

Co-speech gestures in the process of meaning coordination

Judith Holler^{1, 2}, Mark Tutton², Katie Wilkin²

¹ Max Planck Institute for Psycholinguistics, Wundtlaan 1, Nijmegen 6525XD, The Netherlands

² School of Psychological Sciences, Oxford Road, University of Manchester
Manchester, M13 9PL, UK

Judith.holler@manchester.ac.uk, Mark.tutton@manchester.ac.uk, Katie.wilkin@manchester.ac.uk

Abstract

This study uses a classical referential communication task to investigate the role of co-speech gestures in the process of coordination. The study manipulates both the common ground between the interlocutors, as well as the visibility of the gestures they use. The findings show that co-speech gestures are an integral part of the referential utterances speakers produced with regard to both initial references as well as repeated references, and that the availability of gestures appears to impact on interlocutors' referential coordination. The results are discussed with regard to past research on common ground as well as theories of gesture production.

Keywords: co-speech gesture, referential communication, repeated reference, common ground, visibility.

1. Introduction

Clark [1] has described language use as a joint activity, a collaborative process between two or more interactants. Rather than producing utterances independently of each other (much like a speaker in isolation) interactants coordinate their behaviours and conversational contributions to communicate effectively and efficiently.

An essential component of this coordination process is the interactants' common ground [1], [2], i.e., 'their mutual, common, or joint knowledge, beliefs, and suppositions' [1:93]. Common ground can be evaluated based on facts that are known or evident before individuals enter a conversation (such as being of the same nationality, working for the same company, or having attended the same party), but it also accumulates over the course of a conversation (in terms of the knowledge exchanged and the linguistic expressions used to refer to things in the world). In order for common ground to accrue during conversation, interlocutors are required to establish the mutual belief that they have understood what the respective other intended to communicate. That is, they have to 'ground' their conversational contributions [3].

How this grounding process works and how the common ground that accumulates as a consequence of it affects communication has been extensively studied (e.g., [4], [5], [6], [7]). However, these studies have focused almost exclusively on verbal communication. And indeed, in certain contexts (such as on the phone), speech and meta-linguistic cues (intonation, prosody, hesitations and so forth) are the only sources of information available to interlocutors. However, in face-to-face communication, the most common forum of communication [8], speech is ubiquitously accompanied by co-speech gestures [9]. Importantly, these gestural movements are tightly integrated with speech on a semantic, pragmatic and

temporal level both during production (e.g. [10], [11], [12]) and comprehension (e.g. [13], [14], [15]). Co-speech gestures thus ought to be considered an integral component of human language, which, if ignored, may render only partial insights into communication and cognition.

A variety of theories have been developed which attribute different functions to co-speech gestures. Some of these claim that co-speech gestures facilitate speakers' cognitive processes during speech production, such as conceptualisation (e.g. [16], [17]) or lexical retrieval (e.g. [18, 19]). Others attribute predominantly communicative reasons and benefits to these gestures (e.g. [20], [21]). Since there is empirical evidence supporting all of these perspectives, one plausible assumption is that co-speech gestures may be multi-functional.

Here, we aim to explore the role of co-speech gestures in referential coordination; do gestures play a communicative role in this process? If so, how prevalent is their communicative contribution? If it exists during initial coordination, does it vanish with the accumulation of common ground? And does not having access to the gestural modality affect interlocutors' communication and their ability to ground their contributions?

The present study tries to answer these questions by focusing on referential communication, a process core to everyday talk. The data were obtained from a classic referential communication task based on a previous study by Clark and Wilkes-Gibbs [4] but which, in addition to the original condition where interlocutors were prevented from seeing each other, includes a face-to-face condition for comparison. As part of the respective task, interlocutors were required to create a particular sequence of cards showing images of abstract geometrical figures. Importantly, they created new sequences with these same figures over six consecutive trials, thus being required to refer to them repeatedly and to coordinate on mutually understood referring expressions (either with gesture being available or not). We aimed to answer the following research questions:

