Neuropsychologia 95 (2017) 21–29

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues

David Peeters[a,*], Tineke M. Snijders[a,b], Peter Hagoort[a,b], Aslı Özyürek[a,b]

[a] Max Planck Institute for Psycholinguistics, 6500 AH, Nijmegen, The Netherlands
[b] Radboud University Nijmegen, 6525 HP, Nijmegen, The Netherlands

ARTICLE INFO

Keywords:
Multimodal communication
Gesture
Pointing
Conceptual matching
Referential communication
Pragmatics
Mentalizing

ABSTRACT

In everyday communication speakers often refer in speech and/or gesture to objects in their immediate environment, thereby shifting their addressee's attention to an intended referent. The neurobiological infrastructure involved in the comprehension of such basic multimodal communicative acts remains unclear. In an event-related fMRI study, we presented participants with pictures of a speaker and two objects while they concurrently listened to her speech. In each picture, one of the objects was singled out, either through the speaker's index-finger pointing gesture or through a visual cue that made the object perceptually more salient in the absence of gesture. A mismatch (compared to a match) between speech and the object singled out by the speaker's pointing gesture led to enhanced activation in left IFG and bilateral pMTG, showing the importance of these areas in conceptual matching between speech and referent. Moreover, a match (compared to a mismatch) between speech and the object made salient through a visual cue led to enhanced activation in the mentalizing system, arguably reflecting an attempt to converge on a jointly attended referent in the absence of pointing. These findings shed new light on the neurobiological underpinnings of the core communicative process of comprehending a speaker's multimodal referential act and stress the power of pointing as an important natural device to link speech to objects.

1. Introduction

In everyday talk, people often refer to things in their immediate surroundings. In such situations, an important prerequisite for communicative success is for speaker and addressee to establish joint attention to the object, person, or event they are talking about. Imagine you are sitting at the window in a restaurant and your friend says "Look at that car". How do you identify the specific car your friend is talking about? In many such cases, a speaker may connect her communication to the entity she is referring to by manually pointing at it (Bühler, 1934; Clark, 1996; Kita, 2003), helping the addressee to single out the intended referent (one specific car). In other cases a pointing gesture may not be necessary because one object in the environment is clearly perceptually most salient, such that the addressee may infer that the speaker refers to the salient object (Clark et al., 1983). In both cases, the addressee needs to match the visual object that is referred to (the car) to the spoken label by which it is described ("car"). The aim of the current study is to advance our understanding of the neural architecture supporting this everyday communicative process, both when an object is singled out by a pointing gesture and when it is made perceptually salient by non-communicative physical properties.

Comprehending our interlocutors' pointing gestures is a core feature of everyday communication (Baron-Cohen, 1989; Clark, 1996; Kendon, 2004; Tomasello et al., 2007). Previous neuroimaging studies have looked at the neural correlates of observing pointing gestures outside a referential speech context and at their integration with cues such as the gesturer's gaze direction (e.g., Brunetti et al., 2014; Conty et al., 2012; Gredebäck et al., 2010; Materna et al., 2008; Redcay et al., 2015; Sato et al., 2009). Perceiving a pointing hand compared to perceiving a non-directional closed hand elicits enhanced activation in a set of mainly right-hemisphere regions, including right inferior frontal gyrus (IFG), right angular gyrus, right parietal lobule, right thalamus, and bilateral lingual gyri (Sato et al., 2009). Following the direction of someone's pointing finger elicits bilateral posterior superior temporal sulcus (pSTS) activation (Materna et al., 2008). Integrating someone's pointing gestures with their gaze direction recruits parietal and supplementary motor cortices in the right hemisphere (Conty et al., 2012). Together, these findings suggest an extensive right-hemisphere dominant network that is activated when one perceives a manual pointing gesture that shifts one's attention.

[*] Correspondence to: Max Planck Institute for Psycholinguistics, P.O. Box 310, NL-6500 AH, Nijmegen, The Netherlands.
E-mail address: david.peeters@mpi.nl (D. Peeters).

http://dx.doi.org/10.1016/j.neuropsychologia.2016.12.004
Received 5 September 2016; Received in revised form 25 November 2016; Accepted 5 December 2016
Available online 06 December 2016

In everyday communication, however, pointing gestures are not observed in isolation and often shift one's attention toward a visible entity such as an object (Kita, 2003). Pierno et al. (2009) compared the observation of an image of a hand pointing at an object to the observation of an image of a hand grasping an object and to a control condition of an image of a hand resting next to an object. In comparison to the control condition, the perception of the pointing hand and object elicited enhanced activation in left middle temporal gyrus (MTG), left parietal areas (postcentral gyrus and supramarginal gyrus) and left middle occipital gyrus. The pointing condition did not elicit additional activation compared to the grasping condition. Nevertheless these results suggest that, in addition to the right-hemisphere dominant network involved in perceiving a pointing hand that shifts one's attention, a left-lateralized set of cortical areas may subsequently be involved in visually integrating the pointing hand and an object-referent.

The studies described above each contribute valuable information towards a better understanding of the neural architecture involved in observing pointing gestures, but do not reflect the richness of everyday acts of human referential communication. Pointing gestures often occur in a context in which one perceives not only visual information such as an interlocutor's pointing hand and one object, but also the speech that she may concomitantly produce. Furthermore, the pointing gesture may be produced to single out one specific object from a larger set of visible potential referents. In such situations, an addressee needs to combine incoming information from visual (speaker, pointing gesture, and objects) and auditory (speech) modalities to comprehend the referential act. Furthermore, the perceived spoken label needs to be matched to the specific object the speaker intended to refer to for communication to be successful. The current study focuses on the comprehension of pointing gestures in such richer audiovisual contexts.

The main aim of the current study is to get a better understanding of the neural infrastructure involved in the conceptual matching of a spoken word with a visible object as induced by a referential pointing gesture in comprehension. Pointing gestures may single out an object from a larger set of potential referents while speech may concomitantly describe the object (Bühler, 1934; Clark and Bangerter, 2004), as in someone pointing at an apple while saying "I have bought this apple at the market this morning" (Peeters et al., 2015b). Previous work suggests that conceptual matching between auditory and visual information may recruit pMTG (e.g., Dick et al., 2014). It has been found, for instance, that observing a mismatch (versus a match) between a pantomime gesture and a concurrently encountered spoken word leads to enhanced activation in pMTG (Willems et al., 2009). This suggests that pMTG may be involved in mapping different sources of information onto a common memory representation, a process that has also been called semantic integration (Hagoort et al., 2009; Willems et al., 2009). A typical everyday situation in which semantic integration of auditory (the spoken label) and visual (the identified object) information takes place is presumably referential communication via pointing.

