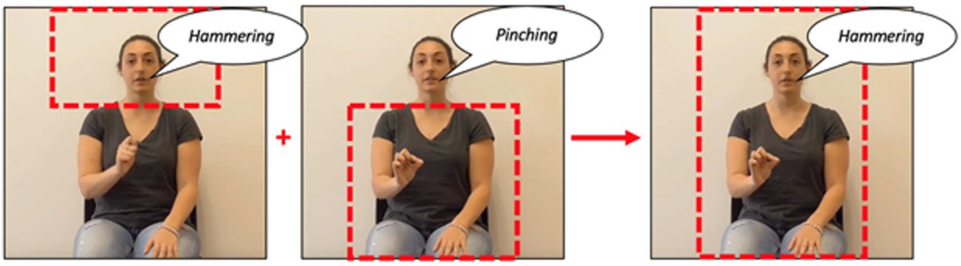


Figure 1. Schematic representation of how the video stimulus materials were created. The example shows the creation of an incongruent speech-gesture combination. The two still frames to the left of the arrow are from the input videos; the still frame to the right of the arrow is from the final stimulus video created through the overlay process. As represented by the dotted red lines, we take the head/face portion of one input video (together with the audio of the spoken word), and overlay it onto the body, depicting the gesture, from the other input video.

Procedure

The experiment, programmed in MATLAB (MathWorks Inc., Natick, USA), consisted of two tasks (speech and gesture tasks). Each participant carried out both tasks, with a break in between. Task order was counterbalanced across participants. In the speech task, each trial began with a picture that stayed on the computer screen for 1.5 s. After an ISI of 0.5 s, a video clip was presented. The video clip could randomly be a unimodal stimulus (speech-only video), or a multimodal stimulus (either a congruent or an incongruent speech-gesture video). Participants were asked to decide if the picture and the word referred to the same object/action and press a yes or no button on the computer keyboard. The gesture task differed from the speech task only in that speech-only videos were replaced with gesture-only videos, and participants were asked to judge if the picture and the gesture referred to the same object/action (see Figure 2 for overview of the procedure and conditions).

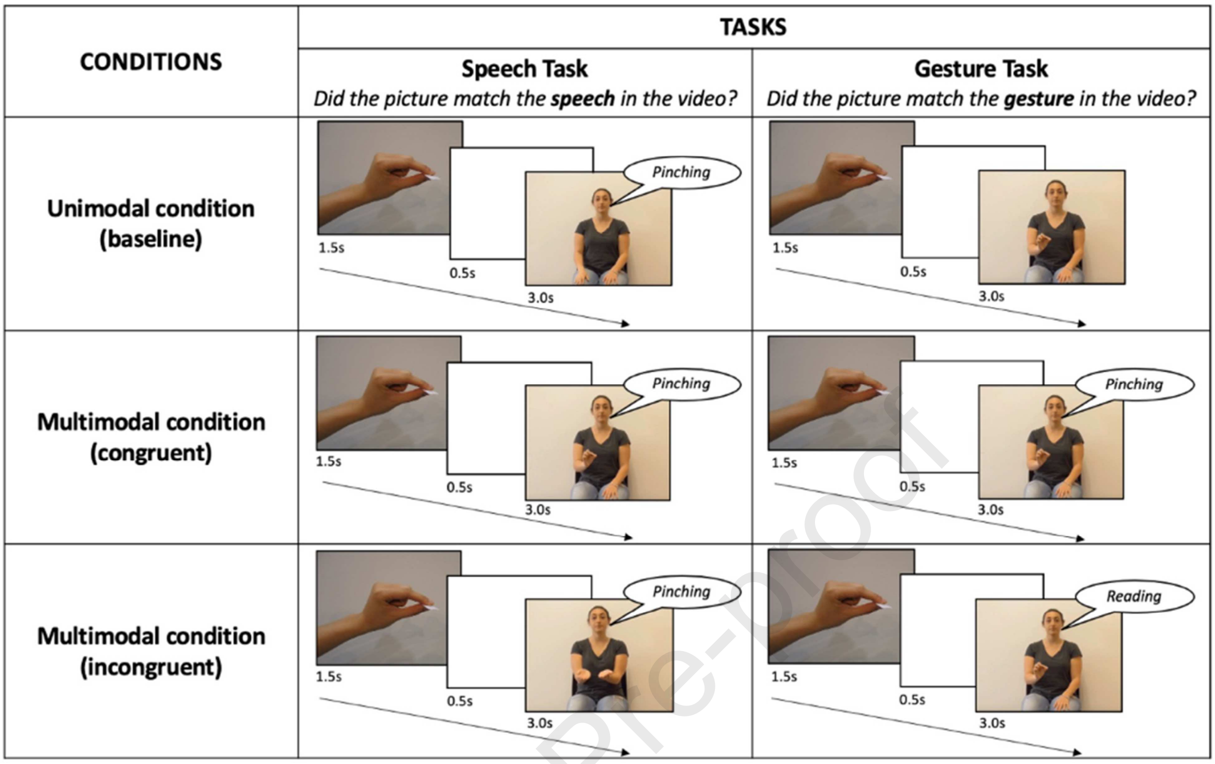


Figure 2. Overview of the conditions and tasks used in the study with an example of matching (requiring a "yes" response) trial sequences. The tasks require participants to assess whether the picture matches the word (speech task) or the gesture (gesture task). Unimodal baseline in the speech task is speech-only, and in the gesture task is gesture-only.

As there were 2 tasks (speech task and gesture task; and 3 possible video types for each task (speech-only, congruent, and incongruent; or gesture-only, congruent, and incongruent), there were 6 experimental conditions[3]. Furthermore, to make the experimental design more robust (see below), each video clip was included twice - once paired with a matching picture, once paired with a mismatching picture. Thus, there were 6 x 2 = 12 types of trials and each participant responded to 96 trials in total.

To make the whole procedure as straightforward as possible, while controlling for order effects, from the standpoint of the participant, the experiment was divided into four main blocks. Each

[3] We included object and action stimuli. However, preliminary analyses showed that there were no differences between object and action trials, therefore the two item-types are collapsed in the analyses reported here.

block was uniquely identified by the pair (task, item type): e.g. in block 1 the participant was instructed to perform the speech task (pay attention to speech) on object items; in block 2 they had to complete the speech task on action items; then they had to do the gesture task on object items, and finally the gesture task on action items. The order of blocks, as well as trials within the blocks, was randomized across participants. Prior to each block, subjects were introduced to the experiment through two training sections and progressed to the experiment proper only if they scored above chance in the second training session. The whole experiment lasted about 1'.45". No part of the study procedures or analyses was pre-registered prior to the research being conducted.

Data Analysis

Behavioral Analysis

Accuracy on the trials where the speech (in the speech task) or gesture (in the gesture task) matched the picture (50% of all trials for both speech and gesture) were analyzed at the trial level through logistic mixed-effect regression models using the R statistical programming environment version 3.5.0 with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The R analysis code can be found at https://github.com/cognition-action-lab/Vigliocco_etal. Following Stadthagen-Gonzales et al. (2009) and in light of the results reported by Perniss, Vinson, and Vigliocco (2020), these trials correspond to "yes" responses which are more reliable than "no" responses in picture-word matching tasks.

In the first set of analysis we tested for a main effect of group (PWA and controls), a main effect of condition (congruent speech-gesture, incongruent speech-gesture, unimodal baseline: speech-or gesture-only). We tested for the interaction between group and condition, with planned follow-up comparisons to assess benefits (congruent speech-gesture vs. unimodal) and costs (incongruent speech-gesture vs. unimodal). The speech and gesture tasks analyses were run in separate models because we do not have any prediction concerning interactions involving task. We included random intercepts for subject, video, and picture in each model.

In the second set of analysis, using only PWA data, we tested for an interaction of the lexical-semantic and gesture recognition tasks and condition on task accuracy. For significant interactions, we then compared whether the relationship between task and accuracy was stronger for benefit/cost effect relative to unimodal. Finally, we followed up significant interactions by examining whether there were simple main effects of the lexical-semantics or gesture recognition task for each condition. We ran separate models for the speech and gesture task and included random intercepts for subject, video, and picture in each model.

All mixed-level models were assessed by mapping the log-likelihood ratio of a full model and a reduced using a chi-square distribution. For models with interaction terms, we removed the interaction term when testing for significant main effects. An alpha threshold of .05 was used to determine statistical significance, and all effects are reported as log odds.