(1) From previous research we know that accumulating common ground results in elliptical speech, such as in a reduced number of words being used for referring expressions (e.g. [4]). However, the issue of how gestures are affected by common ground that accumulates over the course of a conversation remains unexplored. Here, we measure the rate at which interlocutors gesture in the process of creating mutually accepted referring expressions (i.e., with initial attempts to identify a referent, which often involved extended descriptions of these entities, as well as with repeated references to these same entities once a particular interpretation or perspective had been mutually accepted). Does gesture vanish over time during this coordination process? According to cognitive theories of gesture production, they should; not only does conceptualisation become increasingly easy with the repeated use

of certain referential terms (and the move from more extensive descriptions to simple, established noun phrases as is typical for this process), but the fact that participants tend to repeat the same lexical items to construct their referential descriptions should make lexical access considerably easier as interlocutors work through the task. However, if gesture rate does not decrease significantly over the trials, this would be strong support for a retained communicative function in this context.

(2) Performing gestures in visually prominent areas of gesture space has been associated with marking them as communicatively relevant [22]. In the present study, participants were able to perform gestures in visually prominent areas of gesture space, or lower down, behind a cardboard wall that served to obstruct the participants' view of the cards, which would have rendered the gestures invisible to the recipient, too. Therefore, if gestures are of communicative importance in the task, we would expect speakers to make them visually prominent, and if gestures continue to fulfill communicative functions even with increasing common ground, their gestures should remain visually accessible to the recipient.

(3) How does coordination in face-to-face communication compare to coordination in contexts in which vision is obscured and the gestural modality is not available? Do speakers produce any gestures at all when visibility is obscured in the present communicative task? Past research has provided discrepant findings, with some resulting in a reduced gesture rate when gestures are not visible to the addressee [23], [24], while others do not [25]. We therefore compare speakers' gesture rate in the two visibility conditions. Secondly, if speakers do produce gestures, do they differ from those in the face-to-face condition in terms of gesture space? If yes, this would suggest that gesturing in this task is not a simple by-product of the speech production process but influenced by whether speakers are aware that their gestures can be seen or not, meaning that they are communicatively motivated (at least in the face-to-face condition).

(4) Does the availability of gesture influence interlocutors' linguistic coordination? If gestures in this coordination task play a communicative role, we would expect speakers to be more doubtful and uncertain about the efficacy of their verbal referential utterances than interlocutors communicating face-to-face.

(5) If the availability of co-speech gestures benefits coordination in this referential communication task, then we would expect interlocutors to perform better in a face-to-face context (i.e., they should make less errors) than when the gestures cannot be seen by recipients.

2. Method

2.1. Design and Participants

This study is based on a study by Clark and Wilkes-Gibbs [4], which involved participants creating and grounding referring expressions, in pairs, over six consecutive trials. However, in addition to a condition in which participants were separated by an occluding screen we added a face-to-face condition. As in Clark and Wilkes-Gibbs' study, sixteen participants (eight pairs) took part in the screen condition, and a further eight pairs in the face-to-face condition. The number of same gender dyads and opposite gender dyads was the same in both conditions. The participants' mean age was 22.81 years in the screen condition ($SD = 3.82$), and 26.50 years in the face-to-face condition ($SD = 6.37$). All participants were from the University of Manchester population (but unacquainted before the start of the experiment) and were financially compensated for their participation.

2.2. Apparatus and Materials

Participants stood opposite each other, each in front of a table (table distance about 3m). Two horizontal rows of six paper squares were fixed onto each table. The stimuli were two identical sets of twelve cards showing so-called 'tangrams' (see [4]:11). One set was placed onto the squares on one of the tables, the other set of cards was spread along the bottom edge of the other table in one long line. In the screen condition a large (2m×2m) wooden screen was placed between the participants obscuring the view of any part of their body as well as the stimuli. In the face-to-face condition, a small, 20cm high cardboard wall was attached to the table so that participants could see each other but not their cards. In both conditions, participants were filmed split-screen with two wall-mounted cameras.

2.3. Procedure

Participants were told that one person would take on the role of the director (D), and one the role of the matcher (M) (roles were allocated randomly, based on their choice of seats during the instruction phase). D was instructed to guide M in correctly identifying and placing the cards onto the squares in the same order as D's cards. The participants were told that there would be six trials overall, using the same set of cards but arranged in a different order each time. They were allowed to talk to each other about each figure as much as they found necessary in order to allow the matcher to identify the correct card.