Additionally, in the case of complementary or mismatching signals, a novel conceptual representation may have to be construed. Evidence suggests that this process is subserved by left inferior frontal gyrus (LIFG). Observation of images (e.g. of a dog) paired with an incongruent sound (e.g. meowing), for instance, leads to enhanced activation in LIFG compared to observation of images (e.g. of a dog) paired with a congruent sound (e.g. barking; Hein et al., 2007). In the gestural domain, Dick et al. (2014) compared the perception of supplemental iconic gestures with speech to the perception of "redundant" iconic gestures with speech. The former gestures added information to the speech they accompanied (e.g. the verb in the phrase "Sparky attacked" was combined with a "peck" gesture) whereas the latter gestures ("Sparky pecked" combined with a "peck" gesture) did not. An increase in activation was found in LIFG for the gestures that added information to speech. Both such gestures commonly occur in everyday interactions (Holler and Beattie, 2003; Kendon, 2004; McNeill, 1992), suggesting that enhanced activation in LIFG is not restricted to unnatural mismatch situations. Rather, these findings suggest that LIFG is recruited in the online construction of a novel semantic representation, a process that has also been referred to as semantic unification (Hagoort et al., 2009; Willems et al., 2009). Unlike the iconic co-speech gestures used in previous studies, pointing gestures do not convey semantic information. Nevertheless, they do often relate semantic information in speech to (properties of) a physical object in one's immediate environment. Therefore a conceptual mismatch, induced by a pointing gesture, between a spoken word and a visual object might also recruit LIFG. Activation in LIFG and pMTG may be preceded by activation in pSTS linking auditory and visual information at a pre-lexical level (Dick et al., 2014).

A secondary aim of the current study is to investigate the neural underpinnings of referential communication in situations in which a speaker refers to an object that is perceptually salient, in the absence of pointing. In everyday conversations, addressees may identify a particular referent on the basis of its perceptual salience in the absence of a pointing gesture that singles out the object. Clark et al. (1983) showed participants a picture with four types of flower in it and asked "how would you describe the color of this flower? ", without pointing at one of the specific flowers in the picture. When daffodils were perceptually more salient than the other types of flower, participants described the color of the daffodils. Arguably, the addressee in such cases inferred that the speaker was referring to the object that was perceptually most salient. The neural underpinnings subserving such inferential processes in the comprehension of referential communication are unclear. One possibility is that such situations activate the mentalizing system (medial prefrontal cortex, temporo-parietal junction, and possibly precuneus; Frith and Frith, 2006; Schurz et al., 2014; Van Overwalle and Baetens, 2009), because addressees may attribute a belief or intention to the speaker in relation to their common ground. They both know that they both know that, in the absence of a pointing gesture, the most salient object is most likely the intended referent. This mentalizing process may be less necessary in more straightforward cases where a speaker expresses her communicative intent by simply pointing at an object while concurrently describing it in speech.

1.1. The present study

The present study aims to shed more light on the functional roles of different cortical areas recruited in basic communicative situations in which a speaker refers in speech and/or gesture to an object for an addressee in a visual context. In an event-related functional magnetic resonance imaging (fMRI) study, participants were presented with images of a speaker and two different objects while they listened to her speech. In each picture, one of the objects was singled out, either through the speaker's index-finger pointing gesture or through a visual cue that made the object more salient, in the absence of gesture. We employed a mismatch paradigm, such that the object that was singled out was either congruent (on match trials) or incongruent (on mismatch trials) with concurrent speech. In addition we included two separate unimodal runs (audio-only and visual-only; cf. Willems et al., 2009).

The main aim of the study was to get a better understanding of the neural infrastructure involved in the conceptual matching of a spoken word with a visible object as induced by a referential pointing gesture in comprehension. We predicted that brain areas involved in processing combinatorial semantic information through verbal and gestural channels as found in previous studies might also be relevant in the current manipulation. More specifically, we hypothesized that LIFG would be activated more in the case of a mismatch (compared to a match) between speech and the object that was singled out by the pointing gesture (see Dick et al., 2014 and Özyürek, 2014, for overview; Willems et al., 2007). This is in line with a view of LIFG, more specifically its pars triangularis, as involved in semantic unification of information from different input streams (e.g. Hagoort, 2013; Willems et al., 2009). Additionally we predicted that activation levels in pMTG would increase in the case of a mismatch between the spoken label and the object-referent (cf. Hocking and Price, 2008; Dick et al., 2014), as in the case of pantomime gestures that mismatch with concurrent speech (Willems et al., 2009).

It is an open question whether potential LIFG and pMTG activation following a gesture-induced conceptual mismatch between speech and object is specific to cues that are communicatively intended and have a clear referential value (such as pointing gestures) or generalizes to cases where a referent becomes perceptually salient in the absence of a gesture. This question is answered by comparing the condition where speech and the perceptually salient object mismatch to the corresponding match condition. Finally, we investigated whether pSTS is mainly involved in connecting information from visual and auditory modalities in general (Dick et al., 2014), a hypothesis that can be tested by comparing the sum of the unimodal runs to a congruent bimodal condition in a conjunction analysis at the whole-brain level.

A secondary aim of the study was to investigate the neural correlates of the inferential process instantiated in cases in which a speaker refers in speech to a perceptually salient object without pointing at it (Clark et al., 1983). In the absence of a communicative and referential cue such as a pointing gesture, addressees may infer that the speaker is referring to the most salient object. This inferential process elicited in the addressee could be reflected by enhanced activation in the mentalizing system, more specifically medial prefrontal cortex and the temporo-parietal junction (Frith and Frith, 2006; Schurz et al., 2014; Van Overwalle and Baetens, 2009), when speech matches (versus mismatches) a perceptually salient referent.

2. Method

2.1. Participants

Twenty-three right-handed (Oldfield, 1971) native speakers of Dutch (18 female; mean age 23.6, range 18–29) participated in the experiment. Data from three additional participants were discarded due to technical failure (n=2) or drowsiness (n=1). Participants had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided written informed consent and were paid for participation.

2.2. Stimulus materials and experimental design

The experimental materials consisted of 40 spoken items in Dutch of the form “definite article+noun” (e.g., “het kopje”, *the cup*), 80 pictures in which a model (henceforth: the speaker) pointed (index-finger extended; Kendon, 2004) at one of two objects presented at a table in front of her (henceforth “pointing pictures”), and 80 pictures that were the same except that one of the two objects was framed by a green box and that the speaker did not point (henceforth “saliency pictures”). Still pictures were used to allow for full experimental control and to match the amount of movement across the different conditions. The 40 spoken items were spoken at a normal rate by a female native speaker of Dutch, recorded in a sound proof booth, and digitized at a sample frequency of 44.1 kHz. They were equalized in maximum amplitude using *Praat* software (version 5.2.46; Boersma and Weenink, 2009) and had an average duration of 837 ms (*SD*=155 ms). In half of the pointing pictures the speaker pointed at the object at her left and in the other half of the pointing pictures she pointed at the object at her right. Similarly, in half of the saliency pictures the object at her left was framed and in the other half the object at her right. The 40 different table-top objects in the pictures were selected on the basis of a pre-test reported elsewhere (Peeters et al., 2015b) that confirmed that these objects elicited highly consistent labels (i.e. > 90% naming consistency for each object across 16 participants) across individuals from the same participant pool as the current participants.