Lesion Analysis

Lesion-symptom mapping. Support Vector Regression-Lesion Symptom Mapping (SVR-LSM) was performed with the MATLAB toolbox (https://cfn.upenn.edu/-zewang/). SVR-LSM (Zhang et al., 2014) is a multivariate technique that uses machine learning to determine the association between lesioned voxels and behavior when considering the lesion status of all voxels submitted to the analysis. It overcomes several limitations of voxel-based lesion symptom mapping

(VLSM), including inflated false positives from correlated neighboring voxels (Pustina, Avants, Faseyitan, Medaglia, & Coslett, 2018), Type 2 error due to correction for multiple comparisons (Bennett, Wolford, & Miller, 2009), and uneven statistical power due to biased lesion frequency as a function of vascular anatomy (Mah, Husain, Rees, & Nachev, 2014; Sperber & Karnath, 2017). SVR-LSM has been shown to be superior to VLSM when multiple brain areas are involved in a single behavior (Herbet, Lafargue, & Duffau, 2015; Mah et al., 2014). As noted by Zhang et al. (2014), in SVR-LSM the relationship of the behavior to the entire lesion map rather than each isolated voxel is modeled using a non-linear function. The means that inter-voxel correlations are intrinsically considered, resulting in a more sensitive way to examine lesion-symptom relationships. An SVR model is trained to predict a continuous association variable (the behavioral measure) with high accuracy using all voxels' lesion status.

Voxels lesioned in less than 4 patients were excluded. To avoid the concern that patients with larger lesions might drive results, lesion volume was controlled for by using direct total lesion volume control (dTLVC). In this approach, the values of the voxels are divided by the square root of the total volume for each patient (Zhang et al., 2014). Significance values were obtained using 10,000 permutations of the dependent measures, fivefold cross-validation, and a voxel-wise significance threshold of $p < .05$ was applied. Cross-validation of the regression model was done with 5-folds, meaning our sample was divided into 5 sub-groups and the regression model was created using the data from four of the groups. The fifth group was then used to validate the model made with the other four groups. This process was repeated such that each person was in the group that helps validate the model (i.e., the fifth group). After cross-validation and significance of voxels were determined, at a p-value of .05, we also removed any cluster of voxels that was less than 500 (Lacey, Skipper-Kallal, Xing, Fama, & Turkeltaub, 2017), which has been utilized in several other papers from Buxbaum's group to determine significance (Garcea et al., 2019).

Dependent measures for the lesion analysis was patient accuracy on the "benefit of congruent" and "cost of incongruent" trials separately for speech and gesture tasks, with performance on the unimodal speech or gesture trials regressed out.

This methods section reports how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

Results

Comparison between PWA and Controls in the Speech and in the Gesture Tasks.

The first set of analysis was performed in two models, one for the speech and one for the gesture task. Both contained the main effects of group and condition, as well as the two-way interaction between group and condition. Performance was very accurate for both groups, with controls being at or near-ceiling. Figure 3 and Table 2 show the results.

For the speech task, there was a main effect of group ($\chi2(1) = 27.12$, $p < .001$, $-2.35 \pm .4$) with patients performing less accurately than controls. There was also a significant main effect of condition ($\chi2(2) = 93.38$, $p < .001$). Follow-up analyses revealed significant benefits for stimuli in the multimodal congruent condition (vs. unimodal; $\chi2(1) = 12.21$, $p < .001$, $.72 \pm .2$) as well as costs for stimuli in the multimodal incongruent condition (vs. unimodal; $\chi2(1) = 42.28$, $p <$

.001, -1.27 ± .0007). There was no interaction between group and condition ($\chi2(2) = 3.06$, $p = .22$).

For the <u>gesture task</u>, there was a main effect of group ($\chi2(1) = 17.5$, $p < .001$, -1.37 ± .29) with patients performing less accurately than controls. We also found a main effect of condition ($\chi2(2) = 104.88$, $p < .001$). Follow-up contrasts revealed both significant benefits of multimodal congruent condition (vs. unimodal, $\chi2(1) = 7.95$, $p < .005$, .57 ± .19) and costs of multimodal incongruent condition (vs. unimodal, $\chi2(1) = 54.92$, $p < .001$, -1.48 ± .18). Furthermore, the interaction between group and condition was also significant ($\chi2(2) = 10.79$, $p < .01$): there was no significant difference between patients and controls for the multimodal congruent condition ($\chi2(1) = 1.81$, $p = .18$, .42 ± .3), however, patients showed greater costs in multimodal incongruent condition than controls ($\chi2(1) = 3.62$, $p = .057$, -.49 ± .25). We then further tested the interaction by comparing the simple effect of group separately for the three conditions. Patients were worse than controls in all three conditions (congruent ($\chi2(1) = 3.8$, $p = .051$, -.78 ± .39), incongruent ($\chi2(1) = 15.52$, $p < .001$, -1.74 ± .41), and unimodal ($\chi2(1) = 15.33$, $p < .001$, -1.3 ± .0009)).

Table 2. Mixed effect models with Condition and Group

Dependent Variable: Speech Task Accuracy					
	df	χ^2	*Coef.*	*SE.*	*p*
Group					
PWA[a]	1	27.12	-2.35	.4	<.001
Condition	2	93.38	n/a	n/a	<.001
Congruent[b]	1	12.21	.72	.2	<.001
Incongruent[b]	1	42.28	-1.27	.0007	<.001
*Condition*Group*[ᵻ]	2	3.06	n/a	n/a	.22
Dependent Variable: Gesture Task Accuracy					

	df	χ^2	Coef.	SE.	p
Group					
PWA[a]	1	17.5	-1.37	.29	<.001
Condition	2	104.88	n/a	n/a	<.001
Congruent[b]	1	7.95	.57	.19	<.005
Incongruent[b]	1	54.92	-1.48	.18	<.001
*Condition*Group*	2	10.79	n/a	n/a	<.01
Congruent[b]	1	1.81	.42	.3	.18
Incongruent[b]	1	3.62	-.49	.25	.057
Simple effects Group (within condition)					
Congruent[a]	1	3.8	-.78	.39	.051
Incongruent[a]	1	15.52	-1.74	.41	<.001
Unimodal[a]	1	15.33	-1.3	.0009	<.001

Note. *Coef.* = model estimation of the change in response accuracy (in log odds) from the reference category for each fixed effect; *SE* = standard error of the estimate; [a]Reference is Controls; [b]Reference is Unimodal. †Model was not statistically significant and no follow-up analysis was done.

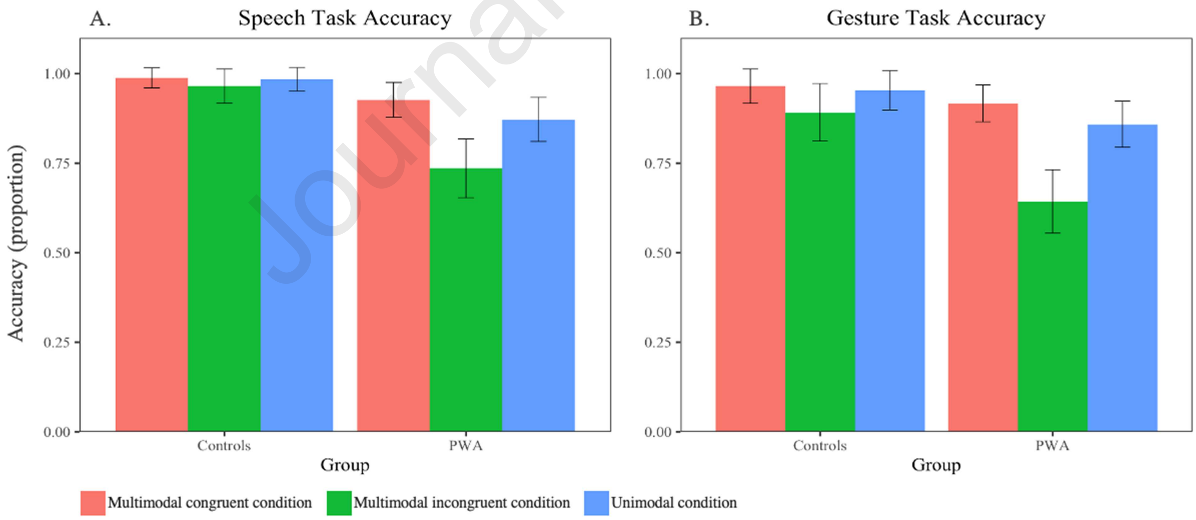


Figure 3. Proportion correct responses in the speech (A) and gesture (B) task for the Controls and PWA groups. Red represents the multimodal congruent condition; green represents the multimodal incongruent condition and blue the unimodal condition. Bars are standard errors.

The Relationship between Task Performance and Background Scores in PWA.

The results of the second set of analysis are presented in Figure 4 and Table 3. For these analyses, we ran several models, which tested for effects of the lexical-semantic and gesture recognition measures on trial accuracy. As before, these were run for the speech and gesture tasks separately.

