When gestures catch the eye: The influence of gaze direction on co-speech gesture comprehension in triadic communication

Judith Holler (judith.holler@mpi.nl)[1,2]
Spencer Kelly (skelly@colgate.edu)[3]
Peter Hagoort (peter.hagoort@mpi.nl)[1,4]
Asli Ozyurek (asli.ozyurek@mpi.nl)[1,5]

[1] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
[2] University of Manchester, School of Psychological Sciences, Coupland Building 1, M13 9PL Manchester, UK
[3] Colgate University, Psychology Department, Center for Language and Brain, Oak Drive 13, Hamilton, NY 13346, USA
[4] Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands
[5] Centre for Language Studies, Radboud University, Erasmusplein 1, 6525HT Nijmegen, The Netherlands

Abstract

Co-speech gestures are an integral part of human face-to-face communication, but little is known about how pragmatic factors influence our comprehension of those gestures. The present study investigates how different types of recipients process iconic gestures in a triadic communicative situation. Participants (N = 32) took on the role of one of two recipients in a triad and were presented with 160 video clips of an actor speaking, or speaking and gesturing. Crucially, the actor's eye gaze was manipulated in that she alternated her gaze between the two recipients. Participants thus perceived some messages in the role of addressed recipient and some in the role of unaddressed recipient. In these roles, participants were asked to make judgements concerning the speaker's messages. Their reaction times showed that unaddressed recipients did comprehend speaker's gestures differently to addressees. The findings are discussed with respect to automatic and controlled processes involved in gesture comprehension.

Keywords: co-speech iconic gesture; eye gaze; recipient status; communicative intent; multi-party communication.

Introduction

When we speak, we frequently move our bodies to supplement what we say with co-speech gestures. A large proportion of these gestures are iconic in nature. Importantly, iconic gestures bear a close link with the speech that they accompany on semantic and temporal levels and have therefore been argued to constitute an integral part of human language (McNeill, 1992; Kendon, 2004) and thus of speaker's utterances (i.e., 'composite utterances', Kendon, 2004). While iconic gestures have been shown to fulfill a variety of cognitive functions which appear to benefit the speaker him or herself (e.g., Chawla & Krauss, 1994; Hostetter, Alibali & Kita, 2007), there is a growing body of evidence that their production is also linked to the speaker's communicative intent (Gerwing & Bavelas, 2004; Holler & Stevens, 2007; Kelly, Byrne & Holler, 2011; Özyürek, 2002).

The comprehension of iconic gestures, especially with respect to the attribution of communicative intentions, has been considerably less well researched. What we do know is that iconic gestures successfully communicate semantic information and that recipients integrate this information with that contained in the accompanying speech (e.g., Holle & Gunter, 2007; Holler, Shovelton & Beattie, 2009; Kelly, Barr, Church & Lynch, 1999; Kelly, Kravitz & Hopkins, 2004; Willems, Özyürek & Hagoort, 2007). However, one limitation of studies on the comprehension of gestures is that many of them have presented stimuli in isolation, that is, video clips showing iconic gestures (and sometimes a torso) but, crucially, no head or facial information, and those studies that have included the face have tended to focus on the lips. In face-to-face communication, however, gestures are not only accompanied by speech and mouth movements, but also by a multitude of additional nonverbal social cues. Instead of focusing our attention solely on speech and gesture when listening to someone speaking we are required to divide our cognitive resources in such a way that allows us to take in and combine all of those cues. How we process and comprehend iconic gestures in more situated contexts that are much closer to real life situations therefore remains a wide-open issue.

Of particular interest in this respect is the influence of eye gaze, one of the most powerful nonverbal social cues (Pelphrey & Perlman, 2009; Senju & Johnson, 2009). Eye gaze is not only an omnipresent contextual cue when observing co-speech gestures, it is also inherently linked to the perception of communicative intent (Kampe, Frith & Frith, 2003; Schilbach et al., 2006) and the regulation of social interaction (Argyle & Cook, 1976; Goodwin 1981; Kendon, 1967). This begs the question of how the co-occurrence of gaze and gesture influences recipients' comprehension. The present study addresses this very question.

