Speaker-gaze Modulates the Inter-personal Repetition of Hand Gestures

Lisette Mol (l.mol@tilburguniversity.edu)[1,2]
[1]Tilburg center for Cognition and Communication (TiCC), Tilburg University
P.O. Box 90135, NL-5000 LE Tilburg, The Netherlands
[2]Cognitive Interaction Technology - Center of Excellence (CITEC), Faculty of Technology,
Bielefeld University, Bielefeld, Germany

Milou Althof (miloualthof@gmail.com)[1]

Abstract

One study found that observers retained more information from hand gestures that speakers gazed at, possibly because speaker-gaze shifted observers' attention covertly. Speaker-gaze may thus modulate the role of gestures in communication. One hypothesized communicative function of gestures, and specifically of the inter-personal repetition of gestures, is to facilitate the process of creating common ground (grounding). Therefore, speaker-gaze may also influence the inter-personal repetition of gestures. In an experimental study, we found that participants were more likely to repeat another speaker's gestures if the original speaker gazed at the gestures. Moreover, speaker-gaze was a better predictor of this repetition than participants' own gaze. This supports the hypothesis that speakers' gaze at their gestures leads to covert attention shifts in observers, causing the gestures to be processed differently. Speaker-gaze could therefore be a valuable cue to the processing and production of gestures by artificial systems that interact with humans.

Keywords: Gesture; Gaze; Perception; Alignment; Adaptation

Introduction

Speech tends to be accompanied by hand gestures (Kendon, 2004; McNeill, 1992, 2005), which can depict aspects of the content we convey (*representational hand gestures*), emphasize certain parts of it (*beats*), or regulate our interaction (*interactive gestures*), (Bavelas, Chovil, Lawrie, & Wade, 1992). Next to several for-speaker functions, representational gestures are likely to serve for-addressee functions (e.g., Alibali, Heath, & Myers, 2001; Bavelas, Gerwing, Sutton, & Prevost, 2008; Jacobs & Garnham, 2007; Mol, Krahmer, Maes, & Swerts, 2011; Özyürek, 2002). Importantly, people can gain semantic information from representational gestures (e.g., Beattie & Shovelton, 1999b; Cassell, McNeill, & McCullough, 1998).

Numerous studies have found that perceiving others' representational hand gestures influences how we shape our own (Holler & Wilkin, 2011; Kimbara, 2006, 2008; Kopp & Bergman, 2013; Mol, Krahmer, Maes, & Swerts, 2012; Parrill & Kimbara, 2006). That is, interlocutors tend to repeat each other's representational hand gestures. This inter-personal repetition of gestures is thought to facilitate grounding (Holler & Wilkin, 2011), analogous to the inter-personal repetition of referring expressions (Clark & Brennan, 1991). This means that as gesture forms converge, the associated concepts converge as well and interlocutors incrementally arrive at common ground.

To draw their addressee's attention to their gestures, speakers might employ gaze (e.g., Goodwin, 1981; Gullberg & Holmqvist, 1999, 2006; Streeck, 1993). Consistent with this hypothesis, addressees gain more information from gestures that speakers gazed at (Gullberg & Kita, 2009). This shows that speaker-gaze can modulate gestures' role in communication. Would speaker-gaze therefore also influence the inter-personal repetition of gestures, thereby potentially modulating gestures' role in grounding?

This study is a first step in testing if and how speaker-gaze modulates the inter-personal repetition of gestures and ultimately grounding. By comparing the role of speaker-gaze and observer-gaze, we also shed light on how gestures are attended to.

Gestures, Gaze, and Information Uptake

In human-human dialogue, addressees mostly gaze at the speaker's face, rather than the speaker's hands (Argyle & Cook, 1976; Kendon, 1990). Studies using eye-tracking report an average percentage of time participants fixated on a speaker's face ranging from 84.9% to 98.4% (Beattie, Webster, & Ross, 2010; Gullberg & Holmqvist, 1999, 2006). Participants were reported to gaze at a speaker's hands only a small percentage of time (<.5% - 2.1%).

