The Role of Left Inferior Frontal Gyrus in the Integration of Pointing Gestures and Speech

David Peeters [1], *Tineke M. Snijders* [2], *Peter Hagoort* [1, 2], *Aslı Özyürek* [1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Radboud University, Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands

David.Peeters@mpi.nl, Tineke.Snijders@let.ru.nl, Peter.Hagoort@mpi.nl, Asli.Ozyurek@mpi.nl

Abstract

Comprehension of pointing gestures is fundamental to human communication. However, the neural mechanisms that subserve the integration of pointing gestures and speech in visual contexts in comprehension are unclear. Here we present the results of an fMRI study in which participants watched images of an actor pointing at an object while they listened to her referential speech. The use of a mismatch paradigm revealed that the semantic unification of pointing gesture and speech in a triadic context recruits left inferior frontal gyrus. Complementing previous findings, this suggests that left inferior frontal gyrus semantically integrates information across modalities and semiotic domains.

Index Terms: pointing gesture, multimodal integration, reference, fMRI

1. Introduction

Pointing gestures are a fundamental part of human communication [1]. By producing them in everyday life we connect our communication to entities in the world around us [2]. In establishing a triadic link between child, caregiver, and referent, they play a crucial role in language acquisition [3] and impairments in the production and comprehension of pointing gestures are an early marker of the neurodevelopmental disorder autism [4]. From a phylogenetic viewpoint, it has been claimed that (declarative) pointing is a uniquely human form of communication in a natural environment [5].

Previous neuroimaging work investigating the comprehension of index-finger pointing gestures has presented the gestures in a context that lacked both a larger visual triadic context and co-occurring speech [6][7]. However, in everyday human referential communication pointing gestures often occur in a context in which one perceives not only the person pointing but also the referent she points at and the speech she may concomitantly produce. It is currently unclear how in such situations input from different modalities (visual: speaker, pointing gesture, referent; auditory: speech) is integrated in the brain. The lack of empirical neurocognitive research in this domain is surprising, because comprehending and integrating our interlocutors' referential (i.e. deictic) gesture and speech in a visual context is often critical to understand what they are talking about and a core feature of everyday communication [8]. The current study therefore investigates the neural mechanisms underlying the semantic integration of manual pointing gestures with speech in a visual, triadic context.

The majority of studies investigating the neural integration of gestures with co-occurring speech have focused on *iconic* co-speech gestures, i.e. hand movements that visually resemble the meaning of the linguistic part of the utterance they accompany [9]. It is relatively uncontroversial that LIFG, more specifically its pars triangularis, plays a role in the integration of speech and iconic gesture, possibly in interplay with MTG [10]. Willems et al. (2007) were the first to study the integration of speech and gesture using fMRI. In an orthogonal design, the ease of integration of linguistic and gestural information into a preceding sentence context was manipulated [11]. An increase in activation in LIFG was found when words and/or gestures were incongruent ("mismatch conditions") compared to when they were congruent ("match condition") with preceding speech. Such findings confirm LIFGs status as a multimodal integration site that plays a crucial role in the semantic unification of information from different modalities [12]. Such accounts argue, however, that LIFG is a node in a larger network that subserves the integration of gesture and speech, and also attribute a role to STS/STG and MTG in the perception and integration of speech-gesture combinations [10] [13].

As outlined above, in the current study we focus on a different type of gesture, namely (deictic) pointing gestures. Unlike iconic gestures, pointing gestures in exophoric use canonically create a vector towards a referent to shift the gaze of an addressee and establish a joint focus of attention [1]. Furthermore, whereas speech and iconic gestures often allow communicating about entities that are not immediately physically present ("displacement", [14]), pointing gestures in exophoric use play a crucial role in referential communication about entities that speaker and addressee may perceive in the immediate extra-linguistic context of a conversation. Therefore, the integration of speech and pointing gestures towards a referent need not necessarily recruit the same neural and cognitive mechanisms as in the integration of speech with iconic or other types of gestures.

