Communicative Gestures and Memory Load

Lisette Mol (L.Mol@uvt.nl)[1]

Emiel Krahmer (E.J.Krahmer@uvt.nl)[1]

Alfons Maes (Maes@uvt.nl)[1]

Marc Swerts (M.G.J.Swerts@uvt.nl)
[1]Faculty of Humanities, Communication and Cognition, Tilburg University
Warandelaan 2, 5037 AB, Tilburg, The Netherlands

Abstract

Previous research has shown that (co-speech) hand gestures sometimes aid cognition and can reduce a speaker's cognitive load. We argue that this is not the case for gestures that are produced primarily to communicate, which we think come at a cognitive cost to speakers instead. In a production experiment with a narrative task, we show that speakers gesture more frequently with a less demanding task, but only if speaker and addressee can see each other. Our results support our theory, without contradicting previous findings.

Keywords: Gesture; Cognitive load; Audience design.

Introduction

In this paper we explore the relationship between memory load and gesturing. First we describe two alternate perspectives on why people gesture, which give rise to different predictions about the relation between gesturing and cognitive load. We then describe previous work supporting the idea that speakers themselves benefit cognitively from gesturing. This is followed by our present study, in which we test a prediction from the perspective that gestures are produced to communicate.

The functional roles of gesturing

Many studies have investigated the functional roles of hand gestures that people spontaneously produce during speech. One function is that gesturing aids speech production. For example by aiding lexical retrieval (Hadar, 1989; Krauss, 1998), helping to hold a mental image while it is verbally expressed (De Ruiter, 1998), or by helping speakers to "organize rich spatio-motoric information into packages suitable for speaking" (Kita, 2000).

There is also a large body of evidence that gesturing serves communicative purposes. It has been shown that addressees can gain information from gesture (Beattie & Shovelton, 1999; Chawla & Krauss, 1994; Cutica & Bucciarelli, 2008; Goldin-Meadow & Sandhofer, 1999; Mol et al., 2009). Speakers also gesture differently depending on many features of the communicative setting, such as whether the addressee can see them or not (i.e. Alibali, Heath & Myers, 2001; Cohen, 1977), where the addressee is located relative to them (Özyürek, 2002), whether information is new or given to the addressee (Enfield, Kita & De Ruiter, 2007;

Jacobs & Garnham, 2006), whether there is dialogue (Bavelas et al., 2008), and whether the addressee is human or artificial (Mol et al., 2009). If speakers adapt their gesturing to such environmental factors, then this could mean that gesturing is sometimes intended for the addressee.

Jacobs and Garnham (2006) point out that the primary functional role of gesturing may depend on the task a speaker is performing (also see Alibali, Kita & Young, 2000). Their study shows that during a narrative task, gestures are produced primarily for the benefit of the addressee. It has also been hypothesized that different kinds or sizes of gestures serve different purposes (i.e. Alibali et al., 2001; Bangerter & Chevalley, 2007; Bavelas et al., 2008; Beattie & Shovelton, 2002; Enfield et al., 2007). Especially larger gestures and gestures depicting some of the content of a speaker's message (*representational gestures*) have been shown to occur more frequently when a speaker is visible to an addressee, and hence are associated with the communicative functions of gesture.

The different functional perspectives on gesturing make opposite predictions about the effect of gesturing on a speaker's total cognitive burden. Gestures that are produced mostly for the benefit of the speaker would ease the process of speech production and may thus lighten a speaker's cognitive load. But gestures produced primarily for the benefit of the addressee would rather come at a cognitive cost to speakers, just as verbal language production does. We next present some evidence for this first effect, which in our view leaves open the possibility for the second.

Gesturing and Cognitive Load

In Cohen (1977) an effect of task difficulty on gesturing is reported. It was found that participants produced more gestures when giving route directions involving four intersections (more demanding task) than when giving directions involving two intersections (less demanding task). The average difference in gesture rate between the more and less demanding task was equal in size when interlocutors could see each other and when they could not. This suggests that the effect resulted from gestures produced primarily for the benefit of the speaker, since such gestures would be produced both when the addressee is visible to the speaker and when not. This is consistent with the direction of the

effect: *more* gestures with the *more demanding* task. However, since the differences in the average gesture rate between the more and less demanding task were close to 2, the effect may relate to the number of intersections that needed to be described quite directly, rather than being a more general result of the cognitive demands of the instruction task.