To do so, it builds on a couple of recent studies that have begun to focus on the issue of perceived communicative intent in conjunction with gesture comprehension. Kelly et al. (2007, 2010) showed that participants integrated co-occurring information from speech and gesture less strongly when the two modalities were perceived as not intentionally coupled (e.g., male hands gesturing accompanied by a female voice speaking) than when they were perceived as intended to form a composite utterance (e.g., male hands gesturing accompanied by a male voice speaking). This is the first empirical evidence that the perceived intentional

stance of a communicator influences recipients' processing of iconic gesture and speech.

However, as many previous studies in this field, these two studies did not present gestures in their natural context but, instead, in isolation of any facial cues (including eye gaze) with the aim to control for the influence of lip movements. In addition, and in line with the predominant gesture comprehension paradigm at the time, both of the studies used mismatching speech-gesture stimuli, that is, stimuli in which the information provided by speech conflicted with that depicted by the accompanying iconic gestures. Whilst this was an ideal test bed for first enquiries into the semantic integration of speech and gesture, it compromises the generalisability of such findings to more natural speech-gesture utterances, something the present study aims to overcome.

Another recent study that has addressed the topic of communicative intent and gesture comprehension was conducted by Straube et al. (2010). In their study, participants watched video clips of a speaker who looked directly at them or who was oriented away from the camera. In contrast to previous studies on co-speech gesture comprehension, Straube et al.'s paradigm did include the speaker's head and eye gaze, and, furthermore, they avoided the use of mismatching gestures. However, the authors manipulated multiple nonverbal social cues simultaneously (body/torso orientation, gesture orientation, as well as gaze direction), preventing us to draw conclusions about the effect of gaze direction specifically on participants' comprehension. In addition, the information depicted by the gestures used as stimuli was redundant with that in speech (e.g., the speaker referred to a 'round bowl' in speech accompanied by a gesture depicting a round, bowl-like shape). It is therefore not possible to identify whether the differences in participants' comprehension (measured in the form of their neural response and memory performance for the stimuli) between the two conditions (frontal/averted) was due to differences in their perception of speech, gesture, or a combination of the two.

The studies by Kelly et al. (2007, 2010) and Straube et al. (2011) are laudable first attempts tapping the issue of communicative intent and gesture comprehension and useful stepping stones for further investigations on this topic. The present study aims to build on this work by investigating the effect of perceived communicative intent, as signaled through the speaker's eye gaze direction, on the comprehension of iconic gestures. Importantly, this study will be presenting gestures in a more natural context (including the head), manipulating social eye gaze as the only social cue of interest, and it will be based on gestures that match the speech but which are complementary in nature. This, in conjunction with the particular experimental paradigm employed in this study, will allow us to tap into the processing of gesture directly, and to zoom into recipients' processing of the verbal and the gestural components of the speaker's messages separately. We will do so by creating a set-up simulating a triadic communicative situation involving one speaker and two recipients, combined with a manipulation of the speaker's eye gaze direction which will indicate to participants when they are an addressed and when they are an unaddressed recipient. Thus, it is the first experimental study looking specifically on the effect of social eye gaze on speech and iconic co-speech gestures.

Method

Participants

Thirty-two female, right-handed German native speakers participated in the experiment (mean age = 21.6yrs) and were compensated with 8€ payment.

Design

The study employed an experimental paradigm simulating multi-party communication involving one speaker-gesturer and two recipients, only one of which was a 'real' participant (the other one was fictive). Participants were made to believe that the other participant taking on the role of the second recipient was located in a different room.

A confederate acted as the speaker-gesturer and produced scripted utterances to a video camera (we used pre-recorded instead of live stimuli to ensure that all participants were presented with identical stimuli). These pre-recorded video clips were presented to participants while making them believe that they were engaging in a live communication with the speaker, who they thought was located in yet a different room to themselves but connected to them via a live camera link. (The second recipient was, allegedly, also connected to the person acting as speaker via a live-camera link, in the same way as they were.)