Gaze at hand gestures has been studied in more detail. The percentage of hand gestures that addressees gaze at, varies as a function of certain properties of the gestures (Beattie, et al., 2010; Gullberg & Kita, 2009), related to the properties of peripheral vision. For example, for iconic gestures from a character viewpoint (CVP)[1], more gestures with a smaller movement span were fixated on (17/30) than gestures with a larger movement span (9/30) (Beattie, et al., 2010). For these iconic CVP gestures[1], participants also gazed at the low span gestures for a longer percentage of their stroke[2], compared to the high span gestures (Beattie, et al., 2010).

[1] Iconic CVP gestures are gestures that depict part of a story, as though the speaker were the character in the story. For example, speakers may pretend to throw a ball if the character was doing so.

[2] The stroke is the most meaningful phase of a gesture and in this case also the most energetic part.

Interestingly, in other studies, iconic CVP gestures with low span were found to be more informative to addressees (Beattie & Shovelton, 1999a, 2002). There may thus be a relation between addressees' information uptake from a gesture and their gaze at a gesture (Beattie, et al., 2010). Yet although these general trends were found for particular types of gestures, correlations between gaze and uptake have not been found within individuals.

Gullberg and Kita (2009) found little evidence for a direct relation between addressees' fixations at speakers' gestures and their information uptake from these gestures. Making use of gestures from a previously collected data set, they found that addressees (i.e. observers of clips) were more likely to fixate on gestures that the speaker had gazed at (*gestures with speaker-gaze*), as well as on gestures that contained a post-stroke hold[3]. Interestingly, onset latencies of addressees' fixations were longer for gestures with speaker-gaze than for gestures with a post-stroke hold. Gullberg and Kita explain this as addressees gazing at gestures with speaker-gaze for top-down, social reasons (social alignment), whereas they gaze at gestures with a post-stroke hold for bottom-up, stimulus driven reasons. Information uptake was not found to be larger for gestures with a post-stroke hold, yet addressees did gain more information from gestures with speaker-gaze. Addressees were more likely to retain non-vital information (direction of movement) when they had observed a gesture with speaker-gaze. However, Gullberg and Kita found little evidence that the effect of speaker-gaze on information uptake was mediated by addressees' own gaze. Rather, there seemed to be a direct effect of speakers' gaze to their gestures on addressees' uptake from these gestures.

Posner (1988) describes that locations of visual stimuli can be attended *covertly*, that is "without any change in eye or head position", and that this covert attention to a location can change the priority of a stimulus in the covertly attended location (Posner & Petersen, 1990, p. 27). It may thus be the case that a speaker's gaze to a gesture guides the addressee's attention to this gesture covertly (i.e. without the addressee gazing at the gesture), resulting in more efficient processing of the gesture and ultimately better information uptake, or recall. Gullberg and Kita (2009, p. 269) speculated that "although overt gaze-following is not automatic, covert attention shift to the target of a speaker's gaze location may well be". Yet they state that their finding, which was not completely replicated in a more controlled experiment in which gaze was manipulated artificially, needs to be consolidated in further studies.

Gaze and the Repetition of Gestures

Suppose that observers gain semantic information from representational gestures (e.g., Beattie & Shovelton, 1999b; Cassell, et al., 1998). Then this information, in the form of one or more semantic representations, will be linked to one or more representations of the observed gesture form. Therefore, when the observer subsequently wants to express this information, the representation(s) of gesture form will get activated and the observed gesture (or a similar one) may be reproduced. This is predicted both by theories that assume automated priming underlying the inter-personal repetition of linguistic behaviors (Pickering & Garrod, 2004) and by theories in which this inter-personal repetition is part of a deliberate grounding process (Clark & Brennan, 1991; Holler & Wilkin, 2011). A similar argument may hold for beat gestures and information on importance/stress.

Now if it is the case that the gestures that a speaker gazes at are attended to more closely by observers, this will lead to higher activations of observers' internal representations. It is therefore expected that observers are *more likely* to repeat gestures that they saw *with speaker-gaze* than those they saw without speaker-gaze. If the associated attention shift indeed happens covertly, speaker-gaze is expected to be a better predictor of the inter-personal repetition of gestures than observer-gaze.