2.4. Analysis

2.4.1. Number of words

The sixteen interactions were transcribed in detail. The number of words used to refer to each individual figure was counted (but not references to the cards themselves or the cards' positions), with compound words counted as two separate words. Full words constituting repairs were also included in the word count. However, 'um' and 'uh' (which speakers use to signal different kinds of delays in speaking; [26]) were here excluded from the word count as they do not constitute part of the referring expression as such. Also excluded from the analysis were words (and any accompanying gestures) used to describe the position of a card within the set, which changed from trial to trial and thus was not part of the referring expression for a particular figure.

2.4.2. Number of gestures and gesture rate

Originally, the gestures were identified and categorised into 'iconic', 'metaphoric', and 'deictic' [9] as well as 'interactive' gestures [8]. Iconic gestures involving more than the arms and hands, used to mime actions and postures, were classed as 'reenactments'. However, for the present analyses, an overall gesture rate was calculated by dividing the number of gestures produced by each speaker per trial by the number of words produced by the respective speaker during the same trial to account for differences between speakers in the number of words produced.

2.4.3. Gesture space

All gestures were classified according to two gesture space categories (adapted from [9]); one category, which we labeled *allocentric* gestures, captured all those gestures performed in the central area of the gesture space, above the cardboard wall that was mounted onto the tables (which corresponded roughly to the hip-line of the speaker), and which were thus visible to the interlocutor in the face-to-face condition (and which would have been visible in the screen condition if the screen was removed). The second category captured all those gestures performed below the upper edge of the cardboard wall, i.e., which were performed directly on, or very close to, the table surface/card surface (this latter category

therefore represents a slightly truncated version of the lower gesture space area defined by McNeill [9]). These gestures were termed *egocentric* as they were visible only to the speaker him- or herself and were not presented to the addressee.

2.4.4. Uncertainty indicators

We coded the interlocutors' utterances for indicators of uncertainty or doubt (see our predictions) for which we created three categories: try markers (rising intonation at the end of declarative phrases; [27]), questions, and other expressions of uncertainty (such as 'I don't know if we've got it right', 'Not entirely sure whether we're talking about the same one').

3. Results

3.1. Analysis of number of words

Our first analysis compared the average number of words used by speakers to refer to each figure. As expected based on previous research [4], the results of the 2×6 (visibility \times CG) mixed ANOVA revealed a significant main effect of common ground (CG), $F_{(1.7, 23.9)} = 46.54$, $MSE = 731.49$, $p = .00001$. However, the difference between visibility conditions was not significant, $F_{(1, 14)} = 0.94$, $MSE = 724.70$, $p = .348$. The interaction between visibility and CG also failed to reach significance, $F_{(1.7, 23.9)} = 2.96$, $MSE = 731.49$, $p = .078$.

3.2. Analysis of gesture frequency and gesture rate

To account for the fact that speakers produced verbal utterances of varying lengths, we calculated the rate at which speakers gestured (gestures/words, see Method). This analysis revealed that, although we found a similar pattern when considering the number of words and number of gestures on their own (i.e., a significant decline over the trials for both modalities), the rate of gesturing *increased* with accumulating common ground. That is, on the whole, speakers tended to use more gestures/words the more common ground had accumulated, $F_{(5, 70)} = 3.62$, $MSE = .001$, $p = .006$. This means that, while speakers produced both fewer words *and* fewer gestures with accumulating common ground, this decrease was more pronounced for the number of words than for the number of gestures, leading to a proportional increase in gesture to words. The difference between visibility conditions was not significant, $F_{(1, 14)} = 0.60$, $MSE = .011$, $p = .451$, and neither was the interaction between CG and visibility, $F_{(5, 70)} = 0.49$, $MSE = .001$, $p = .786$.