Speech Task. There was a main effect of the lexical-semantic measure on accuracy in the speech task, ($\chi2(1) = 10.1$, $p < .005$, 3.89 ± 1.91). We tested for simple effects to assess the relationship between the lexical-semantic task and performance for each condition. Higher lexical-semantic scores predicted higher speech task performance for both multimodal incongruent ($\chi2(1) = 12.19$, $p < .001$, 4.67 ± 1.21) and unimodal ($\chi2(1) = 6.42$, $p < .05$, 4.15 ± .0009) conditions but not multimodal congruent ($\chi2(1) = 1.85$, $p = .17$, 2.3 ± 1.67) conditions. There was no effect of the gesture recognition measure in the experimental speech task ($\chi2(1) = .66$, $p = .42$, 1.98 ± 2.42).

Gesture Task. There was a trend toward a main effect of the gesture recognition measure on accuracy in the gesture task ($\chi2(1) = 2.99$, $p = .08$, -1.56 ± .17). Assessment of simple effects (congruent ($\chi2(1) = .03$, $p = .86$, -.44 ± 2.51; incongruent ($\chi2(1) = 4.97$, $p < .05$, 5.79 ± 2.51); unimodal ($\chi2(1) = 1.09$, $p = .29$, 1.90 ± 1.8) revealed that higher gesture recognition scores predicted higher accuracy only for the multimodal incongruent trials. There was no effect of the lexical-semantic measure in the experimental gesture task ($\chi2(1) = 1.82$, $p = .18$, 1.46 ± 1.07).

Table 3. Mixed effect models with Condition, Lexical-Semantics, and Gesture Recognition Task (PWA only)

Dependent Variable: Speech Task Accuracy

	df	χ^2	*Coef.*	*SE.*	*p*
Lexical-Semantics	1	10.1	3.89	1.91	<.005
Simple effect Condition					
Congruent	1	1.85	2.3	1.67	.17
Incongruent	1	12.19	4.67	1.21	<.001
Unimodal	1	6.42	4.15	.0009	<.05
	df	χ^2	*Coef.*	*SE.*	*p*
Gesture Recognition†	1	.66	1.98	2.42	.42
Dependent Variable: Gesture Task Accuracy					
	df	χ^2	*Coef.*	*SE.*	*p*
Lexical-Semantics†	1	1.82	1.46	1.07	.18
	df	χ^2	*Coef.*	*SE.*	*p*
Gesture Recognition	1	2.99	-1.56	.17	.08
Simple effect Condition					
Congruent	1	.03	-.44	2.51	.86
Incongruent	1	4.97	5.79	2.51	<.05
Unimodal	1	1.09	1.9	1.8	.29

Note. *Coef.* = model estimation of the change in response accuracy (in log odds) from the reference category for each fixed effect; *SE* = standard error of the estimate [a]Reference is Unimodal.†Model was not statistically significant and no follow-up analysis was done.

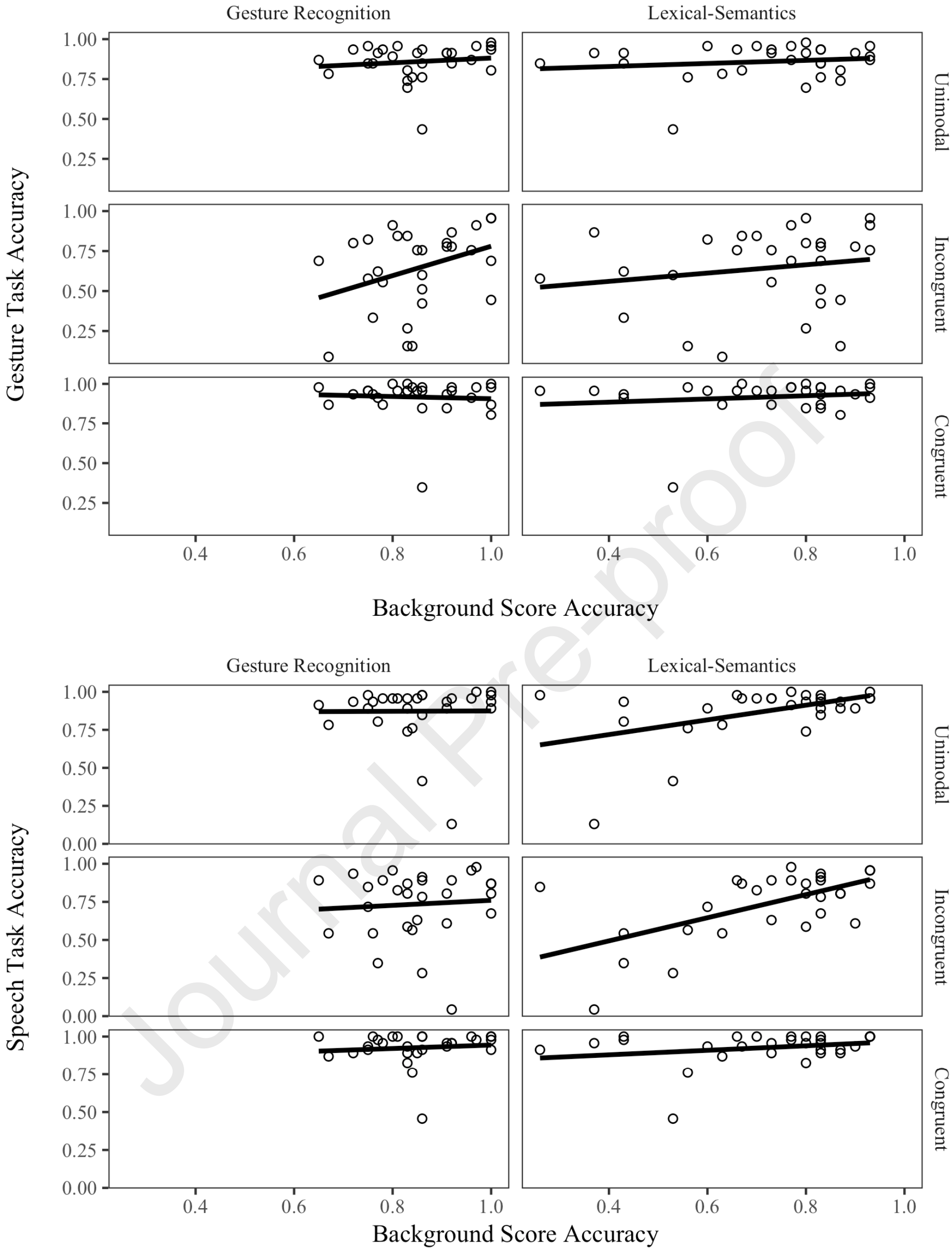


Figure 4. Relationship between Lexical-Semantics or Gesture Recognition scores and proportion correct in the different conditions in the speech (above) and gesture (below) tasks.

Lesion Analyses

Figure 5 depicts the overlap among the 28 participants with high resolution CT or MRI anatomical data. The SVR-LSM analysis revealed several significant clusters where presence of lesions was associated with better performance (benefits) in the multimodal congruent condition relative to the unimodal condition and decrease in performance (costs) in the multimodal incongruent condition compared with the unimodal condition.

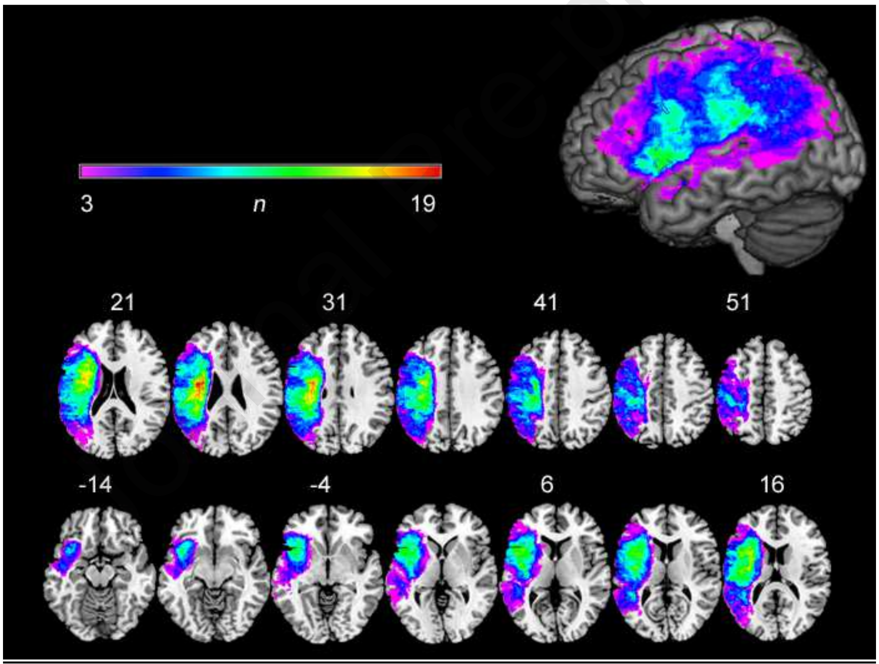


Figure 5. Overlap of all 28 lesions included in the analyses. Only voxels with a minimum of 4 overlapping lesions are displayed. The maximum overlap was 19 lesions. Surface rendering is displayed at a search depth of 8 mm. Z coordinates of axial slices are listed in MNI standardized space.