To prevent participants from realising that the speaker was pre-recorded, they were told that the camera link was a one-way connection in that the two recipients were able to hear and see the speaker but that the speaker was not able to hear or see them (and the two recipients were, of course, not able to see or hear each other). They were also told that the speaker had been asked to stand in front of two different cameras, that she knew of the two recipients' presence, and that she had been told that each camera was hooked up to one of the recipients' computer monitors, allowing them to hear and see what she was communicating. In order to create a more plausible situation and convince subjects that the speaker did not memorise the entire set of scripted sentences by heart, the video clips showed the speaker looking down before each sentence was spoken; participants were told that a laptop had been positioned on a table in front of the speaker displaying a black and white drawing accompanied by a couple of words before each trial, and that the speaker had been instructed to communicate the contents of the information displayed on the screen spontaneously and in a way that felt natural to them (no explicit mention of gesture was made). They (the actual participants) were then informed that the speaker would sometimes address them by looking into the camera linked

to their own on monitor, and sometimes the other, second recipient, by looking into the respective other camera. This created two different views for the actual participant, one in which they were directly gazed at, and one in which they observed the speaker's gaze being averted (see Fig 1). Gaze direction, as implemented through this manipulation, constituted our main IV (within-participants). In addition, as a second IV (modality), we manipulated the occurrence of gesture in association with the sentences spoken by the speaker in the video clips (within-participants).

Stimuli

The experiment set-up described in the preceding section required the creation of four types of video stimuli: a) Direct gaze (speech only), b) Averted gaze (speech only), c) Direct gaze (speech + gesture), d) Averted gaze (speech + gesture) (Fig. 1). In each video vignette, the actor spoke a short sentence (canonical SVO structure), e.g. 'she goes through the list' ('*sie geht durch die Liste*'). Crucially, the verb included in the sentence was always manner *un*specific, i.e., 'to go through' (*durchgehen*). The iconic gestures accompanying these verbs always specified the manner of action, e.g., to tick items on the list (*abhaken*). This manipulation allowed us to measure participants' comprehension of the gestures independently of speech, without using mismatching gestures (see Introduction). Participants watched the videos on a computer screen in a soundproof experimental test booth; the audio signal was presented via closed-back headphones. Materials were presented with Presentation® software (www.neurobs.com).

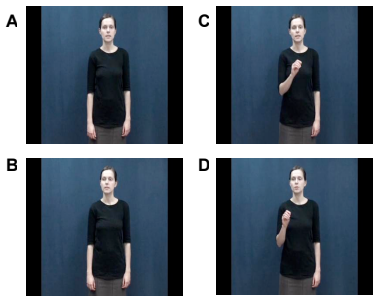

Figure 1: Examples of the four types of video stimuli used, A = Direct gaze (S only), B = Averted gaze (S only), C = Direct gaze (S + G), D = Averted gaze (S + G).

Procedure

First, participants completed six practice trials (showing a different actor). Before the start of the experiment proper, the experimenter made a fake phone call to check whether 'the other participants' were ready to start.

Participants then watched 160 videos (40 stimuli per condition). Each video clip was followed by a written word (presented in capital letters, centre of screen) that matched either the verb contained in the preceding spoken sentence (*speech-related targets* [20 items per condition]; designed to tap into the processing of the verbal component of the speaker's message) or the content of the gesture performed by the speaker in the video (*gesture-related targets* [20 items per condition]; designed to tap into the processing of the gestural component of the speaker's message).

Task

Participants were asked to judge *"whether the word displayed on the screen had been mentioned by the speaker in the preceding video"*, thus requiring 'yes' answers for all speech-related targets, and 'no' answers for all gesture-related targets. Reactions times (RTs) to participants' yes/no answers (delivered via a button box; yes = dominant hand) as well as errors[1] were recorded.

Results

RTs for gesture and speech-related targets were entered into two separate 2 (gaze: direct vs. averted) x 2 (modality: speech only vs. speech+gesture) repeated measures ANOVA, excluding errors (constituting 2% of the total number of trials) and outliers (2 SD).

Speech-related targets

Our first comparison concerned participants' responses to the speech-related targets (e.g., 'to go through' (*durchgehen*)) designed to tap primarily into the processing of the verbal component of the speaker's composite utterances. This analysis revealed a significant main effect of modality (p = .0001), with slower response times in the speech + gesture conditions than in the speech only conditions. The main effect of gaze was not significant (p = .090), and neither was the interaction between gaze and modality (p = .870).

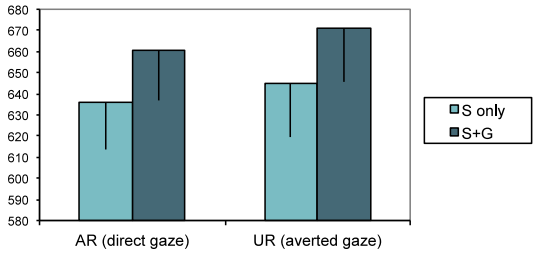

Figure 2: Addressed and unaddressed recipients' (AR/UR) RTs (ms) in the speech-only and speech + gesture conditions for speech-related targets (error bars = *SE*).