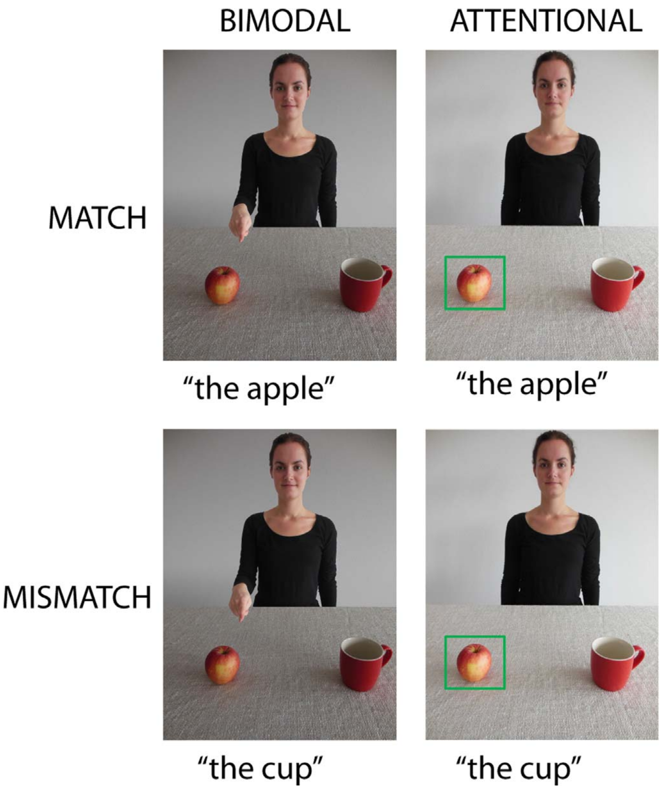

Fig. 1. Overview of the four conditions in the mixed block.

The experiment consisted of three blocks. The *speech-only* block (AUDIO) consisted of the 40 spoken items. The *picture-only* block (VISUAL) consisted of 40 pictures in which the speaker pointed at an object. The *mixed block* consisted of 160 speech-picture pairs that made up four conditions in a 2×2 design contrasting Cue (Pointing versus Saliency) with Congruency (Match versus Mismatch). In the Pointing Match (PM) condition, the spoken stimulus matched the object the speaker pointed at. In the Pointing Mismatch (PMM) condition, the spoken stimulus did not match the object she pointed at but the other object. In the Saliency Match (SM) condition, the spoken stimulus matched the perceptually salient (framed) object. In the Saliency Mismatch (SMM) condition, the spoken stimulus matched the object that was not framed. Each condition consisted of 40 speech-picture pairs. Fig. 1 shows a subset of pictures used in the experiment.

2.3. Procedure

The three blocks were presented sequentially with specific instructions preceding each block. The order of presentation of the blocks was counterbalanced across participants. All stimuli were presented in an event-related design and in a randomized order. Twelve different randomized lists were used. The *speech-only block* consisted of the presentation of the 40 spoken stimuli. A trial in this block consisted of a fixation cross presented for a jittered duration of 2–6 s followed by the presentation of the spoken stimulus. The *picture-only block* consisted of the presentation of 40 pictures in which the speaker pointed at one of the two objects. No speech was presented during this block. A trial in this block consisted of a fixation cross presented for a jittered duration of 2–6 s followed by the presentation of the picture for 2 s. The *mixed block* consisted of 160 target trials in which a fixation cross (jittered duration of 2–6 s) was followed by the presentation of a picture (for 2 s) with a concurrently presented spoken stimulus. The onset of the spoken stimulus was 50 ms after the onset of the picture presentation. In both the picture-only block and the mixed block, the speaker pointed at the object at her left in half of the cases, and at the object at her right in the other half of the cases. In the mixed block, in half of the saliency pictures the object at the speaker's left was framed and in the other half of the saliency pictures it was the object at her right.

Pictures were presented on the screen using *Presentation* software (Neurobehavioral Systems) and speech was presented through non-magnetic headphones that reduced scanner noise. Participants looked at the screen via a mirror mounted to the head coil. The size of the pictures on the screen was determined on the basis of judgments from two pilot subjects that did not participate in the main experiment. They confirmed that all objects, the speaker's gesture, and the saliency markers were clearly visible while focusing on the center of the screen.

Participants in the main experiment were instructed to carefully listen to the speech and look at the pictures. To ensure that participants paid attention to all aspects of the stimuli, they were asked to press a button with the middle finger of their left hand when an item (i.e. a spoken stimulus in the speech-only block, a picture in the picture-only block, and the picture-speech pair in the mixed block) was exactly the same on two subsequent trials (cf. Redcay et al., 2015). In the speech-only block and the picture-only block, four stimuli were repeated on two subsequent trials. In the mixed block 16 stimuli were repeated on two subsequent trials. The second presentations of such items in this 1-back task thus served as catch trials eliciting a button press and were excluded from further MRI analyses.

The experiment was preceded by a practice session that consisted of three blocks of nine trials each (i.e. eight items of which one was repeated and served as a catch trial to familiarize participants with the task). Before the start of the practice block the scanner was switched on and a number of spoken stimuli were played in order to adjust the volume level of the spoken items. Participants were asked to indicate whether the volume should go up or down. The items used in this audio check and the items used in the practice blocks were not used in the main experiment.

2.4. fMRI data acquisition

Participants were scanned with a Siemens 3-T Skyra MRI scanner using a 32-channel head coil. The functional data were acquired in one run using a multi echo-planar imaging sequence (Poser et al., 2006), in which image acquisition happens at multiple echo times (TEs) following a single excitation [time repetition (TR)=2250 ms; TE1=9 ms; TE2=19.5 ms; TE3=30 ms; TE4=40 ms; echo spacing=.51 ms; flip angle=90°]. This procedure broadens T2* coverage and improves T2* estimation (see Poser et al., 2006, for details). Each volume consisted of 36 slices of 3 mm thickness [ascending slice acquisition; voxel size =3.3×3.3×3 mm; slice gap =10%; field of view (FOV) =212 mm]. The first 30 volumes preceded the start of the presentation of the first stimulus and were used for weight calculation of each of the four echoes. Subsequently, the 31st volume was taken as the first volume in preprocessing. The functional run was followed by a whole-brain anatomical scan using a high resolution *T*1-weighted magnetization-prepared, rapid gradient echo sequence (MPRAGE) consisting of 192 sagittal slices (TR =2300 ms; TE =3.03 ms; FOV =256 mm; voxel size =1×1×1 mm) accelerated with GRAPPA parallel imaging.