Benefit from Congruent Trials

In the speech task, the SVR-LSM analysis revealed several significant clusters of lesioned voxels that were associated with greater benefit of congruent gesture (see Figure 6 and Table 4), including the Postcentral Gyrus (PoCG), Rolandic Operculum (ROL), Precentral Gyrus (PreCG), Superior Temporal Gyrus (STG) (anterior), Inferior Parietal Lobule (IPL), Supramarginal Gyrus (SMG), Insula, and IFG (opercular) (IFG opercular).

In the gesture task (see Figure 6 and Table 4) we found clusters of lesioned voxels associated with greater benefit of congruent speech in the Middle Occipital Gyrus (MOG), Angular Gyrus (ANG), IPG, and MTG (posterior).

Table 4. Results of SVR-LSM analysis. *Peak voxels and percent damage to regions with clusters > 500mm^3 voxels associated with greater cross-modal benefit and cost as identified by Automated Anatomical Labeling (AAL).*

	Speech Task, Congruent Gesture Benefit			Gesture Task, Congruent Speech Benefit			Speech Task, Incongruent Gesture Cost			Gesture Task, Incongruent Speech Cost		
Region	# of Voxels mm^3	% of Region	Peak Voxel	# of Voxels mm^3	% of Region	Peak Voxel	# of Voxels mm^3	% of Region	Peak Voxel	# of Voxels mm^3	% of Region	Peak Voxel
STG	555	3.03	-59, 3, 1	-	-	-	4521	24.69	-47, -13, 4	657	3.58	-52, -15, -4
MTG	-	-	-	1250	3.17	-47, -68, 24	2556	6.49	-53, -20, -2	2550	6.47	-46, 1, -26
TPO superior	-	-	-	-	-	-	1277	12.48	-54, 9, -3	-	-	-

Insula	485	3.22	-46, -10, 2	-	-	-	3141	20.9	-45, -13, 4	-	-	-
ROL	1625	20.47	-54, -1, 11	-	-	-	935	11.77	-64, -9, 12	-	-	-
IFG triangular	-	-	-	-	-	-	929	4.62	-56, 21, 3	-	-	-
MOG	-	-	-	3561	13.7	-44, -73, 33	-	-	-	4849	18.66	-27, -62, 41
ANG	-	-	-	3031	32.54	-59, -55, 33	-	-	-	1399	15.02	-41, -56, 35
PoCG	3538	11.39	-43, -24, 46	-	-	-	-	-	-	-	-	-
PreCG	1071	3.8	-35, -6, 48	-	-	-	-	-	-	-	-	-
IPL (Lateral)	912	4.69	-60, -36, 39	-	-	-	-	-	-	-	-	-
IPL (Medial)	-	-	-	1674	8.6	-43, -38, 42	-	-	-	-	-	-
SMG	788	7.95	-62, -37, 33	-	-	-	-	-	-	-	-	-
IFG opercular	404	4.88	-53, 15, 8	-	-	-	-	-	-	-	-	-
Total # of Voxels Lesioned	9378			9516			13359			9455		

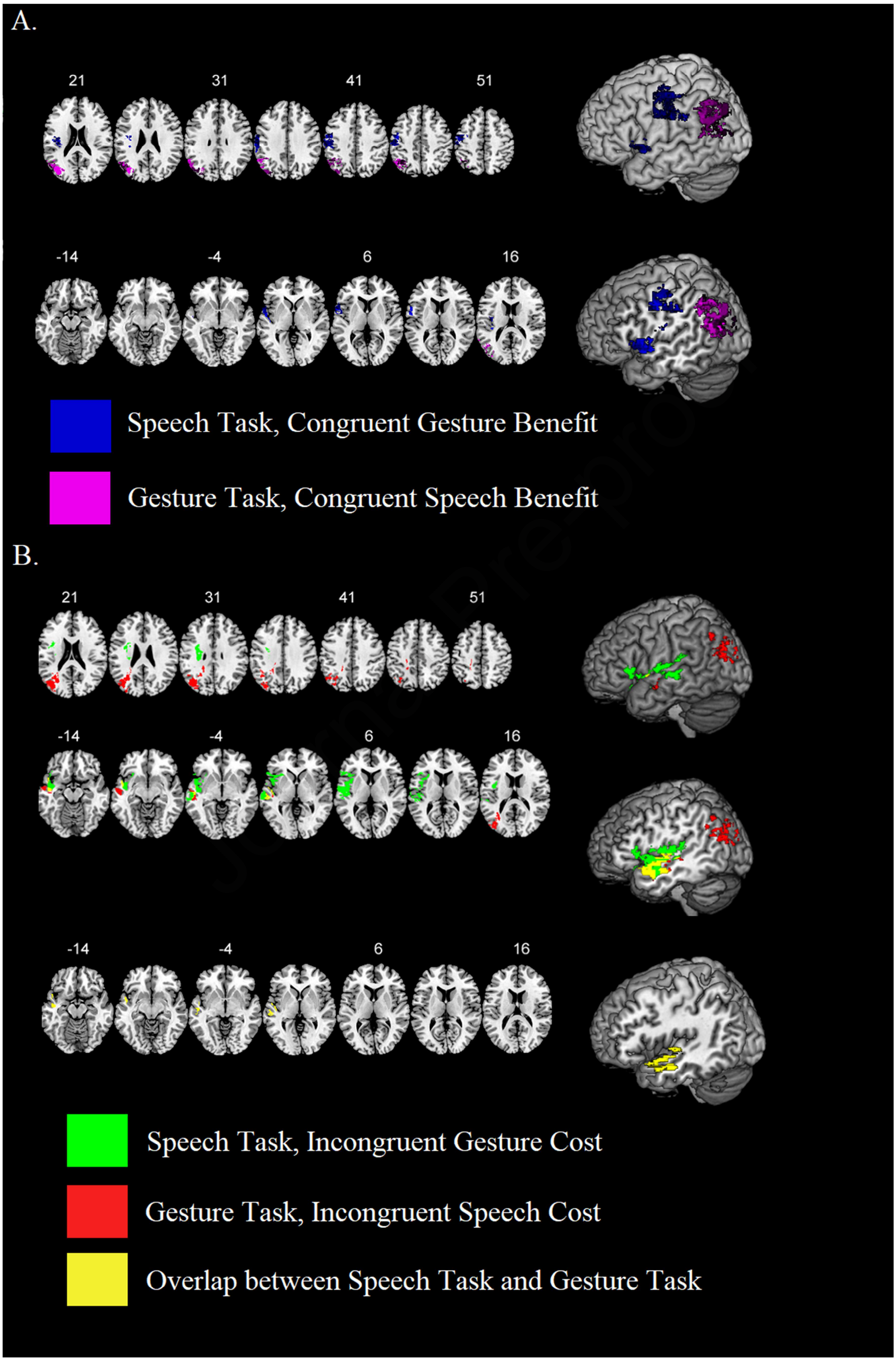

A.
21
31
41
51
-14
-4
6
16
Speech Task, Congruent Gesture Benefit
Gesture Task, Congruent Speech Benefit
B.
21
31
41
51
-14
-4
6
16
-14
-4
6
16
Speech Task, Incongruent Gesture Cost
Gesture Task, Incongruent Speech Cost
Overlap between Speech Task and Gesture Task

Figure 6. A, SVR-LSM analysis showing significant voxels associated with: A. benefit of congruent gesture on the speech task (Blue) and benefit of congruent speech on the gesture task (Pink); as well as B. cost of incongruent gesture on speech task (green) and cost of incongruent speech on gesture task (red). Overlap between costs of incongruence in the other modality is shown in yellow. Whole-brain results are rendered in MNI space in increments of 5 mm. SVR-LSM maps are set to a voxelwise threshold of $p < .05$ with 10,000 iterations of a Monte Carlo style permutation analysis with K-fold cross-validation; cluster size > 500 contiguous 1 mm^3 voxels.