In a picture description task, Melinger and Kita (2007) found that the complexity of the picture influenced gesture rates. For pictures of colored dots along a path with multiple branches gestures were produced more frequently than for pictures of dots along a continuous path, without any choice points. They also found that in a dual-task situation, a spatial secondary task lead to higher gesture rates than a non-spatial secondary task. This is strong evidence that gesturing can help a speaker, especially since speaker and addressee could not see each other in this study. However, since visibility was blocked by a wooden screen, these results may not generalize to all gestures that are produced in face-to-face interaction.

Goldin-Meadow et al. (2001) investigated the relation between cognitive load and gestures by manipulating gesturing rather than task difficulty. They made use of a dual-task paradigm, in which participants performed a memory task while explaining a math problem. They found that children and adults performed better on the memory task if they gestured during their explanation, provided that the memory task was sufficiently challenging. This effect was found both when people were instructed not to gesture and when they refrained from gesturing spontaneously. This indicates that gesturing can reduce the cognitive load of the explanation task, thereby leaving more cognitive resources available for the memory task.

This result was refined in Wagner, Nusbaum, and Goldin-Meadow (2004). Gesturing was found to benefit both spatial and propositional memory, but especially when gestures expressed the same content as the concurrent speech. This leaves open the possibility that this result is driven by gesturing for the speaker as well, rather than by gesturing for the addressee. So the question remains open whether communicative gestures come at a cognitive cost to speakers.

Present Study

In our present study, we are interested in the relationship between cognitive load and gestures that are produced primarily to communicate. We expect such gestures to come at a cognitive cost to speakers, rather than reducing their cognitive load. Therefore, we vary the demands that a narration task makes on speakers' memory. We expect that with a less demanding task, speakers will produce more gestures that primarily aid the addressee than with a more demanding task. We compare a communicative setting in which gestures can be used to communicate, to one in which they can not. Especially in this first setting, we would expect speakers to gesture *more* with a *less demanding* task.

For this study we have asked participants to retell an animated cartoon. To reduce the task's memory demands,

we have cut up the information to be described into smaller parts. We assume that the more demanding task is sufficiently difficult, such that speakers are unable to produce all communicative gestures that an addressee could benefit from. Since the content that addressees need to remember is the same in both tasks, we also assume that addressees can still benefit from gestures in the condition with the less demanding task.

The different theories on the primary function of gesturing make different predictions on how the memory demands of a task influence the gesture rate. For gestures that aid speakers by facilitating lexical access, we would not expect an effect of our manipulation of memory demand on gesture rate. It seems unlikely that finding the words to describe a cartoon would be easier or harder when describing a shorter or longer part of it. Neither would we expect an effect of mutual visibility on such gestures, since they are tied to the speech production process within the speaker. This prediction is depicted in Figure 1a.
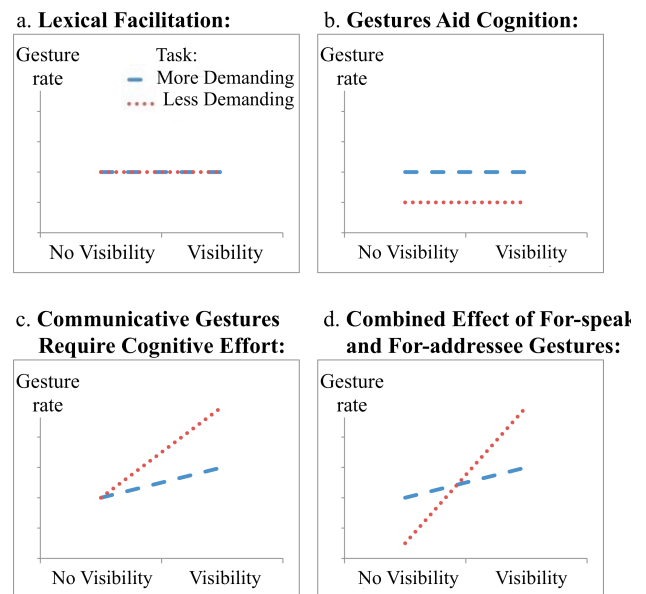

Figure 1: Predictions made by the different theories on how task demand and visibility influence gesture rate.

If gesturing mostly aids cognitive processes within the speaker, such as holding a mental image or organizing information into packages suitable for speaking, then we would expect a lower gesture rate when the task is split up. This less demanding task may give rise to fewer gestures that are needed to facilitate cognition. If so, this effect should be present both when speaker and addressee can see each other, and when they can not (see Figure 1b). Another possibility is that these processes are automated to such an extent, or take place on such a local scope, that there is no effect of our global manipulation of memory demand on gesturing (like in Figure 1a).