2.5. Data Analysis

Data were analyzed using statistical parametric mapping (SPM8; www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab (Mathworks Inc., Sherborn, MA, USA). The four echoes of each volume were combined to yield one volume per TR (Poser et al., 2006), after which standard preprocessing was performed [realignment to the first volume, slice acquisition time correction to time of acquisition of the middle slice, coregistration to T1 anatomical reference image, normalization to Montreal Neurological Institute (MNI) space (EPI template), smoothing with an 8 mm full-width at half-maximum (FWHM) Gaussian kernel, and high-pass filtering (time-constant =128 s)] (Friston et al., 1995).

Statistical analysis was performed in the context of the general linear model (GLM). Stimulus onset (i.e. the onset of the picture in all conditions, except the speech-only condition in which it was the onset of speech) was modeled as the event of interest for each condition. Each condition thus contained 40 events. The 6 condition regression parameters were convolved with a canonical hemodynamic response function. Additionally, 6 motion parameters from the realignment preprocessing step were included in the first-level model.

Whole-brain analyses were performed by entering first-level contrast images of each of the six conditions > baseline for each participant into a flexible factorial model at second-level (with factors Condition [6] and Participant [23]). We tested for the 2×2 interaction (see Fig. 1) between Cue (Pointing versus Saliency) and Congruency (Match versus Mismatch). Two follow-up analyses were performed to compare mismatch to match conditions. First, the Pointing Mismatch condition was compared to the Pointing Match condition (PMM > PM and the reverse contrast). Second, the Saliency Mismatch condition was compared to the Saliency Match condition (SMM > SM and the reverse contrast). Additionally, a conjunction analysis, testing a logical AND (Nichols et al., 2005), was performed to subsequently verify whether any areas were activated more in the bimodal compared to the unimodal presentation of the stimuli. This analysis was implemented as (PM > AUDIO ∩ PM > VISUAL), inclusively masked with the conjunction of the unimodal conditions compared to zero, thus yielding the comparison (0 < AUDIO < PM > VISUAL > 0).

Whole-brain correction for multiple comparisons was applied by combining a significance level of p=.001 (uncorrected at the voxel level) with a cluster extent threshold using the theory of Gaussian random fields (Friston et al., 1996). All clusters are reported at an alpha level of $p < .05$ family-wise error (FWE) corrected across the whole brain (Nichols and Hayasaka, 2003).

Because we had a priori hypotheses related to activation differences in LIFG and bilateral pMTG for the different conditions in the mixed block, region-of-interest (ROI) analyses were performed in these regions. The LIFG ROI was an 8 mm sphere around center coordinates taken from a meta-analysis on a large number of neuroimaging studies of semantic processing (Vigneau et al., 2006; cf. Willems et al., 2009). MNI coordinates were [−42 19 14]. The bilateral pMTG ROIs were 8 mm spheres around center coordinates taken from a meta-analysis of multimodal integration studies (Hein and Knight, 2008; cf. Willems et al., 2009). MNI coordinates were [−49 −55 14] and [50 −49 13] for left and right hemisphere regions respectively. Contrast estimates in the ROI analyses were calculated for each participant at first-level for the four conditions (PM, PMM, SM, SMM) using Marsbar (http://marsbar.sourceforge.net/).

3. Results

3.1. Behavioral performance

Participants detected 91.5% of all catch trials. These data were not further analyzed.

3.2. Whole-brain analyses

The test for a 2×2 interaction effect showed a significant interaction between Cue and Congruency in clusters at the temporo-parietal junction and left superior and medial frontal areas. Follow-up analyses comparing the mismatch conditions to the match conditions for both types of cue were performed to further investigate this interaction. Contrasting PMM with PM (PMM > PM) showed increased activations in LIFG's pars triangularis. The reverse contrast (PM > PMM) did not

Table 1
Results of the whole-brain analyses comparing congruent (match) to incongruent (mismatch) conditions. p-values are at the cluster-level, FWE-corrected.

Contrast	*p*	k (extent)	*t*-value	MNI coordinates			BA	Region/Peak
(PMM - PM) > (SMM - SM)	.000	729	4.28	−6	−64	10	17	Bilateral calcarine cortex, right cuneus
			4.26	4	−62	8		
			4.08	6	−70	22		
	.000	518	4.80	46	−38	16	21/41	Right superior/middle temporal gyrus
			4.46	50	−50	12		
			4.25	44	−54	18		
	.010	217	4.25	−6	46	20	32	Left superior medial gyrus, left anterior cingulate
			3.67	−14	44	10		
			3.50	−4	44	6		
	.027	175	4.62	−32	−20	38	6	Left middle frontal gyrus
			4.23	−20	−14	54		
			3.87	−24	−10	48		
PMM > PM	.010	220	4.01	−46	20	20	45	Left inferior frontal gyrus (pars triangularis)
			3.72	−36	18	20		
			3.69	−50	28	18		
PM > PMM	–	–	–	–				
SMM > SM	–	–	–	–				
SM > SMM	.000	7663	5.88	−6	−64	12	17	Left calcarine cortex, left lingual gyrus, right hippocampus
			5.50	−12	−52	4		
			5.34	38	−10	−20		
	.000	3695	5.19	−38	−18	58	4	Left precentral gyrus, left/right middle cingulate
			5.08	16	−36	46		
			5.02	−8	−30	48		
	.000	1103	4.80	−40	−18	4	48	Left Heschl's gyrus, left putamen
			4.51	−30	−4	−4		
			4.22	−26	12	−12		
	.000	984	4.67	66	−40	22	21/22	Right superior/middle temporal gyrus
			4.25	46	−26	16		
			4.21	52	−48	12		
	.000	880	4.83	−8	46	14	10/32	Left anterior cingulate, left medial superior frontal gyrus
			4.10	−10	52	24		
			4.08	−12	50	2		
	.000	625	5.07	−18	30	44	9/32	Left superior frontal gyrus
			4.57	−18	38	46		
			4.22	−12	42	38		
	.000	487	5.12	42	−14	46	3/6	Right precentral/postcentral gyrus
			4.01	40	−18	38		
			3.78	48	−10	42		
	.003	280	4.21	−54	−30	22	48	Left superior temporal gyrus, left supramarginal gyrus
			3.97	−44	−32	24		
			3.70	−66	−30	20		

Abbreviations: BA, Brodmann Area; PM, Pointing Match; PMM, Pointing Mismatch; SM, Saliency Match; SMM, Saliency Mismatch

