Gestural Alignment in Natural Dialogue

Kirsten Bergmann (kirsten.bergmann@uni-bielefeld.de)
Stefan Kopp (skopp@techfak.uni-bielefeld.de)
Faculty of Technology, Center of Excellence "Cognitive Interaction Technology" (CITEC)
Collaborative Research Center "Alignment in Communication" (SFB 673)
Bielefeld University, P.O. Box 100 131, D-33501 Bielefeld, Germany

Abstract

A well-known phenomenon in natural interaction is that speakers adapt their linguistic and nonverbal behaviors. Research on *gestural* alignment is, however, still in its early stages based on evidence from experimental settings. This paper provides a first systematic study of gesture form convergence based on a large sample of naturalistic dialogue data. We found evidence for gestural alignment, but not all form features of co-speech gestures are subject to this effect. In a detailed analysis of those sensitive features we further address questions of how gestural alignment depends on the temporal distance between gestures, and whether *intra*-speaker or *inter*-speaker influences on gesture form are stronger.

Keywords: Alignment, co-speech gestures, natural interaction

Introduction

Co-speech gesturing is an integral part of human communication, but it is not well understood why and how gestures take on their particular physical form. This holds especially for iconic gestures that apparently communicate by virtue of iconicity, i.e., through a correspondence between their form and geometrical or spatial properties of what they refer to. Empirical studies, however, revealed that similarity with the referent cannot fully account for all occurrences of iconic gesture use (Kopp, Tepper, Ferriman, Striegnitz, & Cassell, 2007). Findings also indicate that a gesture's form is influenced by other contextual constraints such as discourse (Holler & Stevens, 2007) or the linguistic context (Kita & Özyürek, 2003). In addition there are considerable differences in how speakers gesture, partly assumed to be due to different cognitive abilities (Hostetter & Alibali, 2007).

In addition to intra-speaker sensitivities, co-speech gesturing in dialogue may also be influenced by the gestures of the interlocutor. A large body of work has demonstrated inter-personal sensitivities in verbal and nonverbal behavior in natural social interaction, leading often to coordination and alignment between interlocutors (cf. (Kopp, 2010)). For example, linguistic coordination has been reported with respect to words, phrase structures, speech rate, tones of voice, speech rhythms, etc. (cf. Chartrand, Maddux, and Lakin (2005); Branigan, Pickering, Pearson, and McLean (2010)). Pickering and Garrod (2004) ascribed this linguistic alignment largely to an automatic priming of interlocutors lexical, syntactic, or semantic representations and a percolation of activation between adjacent representational levels. Others, e.g. (Brennan & Clark, 1996), suggested that speakers strategically design utterances for an addressee and thereby prefer previously used (grounded) constructions. Regarding nonverbal behavior, interactants can likewise be found to mimic each other, e.g., in posture, body movements like foot shaking, mannerism, or facial expressions (cf. Chartrand and Bargh (1999); Lakin and Chartrand (2003)). This kind of mimicry is assumed to be largely non-conscious and automatic, being mediated by perception-action links that involve the own motor system in the perception of others actions (Dijksterhuis & Bargh, 2001). Only recently, researchers have started to look at whether speakers also align in their *co-speech* gestures, i.e., the spontaneous and meaningful hand movements that accompany speech. Such gestures stand out as they are very closely linked to the speech they accompany, in both content and timing (McNeill, 1992). Investigating whether speakers align in and via their gestures can thus help, first, to understand what shapes these gestures and, second, to shed light on the role of interpersonal coordination in dialogical communication.

Figure 1: Example of alignment of two successive gestures (left: router's assertion; right: follower's acknowledgement).