Gesture-related targets

Our main comparison focused on participants' responses to the gesture-related targets (e.g., 'to tick' (*abhaken*)) intended to tap primarily into the processing of the gestural

[1] Due to restrictions on space, we only report our RT results here. Note, however, that the error rate analysis revealed very few significant differences, and those that did emerge did not relate to the relevant differences in RTs in a meaningful way.

component of the speaker's composite utterances. This analysis revealed no main effect of modality (p = .216) and no main effect of gaze (p = .087). However, the interaction between gaze and modality was significant (p = .045). Independently from the omnibus interaction effect, we carried out two a priori contrasts (UR S+G vs. AR S+G; UR S-only vs. AR S-only). These showed that the interaction is driven by unaddressed recipients taking significantly longer to respond in the S+G condition than addressed recipients in the S+G condition (p = .026). The comparison of unaddressed and addressed recipients' RTs in the speech-only conditions was not significant.

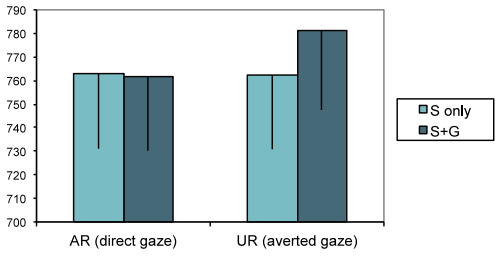

Figure 3: Addressed and unaddressed recipients' (AR/UR) RTs (ms) in the speech-only and speech + gesture conditions for gesture-related targets (error bars = *SE*).

Discussion

The present study has investigated co-speech gesture comprehension in the presence of eye gaze, a powerful social cue integral to human face-to-face communication. Our findings reveal that recipients' gesture comprehension is indeed influenced by the speaker's eye gaze direction. More specifically, we have shown that, when a speaker's eye gaze is used to signal communicative intent in the sense of address, recipients who are currently unaddressed (but ratified participants in a communication, Goffmann, 1981) do process speech-accompanying iconic gestures differently to addressed recipients. This finding advances our understanding of human communication by pointing to an important way in which pragmatic processes shape and influence the comprehension of co-speech gestures, which, to date, has been addressed by only a very small number of studies (Kelly et al., 1999, 2007, 2010; Straube et al., 2011).

There are at least two competing interpretations of our effects of eye gaze direction on gesture comprehension. One is what we have termed the 'Fuzzy Representation Hypothesis'. According to this hypothesis, due to unaddressed recipients perceiving the speaker's gestures as not intended for them (but for the other, gazed at recipient instead), they interpret the gesture to a lesser degree. As a consequence, they take longer to respond to the gesture-related targets because they have constructed merely a partial, or fuzzy, mental representation of the gestural meaning. They are aware that something relating to the meaning of the word displayed on the screen may have been presented to them, but they have a hard time making a quick decision on the modality in which this information was presented (since the gestural component of their mental representation is 'incomplete' or 'fuzzy'). Further, we would argue that the underlying mechanism leading to this fuzzy representation would not simply be one of reduced attention. The reason for this is that, first, addressed and unaddressed recipients do not differ in the number of errors they made, and, second, the modality effect we found for the speech-related targets is as strong for unaddressed as for addressed recipients. Thus, our data suggest that unaddressed recipients do process the gestural information but less strongly so, and that the reason for this is a modulation of perceived communicative intent rather than a pure decrease in attention.

An alternative possibility is what we have called the 'Competing Modalities Hypothesis'. The rationale underlying this account is that, while addressed recipients (in both dyadic and multi-party interactions) are expected to engage in mutual gaze with a speaker (Argyle & Dean, 1976; Kendon, 1967), unaddressed recipients are free to disengage from the process of gazing at the speaker. This means that the default situation requires recipients who are directly addressed through a speaker's gaze to split their attention between information coming from multiple modalities including speech, gesture, and gaze (and additional facial cues). Unaddressed recipients, on the other hand, have fewer visual social cues to process since the speaker's gaze is averted from them. They may therefore zoom in on gesture, instead of processing gesture and gaze simultaneously, and may thus have more cognitive resources available to focus on the processing of gesture. As a consequence, in the current paradigm, unaddressed recipients are taking longer than addressed recipients to respond to the gesture-related targets (requiring a 'no' answer) because the gestural component of the speaker's utterance constitutes a more prominent component of their mental representation of the event described (since they focused on gesture more). To declare something as not having been mentioned by the speaker despite the stronger memory trace of the gesture being at the forefront of their mind appears to be a difficult task.