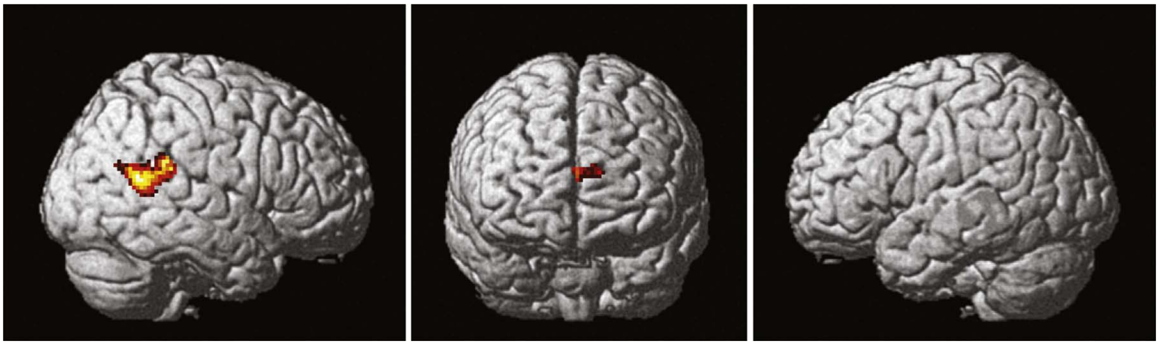

Fig. 2. Results from the whole brain analysis testing for a 2×2 interaction effect: (PMM−PM) > (SMM−SM). The left panel shows the right hemisphere. The middle panel shows a frontal view of the brain. The right panel shows the left hemisphere.

show any significant cluster that survived the statistical threshold. Contrasting SMM with SM (SMM > SM) did not show any areas that survived the statistical threshold. The reverse contrast (SM > SMM) yielded several clusters with increased activation in the Saliency Match compared to the Saliency Mismatch condition, including in medial prefrontal areas, motor areas, and temporo-parietal junction. Table 1 and Figs. 2 and 3 present the results of these analyses.

Second, as a sanity check for the conjunction analysis, we compared the unimodal AUDIO and VISUAL conditions to baseline. These comparisons revealed mainly enhanced activation in unimodal (i.e.

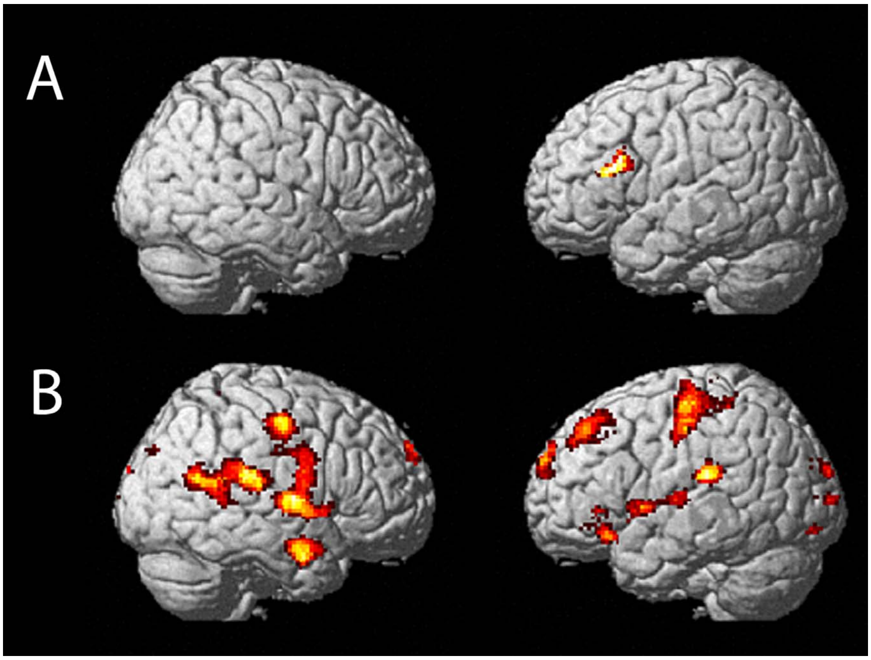

Fig. 3. Results from the whole brain analysis comparing (A) Pointing Mismatch (PMM) > Pointing Match (PM), and (B) Saliency Match (SM) > Saliency Mismatch (SMM).

primary auditory and visual) and motor areas. The results of these analyses can be found in the Supplementary Materials. The conjunction analysis comparing the Pointing Match condition to the unimodal conditions (0 < AUDIO < PM > VISUAL > 0) failed to show any cluster that survived the statistical threshold (no voxels < .001 uncorrected).

3.3. ROI analyses

Fig. 4 presents the contrast estimates for the four conditions in the three ROI analyses. A first ROI analysis was performed comparing mismatch to match conditions in the predefined ROI in LIFG (8 mm sphere around MNI coordinates −42 19 14). The interaction between Cue and Congruency did not reach significance, F (1,22) =2.10, p=.162. However, planned dependent samples t-tests revealed that there was enhanced activation in LIFG in the PMM compared to the PM condition, t (22)=−2.43, p=.024. There was no such difference in activation in the ROI between the SMM and SM conditions, t (22) =.48, p=.637. Because of the discrepancy between the absence of a significant interaction effect in LIFG and simple comparisons that did suggest such an interaction effect, we plotted the contrast estimate differences between mismatch and match conditions for all participants for both the Pointing and the Saliency cue (see Fig. 5). In line with the absence of an interaction effect, no consistent evidence across participants is observed that suggests that the enhanced LIFG activation is specific to pointing gestures.

Second, the ROI analysis in left pMTG (8 mm sphere around MNI coordinates −49 −55 14) revealed a significant Cue×Congruency interaction, F (1,22) =6.30, p=.020. Dependent samples t-tests showed enhanced activation in this region in the PMM compared to the PM condition, t (22) =−2.20, p=.038, but no significant difference in the comparison of the SMM and SM condition, t (22) =1.39, p=.177.

Third, the ROI analysis in the right pMTG (8 mm sphere around MNI coordinates 50 −49 13) showed a significant Cue x Congruency interaction effect, F (1,22) =18.41, p=.001. This interaction reflected a relative increase in activation in this region in the PMM compared to the PM condition, t (22) =−2.44, p=.023, and a relative increase in activation in the SM compared to the SMM condition, t (22) =3.86, p=.001.