On the other hand, eye-contact was found to facilitate the deliberate repetition of hand movements (Wang, Newport, & Hamilton, 2011). Hand movements were repeated faster if the gaze of a person in a video-clip, who performed the movements, was directed towards the person watching the clip, who needed to repeat the movements. Therefore, the (deliberate) repetition of hand gestures may be facilitated by eye-contact. Hence, gestures that are gazed at by a speaker may be *less likely* to be repeated by an observer, since there is less eye-contact. However, in communication the repetition of a hand gesture by another interlocutor can happen at any later time (Holler & Wilkin, 2011). This may reduce the role of eye-contact in the repetition of gestures. Moreover, speakers tend to alternate their gaze between their gesture and the addressee (Streeck, 1993).

Gullberg and Kita (2009) hypothesized that addressees may gaze at gestures with speaker-gaze for social reasons (social alignment). Since social reasons can also underlie the inter-personal repetition of behaviors (Cheng & Chartrand, 2003), gestures that are gazed at by a speaker may be more likely to be copied for social reasons as well. In this case, observer-gaze may be more strongly correlated to the inter-personal repetition of gestures than speaker-gaze, since both gaze-following and the repetition of the gesture would be instances of social alignment.

Present Study

As a first step in testing if and how speaker-gaze modulates the inter-personal repetition of gestures, we tested whether a speaker's gaze at her own gestures affected the likelihood of these gestures being subsequently repeated by another speaker. In this first study, we minimized effects of social processes and of deliberate processes involved in grounding,

[3] During a post-stroke hold, the hands are steady for a bit, before returning to a resting position or moving towards the next stroke.

by having participants see the gestures performed by one person and then talk to another person themselves. This allowed us to first reveal any automated mechanisms at play.

We used life-sized projections of a speaker as stimuli. This way, speaker-gaze could be manipulated reliably and participants' eye-movements could be tracked with a freestanding eye-tracker, allowing us to measure the effect of speaker-gaze and of observer-gaze.

Research Question and Hypotheses

RQ: Are people more likely to repeat gestures that they perceived with than without speaker-gaze?

Hypothesis 1: When participants observe and retell a narration that includes gestures, they are *more likely* to repeat the gestures that the original speaker gazed at.

Hypothesis 2: Speaker-gaze is a stronger predictor of whether a gesture will be repeated than is participants' own gaze while perceiving the narration (observer-gaze).

Figure 1. Stimulus movie snapshots. Top: gesture with speaker-gaze. Bottom: same gesture without speaker-gaze.

Method

Participants

Twenty-five (16 female) Dutch students of Tilburg University participated in this study for course credit (excluding one participant who knew the aim of the study).

Design

Whether the speaker gazed at a gesture or not was manipulated within participant. The dependent variable consisted of the number of gesture that participants repeated in their own retellings of the stimuli.

To control for any factors related to the gestures as such, two versions of the stimulus movie were created and each was shown to half of the participants who did the retelling. In either version, the same speaker performed the same narration with the same gestures. However, in one movie, she gazed at one half of her gestures and in the other she gazed at the other half of her gestures (Figure 1). This way, if participants were more likely to repeat the gestures that the speaker gazed at, this could not be due to intrinsic properties of the gestures, such as movement span or perspective.

Additionally, participants' eye-movements were tracked with a free-standing eye-tracker, to test if participants' own gaze was a better predictor of the repetition of gestures than was speaker-gaze.

Material

The gestures in the stimulus movie were taken from previously collected retellings of a Tweety and Sylvester cartoon (Mol, Krahmer, Maes, & Swerts, 2009; Mol, et al., 2011). The speaker in the stimulus movies produced 24 iconic gestures and 10 beats, half of which she gazed at (alternating). The rest of the time, she looked into the camera. One group of participants saw the speaker gaze at one half of her gestures (stimulus movie 1) and the other group at the other half (stimulus movie 2).

Out of the 24 iconic gestures, 15 were from a character viewpoint (CVP), e.g. pretending to grab Tweety, and 9 were from an observer viewpoint (OVP), e.g. outlining manner and path of how Sylvester rolled down the street. The number of CVP and OVP gestures with speaker-gaze in stimulus movie 1 and 2 was balanced. All beats were performed in the same manner: lifting the joined hands briefly from their position in the speaker's lap. They mostly accompanied character names.