If most gestures are produced for the benefit of the addressee rather than the speaker, then a different pattern of results is predicted. In that case we would expect speakers

needing to put cognitive effort into their gesture production, rather than experiencing a reduction in memory load from it. Therefore, we would expect a higher gesture rate if the narration task is split up into smaller parts, and its demands on memory are lower. However, such communicative gestures are expected especially when speaker and addressee can see each other. So we would predict an interaction between mutual visibility and task demand: The gesture rate will be higher with the less demanding task, but only if speaker and addressee can see each other. No effect of memory demand on gesture rate is predicted for the condition in which speaker and addressee cannot see each other. This prediction is depicted in Figure 1c.

If we assume that with this narration task speakers will produce both gestures that aid their own cognition or speech production and gestures that are intended primarily for the addressee, we need to combine the predictions for both types of gestures. We would expect for-speaker gestures to form the majority in the no visibility condition. Therefore, in this condition the gesture rate will be lower with the less demanding task. However, in the visibility condition, where we expect that most gestures are produced primarily for the benefit of the addressee (Jacobs & Garnham, 2006), the gesture rate will be higher with the less demanding task. This is depicted in Figure 1d.

Method

Design

We have used a 2 x 2 between subjects design. The two independent variables are whether an animated cartoon is seen and retold all eight episodes at once, or one episode at a time, and whether speaker and addressee can see each other or not. We have used a narrative task, because we would expect communicatively intended gestures to occur frequently with such a task (Jacobs & Garnham, 2006).

Participants

39 first year students of Tilburg University took part as speakers in this study, as part of their first year curriculum. They were all native speakers of Dutch. Addressees were first year students and native Dutch speakers as well.

Procedure

The two conditions with the more demanding task of retelling the cartoon all at once were actually taken from an earlier study, which looked at the effect of the addressee being human or artificial (Mol et al., 2009). The procedure was similar to the one described below, except that speakers first saw the animated cartoon in its entirety, and then retold it to an addressee.

For the less demanding task, we randomly assigned participants to the no visibility (Screen) or visibility (Face-to-Face) condition (after Alibali et al., 2001). The Screen condition differed from the FtF condition in that a wooden screen separated speaker and addressee, such that there was

no mutual visibility. The experimenter randomly assigned participants the role of narrator or listener.

Participants first read the instructions and could ask any questions they had on the task. The instructions focused on the task of the addressee, namely summarizing the speaker's narration. This way we implied that the study was on summarizing. Speakers were explicitly asked not to summarize, but to just retell the story. The instructions stated that they were videotaped with the purpose of comparing the addressee's summary to their narration afterwards. Addressees were instructed not to interrupt the speaker.

The animated cartoon we used was "Canary Row", by Warner Brothers. This is a seven-minute animated cartoon in which a cat tries to capture a bird in eight different ways. The cartoon was separated into its eight episodes, by inserting 10 seconds of blank video after each episode.

Speakers watched the cartoon on a computer screen that only they could see. While the speaker watched the cartoon, the addressee was seated across from the speaker, as depicted in Figure 2, and was asked to listen to music through headphones. Once the episode had finished, the speaker paused the movie and signaled to the addressee that the episode had ended. The addressee then took off the headphones, and speaker and addressee moved into chairs facing each other (see Figure 2). In the FtF condition, there was nothing in between speaker and addressee. The camera was right behind the addressee, slightly to the side. In the Screen condition, speakers were facing the camera, which was right in front of the Screen. The addressee was seated at the same distance from the speaker as in the FtF condition.

Before the experiment started, participants were seated as in positions 3 and 4 in Figure 2. The experimenter explained that the speaker was to address the addressee rather than the camera. Then followed a practice run, to make sure both participants understood the procedure. The experimenter then started the camera and left the room.

After the retelling of the cartoon, the experimenter took the addressee to another room to write the summary using a common word processor. To exclude the possibility that the task with a screen separating speaker and addressee was experienced as more difficult than the FtF task or vice versa, narrators were then asked to complete the NASA Task Load Index (Hart & Staveland, 1988), in which subjective workload is assessed on six scales. They next completed a second questionnaire, which included questions on how they had experienced the communication. We fully debriefed all participants and asked their consent for using the recordings and summaries. All participants agreed to the use of their data for scientific purposes.