4. Discussion

The present study investigated the neural architecture involved in the core communicative process of comprehending a speaker's reference in speech and gesture to a visible object. We looked at situations in which a speaker's pointing gesture singled out an object as the intended referent and at situations in which an object could be identified as the intended referent because it was perceptually the most salient alternative in the absence of gesture. In both cases

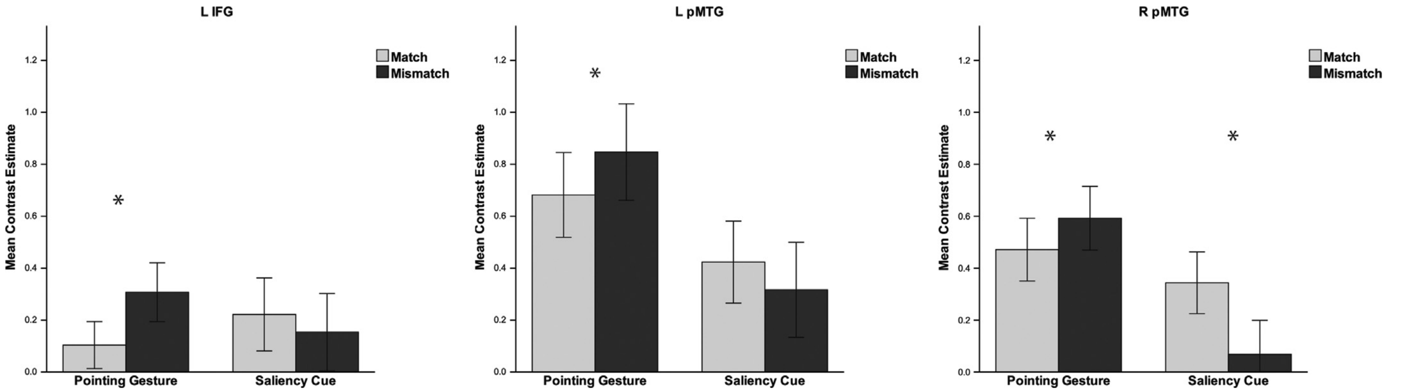

Fig. 4. ROI results. Mean contrast estimates for the four conditions in the mixed block in the three ROI analyses. Error bars represent standard errors around the mean.

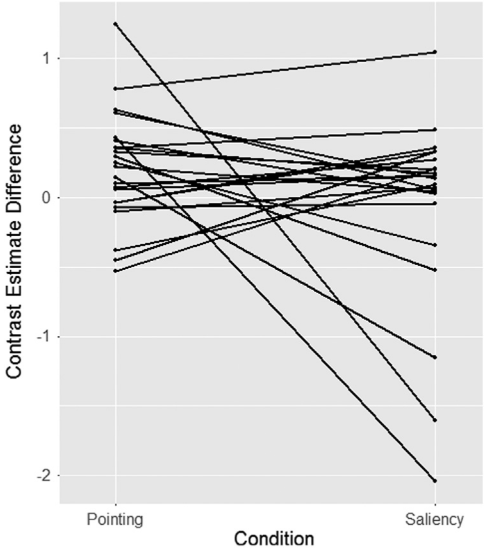

Fig. 5. Contrast estimate differences in LIFG (mismatch minus match) for individual participants in the case of Pointing and Saliency cues. For both the Pointing manipulation and the Saliency manipulation, a positive contrast estimate difference indicates a higher activation in LIFG for the mismatch condition compared to the match condition. A negative slope indicates that the activation difference in LIFG between mismatch and match conditions was higher in the case of Pointing compared to the Saliency cue.

concurrent speech either matched or mismatched the intended referent. Enhanced activation was found in LIFG and bilateral pMTG when the spoken label did not match the visible object that was singled out by the speaker's pointing gesture. In the absence of gesture, a match between a spoken label and a perceptually salient object elicited enhanced activation in several brain structures including medial frontal areas, temporo-parietal junction, and motor areas. We will now relate these findings to previous work investigating the neurobiological underpinnings of speech-gesture comprehension and referential communication more broadly.

Earlier neuroimaging studies investigating the comprehension of gestures in a speech context have mainly focused on iconic gestures. It has been shown that addressees process a speaker's iconic co-speech gestures and on-line integrate their meaning with concurrent speech (e.g., Dick et al., 2014; Habets et al., 2011; Holle et al., 2008; Özyürek et al., 2007). A common denominator in neuroimaging studies looking at the processing of gestures (e.g. iconic gestures, pantomimes, or metaphoric gestures) and concurrent speech is that an increase in semantic unification load leads to an increase in LIFG activation (cf. Andric and Small, 2012; Dick et al., 2014; Hagoort et al., 2009; Kircher et al., 2009; Nagels et al., 2013; Özyürek, 2014; Skipper et al., 2007; Willems et al., 2007, 2009). For instance, iconic gestures that are unrelated to concurrently perceived speech require additional processing compared to iconic gestures that relate to the concurrently presented speech because building a conceptual representation on the basis of the different streams of information is more effortful in the former compared to the latter case. The same holds for metaphoric co-speech gestures compared to iconic co-speech gestures (Straube et al., 2011). Similarly, iconic gestures or pantomimes that are incongruent with concurrent speech activate LIFG more than iconic gestures and pantomimes that match the speech they accompany (Willems et al., 2007, 2009). In such cases, arguably, LIFG is recruited in the online construction of a novel conceptual representation that is not already available in long-term memory (Hagoort et al., 2009).

The current study focuses on purely deictic pointing gestures and thereby differs from the speech-gesture comprehension studies described above in that there was no semantic information available in the gesture itself. Hence, it was not the semantic relation between speech and gesture that induced LIFG activation, but rather the degree of conceptual match between speech and a visible object as induced by the gesture. LIFG is found to play a role not only in semantic unification of speech and gesture, but also in the semantic unification of word meaning and world knowledge into a preceding context in speech itself (Hagoort, 2013; Hagoort et al., 2004, 2009; Zhu et al., 2012). Taken together, these findings suggest that semantic unification recruits LIFG across different semiotic domains. LIFG may play a role in the case of an indexical semiotic relation between gesture, speech, and a referent (the current study), in addition to symbolic and iconic manners of signification (as in arbitrary word-meaning mappings and resemblance between iconic gestures / pantomimes / pictures and referents respectively). Furthermore, a core property of language (including gesture) is that is allows for displacement, i.e. the ability to refer to entities that are not immediately present (Gunter et al., 2015; Hockett, 1960; Perniss and Vigliocco, 2014). The current study suggests that also when a referent is physically present in the immediate visual context, LIFG may subserve the semantic unification of auditory and visual information at a higher-order semantic level.

Region of interest analyses additionally revealed enhanced activation in bilateral pMTG when the object that was singled out by a pointing gesture was incongruent with information concurrently conveyed in speech. A similar pattern of results has been observed in the comprehension of pantomime gestures that mismatched concurrent speech (Willems et al., 2009). One proposal in line with the current findings is that activation levels in pMTG increase in the service of higher-order unification processes subserved by LIFG (Hagoort et al., 2009). In addition to LIFG and pMTG, another node in a network subserving the online processing of concurrent speech and gesture may be pSTS. Research suggests that this region is involved in connecting information from visual and auditory modalities at a pre-lexical level (Dick et al., 2014; cf. Hocking and Price, 2008). The fact that we did not find pSTS activation in the conjunction analysis comparing a bimodal (i.e., speech + pointing) condition to the sum of the unimodal conditions may be due to the absence of motion in our visual stimuli (cf. Dick et al., 2009). The current study may serve as a baseline for future studies investigating the processing of pointing gestures and speech in more dynamic and interactive situations (cf. Holler et al., 2014).