Procedure and Task

Two participants came to the lab and were each assigned to the role of narrator or listener. It was taken into account that the narrator's gaze needed to be tracked. Therefore, if one participant wore glasses, they would be the listener.

The narrator took place in a seat with a freestanding eye-tracker placed in front of it, on a laptop stand. This seat was facing a white wall, on which stimulus movie 1 or 2 was projected life-sized. Before each episode, the eye-tracker was calibrated. Then the narrator watched the episode (sound was played over speakers), while the listener listened to music over headphones, in a chair that was facing away from the projection.

After watching an episode, the narrator took place in a chair that was aligned with the chair of the speaker in the stimulus movie and the listener sat across (Figure 2, next page). The experimenter switched on the camera capturing the narrator and the narrator related the story of the cartoon episode to the listener. Afterward, the experimenter switched off the camera. Then the listener answered some questions on the story, while listening to music and facing

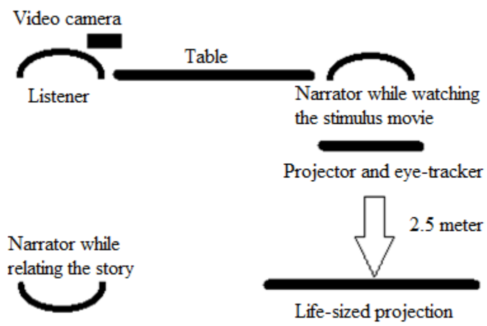

Figure 2: Setting

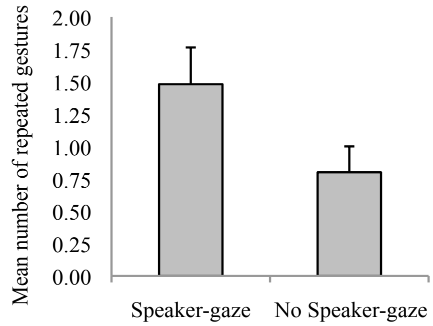

Figure 3: Mean number of repeated gestures with and without speaker-gaze. Error bars represent *SEM*.

away from the projection. At the same time, the eye-tracker was calibrated and the narrator watched the next episode. This was repeated for all five episodes and all participants.

Coding & Analyses

The resulting videos were coded using Elan (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). For each content-unit that had occurred with a gesture in the stimulus-movie (e.g. throwing a bowling ball, swinging across), it was determined whether participants mentioned it verbally and if so, whether they simultaneously produced: *a repetition* of the observed gesture*, a partial repetition, another gesture,* or *no gesture.* This decision was based on: hand shape, hand orientation, location and movement (speed, direction, size). Twenty percent of the data was double coded. Percentage agreement on whether a content-unit was mentioned was 95%. Cohen's kappa on the original four gesture labels was .61. The coders disagreed most on partial repetitions. Cohen's kappa for whether there was a full repetition was .80, indicating substantial agreement (Landis & Koch, 1977). We report full repetitions.

Unfortunately, it turned out that one beat was missing in stimulus movie 1. Since there cannot be gaze to a gesture that was not performed, this data point is not informative to our hypothesis. It was therefore treated as missing data for participants who saw stimulus movie 1.

The eye-tracking data was pre-processed with BeGaze by SMI. This rendered a movie clip for each participant, in which the participants' fixations were shown as a small circle projected onto the stimulus movie. Elan was used to manually code whether participants gazed at each gesture.

Unfortunately, our video-recordings did not allow us to code participants' gaze to their own gestures.

Results

A paired-samples *t*-test revealed that, consistent with hypothesis 1, participants repeated more gestures that the speaker gazed at (M=1.48, SD=1.42) than that she did not gaze at (M=.80, SD=1.00), $t(24)$=2.37, p=.026, 95% CI of difference = (.09, 1.27), see Figure 3. Results were similar for iconics (p=.061) and beats (p=.11). Similar patterns were also observed when controlling for the number of content-units participants mentioned. No significant differences between the gaze and no-gaze condition were found for the other categories (partial repetition p = .47, other iconic gesture p = .58, other beat gesture p = 1, no gesture p = .59, all these categories combined p = .17).