Cost from Incongruent Trials

In the speech task, the SVR-LSM showed lesioned voxels associated with greater cost of incongruent gesture in STG (anterior/middle), MTG (middle), TPO superior (superior temporal pole), Insula, Inferior Frontal Gyrus (triangular) (IFG triangular) and ROL (see Figure 6 and Table 4). In the gesture task, greater cost of incongruent speech was associated with clusters of lesioned voxels in the MOG, MTG (anterior), ANG, and STG (posterior) (see Figure 6 and Table 4). Finally, as also shown in Figure 6 and Table 5, there was an overlap between the clusters associated with costs of incongruent gestures or speech in MTG (anterior), STG (anterior), and TPO superior. Note here that for the overlap between Gesture and Speech cost effects (Table 5), we report all overlapping voxels as long as results were above the 500 voxel threshold in either or both conditions.

Table 5. Overlap of gesture and speech cost effects *identified by Automated Anatomical Labeling (AAL).*

Region	# of Voxels mm3	% of Region	Peak Voxel
MTG	955	2.42	-45, 1, -26
STG	593	3.23	-46, 10, -15
TPO superior	295	2.88	-46, 6, -22

Discussion

This study is the first investigation of the neural systems engaged in comprehending words accompanied by gestures and gestures accompanied by words in aphasic patients. Moreover, we considered for the first time the influence of the ability to derive meaning from lexical and gestural input on the pattern of benefits and costs of multimodal vs. unimodal processing in PWA. Overall, PWAs showed larger effects of multimodal congruency and incongruency than controls, although both groups showed costs associated to having multimodal incongruent speech-gesture pairings both when the task focused on speech as well as when it focused on gestures (replicating previous studies, e.g., Eggenberger et al., 2016).

We contrasted an integration account arguing that nodes such as IFG and pTC/LTO play an integration role in the processing of speech-gesture combinations with a complementary processing account, according to which these nodes play a key role in extracting meaning from speech (IFG) and gesture (pTC/LTO) but not in integrating information from the two channels.

The integration account predicts that lesions to IFG and pTC/LTO should be related to performance in both the speech and in the gesture tasks. In contrast, the complementary processing account predicts that performance in the speech task is associated with lesions in IFG with sparing of pTC/LTO, while performance in the gesture task is associated with lesions in pTC/LTO with sparing of IFG. Note that this latter account does not preclude integration between the two channels: some form of integration or matching may occur outside the network discussed above in multimodal semantic hubs involved in combining conceptual information from different sources, such as ATL (Holland & Lambon Ralph, 2010).

Our results support the complementary processing view:

(a) For multimodal congruent stimuli, we found dissociations between the speech and the gesture tasks. The amount of benefit from having the additional channel for each patient was associated with lesions affecting largely distinct nodes. When the task focused on speech, lesions to frontal (including IFG), parietal, and anterior temporal regions, and sparing of posterior pTC/LTO regions were associated with the largest advantage of congruent gestures. When the task focused on gestures, instead, lesions involving more posterior temporal, parietal, and occipital regions including pTC/LTO, and sparing of anterior regions were associated with the largest advantage of congruent speech.

(b) For multimodal incongruent stimuli, we found that for the speech task, greater costs from incongruent gesture were associated with lesions in IFG as well as anterior and middle STG and MTG. For the gesture task, greater costs were associated with lesions in posterior temporal, parietal and occipital regions including pTC/LTO. Thus, just as we discussed for benefits, IFG and pTC/LTO do not appear to be critical in integration between modalities as their role is specific for speech (IFG) or gesture (pTC/LTO). Importantly, we also found overlap between the regions associated with greater costs in the speech task, and those in the gesture task in regions comprising anterior STG and MTG. Such overlap is indicative of involvement of these regions in genuine integration across modalities.

(c) In the gesture task, higher gesture recognition scores predicted higher accuracy for incongruent trials. For both tasks, the other predictor (lexical-semantic for gesture and gesture recognition for speech) was not significant. This further supports reliance on the unimpaired modality in dealing with the multimodal stimuli. When information from each of the two modalities is congruent, the use of the other modality leads to benefits (and - nearly - at ceiling performance). When the information is incongruent, the extent to which the patient is disrupted by the other modality depends on their ability to extract meaning from words (in the speech task) or from gestures (in the gesture task).

It is interesting to note here that a complementary processing view has also been argued to account for the dynamically changing weight given to gestures in the comprehension of more naturalistic audio-visual narratives (Skipper, Goldin-Meadow, Nusbaum, & Small, 2009; Zhang, Frassinelli, Tuomainen, Skipper, & Vigliocco, preprint).

Benefits of Multimodal Language

We found clear dissociations between voxels associated with benefit from multimodal stimuli in the speech and in the gesture tasks. Lesions to ROL, middle portions of the STG, inferior parietal lobe and pars opercularis of the IFG (and sparing of posterior regions) uniquely were associated with larger benefits of congruent gestures when the task focused on speech, whereas lesions to voxels that were generally more posterior, including middle occipital, inferior and superior parietal, and posterior temporal regions (and sparing of anterior regions) were associated with larger benefits of congruent speech when the task focused on gesture. These dissociations can be understood in terms of the vulnerability to deficits in extracting semantic information from words and gestures that are associated with lesions to peri-sylvian temporal and IFG regions (e.g., Dronkers, Wilkins, Van Valin, Redfern, & Jaeger, 2004) versus temporo-occipital regions (e.g., Tarhan et al., 2015), respectively. Thus, in the context of deficient semantic comprehension in a given modality, residual processing in the other modality may be used in a compensatory manner.

In addition to left IFG and superior temporal regions, pre- and post-central regions (motor and sensory cortices), when lesioned, are also associated with greater benefit of gesture when the task focused on speech. The latter are not regions traditionally associated with difficulties with language comprehension or in extracting semantic information from linguistic stimuli. However, abundant recent evidence indicates that sensory-motor regions play a role in word processing, especially of words referring to actions. For example, understanding action verbs activates premotor and parietal (Rueschemeyer, Ekman, van Aceren, Kilner, 2014) as well as primary motor cortices (e.g., Garcia, Moguilner, Torquati, Garcia-Marco et al., 2019; Vigliocco, Warren, Siri, Arciuli, Scott, & Wise, 2005). Disrupting motor cortex with TMS slows action word processing (Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015; Vukovic, Feurra, Shpektor, Myachykov, & Shtyrov, 2017) and excitatory tDCS to motor cortex facilitates gesture-verb matching tasks (Hayek, Floel, & Antonenko, 2018). Conceptual processing of action words is deficient in patients with IFG lesions as well as hand-related premotor and motor cortices (Kemmerer, Rudrauf, Manzel, & Tranel, 2012, see also Vigliocco et al., 2011).

Although the data with respect to the role of sensory-motor regions in noun processing is less abundant, there is some evidence that the motor system may be involved in concrete noun processing (Marino, Gough, Gallese, Riggio, & Buccino, 2013), and IFG has long been involved in language comprehension, broadly speaking (Dronkers et al., 2004; Turken & Dronkers, 2011). The present data suggest that limitations in word comprehension associated with frontal and parietal lesions may be at least partly mitigated by a compensatory reliance on gesture processing.

Costs Associated with Mismatching Speech and Gesture

The pattern of SVR-LSM results with respect to the costs of mismatching cross-modal information was similar in some respects to that seen for the benefit of multimodal congruent information. Specifically, peri-sylvian regions in the IFG (pars triangularis in this case) and superior temporal lobe, when lesioned (and sparing of posterior regions), were associated with greater costs of mismatching gesture in the speech task, whereas lesions to more posterior regions including occipital, posterior temporal, and inferior parietal cortices (and sparing of more anterior regions) were associated with greater costs of mismatching speech in the gesture task. Similar to the account we proposed for the benefit of congruent pairings, the cost of mismatched cross-modal information is particularly strong when there is a vulnerability in a given modality. Thus, lesions affecting the extraction of semantic meaning from language render particular sensitivity to mismatching gestural information, and vice versa.

On the basis of results of previous fMRI studies that contrasted incongruent to congruent speech gesture pairings (Willems et al., 2007; 2009), Özyürek (2014) suggests that IFG and pMTG may play different roles in the semantic integration of information from speech and gesture. Specifically, IFG is argued to be sensitive to the degree of semantic processing required to integrate somewhat ambiguous information from speech and gesture (which is greater when the two are incongruent). In contrast, pMTG is considered to be involved in matching two input streams (gestural and verbal) when each is providing unambiguous semantic information. Although our study was not designed to assess this hypothesis, the lack of involvement of IFG regions in the processing of incongruent speech-gesture pairings when the task focuses on gesture (in contrast to a focus on speech, as in the previous studies), indicates that involvement of IFG is asymmetrical between the two modalities.