3.3. Gesture space

A $2 \times 2 \times 6$ (visibility \times gesture space \times CG) mixed ANOVA showed that, overall, a statistically equal number of gestures were produced in the two gesture space areas, $F_{(1, 14)} = 0.28$, $MSE = 897.45$, $p = .608$. The main effect of CG was, again, significant, $F_{(1.5, 20.6)} = 23.33$, $MSE = 932.07$, $p = .0001$. The interaction between CG and gesture space was not significant, $F_{(1.6, 22.2)} = 1.34$, $MSE = 631.61$, $p = .278$. That is, even with increasing common ground, the number of allocentric gestures (i.e., gestures in the central and upper areas of the gesture space, visible by their addressee and often oriented towards the addressee) in relation to the number of egocentric gestures (i.e., in the lower area of the gesture space, only visible to themselves irrespective of the big screen, and not oriented towards their addressee) did not change. However, the interaction between space and visibility was significant, $F_{(1, 14)} = 18.14$, $MSE = 897.45$, $p = .001$. Interactants in the face-to-face condition produced mainly allocentric gestures, while participants in the screen condition produced predominantly egocentric gestures.

3.4. Analysis of uncertainty expressions

Our analysis of expressed uncertainty revealed that interlocutors in the screen condition used significantly more try markers ($M = 29.00$, $SD = 19.46$) than interlocutors in the face-to-face condition ($M = 6.50$, $SD = 4.28$), $t(14) = 3.20$, $p = .006$. No differences were found for the number of questions asked or for other expressions of uncertainty.

3.5. Accuracy analysis

Each pair was able to achieve a maximum of 12 correct matches per trial (i.e., one per figure), and hence 72 matches in total. Overall, the maximum number of correct matches summed for the eight participant pairs in each condition was 576. Although the data show that, in line with previous research, the error rate was fairly low, participants in the screen condition made almost four times as many errors as those in the face-to-face condition, namely 38 (6.6%) as compared to 10 (1.7%) errors. However, a *U*-test revealed this difference in accuracy between the two groups not to be a reliable one, $U = 27.50$, $n_1 = 8$, $n_2 = 8$, $p = .323$, one-tailed. But when considering only those pairs who did make errors (namely, three out of eight in each condition) it was shown that the pairs in the screen condition made significantly more errors than pairs in the face-to-face condition, $U = 0.00$, $n_1 = 3$, $n_2 = 3$, $p = .050$, one-tailed, $r = .50$ (face-to-face: $Mdn = 3.00$, $Range = 1.00$; screen: $Mdn = 13.00$, $Range = 3.00$).

4. Discussion

The present study provides us with a range of insights into the role of co-speech gestures in referential coordination. One of our main findings is that with increasing common ground, speakers' gesture rate also increased. That is, the proportional gestural contribution towards the speakers' utterances increased over the trials. This shows that speakers considered gesture an important modality in the process of coordinating their referential communication, even when they knew that their partner already shared common ground about the respective referents with them.

That gesture rate increases with repeated references (or descriptions) cannot be explained by cognitive theories of gesture production. Since mutually shared knowledge of how to conceptualise the individual referents increases over the trials, the conceptualisation process underlying the verbal utterances used to refer to the figures (or to describe them) should become considerably easier over the trials, too. According to the conceptualisation hypothesis [16], [17], we should therefore see a decrease in gesture rate. Likewise, since the interactants created shared referential expressions over the successive trials, thus accessing the same lexical items repeatedly and becoming narrower in the number of lexical items that have to be accessed for each referent, the lexical access theory [18], [19] would also predict a decrease in gesture rate.

Instead, the results are in strong support of the notion that the gestures were used communicatively in this task. This is not to say that the gestures' communicative role did not change over time (e.g., from collaborating with the addressee to create an initial, mutually agreed interpretation of the figures on earlier trials to the re-activation of already established and mutually agreed interpretations during later trials), but that whatever function they may have fulfilled at any given moment, this function appears to have been communicatively motivated.

The findings also advance our knowledge of gesture in the context of common ground. Although some studies have already started to explore this issue, studies on this topic remain sparse,

and the pattern of results seems complex. While some studies have found evidence suggesting that gestures become more elliptical when more common ground exists [28], [29], others have yielded results that support the idea that gestures carry a greater communicational weight when more common ground exists. Holler and Wilkin [30] compared speakers' gesture rate during narrative when common ground between speaker and addressee existed and when it did not. They, too, found that speakers in the common ground condition gestured at a higher rate than speakers in the no common ground condition. The present findings are in line with this. Overall, it appears that the probably most intuitive assumption, namely that gesture tends to whither away the easier communication gets due to increasing common ground, does not seem provide us with the complete picture. While this appears to happen in certain contexts, others seem to encourage an enhanced use of gesture. Future studies are needed to explore the details of the kinds of functions that gestures fulfill in these various contexts.