Although it is currently unknown which cortical areas are involved in integrating pointing gestures and speech, a number of studies have looked at the neural correlates of comprehending pointing gestures in isolation and at their integration with other cues such as the gesturer's gaze direction. Sato et al. (2009), for instance, showed that the perception of a (meaningless) pointing hand, compared to a non-directional closed hand, elicits enhanced activation in a network of mainly right-hemisphere regions, including right IFG, right angular gyrus, right parietal lobule, right thalamus, and bilateral lingual gyri [7]. Materna et al. (2008) suggest that bilateral posterior STS is involved in following the direction of a pointing finger [6]. Conty et al. (2012) show that integration of pointing gestures and gaze direction in comprehension recruits parietal and supplementary motor cortices in the right hemisphere [15]. All in all, these findings suggest an extensive right-hemisphere dominant network that is activated when one perceives a manual pointing gesture that shifts one's attention.

Finally, Pierno et al. (2009) compared the observation of a static image of a hand pointing at an object to the observation of a hand grasping an object and to a control condition of a

hand resting next to an object [16]. Compared to the control condition, the perception of the pointing hand and object elicited enhanced activation in left MTG, left parietal areas (postcentral gyrus and supramarginal gyrus) and left middle occipital gyrus. However, the pointing condition did not recruit significant differential activity compared to the grasping condition. Nevertheless these results suggest that, in addition to the right-lateralized network involved in perceiving a pointing hand, a left-lateralized set of cortical areas may be involved in visually integrating a pointing hand and an object.

1.1. The present study

In the present study, we investigated which cortical regions subserve the integration of pointing gestures with speech in a visual, everyday context. In an event-related functional magnetic resonance imaging (fMRI) study, participants were presented with images of a speaker who pointed at one of two different objects as they listened to her speech. We employed a mismatch paradigm, such that speech either referred to the object the speaker pointed at or to the other visible object. As such, speech and gesture were individually always correct, but there was congruence or incongruence when semantically integrated in the larger visual context. Thus, the match-mismatch comparison taps into the semantic integration/unification of pointing gestures and speech. Mismatch paradigms have been successfully used in the past to study the integration of iconic gestures and speech [13].

Because this is the first study investigating the neuronal integration of pointing gestures with speech in comprehension, predictions were derived on the basis of previous speech-gesture integration studies that used *iconic* gestures in their stimulus materials. If LIFG plays a key role in the semantic integration of gesture and speech [10] [13], it should show enhanced activation in the mismatch compared to the match condition. This is in line with a view of LIFG as a modality-independent multimodal integration site, with its pars triangularis specifically involved in semantic unification of information from different input streams [11] [12]. Conversely, if multimodal semantic integration of gesture and speech recruits the posterior part of the STS region [17], then this region should show enhanced activation in the mismatch-match comparison.

Finally, we included two conditions in which one of the two objects in the images was highlighted by an attentional cue in the absence of gesture. This allowed investigating whether the possible role of LIFG in semantic unification of speech and pointing gesture in a triadic context was dependent on the perceived communicative intentions of the gesturer. Research by Kelly and colleagues suggests that speech-gesture integration differs from the integration of gestures with actions more broadly because the former are generally viewed as more intended to accompany the speech signal compared to the latter [18]. Pointing gestures are shaped by the communicative intentions of the gesturer [19], and in that sense differ from other cues in the environment that may shift our attention. Therefore the integration of pointing gestures with speech may differ from the integration of other attentional cues with concurrently perceived speech. In sum, the current study thus aims to shed more light on the functional roles of different cortical areas involved in speech-gesture integration by investigating the integration of speech with a novel type of gesture, namely index-finger pointing.

2. Method

2.1. Participants

Twenty-three right-handed 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. Stimuli 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, [9]) at one of two objects presented at a table in front of her (henceforth "target 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 "attentional pictures"). 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 had an average duration of 837 ms (SD = 155 ms). In half of the target pictures the speaker pointed at the object at her left and in the other half of the target pictures she pointed at the object at her right. Similarly, in half of the attentional 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 [20] 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.