The present study was a fruitful undertaking as it has shown that recipient status can influence gesture processing, but also because it allowed us to formulate two possible accounts of a potential process model explaining those effects. Currently, we are unable to unequivocally declare one of them as the more appropriate one, but further studies are currently underway tackling this issue (this on-going research will also add further insights into participants' visual fixations in the two recipient roles, and it tests recipients' gesture comprehension avoiding the suppression of gestural information/no responses). That said, we believe that the Competing Modalities Hypothesis provides the more intuitive account, and some additional data we have collected speak to this preliminary conclusion, too: as a follow-up analysis, we obtained ratings for all of our stimuli from an independent set of participants which gave us an

insight into the degree of ambiguity/clarity of the individual gesture stimuli in the absence of speech. If the Fuzzy Representation Hypothesis should hold, gestures that are more ambiguous in the absence of speech (i.e., less pantomimic) should cause unaddressed recipients particular problems of interpretation (since they require more interpretation effort and more integration work). Unaddressed recipients should therefore have taken especially long to respond to those iconic gestures. However, when we correlated unaddressed recipients' RTs with the ambiguity ratings of the individual gestures, no relationship of this sort was found. One could argue, of course, that more pantomimic gestures should slow unaddressed recipients down also if we assume that the Competing Modalities Hypothesis holds, since they might 'stick' particularly well in the participant's mind; that is, the gestural components of utterances accompanied by more pantomimic gestures might become particularly prominent parts of unaddressed recipients' mental representations. However, this argument only holds if we assume that unaddressed recipients also integrate the verbal and the gestural information less than addressed recipients. According to the Fuzzy Representation Hypothesis, this would be the case, since processing the gestural information less well will also affect the integration of this information with speech. According to the Competing Modalities Hypothesis, however, unaddressed recipients may integrate speech and gesture to the same extent as addressed recipients, with the addition that they process the gestural information more strongly than them. Therefore, our favoured interpretation is the Competing Modalities Hypothesis, but further research is needed before we can draw firm conclusions.

Automatic and controlled processes in gesture comprehension

With regard to the difference in response patterns for the speech-related and gesture-related targets, our results also relate to the distinction between automatic and controlled processes in co-speech gesture comprehension (Kelly et al., 2010). Bear in mind that these two sets of targets were designed to tap the processing of the speaker's information in different ways. What is striking is that, in the case of speech-related targets - a comparatively easy task - we observed a clear modality effect, as has been demonstrated by previous comprehension studies. The semantic integration of gesture and speech has been argued to be automatic in the sense of a low-level, fast and obligatory process (Kelly et al., 2010; see also Kelly, Özyürek & Maris, 2010). This explanation would account for the intrusion of the gestural information (longer RTs in the speech + gesture conditions) despite participants here having been asked to judge the content of speech only. At the same time, the results show a lack of an effect of our gaze manipulation, indicating that this task (saying 'yes' in response to a visually presented word that was presented auditorily immediately prior to this) might have been so easily and quickly accomplished that higher-order processes involved in the processing of pragmatic information, and judgements of speaker-intentions in particular, may not have come into play.

Responses to the gesture-related targets tell a different story, however. Here, participants were slower in general, indicating that this task might have been perceived as more difficult (i.e., saying 'no' to indicate that a certain meaning had *not* been mentioned by the speaker, despite the meaning of the word displayed on screen being related to the meaning in the speaker's gesture). This seems plausible since participants here had to consult their mental representation more carefully by actively teasing apart what they heard and what they saw to arrive at a decision. In order to answer accurately, they were essentially required to suppress the gestural information they had received. This contrasts with the speech-target situation where intrusion of the gestural information may have slowed participants down since there was more information to process, but the gestural information did not interfere as such; after all, 'ticking' items on a list is part of the event of 'going through' a list. To answer 'yes', which participants were required to do for speech-related targets, is still correct, even if the gestural information is taken into account. Answering 'no' to the gesture-related targets involved a very different process, as it required the temporary suppression of the gestural information. Consequently, slower and more difficult information processing may have led to the involvement of more controlled, higher-order cognitive operations, which do take into consideration the intentional stance of a speaker.