The stimulus movie contained 24 iconic gestures and 9 (version 1) or 10 (version 2) beats. Thirteen participants saw version 1 and twelve saw version 2, rendering 837 gesture tokens. In total, participants repeated 57 observed gestures in their own retelling (6.8 %), including 14 out of 237 beats (5.9%) and 43 out of 600 iconics (7.2%).

For each of these repeated gestures, it is known whether the speaker in the stimulus clip gazed at it (*speaker-gaze*) and it was measured whether the participant fixated on the gesture while watching the stimulus clip (*participant-gaze*). Therefore, using the binomial distribution, we can compute whether gestures seen with speaker-gaze or being produced with participant-gaze were over-represented in the set of repeated gestures.

The speaker in the stimulus clip gazed at 425 out of 837 gestures, that is, with a chance (p) of .508. Out of the 57 repeated gestures (n), 27 iconics and 10 beats had been seen with speaker gaze, making for 37 gestures with speaker-gaze (k). This renders a binomial z-ratio of 2.00, indicating that the number of gestures with speaker-gaze in the set of gestures that were repeated is higher than chance (one-tailed test: p = .023).

In sum, participants gazed at 159 out of 837 gestures. However, in 201 cases (24%), there was missing data, because calibration was off, or no circle appeared on the clip, leaving it unclear whether the participant fixated on the gesture. The latter could either mean the participant looked elsewhere (not on the clip), or the tracker lost track of the participant's gaze. Thus, participants gazed at a gesture in 159 out of 636 observed cases, that is, with a chance (p) of .25. Out of the 57 (n) repeated gestures 12 (k) iconics were fixated on by participants (4 cases missing) and no beats (3 cases missing). The number of gestures with participant-gaze in the set of repeated gestures did not differ from chance (binomial z-ratio = -.54, one-tailed test: p = .30), not even when assuming all cases of missing data were gestures with observer-gaze. The data therefore support hypothesis 2.

Discussion and Conclusion

This study is first to show that speakers' gaze towards their gestures can increase the repetition of these gestures by another speaker. This is consistent with the hypothesis that speaker-gaze can signal the communicative import of a gesture (Goodwin, 1981; Gullberg & Holmqvist, 1999, 2006; Streeck, 1993) and it shows that addressees are sensitive to this. This is relevant to work on embodied conversational agents.

Consistent with findings on gaze and information uptake (Gullberg & Kita, 2009), we found that speaker-gaze was a better predictor of whether a gesture would be repeated than was the observer's own gaze. This is in line with the explanation that speaker-gaze leads to a covert attention shift in the observer. This shift may be to the location of the gesture, causing the gesture to be processed differently, analogous to the way in which other stimuli in attended locations are processed differently from those in unattended locations (Posner, 1988). Apparently, this difference in processing caused the gesture to be more prone to repetition (this study) and for the information from the gesture to be more likely retained (Gullberg & Kita, 2009).

Interestingly, our results were obtained without dialogue between the speaker who originally performed the gesture and the speaker repeating it. Thus, the repetitions we found cannot be instances of deliberate social alignment, nor could they be part of a deliberate grounding process between the original performer and the repeater. Rather, a link between a representation of gesture form and a representation of the associated meaning seems to have been formed as a result of automated processes, causing the observer to be more likely to reproduce the observed gesture when later expressing the same meaning (cf. Pickering & Garrod, 2004).

Given that we found similar patterns for iconic gestures and beats, it seems that it does not matter whether this meaning is semantic, or pragmatic (stress/import) in nature. It is somewhat surprising though that beat gestures were repeated across individuals too. To our knowledge, this has not been shown or tested before. Even though the original speaker always performed the same beat gesture, this gesture was more likely to be repeated with parts of the story in which she had gazed at it. However, it may be the case that participants interpreted these beats as deictic gestures, indicating the location of the (usually) concurrently mentioned character. An informal analysis of an existing gesture corpus showed that speakers hardly gaze at their beats. Hence, speaker-gaze may have affected the interpretation of the beat gestures. Therefore, the finding that beats were repeated inter-personally too needs to be interpreted with caution, until it is replicated for different beats and supported by observational studies.