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 the Bimodal Match (BM) condition, the spoken stimulus matched the object the speaker pointed at. In the Bimodal Mismatch (BMM) condition, the spoken stimulus did not match the object she pointed at but the other object. In the Attentional Match (AM) condition, the spoken stimulus matched the framed object. In the Attentional Mismatch (AMM) condition, the spoken stimulus matched the object that was not framed. Each condition consisted of 40 speech-picture pairs. The speech-only block and the picture-only block were included for a bimodal enhancement analysis that will be reported elsewhere. Figure 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 2-6s 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 pre-

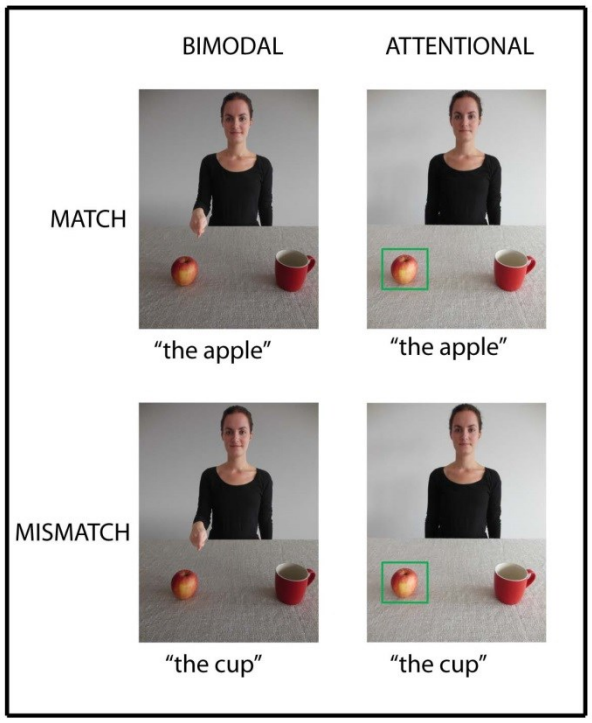

Figure 1: *Overview of the experimental conditions.*

sented during this block. A trial in this block consisted of a fixation cross presented for a jittered duration of 2-6s followed by the presentation of the picture for 2s. The *mixed block* consisted of 160 target trials in which a fixation cross (jittered duration of 2-6s) was followed by the presentation of a picture (for 2s) 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 attentional pictures the object at the speaker's left was framed and in the other half of the attentional pictures the object at her right.

Pictures were presented on the screen using *Presentation* software (Neurobehavioral Systems) and speech was presented through nonmagnetic 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 attentional 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. 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. 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 thus served as catch trials eliciting a button press and were excluded from further MRI analyses. The experiment was preceded by a practice session.

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 multiecho echo-planar imaging sequence, 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 = 0.51 ms; flip angle = 90 °]. This procedure broadens T2* coverage and improves T2* estimation. Each volume consisted of 36 slices of 3 mm thickness [ascending slice acquisition; voxel size = 3.3 x 3.3 x 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 x 1 x 1 mm) accelerated with GRAPPA parallel imaging.

2.5. fMRI 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, after which standard pre-processing 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)].

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.

A whole-brain analysis was 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)]. Two analyses were performed to compare semantic mismatch to semantic congruency. First, the bimodal mismatch condition was compared to the bimodal match condition (BMM > BM). Second, the attentional mismatch condition was compared to the attentional match condition (AMM > AM).

Whole-brain correction for multiple comparisons was applied by combining a significance level of $p = 0.001$ (uncorrected at the voxel level) with a cluster extent threshold using the theory of Gaussian random fields. All clusters are reported at an alpha level of $p < 0.05$ family-wise error (FWE) corrected across the whole brain.