Interestingly, we did not find enhanced activation in pMTG in cases in which one object could be identified as the intended referent because it was made perceptually more salient by a visual cue and mismatched concurrent speech. Unlike intrinsic object properties that grasp one's attention, pointing gestures have referential force and are often communicatively motivated (Enfield et al., 2007). People tailor the specific kinematic properties of their pointing gestures to the informational needs of their addressees (De Langavant et al., 2011; Peeters et al., 2015a). In the absence of such a clear communicative and referential cue, perceptual saliency alone arguably may not naturally lead to the conceptual matching between speech and a perceptually salient object. This is reflected by the absence of enhanced activation in pMTG in the saliency mismatch condition compared to the saliency match condition (see Fig. 4), which may suggest that the mismatching salient object is not considered to be in common ground between speaker and addressee to the same extent as an object that is singled out by a pointing gesture.

Research in the domain of co-speech iconic gestures suggests that speech-gesture integration differs from the integration of speech with concurrently performed instrumental actions on objects because the former are generally viewed as more intended to accompany the speech signal compared to the latter (e.g., Kelly et al., 2015). A similar conclusion was reached in the domain of beat gestures. It has been found that more complex syntactic structures are easier to process when encountered in the accompaniment of such a rhythmic hand movement. This processing advantage is not observed, however, when the same syntactically complex sentence is accompanied by a moving visual stimulus, arguably because only the beat gesture (and not the visual stimulus) is produced with a communicative intention (Holle et al., 2012). In line with these studies on iconic co-speech gestures and co-speech beat gestures, the conceptual matching process induced by a pointing gesture in the current study, as reflected by pMTG activation, seems to differ qualitatively from cases in which objects attract attention during concurrently perceived speech via non-referential and non-communicative means. LIFG did not consistently across participants differentiate between pointing gestures and the non-communicative saliency cue (see Fig. 5).

Nevertheless, in the absence of a pointing gesture, addressees may understand that a speaker is referring to a particular object when her speech matches a perceptually salient referent. This process requires a complex (metacognitive) inferential process in which an addressee may infer that the speaker assumes that the addressee understands that she refers to the most salient object against their common ground (Clark et al., 1983). Brain structures that are commonly recruited when people think about the mental states of others ("mentalize") are medial prefrontal cortex and bilateral temporo-parietal junction (Frith and Frith, 2006; Schurz et al., 2014; Van Overwalle and Baetens, 2009). Our results suggest that these areas, more precisely in superior frontal cortex and in superior temporal gyrus and supramarginal gyrus, were indeed activated in situations in which participants comprehended a speaker's referential spoken label for a perceptually salient object in the absence of pointing. This mentalizing process was arguably induced in cases where speech matched the perceptually salient object. In the absence of a match between speech and the salient visual information on the object, the inference that the speaker referred to the most salient object was arguably not made. Activation in areas supporting mentalizing may thus explain why interlocutors successfully converge on a jointly attended referent in the absence of pointing. Clark et al. (1983) showed that addressees identify the perceptually most salient of four types of flowers as the intended referent when a speaker asked "how would you describe the color of this flower?". The current study suggests that in such cases the mentalizing system contributed to the successful identification of the intended referent. The identification of an intended referent will generally be more straightforward in the presence of overt cues that may even automatically direct one's attention in the right direction, such as pointing gestures (Langton and Bruce, 2000). Mentalizing may also be necessary however in metonymic situations (not investigated in the current study) where a speaker points at an object (e.g. an empty chair) that is different from the entity she intends to refer to (e.g. the director that always sits in that chair).

In addition to mentalizing areas, we observed increased activation in motor and somatosensory areas in the case of a match between speech and a perceptually salient object. These findings are reminiscent of the parietal (e.g., postcentral gyrus) activations elicited in the comparison of the observation of an image of a hand pointing at an object to an image of a hand resting next to an object (Pierno et al., 2009), here also confirmed in the comparison of our visual-only condition compared to baseline (see Suppl. materials). Such activations may reflect the preparation of a manipulative action toward an object (Pierno et al., 2009). In everyday situations, people point at objects in the immediate context of their conversation not only to shift their addressee's attention to an object, but often also to subsequently or concurrently indicate that they want their addressee to do something with the object (e.g., Southgate, Van Maanen, and Csibra, 2007). For instance, one may point at the cheese at one's breakfast table to request one's addressee to pass it. Similarly, in case of established joint attention to a specific object in the absence of a pointing gesture, addressees may usually assume that a speaker named the object with a particular directive illocutionary force (Searle, 1975), as in the case of a request, and therefore prepare a motor response (see Kelly et al., 1999).

In sum, the current study aimed to shed more light on the functional roles of different cortical areas recruited in comprehending basic communicative situations in which a speaker refers in speech and/or gesture to an object for an addressee in a visual context. LIFG and bilateral pMTG were found to play a role in the conceptual matching process between speech and visual information. Only for pMTG this activation was unique to communicatively intended cues (pointing gestures). In the absence of pointing, the mentalizing system was recruited in the comprehension of a speaker's verbal reference to a perceptually salient object. This study can be informative as a starting point for studies investigating specific populations with impairments in the comprehension of referential speech and gesture and the subsequent establishment of joint attention in everyday life (e.g., as in autism spectrum disorders). It also has implications for the processing of multimodal educational materials in which objects may be made salient through communicative and non-communicative cues.

Acknowledgments

We thank Paul Gaalman for assistance during the scanning sessions and Doris Deckers for help in creation of stimuli. We thank Anthony Dick and Thomas C. Gunter for valuable comments on an earlier version of this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuropsychologia.2016.12.004.

References

Andric, M., Small, S.L., 2012. Gesture's neural language. Front. Psychol., 3.

Baron-Cohen, S., 1989. Perceptual role taking and protodeclarative pointing in autism. Br. J. Dev. Psychol. 7 (2), 113–127.

Boersma, P., Weenink, D., 2009. Praat: doing phonetics by computer (Version 5.1.05) [Computer program].

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M.G., Cirillo, S., Romani, G.L., Pizzella, V., Aureli, T., 2014. Do you know what I mean? Brain oscillations and the understanding of communicative intentions. Front. Hum. Neurosci. 8, 36.

Bühler, K., 1934. Sprachtheorie. Fischer, Jena.

Clark, H.H., 1996. Using Language. Cambridge University Press, Cambridge.

Clark, H.H., Bangerter, A., 2004. Changing ideas about reference. In: Noveck, I.A., Sperber, D. (Eds.), Experimental Pragmatics. Palgrave Macmillan, Basingstoke, 25–49.

Clark, H.H., Schreuder, R., Buttrick, S., 1983. Common ground at the understanding of demonstrative reference. J. Verbal Learn. Verbal Behav. 22 (2), 245–258.

Conty, L., Dezecache, G., Hugueville, L., Grèzes, J., 2012. Early binding of gaze, gesture, and emotion: neural time course and correlates. J. Neurosci. 32 (13), 4531–4539.