Crucially, in the SVR-LSM analysis for multimodal incongruent pairings, in contrast to congruent pairings, we found relatively more (and larger) regions of overlap associated with greater costs of mismatching cross-modal information in both speech and gesture tasks. These regions are in anterior superior and middle temporal lobe, and temporal pole (i.e., ATL) as well as in posterior temporal-occipital cortex. We take the overlap in these regions to indicate genuine integration across modalities.

Left posterior STG and, especially, MTG have been associated with semantic integration of speech and gesture in a number of previous studies (Green, Straube, Weis, Jansen, Willmes, Konrad, & Kircher, 2009; Holle et al., 2008; 2010; Willems et al., 2009). There is also clear evidence for sensory-level audio-visual integration in left pSTS/STG (Calvert, Campbell, & Brammer, 2000). STS has been shown to play a role in the sensory integration of visual objects with their associated sounds (Beauchamp, Lee, Argall, & Martin, 2004), and auditory speech with its accompanying mouth movements (Calvert et al., 2000). However, our results do not converge on this picture. First, we showed that more posterior regions (MTG and adjacent pTC/LTO) do play a role in comprehension of speech-gesture combinations, but crucially, not as integration zones. Second, the regions of STG/MTG which instead we found to be critical in the integration of speech and gestures are more anterior extending into ATL.

ATL (bilaterally) has been shown to be associated with the representation of semantic knowledge. ATL involvement in multimodal conceptual knowledge has been observed in PET studies (Sharp, Scott, & Wise, 2004; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996), distortion-corrected fMRI (Binney, Embleton, Jefferies, Parker, & Lambon Ralph, 2010; Visser & Lambon Ralph, 2011), MEG (Marinkovic, Dhond, Dale, Glessner, Carr, & Halgren, 2003) and TMS (Pobric, Jefferies, & Lambon Ralph, 2007; Pobric, Jefferies, & Lambon Ralph, 2010). It has also been shown in the syndrome of semantic dementia (SD), in which atrophy to this area results in progressive impairment to verbal and non-verbal semantic knowledge (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000; Patterson, Nestor, & Rogers, 2007) and in PWA (Mirman et al., 2015).

Most previous imaging studies concerning speech and gesture have not reported ATL involvement in processing speech-gesture pairings and, therefore, in their integration. This may be due to susceptibility artefacts that make it difficult to obtain reliable signal in this area with standard, gradient-echo fMRI (Devlin, Russell, Davis, Price, Wilson, Moss, Matthews, & Tyler, 2000; Visser, Jefferies, & Lambon Ralph, 2010). While it has been shown that these problems can be ameliorated using specific steps (e.g., Embleton, Haroon, Morris, Lambon Ralph, & Parker, 2010; Halai, Welbourne, Embleton, & Parkes, 2014), most fMRI studies do not do so and, therefore, have reduced sensitivity to activation especially in the ventral ATL. Interestingly, a recent study of speech-gesture comprehension showed decreased activity in ATL (more specifically STG/MTG) when more semantically demanding passages were accompanied by a larger number of gestures (Cuevas, Steines, He, Nagels, Culham, & Straube, 2018). This study investigated differences in the comprehension of naturalistic stimuli that differ in their semantic complexity as well as in the number of gestures for each segment of the story. The interaction between semantic complexity of the verbal materials and number of gestures was further accompanied by a general reduction of activation in left IFG for segments accompanied by representational gestures compared to those with no gestures.

We report here initial evidence for a causal role of ATL in speech gesture integration. The finding of greater costs for incongruent speech-gesture pairings in PWA with ATL lesions in both the speech and gesture tasks strongly suggest that this region, part of a multimodal "semantic hub", further participates in genuine integration of the two modalities.

Implications for Clinical Applications

A strength of our study in comparison to previous studies with stroke populations is that we have brought together PWA's performance in the speech-gesture study with their lesion profile as well as their psycholinguistic and gesture recognition profiles. This allowed us to assess the characteristics of PWAs who benefitted from co-speech gesture. Our behavioral analysis comparing PWA and control participants showed that in general, our patient group benefited more from congruent speech-gesture pairings than controls. The lesion analysis provides a key to understanding why this is the case: PWAs with IFG lesions and sparing of pMTG often have intact gesture recognition, and can use gesture to compensate for their impairments in extracting semantic information from speech; and PWAs with pMTG lesions and sparing of IFG can use speech to compensate for their impairment in extracting semantic information from gestures. Thus, both patient groups can benefit from multimodal stimuli although in different ways. These results are an important step in the development of future treatment studies that may prospectively assign participants to treatments on the basis of lesion loci. Our analysis of correlations with lexical-semantic and gesture recognition tests reinforce the link between lexical-semantic problems and costs of incongruent gesture in speech comprehension on one hand, and gesture recognition problems -- facet of the limb apraxia syndrome-- and costs of incongruent speech in gesture comprehension on the other hand. Although incongruent speech-gesture pairings are arguably nearly absent in real-world communication, it remains an open question whether PWA with lexical-semantic deficits and lesions in IFG and/or anterior

STG/MTG would suffer from other types of less meaningful, but potentially distracting gestures (such as beats or pragmatic gestures) which are, instead, well represented in everyday communication.

Conclusions

In the first lesion study of people with aphasia (PWA) – accompanied by different degrees of deficits in lexical-semantics and gesture recognition – that investigates multimodal word comprehension we have provided new insight into the role of specific nodes (IFG, pTC/LTO and anterior STG/MTG), part of the language and/or action networks, in the semantic processing of spoken words and gestures.

Acknowledgments: This research was supported by NIH R01-NS099061 awarded to Laurel Buxbaum, the Economic and Social Research Council (ESRC) of Great Britain: grant no. RES-620-28-6002 and European Research Council 743035 awarded to Gabriella Vigliocco, and by the Moss Rehabilitation Research Institute. We thank H. Branch Coslett and Olu Faseyitan for assistance with lesion image segmentation and warping.

References

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology*, *89*(1), 183–193. https://doi.org/10.1037/0022-0663.89.1.183

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable Controlled Retrieval and Generalized Selection Mechanisms in Ventrolateral Prefrontal Cortex. *Neuron*, *47*(6), 907–918. https://doi.org/10.1016/j.neuron.2005.07.023

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. Neuron, 41(5), 809–823. https://doi.org/10.1016/S0896-6273(04)00070-4

Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic Adaptation and Competition during Word Comprehension. *Cerebral Cortex (New York, NY)*, *18*(11), 2574–2585. https://doi.org/10.1093/cercor/bhn018

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, *4*(4), 417–422. https://doi.org/10.1093/scan/nsp053

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex (New York, NY)*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon Ralph, M. A. (2010). The Ventral and Inferolateral Aspects of the Anterior Temporal Lobe Are Crucial in Semantic Memory: Evidence from a Novel Direct Comparison of Distortion-Corrected fMRI, rTMS, and Semantic Dementia. *Cerebral Cortex*, *20*(11), 2728–2738. https://doi.org/10.1093/cercor/bhq019

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*(9), 1207–1215. https://doi.org/10.1016/S0028-3932(00)00034-8

Buxbaum, L. J., Kyle, K. M., & Menon, R. (2005). On beyond mirror neurons: Internal representations subserving imitation and recognition of skilled object-related actions in humans. *Brain Research. Cognitive Brain Research*, *25*(1), 226–239. https://doi.org/10.1016/j.cogbrainres.2005.05.014

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology: CB*, *10*(11), 649–657. https://doi.org/10.1016/s0960-9822(00)00513-3

Cocks, N., Byrne, S., Pritchard, M., Morgan, G., & Dipper, L. (2018). Integration of speech and gesture in aphasia: Integration of speech and gesture in aphasia. *International Journal of Language & Communication Disorders*, *53*(3), 584–591. https://doi.org/10.1111/1460-6984.12372

Cocks, N., Sautin, L., Kita, S., Morgan, G., & Zlotowitz, S. (2009). Gesture and speech integration: An exploratory study of a man with aphasia. *International Journal of Language & Communication Disorders*, *44*(5), 795–804. https://doi.org/10.1080/13682820802256965

Cuevas, P., Steines, M., He, Y., Nagels, A., Culham, J., & Straube, B. (2019). The facilitative effect of gestures on the neural processing of semantic complexity in a continuous narrative. *NeuroImage*, *195*, 38–47. https://doi.org/10.1016/j.neuroimage.2019.03.054

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., … Tyler, L. K. (2000). Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic Task. *NeuroImage*, *11*(6), 589–600. https://doi.org/10.1006/nimg.2000.0595

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gesture in the developing brain. *Developmental Science*, *15*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2011.01100.x

Dick, A. S., Mok, E. H., Beharelle, A. R., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human Brain Mapping*, *35*(3), 900–917. https://doi.org/10.1002/hbm.22222