We had the a priori hypothesis that LIFG would be recruited more in the BMM condition compared to the BM condition as this comparison arguably taps into semantic integration/unification of speech and gesture. However, it is unclear whether such a potential involvement of LIFG is specific to communicatively intended gestures and speech or, instead, generalizes to any semantic speech-referent relation as induced by an attentional cue (i.e. it would also show up in the AMM-AM comparison). Therefore, a region-of-interest (ROI)

analysis was performed in LIFG. The ROI was an 8 mm sphere around centre voxels in LIFG taken from a meta-analysis on a large number of neuroimaging studies of semantic processing [13][21]. MNI coordinates were [-42 19 14]. Contrast estimates were calculated for each participant at first-level for the four conditions (AM, AMM, BM, BMM) 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 analysis

We first compared the mismatch conditions to the match conditions at whole-brain level. Contrasting BMM with BM showed increased activations in left inferior frontal gyrus (Fig. 2 and Table 1). The reverse contrast (BM > BMM) did not show any significant cluster that survived the statistical threshold. Also contrasting AMM with AM did not show any areas that survived the statistical threshold (Table 1).

Figure 2: *Results from the whole brain analysis comparing Bimodal Mismatch (BMM) > Bimodal Match (BM). Results are displayed at p < .05, family-wise error corrected at the cluster-level.*

3.3. ROI analysis

An ROI analysis was performed comparing mismatch to match conditions in the predefined ROI (8 mm sphere around MNI coordinates -42 19 14) in LIFG. The interaction between cue (pointing gesture / attentional cue) and congruency (match / mismatch) failed to reach significance, F (1,22) = 2.10, p = .162. However, dependent samples t-tests revealed that there was enhanced activation in LIFG in mismatch vs. match conditions when the speaker's pointing gesture indicated the referent object, $t(22) = -2.43$, $p = .024$. There was no difference in activation in the ROI between the attentional mismatch and match conditions, $t(22) = .48$, $p = .637$. Figure 3 presents the contrast estimates for the four conditions.

4. Discussion

The present study investigated the neural integration of pointing gestures and speech in a visual, triadic context in comprehension. A mismatch analysis revealed that LIFG was sensitive to the congruence between speech and a concurrently presented pointing gesture towards a referent, whereas the posterior STS region was not.

Enhanced activation in LIFG has been found in previous studies that investigated the integration of iconic gestures with speech [10][11][13], pantomimes with speech [13], and metaphoric gestures with speech [22]. The common

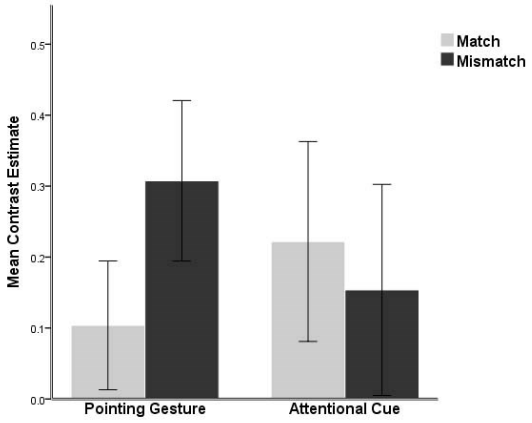

Figure 3: *ROI results. Mean contrast estimates for AM, AMM, BM, and BMM. Error bars represent standard errors around the mean.*

denominator in these studies is that an increase in semantic unification load led to an increase in LIFG activation [10]. For instance, gestures that are unrelated to concurrently presented speech require additional semantic processing because they are harder to semantically integrate with speech compared to iconic gestures that relate to the concurrently presented speech. Therefore, the former lead to enhanced LIFG activation compared to the latter [23]. The same holds for metaphoric co-speech gestures compared to iconic co-speech gestures [22]. Similarly, iconic gestures or pantomimes that are incongruent with speech activate LIFG more than iconic gestures and pantomimes that match the speech they accompany [11][13]. Confirming such previous findings, in the current study incongruence between speech and a visible object, as induced by a pointing gesture, led to enhanced activation in LIFG compared to a matched congruent condition.