One could argue that for some iconic character-viewpoint-gestures, it does not make sense to gaze at the gesture and therefore these gazed-at gestures are more likely repeated. For example, when climbing up a drainpipe, Sylvester may look at his goal (Tweety), rather than his paws. The repeated CVP-gestures were: having a ball in one's stomach, lifting a rug, hitting with an umbrella, carrying luggage, drawing, throwing a weight, grabbing Tweety, holding Tweety (while shooting up). The following CVP gestures were never repeated: climbing up, throwing a ball, playing an organ, giving a coin, throwing away a suitcase, rubbing hands, holding Tweety (while falling down). From this, we see no evidence that the effect was caused by unnatural gaze.

Future studies need to assess whether speaker-gaze plays a larger role when interacting in dialogue, with the same partner being present throughout the conversation. Since dialogue allows for (deliberate) grounding, both automated and flexible processes may influence the inter-personal repetition of gestures in dialogue (Kopp & Bergman, 2013). Hence, more inter-personal repetition of gestures is expected and possibly a larger role for speaker-gaze. Yet although the percentage of repeated gestures in our data (7%) may seem small, it is not too far from rates found in natural interaction (Mdn = .04, Range = .11), (Holler & Wilkin, 2011). Holler and Wilkin used a very different task, involving the description of tan gram figures. It would be highly interesting to compare the inter-personal repetition of gestures and the effect of speaker-gaze on it between a dialogue setting and a setting like in the current study, yet with similar tasks. Since dialogue also allows for social alignment, the correlation between an observer's gaze to a speaker's gestures and their repetition of these gestures may also be larger in dialogue, as observers may both follow a speaker's gaze and copy their gesture for social reasons.

Acknowledgements

We thank Karin van Nispen for being the speaker in our stimulus clips, Rein Cozijn for assistance with the eye-tracking equipment and data-processing, Stefan Kopp, Fons Maes, Jorrig Vogels and the anonymous reviewers for their helpful comments to earlier versions. This research was enabled by a Veni Grant from the Netherlands Organisation for Scientific Research (NWO), awarded to the first author.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169-188.

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

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469-489.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520.

Beattie, G., & Shovelton, H. (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental. investigation. *Semiotica, 123*, 1-30.

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438-462.

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that affect their communicative power. *British Journal of Psychology, 93*(2), 179-192.

Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology, 29*(2), 194-213.

Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic & nonlinguistic information. *Pragmatics & Cognition, 6*(2), 1-33.

Cheng, C. M., & Chartrand, T. L. (2003). Self-monitoring without awareness: Using mimicry as a nonconscious affiliation strategy. *Personality and Social Psychology, 85*(6), 1170-1179.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* Washington, DC: APA.

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

Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition, 7*, 35-63.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what listeners look at. Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition, 14*, 53-82.

Gullberg, M., & Kita, S. (2009). Attention to speech-accompanying gestures: eye movements and information uptake. *Journal of Nonverbal Behavior, 33*(4), 251-277.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior, 35*, 133-153.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*(2), 291-303.

Kendon, A. (1990). *Conducting Interaction*. Cambridge: Cambridge University Press.

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

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6*(1), 39-61.

Kimbara, I. (2008). Gesture form convergence in joint description. *Journal of Nonverbal Behavior, 32*(2), 123-131.

Kopp, S., & Bergman, K. (2013). Automatic and strategic alignment of co-verbal gestures in dialogue. In I. Wachsmut & J. P. De Ruiter (Eds.), *Alignment in Communication: Towards a New Theory of Communication.* Amsterdam: John Benjamins Publishing Company.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

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

McNeill, D. (2005). *Gesture and Thought*. Chicago and London: University of Chicago Press.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human-human and human-machine interactions. *Gesture, 9*(1), 97-126.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication, 17*(1), 77-100.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language, 66*(1), 249-264.

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*(4), 688-704.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior, 30*(4), 157-166.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-225.

Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), *Master Lectures in Clinical Neuropsychology.* Waschington, DC: Am. Pshych. Assoc.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25-42.

Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs, 60*, 275-229.

Wang, Y., Newport, R., & Hamilton, A. F. (2011). Eye contact enhances mimicry of intransitive hand movements. *Biol Lett, 7*, 7-10.

Wittenburg, P., Brugman, H., Russel, H., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In Proceedings of the LREC 2006, Fifth International Conference on Language Resources and Evaluation (pp. 1556-1559).