The conclusion that, in the current context, speakers in the face-to-face condition used the gestures communicatively throughout the task is corroborated by our data on gesture space. Speakers produced the large majority of their gestures in visually prominent areas of gesture space, thus making them visually accessible for their addressee. Not only did this happen during trial one but throughout the task. This suggests that speakers considered their gestures to be an important part of the referential expressions and descriptions they provided, even when common ground had been established.

In terms of comparing the group that communicated face-to-face with that of interlocutors being separated by a screen, a number of interesting observations could be made, too. Firstly, we found that the two conditions did not differ in terms of the rate at which speakers gestured. This stands in contrast to some earlier studies that have manipulated visibility between speaker and addressee and shown that speakers gesture less when their gestures cannot be seen (e.g., [23], [24]). There could be at least two explanations for this. One the one hand, it could be that speakers in the screen condition gestured for a different reason than speakers communicating face-to-face, such as to aid their conceptualisation or their lexical retrieval. Such a conclusion would receive support from the finding that speakers in this condition differed in how they performed the gestures – namely in low gesture space, close to or on the table. The gestures also differed in their orientation as a consequence of this; even if they had not been occluded by the screen, the gestures would have been visibly accessible by the speaker only, not by the addressee. However, as mentioned above, these two cognitive theories of gesture production would have predicted that gesture rate decreases over the trials, whereas we observed an increase over the trials, also in the screen condition. This renders this particular explanation an unlikely contender.

An alternative interpretation of the findings is that speakers in the screen condition used the gestures communicatively despite their awareness that they would not be communicatively effective. This would be in line with the results from a recent study by Bavelas, Gerwing, Sutton and Prevost [25] which compared speakers' gesture rate in a face-to-face context, one in which they were visually separated but still engaged in dialogue, and one where speakers talked into a tape recorder in a monologue-type context. Whereas this study showed no significant difference between the first two conditions, they did find that gesture rate was significantly reduced in the tape recorder context. Their conclusion was that gesture use is an integral element of dialogue, and hence occurs even in dialogic situations in which the gestures are visually not available to the addressee. The present study employed a strongly dialogic, collaborative task, and thus makes an

interpretation of the data along the same lines very plausible. In addition, other aspects of the speakers' behaviour not systematically analysed here suggest that the dialogue-linked interpretation may capture our data most aptly; for example, we observed that speakers in the screen condition tended to raise their gestures and perform them in central gesture space (i.e., what mapped onto visually prominent areas of gesture space in the face-to-face condition) in response to coordination difficulties, such as when it was evident that the interlocutors had miscommunicated or the partner was just not understanding what the other tried to convey. According to this social and communicative interpretation, speakers could not help but gesture due to the strong dialogic nature of the task, but their awareness that their gestures would not be seen affected the orientation of the gestures (and when understanding problems were signaled, even this was overridden and the gestures then looked exactly like those performed in the face-to-face condition).

In short, while we cannot exclude the possibility that the gestures produced by speakers in the screen condition may have fulfilled some cognitive functions benefitting the speakers' memory or speech production, we would argue that our data provide stronger evidence for a social-interactional interpretation [20], [21].

We also carried out a comparison of the number of words used by speakers in the two experimental conditions. While it was not surprising that we replicated the significant decrease in the number of words used over trials [4], it was somewhat surprising to see that speakers who did not have the gestural modality at their disposal to convey information did not create longer, more complex verbal utterances. Their referential utterances comprised the same number of words than those accompanied by co-speech gestures. It appears that they were able to communicate what they deemed good enough for current purposes as efficiently as speakers who were able to communicate gesturally. One possible interpretation is that they simply did not know how to compensate for the lack of gesture and settled for less ideal but nevertheless concise and simple s which they considered standing a chance in being successful under the given circumstances – in other words, they might have sacrificed certainty regarding the effectiveness of their referential expressions and descriptions in order to avoid prolixity and lack of efficiency.