Previous studies have criticized the use of mismatch paradigms in gesture-speech integration studies, for instance arguing that "mismatches, which are rarely encountered in spontaneous discourse, may trigger additional integration processes which are not normally part of multimodal language comprehension"[17, p. 876], such that activations in LIFG may be a result of "the processing of unnatural stimuli and rather relate to error detection processes" [23, p. 3317]. There are convincing reasons to believe, however, that gesture-speech mismatch manipulations tap into semantic speech-gesture integration. For instance, LIFG activation is often also present in the "match" condition compared to baseline [13]. Furthermore, enhanced LIFG activation has also been found in speech-gesture integration studies that manipulated semantic load in a different way, not using a mismatch paradigm [10][24]. Dick et al. (2014), for instance, compared the integration of supplemental iconic gestures with speech to the integration 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) and therefore increased semantic processing and unification load compared to the latter gestures ("Sparky pecked" combined with a "peck" gesture). Indeed, a robust increase in activation was found in LIFG for the gestures that added information to the speech and therefore required additional semantic processing compared to the "redundant" gestures [10]. Crucially, both such gestures commonly occur in everyday interactions [9][25].

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	*t*-value	MNI coordinates			Region/Peak
BMM - BM	.01	220	4.01	-46	20	20	LIFG (pars triangularis)
			3.72	-36	18	20	
			3.69	-50	28	18	
AMM - AM	-			-	-	-	

Abbreviations: AM, Attentional Match; AMM, Attentional Mismatch; BM, Bimodal Match; BMM, Bimodal Mismatch; k, extent (voxels).

LIFG plays 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 [26]. The current study extends previous work in showing that semantic unification recruits LIFG across semiotic domains. LIFG thus plays a crucial 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 iconic gestures) is that is allows for displacement, i.e. the ability to refer to entities that are not immediately present [14]. The current study shows that also when a referent is physically present in the immediate visual context, LIFG subserves the semantic unification of auditory and visual information at a higher-order semantic level. The involvement of LIFG in the case of pointing-speech integration may be dependent on whether transmitted information is semantic and/or communicatively intended, as it was not sensitive to the congruence between speech and an attentional cue around a visual object.

Finally, previous studies investigating the neural mechanisms involved in the perception of pointing gestures have focused on the gesture as a directional cue outside a speech context. Pierno et al. (2009), for instance, compared the observation of a static image of a hand pointing at an object to the observation of a hand grasping that object and to a control condition of a hand resting next to that object. Compared to the control condition, both types of actions activated a left-lateralized network that included parietal areas (postcentral gyrus and supramarginal gyrus) and left middle occipital gyrus [16]. Here we find that, when pointing gestures are produced with speech, LIFG is recruited and may be part of a larger network that comprises the areas found by Pierno et al. (2009). Furthermore, in that study no area was activated significantly more in the pointing condition compared to the grasping condition. Future work may therefore investigate whether the results of the current study generalize to situations in which a speaker grasps an object while concurrently producing speech. After all, in everyday life speakers may both point at an object and grasp and hold up or place an object to bring it into their addressee's attention [2]. It is not unlikely that the extent of overlap between pointing-speech integration and grasping-speech integration might differ as a function of the perceived communicative intentions of the speaker (see [18]).

5. Conclusion

In sum, the current study investigated the neural integration of pointing gestures and speech in a visual, triadic context. We found that LIFG subserved the semantic unification of referential gesture and speech in a triadic context. This study can be informative as a starting point for studies investigating specific populations with impairments in the comprehension and integration of deictic speech and gesture and the subsequent establishment of joint attention in everyday life, as in autism spectrum disorders.

6. Acknowledgments

We would like to thank Paul Gaalman for assistance during the scanning sessions and Doris Deckers for help in creation of the stimuli.