Data Analysis

The first author transcribed each narration from the video-tape. Repairs, repeated words, false starts, and filled pauses were included. The annotation of hand gestures was initially done by the first author. Difficult cases were resolved by discussion among all authors. We first discriminated between gestures and other movements such as self-adjustment. Then we labeled gestures that seemed to depict
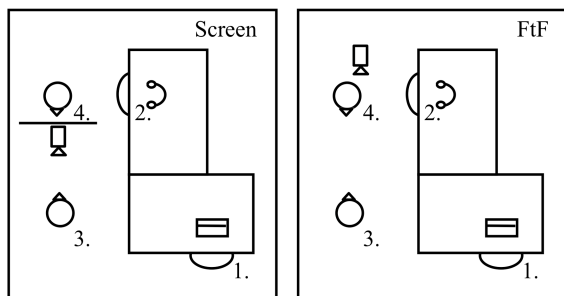
Figure 2: Experimental Setup. 1 = position of speaker while watching video, 2 = position of addressee while listening to music on headphones, 3 = position of speaker during narration, 4 = position of addressee during narration.

some of the content of the cartoon as *representational gestures* (McNeill, 1992). All other gestures, including simple biphasic movements of the hands (*beats)* and other *interactive gestures* (Bavelas, 1992), were labeled as *non representational gestures*.

In a separate round of gesture coding, we coded for gesture size. Gestures that were produced using only the fingers received a score of 1. If there was significant wrist movement, the gesture received a score of 2. Gestures that also involved significant movement of the elbow or lower arm received a score of 3, and gestures in which the upper arm was also used in a meaningful way or that involved movement of the shoulder received a score of 4. This allowed an average gesture size to be computed for each participant.

For all tests for significance we have used univariate analysis of variance (ANOVA), with mutual visibility (2 levels: yes, no) and task demand (2 levels: high, low) as the fixed factors and a significance threshold of .05. Where needed we have performed pairwise comparisons between all four conditions using the LSD method with a significance threshold of .05. We have used partial Eta square as a measure of effect size.

Results

More (different) words with less demanding task

Speakers doing the less demanding task produced more words (M = 1204, SD = 404) than speakers doing the more demanding task (M = 610, SD = 189), $F(1,35) = 34.250$, $p < .001$, $\eta_p^2 = .50$, regardless of whether visibility was blocked (p = .22). The effect of visibility on the total number of words used was not significant (p = .36). We found no significant effects of visibility or task demand on the number of words per second, or the number of filled pauses per word.

The type-token-ratio (TTR) is a measure of repetition of words in a text. It is computed by dividing the number of unique words in a text or corpus by the total number of words. The TTR was higher with the less demanding task (33% vs. 26%), indicating greater word variety. We found

no effect of visibility on the TTR, neither with the more (p = .44) nor with the less demanding task (p = .7).

More gestures with less demanding task in FtF setting only

Speakers produced more gestures per 100 words when speaker and addressee could see each other $F(1,35) = 28.804$, $p < .001$, $\eta_p^2 = .45$. The main effect of task demand was not significant (p = .42), but there was a significant interaction of visibility and task demand, $F(1,35) = 4.643$, $p < .05$, $\eta_p^2 = .12$, see Figure 3. Pairwise comparisons showed that in the FtF condition, gestures were significantly more frequent with the less demanding task. Similar results were obtained for the number of gestures per second.
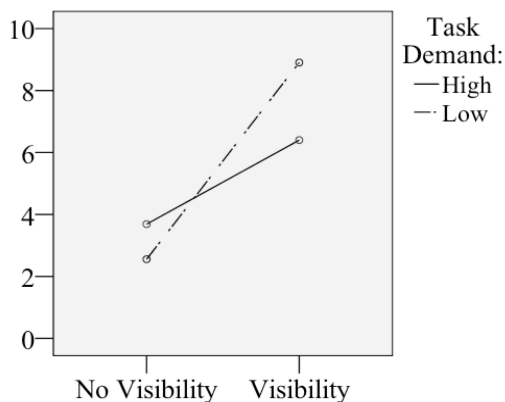

Figure 3: Means of the average number of gestures produced per 100 words.