Our results from the uncertainty analysis support such an interpretation. While speakers did not ask more questions nor utter more expressions explicitly indicating uncertainty or doubt (both of which would make the communication less efficient and more cumbersome), they did produce significantly more utterances marked by try markers. Try markers [27] are used by speakers to signal uncertainty as to whether the marked utterance will be understood by the addressee or not [4], [31], and in this task whether the addressee would be able to identify the correct referent on the basis of it. By using try markers (i.e., simply altering intonation) instead of formulating questions and additional words expressing uncertainty interactants were able to adhere to the principle of 'least collaborative effort' [4]. Thus, the unavailability of the gestural modality appears to have a direct effect on how interlocutors create mutually shared referential expressions and how they ground information.

As a final analysis, we looked at how accurate interlocutors were in completing the task. Here, we found that the relatively small sample size did not allow us to draw definite conclusions. While the same number of people made mistakes in the two conditions (three pairs in each), those that did make errors made significantly more of them in the screen condition. It appears that interlocutors did not have more difficulty to complete the task

without gesture per se, but that, if they did miscommunicate, they had considerably more trouble in correcting, or possibly even noticing their errors. It has to be noted though that in this task, participants performed at ceiling level, meaning that also effects on the overall accuracy might be revealed by a more fine-grained measure.

Despite revealing interesting insights, the present study has certain limitations, of course. One is that the stimuli are rather spatial in nature, and therefore we cannot necessarily generalize our findings to dialogue about other types of referents. However, speakers often likened the figures they described to real world referents (e.g., ‘looks like a man sitting down’, looks like a woman praying’), thus establishing a clear relevance for everyday discourse. Further, interlocutors in our study were not able to see each other at all, i.e., neither their gestures, nor any other aspect of their visual behaviour (such as gaze). While it is unlikely that any visual behaviour other than gestures would have communicated semantically relevant information, other behaviours that might have also been important in this context, such as signaling understanding through head nodding or placing a card were easily compensated for verbally (e.g., through back-channelling cues) or through audible actions (e.g., by enhancing the noise of placing the cards), and this did indeed occur.

In conclusion, the present study demonstrates that co-speech gestures are an integral component of speakers’ referential utterances in face-to-face dialogic interaction, at least in the referential context examined here (for findings along those lines relating to pointing gestures, see [32], [33]). Particularly interesting is the fact that speakers continue to employ gesture even when they mutually share knowledge about the referents and their conceptualisations with their addressees. The exact functions of the gestures used in this context, as well as in those cases where speakers gestured without their gestures being visible, remain to be explored in future studies. Further, while we found evidence that not being able to communicate gesturally does appear to impact on speakers’ coordination, in particular their certainty, future studies may be able to shed more light on whether this consistently affects how successful they are in their referential communication.

5. Acknowledgements

We would like to thank the Economic and Social Research Council, UK, for funding this research (RES-061-23-0135 and RES-000-22-3748), Rachel Poorhady for help with the data collection, as well as the participants who took part in our study.