7. References

[1] Kita, S. "Pointing. Where language, culture, and cognition meet", Hillsdale, NJ: Erlbaum, 2003.
[2] Clark, H. H., "Pointing and placing", In S. Kita [Ed], Pointing. Where language, culture, and cognition meet, 243-268, Hillsdale, NJ: Erlbaum, 2003.
[3] Carpenter, M., Nagell, K. and Tomasello, M., "Social cognition, joint attention, and communicative competence from 9 to 15 months of age", Monographs of the Society for Research in Child Development, 255: Vol. 63, 1-174, 1998.
[4] Baron-Cohen, S., "Perceptual role taking and protodeclarative pointing in autism", British Journal of Developmental Psychology, 7(2): 113-127, 1989.
[5] Call, J. and Tomasello, M., "Production and comprehension of referential pointing by orangutans (Pongo pygmaeus)", Journal of Comparative Psychology, 108(4): 307-317, 1994.
[6] Materna, S., Dicke, P. W. and Thier, P., "The posterior superior temporal sulcus is involved in social communication not specific for the eyes", Neuropsychologia, 46(11): 2759-2765, 2008.
[7] Sato, W., Kochiyama, T., Uono, S. and Yoshikawa, S., "Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols", NeuroImage, 45(3): 984-992, 2009.
[8] Bühler, K., "Sprachtheorie", Jena: Fischer, 1934.
[9] Kendon, A., "Gesture: Visible action as utterance", Cambridge: Cambridge University Press, 2004.
[10] Dick, A. S., Mok, E. H., Beharelle, A. R., Goldin-Meadow, S. and Small, S. L., "Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech" Human brain mapping, 35: 900-917, 2014.

[11] Willems, R. M., Özyürek, A. and Hagoort, P., "When language meets action: the neural integration of gesture and speech", Cerebral Cortex, 17(10): 2322-2333, 2007.
[12] Hagoort, P., "On Broca, brain, and binding: a new framework", Trends in cognitive sciences, 9(9): 416-423, 2005.
[13] Willems, R. M., Özyürek, A. and Hagoort, P., "Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language", Neuroimage, 47(4): 1992-2004, 2009.
[14] Hockett, C. D., "The origin of speech", Scientific American, 203(3): 88-96, 1960.
[15] Conty, L., Dezecache, G., Hugueville, L. and Grèzes, J, "Early binding of gaze, gesture, and emotion: neural time course and correlates", The Journal of Neuroscience, 32(13): 4531-4539, 2012.
[16] Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P. and Castiello, U., "Neurofunctional modulation of brain regions by the observation of pointing and grasping actions", Cerebral Cortex, 19(2): 367-374, 2009.
[17] Holle, H., Obleser, J., Rueschemeyer, S. A. and Gunter, T. C., "Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions", Neuroimage, 49(1): 875-884, 2010.
[18] Kelly, S., Healey, M., Özyürek, A. and Holler, J., "The processing of speech, gesture, and action during language comprehension", Psychonomic bulletin & review: 1-7, 2014.
[19] Peeters, D., Chu, M., Holler, J., Özyürek, A. and Hagoort, P., "Getting to the point: The influence of communicative intent on the kinematics of pointing gestures", In M. Knauff, M. Pauen, N. Sebanz and I. Wachsmuth [Eds], Proceedings of the 35th Annual Meeting of the Cognitive Science Society, 1127-1132, Austin, TX : Cognitive Science Society, 2013.
[20] Peeters, D., Hagoort, P. and Özyürek, A., "Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives", Cognition, 136: 64-84, 2015.
[21] Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B. and Tzourio-Mazoyer, N., "Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing", Neuroimage, 30(4): 1414-1432, 2006.
[22] Straube, B., Green, A., Bromberger, B. and Kircher, T., "The differentiation of iconic and metaphoric gestures: Common and unique integration processes", Human brain mapping, 32(4): 520-533, 2011.
[23] Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K. and Kircher, T., "Neural integration of iconic and unrelated coverbal gestures: a functional MRI study", Human brain mapping, 30(10): 3309-3324, 2009.
[24] Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C. and Small, S. L., "Speech-associated gestures, Broca's area, and the human mirror system", Brain and language, 101(3): 260-277, 2007.
[25] Holler, J. and Beattie, G., "Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?", Gesture, 3(2): 127-154, 2003.
[26] Hagoort, P., Hald, L., Bastiaansen, M. and Petersson, K. M., "Integration of word meaning and world knowledge in language comprehension", Science, 304(5669): 438-441, 2004.