When looking at the mean total number of gestures produced, without normalizing by words or seconds, more gestures were produced when visibility was not blocked, $F(1, 35) = 31.98$, $p < .001$, $\eta_p^2 = .48$, as well as with the less demanding task, $F(1, 35) = 20.36$, $p < .001$, $\eta_p^2 = .37$. There was an interaction between visibility and task demand, $F(1, 35) = 12.94$, $p < .01$, $\eta_p^2 = .27$, see Figure 4.
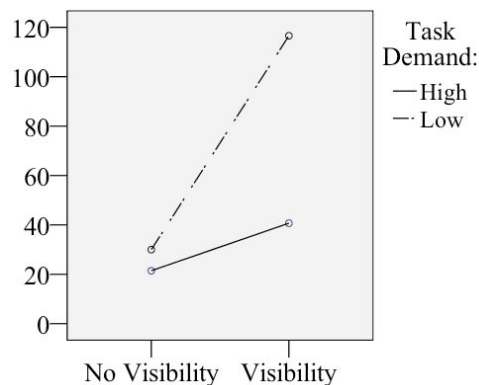

Figure 4: Means of the number of gestures produced.

More representational gestures per 100 words were produced when visibility was not blocked, $F(1, 35) = 32.461$, p

< .001, η_p^2 = .48. The main effect of task demand was not significant (p = .28), but in the FtF setting, representational gestures were more frequent with the less demanding task, $F(1, 35) = .4419$, $p < .05$, $\eta_p^2 = .11$. For non representational gestures, only the effect of visibility reached significance, $F(1, 35) = 5.605$, $p < .05$, $\eta_p^2 = .14$. Gestures were more frequent when visibility was not blocked, and like for representational gestures, the difference was larger with the less demanding task.

Different gestures in FtF conditions

Gestures were larger when speaker and addressee could see each other (M = 2.8, SD = .4) compared to when they could not (M = 2.2, SD = .5), $F(1, 35) = 14.669$, $p < .01$, $\eta_p^2 = .30$. We also found a higher percentage of gestures made from a character viewpoint[1] when visibility was not blocked, $F(1, 35) = 8.866$, $p < .01$, $\eta_p^2 = .21$.

The NASA Task Load Index indicated that the narration task was considered equally difficult when there was a screen in between speaker and addressee and when there was not. There was no significant difference between these conditions on any of the six scales.

Discussion

The memory demand of a narration task had an effect on both verbal language production and gesture production. With a less demanding task, speakers produced longer narrations and used a higher percentage of unique words. Though further linguistic analysis is needed, this suggests that speakers added extra information to their narrations. We did not find a significant effect of visibility on the length of a narration or the percentage of unique words. Neither did visibility influence two measures of speech fluency: speech rate and the number of filled pauses per 100 words. This indicates that narrations were broadly similar when visibility was blocked and when it was not, though more detailed linguistic analysis is needed to support this claim.

As expected, gestures were more frequent when visibility was not blocked. But more importantly, this difference was larger with the less demanding task. Though narrations were longer with the less demanding task regardless of visibility, the gesture rate was higher only when visibility was not blocked. This suggests that the higher gesture rate resulted from gestures produced mostly for the addressee, rather than from gestures that the speaker needed in order to produce the verbal narration (especially since we found no significant effects of visibility on any of the variables we used to measure verbal behavior).

Our results support the theory that speakers need to put cognitive effort into the production of gestures that mostly

---

[1] In the character viewpoint, body parts of the speaker map directly onto the same body parts of a character that is being described, rather than part of the speaker representing a different part of a character or situation (see McNeill, 1992). For example, if speakers use their hands and arms to depict the cat's paws and forelegs, this is considered character viewpoint, whereas using one's fingers to represent the cat's legs is not.

serve a communicative purpose. This theory predicts more gestures with a less demanding narration task, but only when gestures have the potential to communicate. This is indeed what we found.

Clearly, the two tasks that we have used resulted in different verbal and non-verbal behavior. But is this a result of the different memory demands of retelling the cartoon all at once or one episode at a time? A difference in pragmatics of these two tasks may have lead to different verbal behavior, which in turn resulted in different gesture rates. For this explanation one has to add that these different pragmatics only influence gestures that have the potential to communicate, since the difference was not found when visibility was blocked. To further address this issue, we are planning to do an additional analysis in which only utterances encoding certain key elements of information are included. This should lead to more comparable language use in the more demanding and less demanding task.

In addition to more qualitative linguistic analysis, a dual-task paradigm may help to discriminate between whether speakers produced communicative gestures more frequently because of a lower cognitive load, or whether this resulted from another factor in our less demanding task. We are planning to use such a design in our future research.