6. References

- [1] Clark, H.H., “Using language”, Cambridge: Cambridge University Press, 1996.
- [2] Stalnaker, R., “Assertion”, in P. Cole [Ed.], *Pragmatics: Syntax and semantics* (Vol. 9), 315-322, New York: Academic Press, 1978.
- [3] Clark, H.H. and Brennan, S.A., “Grounding in communication”, in L.B. Resnick, J.M. Levine, and S.D. Teasley [Eds.], *Perspectives on socially shared cognition*, 127-149, Washington: APA Books, 1991.
- [4] Clark, H.H., and Wilkes-Gibbs, D., “Referring as a collaborative process”, *Cognition*, 22:1-39, 1986.
- [5] Fussell, S.R. and Krauss, R.M., “The effects of intended audience on message production and comprehension: Reference in a common ground framework”, *Journal of Experimental Social Psychology*, 25:203-219, 1989.
- [6] Fussell, S.R. and Krauss, R.M., “Coordination of knowledge in communication: Effects of speakers’ assumptions about others’ knowledge”, *Journal of Personality and Social Psychology*, 62:378-391, 1992.
- [7] Isaacs, E.A. and Clark, H.H., “References in conversations between experts and novices”, *Journal of Experimental Psychology: General*, 116:26-37, 1987.
- [8] Bavelas, J.B., Chovil, N., Lawrie, D. and Wade, A., “Interactive gestures”, *Discourse Processes*, 15:469-489, 1992.
- [9] McNeill, D., “Hand and mind: What gestures reveal about thought”, Chicago: University of Chicago Press, 1992.
- [10] Holler, J., and Beattie, G., “Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?”, *Gesture*, 3:127-154, 2003.
- [11] Holler, J., & Beattie, G., “How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process?”, *Semiotica*, 146, 81-116, 2003.
- [12] Loehr, D., “Gesture and Intonation”, unpublished doctoral thesis. Georgetown University, 2004.
- [13] Habets, B., Kita, S., Shao, Z., Ozyurek, A. and Hagoort, P., “The role of synchrony and ambiguity in speech: Gesture integration during comprehension”, *Journal of Cognitive Neuroscience*, 23: 1845-1854, 2010.
- [14] Kelly, S. D., Ozyurek, A., Maris, E., “Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension”, *Psychological Science*, 21:260-267, 2010.
- [15] Kelly, S. D., Barr, D., Church, R. B., and Lynch, K., “Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory”, *Journal of Memory and Language*, 40:577-592, 1999.
- [16] Hostetter, A. B., Alibali, M. W., Kita, S., “I see it in my hands’ eye: Representational gestures reflect conceptual demands”, *Language and Cognitive Processes* 22:313–336, 2007.
- [17] Kita, S., Davies, T. S., “Competing conceptual representations trigger co-speech representational gestures”, *Language and Cognitive Processes* 24:761-775, 2009.
- [18] Krauss, R.M., Chen, Y. and Gottesman, R.F., “Lexical gestures and lexical retrieval: A process model”, in D. McNeill [Ed.], *Language and gesture*, 261-283, Cambridge: Cambridge University Press, 2000.
- [19] Pine, K. J., Bird, H., Kirk, E., “The effects of prohibiting gestures on children’s lexical retrieval ability”, *Developmental Science*, 10:747-754, 2007.
- [20] Bavelas, J.B., and Chovil, N., “Visible acts of meaning: An integrated message model of language in face-to-face dialogue”, *Journal of Language and Social Psychology*, 19:163-194, 2000.
- [21] Kendon, A., “Gesture: Visible action as utterance”, Cambridge: Cambridge University Press, 2004.
- [22] Streeck, J., “Gestures as communication II: The audience as co-author”, *Research on Language and Social Interaction*, 27:239-267, 1994.
- [23] Alibali, M.W., Heath, D.C., and Myers, H.J., “Effects of visibility between speakers and listeners on gesture production: Some gestures are meant to be seen”, *Journal of Memory and Language*, 44:169-188, 2001.
- [24] Bavelas, J. B., Kenwood, C., Johnson, T., and Phillips, B., “An experimental study of when and how speakers use gestures to communicate”, *Gesture*, 2:1–17, 2002.
- [25] Bavelas, J.B., Gerwing, J., Sutton, C., and Prevost, D., “Gesturing on the telephone: Independent effects of dialogue and visibility”, *Journal of Memory and Language*, 58:495-520, 2008.
- [26] Clark, H.H. and Fox Tree, J. E., “Using uh and um in spontaneous speech”, *Cognition*, 84:73-111, 2002.
- [27] Sacks, H. and Schegloff, E.A., “Two Preferences in the Organization of Reference to Persons in Conversation and Their Interaction”, in G. Psathas [Ed.], *Everyday Language: Studies in Ethnomethodology*, Irvington Press, New York, NY, 15–21, 1979.
- [28] Gerwing, J. and Bavelas, J.B., “Linguistic influences on gesture’s form”, *Gesture*, 4, 157-195, 2004.
- [29] Holler, J. and Stevens, R., “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, 2007.

- [30] Holler, J. and Wilkin, K., "Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task", *Language and Cognitive Processes*, 24:267-289, 2009.
- [31] Clark, H.H. and Schäfer, E.F., "Concealing one's meaning from overhearers" *Journal of Memory and Language*, 26:209-225, 1987.
- [32] Bangerter, A., "Using pointing and describing to achieve joint focus of attention in dialogue", *Psychological Science*, 15: 415-419, 2004.
- [33] Van der Sluis, I. and Kraemer, E., "The influence of target size and distance on the production of speech and gesture in multimodal referring expressions", *Proceedings of the International Conference on Spoken Language Processing*, 2004.