Effects of eye gaze direction on speech processing

Our results reveal that speech comprehension was not different for addressed and unaddressed recipients[2]. One possibility therefore is that gesture comprehension processes are more sensitive to perceived communicative intent as signalled through a speaker's gaze than the processing of speech is. However, we have to remain cautious with drawing such a conclusion from the present data. This is because the current paradigm required participants to judge speech (i.e., whether a certain word had been *mentioned* in the preceding clip). As a consequence, participants will have devoted much attention to the processing of speech, irrespective of their recipient status, since they were always required to respond. In other words, a paradigm that does not ask participants to explicitly devote attention to the verbal modality might reveal a modulation of eye gaze direction for the processing of speech only utterances also. In fact, based upon the Competing Modalities Hypothesis,

[2] The lack of an effect of recipient status on the processing of speech stands, at first sight, in slight contrast to a study by Schober & Clark (1989) who found overhearers to be slower and less accurate in their understanding of verbal references than addressees. However, in their study, overhearers were not official recipients of the communication, which distinguishes them from unaddressed recipients (Goffman, 1981).

one might expect to see such a modulation, since unaddressed recipients have more cognitive resources available that they are able to devote to the processing of speech. The present study was designed to mainly measure the processing of gesture, and future research is needed to provide more conclusive answers to the question of how speakers' eye gaze direction influences the comprehension of speech.

Conclusion

In sum, our study suggests that recipients keep an eye on where speakers are looking, and this subtle piece of information has a significant impact on the extent to which co-speech gestures are processed.

Acknowledgments

We thank four anonymous reviewers for their helpful feedback on an earlier version of this paper. We also thank the European Commission for funding this research (Marie Curie Fellowship, EU Project number: 255569), Manuela Schuetze and Nick Wood for extensive help with creating the experimental materials, Ronald Fischer and Idil Kokal for their help with programming, the participants who took part in our study, as well as the NBL and GSL lab groups, Ivan Toni, Natalie Sebanz and Guenther Knoblich, for valuable feedback in discussions of the design of our study.

References

Argyle, M. and Cook, M. (1976). *Gaze and mutual gaze.* Cambridge University Press.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30,* 580-601.

Gerwing, J. & Bavelas, J.B. (2004). Linguistic influences on gesture's form. *Gesture, 4,* 157–195.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience, 19,* 1175-1192.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic gestures really contribute to the semantic information communicated in face-to-face interaction? *Journal of Nonverbal Behavior, 33,* 73-88.

Holler, J., & Stevens, R. (2007). An experimental investigation into the effect of common ground on how speakers use gesture and speech to represent size information in referential communication. *Journal of Language and Social Psychology, 26,* 4-27.

Hostetter, A. B., Alibali, W. M., & Kita, S. (2007). I see it in my hand's eye: Representational gestures are sensitive to conceptual demands. *Language and Cognitive Processes, 22,* 313-336.

Kampe, K., Frith, C.D., & Frith, U. (2003). "Hey John": Signals conveying communicative intention towards the self activate brain regions associated with mentalising regardless of modality. *Journal of Neuroscience, 23,* 5258-5263.

Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40,* 577-592.

Kelly, S. D., Byrne, K., & Holler, J. (2011). Raising the ante of communication: Evidence for enhanced gesture use in high stakes situations. *Information, 2,* 579-593.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89,* 253-260.

Kelly, S. D., Özyürek , A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21,* 260-267.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language, 101,* 222-233.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica, 26,* 22-63.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language, 46,* 688–704.

Pelphrey, K. A., & Perlman, S. B. (2009). Charting brain mechanisms for the development of social cognition. In J. M. Rumsey & M. Ernst (Eds.), Neuroimaging in Developmental Clinical Neuroscience. Cambridge University Press.

Schilbach, L, Wohlschläger, AM, Newen, A, Krämer, N, Shah, NJ, Fink, GR, Vogeley, K (2006). Being With Others: Neural Correlates of Social Interaction. *Neuropsychologia, 44,* 718-30.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21,* 211-232.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures orchestrate brain networks for language understanding. *Current Biology, 19,* 661-667.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences, 13,* 127-134.

Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia, 48,* 382–393.

Willems, R. M., Ozyurek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17,* 2322-2333.