Journal of Neurolinguistics 30 (2014) 69–77

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

# The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies

Lars Marstaller [a,b,c,*], Hana Burianová [b,c]

[a] Department of Cognitive Science, Macquarie University, Sydney, Australia
[b] ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney, Australia
[c] Centre for Advanced Imaging, University of Queensland, Brisbane, Australia

A R T I C L E I N F O

*Article history:*
Received 21 January 2014
Received in revised form 21 March 2014
Accepted 7 April 2014

*Keywords:*
Co-speech gestures
Multisensory perception
Meta-analysis
Action understanding

A B S T R A C T

Co-speech gestures constitute a unique form of multimodal communication because here the hand movements are temporally synchronized and semantically integrated with speech. Recent neuroimaging studies indicate that the perception of co-speech gestures might engage a core set of frontal, temporal, and parietal areas. However, no study has compared the neural processes during perception of different types of co-speech gestures, such as beat, deictic, iconic, and metaphoric co-speech gestures. The purpose of this study was to review the existing literature on the neural correlates of co-speech gesture perception and to test whether different types of co-speech gestures elicit a common pattern of brain activity in the listener. To this purpose, we conducted a meta-analysis of neuroimaging studies, which used different types of co-speech gestures to investigate the perception of multimodal (co-speech gestures) in contrast to unimodal (speech or gestures) stimuli. The results show that co-speech gesture perception consistently engages temporal regions related to auditory and movement perception as well as frontal-parietal regions associated with action understanding. The results of this study suggest that brain regions involved in multisensory processing and action understanding constitute the general core of co-speech gesture perception.



* Corresponding author. Department of Cognitive Science, Australian Hearing Hub, 16 University Avenue, Macquarie University, NSW 2109, Australia.

*E-mail address:* l.marstaller@uq.edu.au (L. Marstaller).

http://dx.doi.org/10.1016/j.jneuroling.2014.04.003
0911-6044/

## 1. Introduction

When we speak, we typically gesture. Co-speech gestures (CSGs) are hand movements that accompany speech and allow the speaker to effectively communicate thoughts and ideas in two separate modalities (for a review, see Hostetter, 2011), i.e., linguistic content in the auditory domain and imagistic content in the visual domain. Neuroimaging studies of the neural correlates of CSG perception have found that during the observation of CSG the brain shows increased activity in areas that are involved in auditory and semantic processing of language. Several studies found that the observation of a person who produces gestures while speaking engages a core set of frontal, parietal, and temporal areas including superior temporal gyrus (STG), middle temporal gyrus (MTG), intraparietal sulcus (IPS), as well as inferior frontal gyrus (IFG) when compared to watching a video of a speaker who does not gesture or a speaker producing gestures without sound (e.g., Green et al., 2009; Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008, 2010; Hubbard, Wilson, Callan, & Dapretto, 2009; Willems, Özyürek, & Hagoort, 2009).

Evidence shows that the perception of CSGs activates STG when compared to unimodal speech or gesture (Hubbard et al., 2009) especially under increased noise conditions (Holle, Obleser, Rueschemeyer, & Gunter, 2010). In addition, there is evidence that the reduced ability to understand CSGs in younger children is related to the level of activity in left STG (Dick, Goldin-Meadow, Solodkin, & Small, 2012). These findings suggest that STG forms an important area for matching speech sounds and gesture movements. Furthermore, studies show that left MTG is activated when CSGs relate to the linguistic meaning metaphorically (Kircher et al., 2009; Straube, Green, Bromberger, & Kircher, 2011) or when gestures provide additional information that is not present in speech, e.g., when the manner of the movement referred to as “falling” in speech is visually expressed in gesture (Dick, Mok, Raja Beharelle, Goldin-Meadow, & Small, 2014). In contrast, activation in left IPS is found for CSGs that are not related to speech, i.e., CSGs where speech audio and gesture video from different trials have been combined (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Green et al., 2009), for gestures whose meaning does not match the meaning of the accompanying speech (Willems, Özyürek, & Hagoort, 2007, 2009), and for CSGs in contrast to self-grooming movements (Holle et al., 2008). Together, these findings suggest that MTG and IPS are crucial for processing the semantic-communicative dimension of CSGs. Finally, evidence shows that activity in left IFG is highly sensitive to the relationship between speech and gestures (Skipper, Goldin-Meadow, Nusbaum, & Small, 2007, 2009) and that IFG is active when the relationship between speech and gesture is semantically complex (Dick, Mok, et al., 2014; Kircher et al., 2009; Straube et al., 2011; Willems et al., 2007, 2009) or when the communicative intention of hand movements is ambiguous (Dick et al., 2009; Green et al., 2009). All in all, the findings from these studies strongly suggest that during the observation of CSGs, frontal and temporal regions are engaged in semantic processing, whereas frontal and parietal areas are associated with assessing whether the movements are communicatively intended or accidentally produced together with speech.

In addition to fMRI studies, electroencephalography has also been used to investigate CSG perception, especially whether viewing of CSGs has modulatory effects on the N400 (Cornejo et al., 2009; Gunter & Bach, 2004; Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Holle & Gunter, 2007; Kelly, Ward, Creigh & Bartolotti, 2007; Obermaier et al., 2011; Wu & Coulson, 2005, 2007a, 2007b). The N400 is an event-related component with increased negativity that peaks approximately 400 ms after a semantic violation is perceived and whose strength is related to contextual expectations, such that a reduction of the N400 indicates a reduction in semantic violation (Lau, Phillips, & Poeppel, 2008). Studies using this approach have found an increased N400 for meaningless vs. meaningful emblem gestures (Gunter & Bach, 2004), incongruent vs. congruent word–gesture pairs (Holle & Gunter, 2007), incongruent vs. congruent metaphorical expression–gesture pairs (Cornejo et al., 2009), incongruent vs. congruent gesture-cartoon pairs (Wu & Coulson, 2005), unimodal (speech) vs. cross-modal (speech and gesture) related word–probe pairs (Wu & Coulson, 2007a), and unrelated vs. related word–gesture pairs (Wu & Coulson, 2007b). The strength of the N400 for CSGs depends on the temporal coordination of speech and gesture (Habets et al., 2011; Obermaier et al., 2011) and on the listener's expectations that the gesture is communicatively intended (Kelly et al., 2007). Taken together, these findings suggest that co-speech gestures shape the semantic context in which the meaning of speech is interpreted.