De Langavant, L.C., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., Bachoud-Lévi, A.C., 2011. Behavioral and neural correlates of communication via pointing. PLoS One 6 (3), e17719.

Dick, A.S., Goldin-Meadow, S., Hasson, U., Skipper, J.I., Small, S.L., 2009. Co-speech gestures influence neural activity in brain regions associated with processing semantic information. Hum. Brain Mapp. 30 (11), 3509–3526.

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. Hum. Brain Mapp. 35, 900–917.

Enfield, N.J., Kita, S., De Ruiter, J.P., 2007. Primary and secondary pragmatic functions of pointing gestures. J. Pragmat. 39 (10), 1722–1741.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S., Turner, R., 1995. Analysis of fMRI time-series revisited. Neuroimage 2 (1), 45–53.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. Neuroimage 4 (3), 223–235.

Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50 (4), 531–534.

Gredebäck, G., Melinder, A., Daum, M., 2010. The development and neural basis of pointing comprehension. Soc. Neurosci. 5 (5–6), 441–450.

Gunter, T.C., Weinbrenner, J.D., Holle, H., 2015. Inconsistent use of gesture space during abstract pointing impairs language comprehension. Front. Psychol. 6, 80.

Habets, B., Kita, S., Shao, Z., Özyurek, A., Hagoort, P., 2011. The role of synchrony and ambiguity in speech–gesture integration during comprehension. J. Cogn. Neurosci. 23 (8), 1845–1854.

Hagoort, P., 2013. MUC (Memory, Unification, Control) and beyond. Front. Psychol., 4.

Hagoort, P., Baggio, G., Willems, R.M., 2009. Semantic unification. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences4th ed.. MIT Press, Cambridge, MA, 819–836.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304 (5669), 438–441.

Hein, G., Doehrmann, O., Müller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J. Neurosci. 27 (30), 7881–7887.

Hein, G., Knight, R.T., 2008. Superior temporal sulcus—it's my area: or is it? J. Cogn. Neurosci. 20 (12), 2125–2136.

Hockett, C.D., 1960. The origin of speech. Sci. Am. 203 (3), 88–96.

Hocking, J., Price, C.J., 2008. The role of the posterior superior temporal sulcus in audiovisual processing. Cereb. Cortex 18 (10), 2439–2449.

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.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A.D., Ward, J., Gunter, T.C., 2012. Gesture facilitates the syntactic analysis of speech. Front. Psychol. 3, 74.

Holler, J., Beattie, G., 2003. Pragmatic aspects of representational gestures: do speakers use them to clarify verbal ambiguity for the listener? Gesture 3 (2), 127–154.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., Özyürek, A., 2014. Social eye gaze modulates processing of speech and co-speech gesture. Cognition 133 (3), 692–697.

Kelly, S.D., Barr, D.J., Church, R.B., Lynch, K., 1999. Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. J. Mem. Lang. 40 (4), 577–592.

Kelly, S., Healey, M., Özyürek, A., Holler, J., 2015. The processing of speech, gesture, and action during language comprehension. Psychon. Bull. Rev. 22 (2), 517–523.

Kendon, A., 2004. Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., Green, A., 2009. Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. Neuropsychologia 47 (1), 169–179.

Kita, S., 2003. Pointing. Where language, culture, and cognition meet. Erlbaum, Hillsdale, NJ.

Langton, S.R., Bruce, V., 2000. You must see the point: automatic processing of cues to the direction of social attention. J. Exp. Psychol.: Hum. Percept. Perform. 26 (2), 747–757.

Materna, S., Dicke, P.W., Thier, P., 2008. The posterior superior temporal sulcus is involved in social communication not specific for the eyes. Neuropsychologia 46 (11), 2759–2765.

McNeill, D., 1992. Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago, IL.

Nagels, A., Chatterjee, A., Kircher, T., Straube, B., 2013. The role of semantic abstractness and perceptual category in processing speech accompanied by gestures. Front. Behav. Neurosci. 7, 181.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25 (3), 653–660.

Nichols, T., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat. Methods Med. Res. 12 (5), 419–446.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.

Özyürek, A., 2014. Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philos. Trans. R. Soc. B: Biol. Sci. 369 (1651), 20130296.

Özyürek, A., Willems, R., Kita, S., Hagoort, P., 2007. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. J. Cogn. Neurosci. 19 (4), 605–616.

Peeters, D., Chu, M., Holler, J., Hagoort, P., Özyürek, A., 2015a. Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. J. Cogn. Neurosci. 27 (12), 2352–2368.

Peeters, D., Hagoort, P., Özyürek, A., 2015b. Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. Cognition 136, 64–84.

Perniss, P., Vigliocco, G., 2014. The bridge of iconicity: from a world of experience to the experience of language. Philos. Trans. R. Soc. B: Biol. Sci. 369 (1651), 20130300.

Pierno, A.C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., Castiello, U., 2009. Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. Cereb. Cortex 19 (2), 367–374.

Poser, B.A., Versluis, M.J., Hoogduin, J.M., Norris, D.G., 2006. BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. Magn. Reson. Med. 55 (6), 1227–1235.

Redcay, E., Ludlum, R.S., Velnoskey, K.R., Kanwal, S., 2015. Communicative Signals Promote Object Recognition Memory and Modulate the Right Posterior STS. J. Cogn. Neurosci. 28 (1), 8–19.

Sato, W., Kochiyama, T., Uono, S., Yoshikawa, S., 2009. Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. NeuroImage 45 (3), 984–992.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34.

Searle, J.R., 1975. Indirect speech acts. In: Cole, P., Morgan, J.L. (Eds.), Syntax and Semantics 3: Speech acts. Academic Press, New York, 59–82.

Skipper, J.I., Goldin-Meadow, S., Nusbaum, H.C., Small, S.L., 2007. Speech-associated gestures, Broca's area, and the human mirror system. Brain Lang. 101 (3), 260–277.

Southgate, V., Van Maanen, C., Csibra, G., 2007. Infant pointing: Communication to cooperate or communication to learn? Child Dev. 78 (3), 735–740.

Straube, B., Green, A., Bromberger, B., Kircher, T., 2011. The differentiation of iconic and metaphoric gestures: common and unique integration processes. Hum. Brain Mapp. 32 (4), 520–533.

Tomasello, M., Carpenter, M., Liszkowski, U., 2007. A new look at infant pointing. Child Dev. 78, 705–722.

Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48 (3), 564–584.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage 30 (4), 1414–1432.

Willems, R.M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. Cereb. Cortex 17 (10), 2322–2333.

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.

Zhu, Z., Hagoort, P., Zhang, J.X., Feng, G., Chen, H.C., Bastiaansen, M., Wang, S., 2012. The anterior left inferior frontal gyrus contributes to semantic unification. NeuroImage 60 (4), 2230–2237.