When visibility was blocked, there was no significant difference in gesture rate between the more and less demanding task. This confirms that gestures produced in this setting differ from gestures produced when visibility was unobstructed. It could very well be that most gestures produced when visibility was blocked were produced for the benefit of the speaker. Our current study cannot clarify how exactly these gestures aid speakers. It has only shown that a global manipulation of task demand does not seem to have a large effect on the frequency with which such gestures are produced (despite the more elaborate verbal descriptions with the less demanding task).

Our study did show several qualitative differences between gestures produced when speaker and addressee could not see each other vs. when they could. Gestures produced when visibility was blocked were smaller and characters were imitated less often. This supports the idea that gestures (solely) serving a for-speaker purpose look markedly different from gestures that have communicative potential.

At first glance our results may seem contradictive to the results found by Cohen (1977). However, Cohen did not find an effect of whether participants were familiar or unfamiliar with the routes to be described. This suggests that both the complex and easy tasks were well within the cognitive capacities of the participants. This could mean that in both tasks, speakers had enough capacity left to produce any appropriate communicative gestures, and the difference in gesture rate was caused by gestures produced primarily for the speaker. As explained earlier, the pattern of results found matches this hypothesis.

It may also be that in giving route directions, or explaining math problems, such as in the study by Goldin-Meadow et al. (2001), gestures produced mostly for the benefit of the

speaker form the majority, as opposed to in narrative tasks (Jacobs & Garnham, 2006). Wagner et al. (2004) did not find a beneficial effect on memory of gestures that contained information that was not expressed in speech. It may very well be that most gestures produced to communicate fell into this category. Therefore, our results do not conflict with these latter two studies either. Rather, our study identifies an important distinction that needs to be made when studying the relationship between gesturing and cognitive load, namely between gesturing primarily for producer-internal purposes and gesturing primarily for producer-external purposes.

Conclusion

Our study has shown that on average, gestures are produced at a higher rate when people perform a narration task that places lower demands on memory, but only when visibility between speaker and addressee is not blocked. This supports the ideas that some gestures are produced primarily to communicate, and that speakers need to put cognitive effort into producing such gestures.

Acknowledgements

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.

Alibali, M.W., Kita, S., & Young, A.J. (2000) Gesture and the process of speech production: We think, therefore we gesture, *Language and Cognitive Processes, 15(6)*, 593-613.

Bangerter, A., & Chevalley, E. (2007) Pointing and describing in referential communication: When are pointing gestures used to communicate? In: I. Van der Sluis, M. Theune, E. Reiter & E. Krahmer (Eds.), *CTIT Proceedings of the Workshop on Multimodal Output Generation (MOG),* Aberdeen, Scotland, January 2007.

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. (1999) 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)*, 179-72.

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

Cohen, A.A. (1977) The communicative functions of hand illustrators, *Journal of Communication*, *27 (4)*, 54-63.

Cutica, I., & Bucciareli, M. (2008) The deep versus the shallow: Effects of co-speech gestures in learning from discourse, *Cognitive Science*, *32*, 921-935.

De Ruiter, J.-P. (1998) Gesture and Speech production. Unpublished dissertation, University of Nijmegen.

Enfield, N.J., Kita, S., & De Ruiter, J.-P. (2007) Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*, 1722-1741.

Goldin-Meadow, S., Nusaum, H., Kelly, S.D., & Wagner, S. (2001) Explaining Math: Gesturing Lightens the Load. *Psychological Science*, *12 (6)*, 516-522.

Goldin-Meadow, S., & Sandhofer, C.M. (1999) Gestures convey substantive information about a child's thoughts to ordinary listeners. *Developmental Science*, *2*, 67-74.

Hadar, U. (1989) Two types of gesture and their role in speech production. *Journal of Language and Social Psychology, 8*, 221-228.

Hart, S.G., & Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: P. Hancock, N. Meshkati (Eds.), *Human Mental Workload*. Amsterdam: Elsevier Science.

Jacobs, N., & Garnham, A. (2006) The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 26*, 291-303.

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

Keysar, B., Lin, S., & Barr, D. (2003) Limits on theory of mind use in adults. *Cognition, 89*, 25-41.

Kita, S. (2000) How representational gestures help speaking. In: D. McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge University Press.

Krauss, R. (1998) Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54-60.

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

Meligner, A & Kita, S. (2007) Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22 (4),* 473-500.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009) The communicative import of gesture: evidence from a comparative analysis of human-human and human-machine interactions. *Gesture 9 (1),* 98-126.

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.

Wagner, S.M., Nusbaum, H, & Goldin-Meadow, S. (2004) Probing the mental representation of gesture: is handwaving spatial? *Journal of Memory and Language*, *50*, 395-407.