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1), 145–177. https://doi.org/10.1016/j.cognition.2003.11.002

Eggenberger, N., Preisig, B. C., Schumacher, R., Hopfner, S., Vanbellingen, T., Nyffeler, T., … Müri, R. M. (2016). Comprehension of Co-Speech Gestures in Aphasic Patients: An Eye Movement Study. *PLOS ONE*, *11*(1), e0146583. https://doi.org/10.1371/journal.pone.0146583

Embleton, K. V., Haroon, H. A., Morris, D. M., Ralph, M. A. L., & Parker, G. J. M. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping*, *31*(10), 1570–1587. https://doi.org/10.1002/hbm.20959

Finkelnburg, D.C. (1870). Niederrheinische Gesellschaft, Sitzung vom 21. Marz 1870 in Bonn (Lower Rhine Society, meeting of 21 March 1870). *Berlin Klin. Wochenschr*. 7, 449–450, 460–462.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Friederici, A. D. (2009). Pathways to language: Fiber tracts in the human brain. *Trends in Cognitive Sciences*, *13*(4), 175–181. https://doi.org/10.1016/j.tics.2009.01.001

Gainotti, G., & Lemmo, M. A. (1976). Comprehension of symbolic gestures in aphasia. *Brain and Language*, *3*(3), 451–460. https://doi.org/10.1016/0093-934X(76)90039-0

Garcea, F., Stoll, H., & Buxbaum, L. (2019). Reduced competition between tool action neighbors in left hemisphere stroke. *BioRxiv*, 547950. https://doi.org/10.1101/547950

García, A. M., Moguilner, S., Torquati, K., García-Marco, E., Herrera, E., Muñoz, E., … Ibáñez, A. (2019). How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. *NeuroImage*, *197*, 439–449. https://doi.org/10.1016/j.neuroimage.2019.05.002

Goldenberg, G., & Randerath, J. (2015a). Shared neural substrates of apraxia and aphasia. *Neuropsychologia*, *75*, 40–49. https://doi.org/10.1016/j.neuropsychologia.2015.05.017

Goldenberg, G., & Randerath, J. (2015b). Shared neural substrates of apraxia and aphasia. *Neuropsychologia*, *75*, 40–49. https://doi.org/10.1016/j.neuropsychologia.2015.05.017

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., & Kircher, T. (2009). Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping*, *30*(10), 3309–3324. https://doi.org/10.1002/hbm.20753

Gunter, T. C., Weinbrenner, J. E. D., & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00080

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2010). The Role of Synchrony and Ambiguity in Speech–Gesture Integration during Comprehension. *Journal of Cognitive Neuroscience*, *23*(8), 1845–1854. https://doi.org/10.1162/jocn.2010.21462

Halai, A. D., Welbourne, S. R., Embleton, K., & Parkes, L. M. (2014). A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe. *Human Brain Mapping*, *35*(8), 4118–4128. https://doi.org/10.1002/hbm.22463

Hayek, D., Flöel, A., & Antonenko, D. (2018). Role of Sensorimotor Cortex in Gestural-Verbal Integration. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00482

Herbet, G., Lafargue, G., & Duffau, H. (2015). Rethinking voxel-wise lesion-deficit analysis: A new challenge for computational neuropsychology. *Cortex*, *64*, 413–416. https://doi.org/10.1016/j.cortex.2014.10.021

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hoeren, M., Kümmerer, D., Bormann, T., Beume, L., Ludwig, V. M., Vry, M.-S., … Weiller, C. (2014). Neural bases of imitation and pantomime in acute stroke patients: Distinct streams for praxis. *Brain*, *137*(10), 2796–2810. https://doi.org/10.1093/brain/awu203

Holland, R., & Lambon Ralph, M. A. (2010). The Anterior Temporal Lobe Semantic Hub Is a Part of the Language Neural Network: Selective Disruption of Irregular Past Tense Verbs by rTMS. *Cerebral Cortex*, *20*(12), 2771–2775. https://doi.org/10.1093/cercor/bhq020

Holle, H., & Gunter, T. C. (2007). The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. *Journal of Cognitive Neuroscience*, *19*(7), 1175–1192. https://doi.org/10.1162/jocn.2007.19.7.1175

Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *NeuroImage*, *39*(4), 2010–2024. https://doi.org/10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, *49*(1), 875–884. https://doi.org/10.1016/j.neuroimage.2009.08.058

Kalénine, S., & Buxbaum, L. J. (2016). Thematic knowledge, artifact concepts, and the left posterior temporal lobe: Where action and object semantics converge. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *82*, 164–178. https://doi.org/10.1016/j.cortex.2016.06.008

Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010a). Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. *Brain*, *133*(11), 3269–3280. https://doi.org/10.1093/brain/awq210

Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010b). Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. *Brain*, *133*(11), 3269–3280. https://doi.org/10.1093/brain/awq210

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. Psychological Science, 21(2), 260–267. https://doi.org/10.1177/0956797609357327

Kelly, S. D., Hirata, Y., Manansala, M., & Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00673

Kemmerer, D., Rudrauf, D., Manzel, K., & Tranel, D. (2012). BEHAVIORAL PATTERNS AND LESION SITES ASSOCIATED WITH IMPAIRED PROCESSING OF LEXICAL AND CONCEPTUAL KNOWLEDGE OF ACTIONS. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 826–848. https://doi.org/10.1016/j.cortex.2010.11.001

Kertesz, A. (1985). Apraxia And Aphasia. Anatomical And Clinical Relationship. In E. A. Roy (Ed.), *Advances in Psychology* (Vol. 23, pp. 163–178). https://doi.org/10.1016/S0166-4115(08)61140-1

Kertesz, A., Kertesz, A., Raven, J. C., & PsychCorp (Firm). (2007). *WAB-R: Western Aphasia Battery-Revised*. San Antonio, TX: PsychCorp.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review*, *124*(3), 245–266. https://doi.org/10.1037/rev0000059

Lacey, E. H., Skipper-Kallal, L., Xing, S., Fama, M., & Turkeltaub, P. (2017). Mapping common aphasia assessments to underlying cognitive processes and their neural substrates. Neurorehabilitation and Neural Repair, 31(5), 442–450. https://doi.org/10.1177/1545968316688797

Lingnau, A., & Petris, S. (2013). Action Understanding Within and Outside the Motor System: The Role of Task Difficulty. *Cerebral Cortex*, *23*(6), 1342–1350. https://doi.org/10.1093/cercor/bhs112

Mah, Y.-H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain*, *137*(9), 2522–2531. https://doi.org/10.1093/brain/awu164

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. Neuron, 38(3), 487–497.

Marino, B. F. M., Gough, P. M., Gallese, V., Riggio, L., & Buccino, G. (2013). How the motor system handles nouns: A behavioral study. *Psychological Research*, *77*(1), 64–73. https://doi.org/10.1007/s00426-011-0371-2

Martin, A., Kyle Simmons, W., Beauchamp, M. S., & Gotts, S. J. (2014). Is a single 'hub', with lots of spokes, an accurate description of the neural architecture of action semantics?: Comment on "Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge" by van Elk, van Schie and Bekkering. *Physics of Life Reviews*, *11*(2), 261–262. https://doi.org/10.1016/j.plrev.2014.01.002

Martin, N., Schwartz, M. F., & Kohen, F. P. (2006). Assessment of the ability to process semantic and phonological aspects of words in aphasia: A multi□measurement approach. *Aphasiology*, *20*(2–4), 154–166. https://doi.org/10.1080/02687030500472520

MATLAB and Statistics Toolbox Release (2012b). The MathWorks, Inc., Natick, Massachusetts, United States.

Metzler, C. (2001). Effects of left frontal lesions on the selection of context-appropriate meanings. Neuropsychology, 15(3), 315–328.

Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O. K., Coslett, H. B., & Schwartz, M. F. (2015). Neural Organization of Spoken Language Revealed by Lesion-Symptom Mapping. *Nature Communications*, *6*, 6762. https://doi.org/10.1038/ncomms7762

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 615–622. https://doi.org/10.1037/0278-7393.18.3.615

Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2009). Elucidating the Nature of Deregulated Semantic Cognition in Semantic Aphasia: Evidence for the Roles of Prefrontal and Temporo-parietal Cortices. *Journal of Cognitive Neuroscience*, *22*(7), 1597–1613. https://doi.org/10.1162/jocn.2009.21289

Obermeier, C., Dolk, T., & Gunter, T. C. (2012). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex*, *48*(7), 857–870. https://doi.org/10.1016/j.cortex.2011.02.007

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296. https://doi.org/10.1098/rstb.2013.0296

Özyürek, A., & Kelly, S. D. (2007). Gesture, brain, and language. *Brain and Language*, *101*(3), 181–184. https://doi.org/10.1016/j.bandl.2007.03.006

Papagno, C., Della Sala, S., & Basso, A. (1993). Ideomotor apraxia without aphasia and aphasia without apraxia: The anatomical support for a double dissociation. *Journal of Neurology, Neurosurgery, and Psychiatry*, *56*(3), 286–289.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews. Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Perniss, P., Vinson, D., & Vigliocco, G. (2020). Making Sense of the Hands and Mouth: The Role of “Secondary” Cues to Meaning in British Sign Language and English. *Cognitive Science*, *44*(7), e12868. https://doi.org/10.1111/cogs.12868

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Category-Specific versus Category-General Semantic Impairment Induced by Transcranial Magnetic Stimulation. *Current Biology*, *20*(10), 964–968. https://doi.org/10.1016/j.cub.2010.03.070

Pobric, G., Jefferies, E., & Ralph, M. A. L. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20137–20141. https://doi.org/10.1073/pnas.0707383104

Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*, *115*, 154–166. https://doi.org/10.1016/j.neuropsychologia.2017.08.027

Robinson, G., Shallice, T., Bozzali, M., & Cipolotti, L. (2010). Conceptual proposition selection and the LIFG: Neuropsychological evidence from a focal frontal group. *Neuropsychologia*, *48*(6), 1652–1663. https://doi.org/10.1016/j.neuropsychologia.2010.02.010

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: FMRI studies of semantic ambiguity. *Cerebral Cortex (New York, N.Y.: 1991)*, *15*(8), 1261–1269. https://doi.org/10.1093/cercor/bhi009

Rueschemeyer, S.-A., Ekman, M., van Ackeren, M., & Kilner, J. (2014). Observing, Performing, and Understanding Actions: Revisiting the Role of Cortical Motor Areas in Processing of Action Words. *Journal of Cognitive Neuroscience*, *26*(8), 1644–1653. https://doi.org/10.1162/jocn_a_00576

Schegloff, E. A. (1984). On Some Questions and Ambiguities in Conversation. In J.M. Atkinson and J.C. Heritage (eds.), Structures of Social Action. *Cambridge: Cambridge University Press*. 28–52.

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(1), 322-327. doi:10.1073/pnas.0805874106

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. *Cerebral Cortex (New York, N.Y.: 1991)*, *25*(10), 3894–3902. https://doi.org/10.1093/cercor/bhu274

Schwartz, M. F., Brecher, A. R., Whyte, J., & Klein, M. G. (2005). A patient registry for cognitive rehabilitation research: A strategy for balancing patients' privacy rights with researchers' need for access. *Archives of Physical Medicine and Rehabilitation*, *86*(9), 1807–1814. https://doi.org/10.1016/j.apmr.2005.03.009

Sharp, D. J., Scott, S. K., & Wise, R. J. S. (2004). Retrieving meaning after temporal lobe infarction: The role of the basal language area. *Annals of Neurology*, *56*(6), 836–846. https://doi.org/10.1002/ana.20294

Skipper, Jeremy I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures Orchestrate Brain Networks for Language Understanding. *Current Biology*, *19*(8), 661–667. https://doi.org/10.1016/j.cub.2009.02.051

Sperber, C., & Karnath, H.-O. (2017). Impact of correction factors in human brain lesion-behavior inference. *Human Brain Mapping*, *38*(3), 1692–1701. https://doi.org/10.1002/hbm.23490

Spunt, R. P., & Lieberman, M. D. (2012). Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *Journal of Neuroscience*, *32*(10), 3575–3583. https://doi.org/10.1523/JNEUROSCI.5715-11.2012

Stadthagen-Gonzales, H., Damian, M.F., Perez, M.A., Bowers, J.S. & Marin, J (2009). Name-Picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology*, *62(8)*, 1581-1597.

Steinthal, H. (1871). *Abriss der Sprachwissenschaft*. Berlin: F. Dümmlers Verlagsbuchhandlung Harrwitz und Gossmann.

Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A Supramodal Neural Network for Speech and Gesture Semantics: An fMRI Study. *PLOS ONE*, *7*(11), e51207. https://doi.org/10.1371/journal.pone.0051207

Tarhan, L. Y., Watson, C. E., & Buxbaum, L. J. (2015). Shared and Distinct Neuroanatomic Regions Critical for Tool-related Action Production and Recognition: Evidence from 131 Left-hemisphere Stroke Patients. *Journal of Cognitive Neuroscience*, *27*(12), 2491–2511. https://doi.org/10.1162/jocn_a_00876

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(26), 14792–14797.

Turken, A. U., & Dronkers, N. F. (2011). The Neural Architecture of the Language Comprehension Network: Converging Evidence from Lesion and Connectivity Analyses. *Frontiers in Systems Neuroscience*, *5*. https://doi.org/10.3389/fnsys.2011.00001

van Elk, M., van Schie, H., & Bekkering, H. (2014a). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews*, *11*(2), 220–250. https://doi.org/10.1016/j.plrev.2013.11.005

van Elk, M., van Schie, H., & Bekkering, H. (2014b). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews*, *11*(2), 220–250. https://doi.org/10.1016/j.plrev.2013.11.005

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*(6597), 254–256. https://doi.org/10.1038/383254a0

Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex,* 16, 1790-1796.

Vigliocco, G., Vinson, D.P., Druks, J. & Cappa, S.F. (2011). Nouns and Verbs in the Brain? A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience and Biobehavioural Reviews, 35,* 407-426.

Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*(6), 1083–1094. https://doi.org/10.1162/jocn.2009.21309

Visser, M., & Lambon Ralph, M. A. (2011). Differential Contributions of Bilateral Ventral Anterior Temporal Lobe and Left Anterior Superior Temporal Gyrus to Semantic Processes. *Journal of Cognitive Neuroscience*, *23*(10), 3121–3131. https://doi.org/10.1162/jocn_a_00007

Vukovic, N., Feurra, M., Shpektor, A., Myachykov, A., & Shtyrov, Y. (2017). Primary motor cortex functionally contributes to language comprehension: An online rTMS study. *Neuropsychologia*, *96*, 222–229. https://doi.org/10.1016/j.neuropsychologia.2017.01.025

Watson, C. E., Cardillo, E. R., Ianni, G. R., & Chatterjee, A. (2013). Action Concepts in the Brain: An Activation Likelihood Estimation Meta-analysis. *Journal of Cognitive Neuroscience*, *25*(8), 1191–1205. https://doi.org/10.1162/jocn_a_00401

Weiss, P. H., Ubben, S. D., Kaesberg, S., Kalbe, E., Kessler, J., Liebig, T., & Fink, G. R. (2016). Where language meets meaningful action: A combined behavior and lesion analysis of aphasia and apraxia. *Brain Structure and Function*, *221*(1), 563–576. https://doi.org/10.1007/s00429-014-0925-3

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex (New York, NY)*, *21*(5), 1066–1075. https://doi.org/10.1093/cercor/bhq180

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When Language Meets Action: The Neural Integration of Gesture and Speech. *Cerebral Cortex*, *17*(10), 2322–2333. https://doi.org/10.1093/cercor/bhl141

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, *47*(4), 1992–2004. https://doi.org/10.1016/j.neuroimage.2009.05.066

Wu, Y. C., & Coulson, S. (2007). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, *14*(1), 57–63. https://doi.org/10.3758/BF03194028

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences*, *106*(49), 20664–20669. https://doi.org/10.1073/pnas.0909197106

Yang, J., Shu, H., Bi, Y., Liu, Y., & Wang, X. (2011). Dissociation and association of the embodied representation of tool-use verbs and hand verbs: An fMRI study. *Brain and Language*, *119*(3), 167–174. https://doi.org/10.1016/j.bandl.2011.06.001

Zempleni, M.-Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage*, *34*(3), 1270–1279. https://doi.org/10.1016/j.neuroimage.2006.09.048

Zhang, Y., Kimberg, D. Y., Coslett, H. B., Schwartz, M. F., & Wang, Z. (2014). Multivariate Lesion-Symptom Mapping Using Support Vector Regression. *Human Brain Mapping*, *35*(12), 5861–5876. https://doi.org/10.1002/hbm.22590

Zhang, Y., Frassinelli, D., Tuomainen, J., Skipper, J. I., & Vigliocco, G. (2020). *More than words: Word predictability, prosody, gesture and mouth movements in natural language comprehension* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.01.08.896712

Zhao, W., Riggs, K., Schindler, I., & Holle, H. (2018). Transcranial Magnetic Stimulation over Left Inferior Frontal and Posterior Temporal Cortex Disrupts Gesture-Speech Integration. *The Journal of Neuroscience*, *38*(8), 1891–1900. https://doi.org/10.1523/JNEUROSCI.1748-17.2017