In this paper we present results from the first large-scale investigation of gestural alignment in natural dialogue. Fig. 1 shows an example, in which some properties of the first gesture (left) are being mimicked (e.g., handshape, trajectory) while others are not (e.g. relative movement direction). This suggests a feature-based, multi-level analysis of gestural alignment. We will thereby focus on the *form-based aspects* of gestural alignment here. In the next section we review the few existing studies that have looked at occurrences of form convergence in co-speech gestures, so far. Then we present the corpus data and the approach taken to investigate the phenomenon of gestural alignment in it, and present results of three analyses meant to answer the following questions: (1) Is there gestural alignment, compared against a baseline, and are there differences between different form features of a gesture? (2) What is stronger, the influence of the interlocutor's gestures or of one's own previous gesturing? (3) How does gestural alignment depend on (temporal) distance between the gestures? Finally, we discuss our findings in light of these questions and draw conclusions.

Related Work

Based on initial evidence by Kimbara (2006), who reported a couple of examples of gesture form convergence among interlocutors, some recent studies addressed the phenomena of gestural alignment (in this context often termed 'mimicry') more deeply. Parrill and Kimbara (2006) investigated the question to what extent observing mimicry affects people's behavior. They found that participants who observed mimicry in a video-recorded interaction were subsequently more likely to reproduce the mimicked behavior in their own descriptions, whereby a gesture was assessed as a reproduction if it corresponded with the stimulus gesture in handshape, motion and location.

In a similar setting, Mol, Krahmer, Maes, and Swerts (2012) provided evidence for the alignment of handshapes in co-speech gestures: Participants who saw a speaker in a video stimulus using gestures with a particular handshape were more likely to produce gestures with these handshapes later on, while retelling the story. This evidence is, however, limited to a particular kind of gestures ('path gestures' in directions), distinguishing between two different handshape classes (index finger extended vs. more than one finger extended). Mol et al. further addressed the role of meaning in this context. They found that gesture forms were only repeated across speakers if they had occurred in a meaningful context as expressed in concurrent speech. It is concluded that gesture form adaptation resembles adaptation in speech, rather than it being an instance of automated motor mimicry.

Kimbara (2008) studied triadic interaction with two co-narrators providing a joint narration to a third person, while manipulating the mutual visibility between co-narrators. Greater convergence in one gesture form feature (handshape) was found when participants could see each other. However, in this setting the two narrators were required to provide a coherent description for the recipient which might enhance the likability for gesture form convergence. Holler and Wilkin (2011) showed that gesture mimicry also occurs in face-to-face dialogue. In repeated references to the same figure-like stimuli, participants were found to be more likely to use similar gestures when they could see each other (vs. a non-visible condition). Holler & Wilkin concluded that gestures seem to play an active role in the process of grounding, because the vast majority of mimicked gestures occurred in phrases devoted to the presentation or acceptance of information.

In sum, existing studies lend considerable evidence that a speaker's gesture use is influenced by others' gestures. However, this quantitative empirical evidence is limited to experimental settings with video-based stimuli or elicited repeated references–an caveat often put forward against studies in linguistic alignment (cf. Howes, Healey, and Purver (2010)). To date, there is no analysis of gestural alignment based on a large sample of naturalistic dialogue data. The present study aims to close this gap.

Present Study

We have conducted statistical analyses on a large data corpus of spontaneous speech and gesture in dialogue (SaGA corpus (Lücking, Bergmann, Hahn, Kopp, & Rieser, 2010)). With these analyses we aimed for a systematic investigation of gesture form convergence going beyond previous studies in several respects. First of all, our corpus provides a detailed coding of the gestures' physical form including handshapes, palm and finger orientation, wrist movement, and position. This allows for addressing the degree to which *single* gesture form features are sensitive to influences of an interlocutor's gestures, instead of considering the "same overall form" (Holler & Wilkin, 2011), one particular form feature only (Kimbara, 2008), or the sum of several form features (Parrill & Kimbara, 2006). Second, some of the above mentioned experimental studies manipulated the visibility between interactants to create a baseline for gestural alignment occurring by chance. Investigating natural dialogues allows for an alternative baseline by creating artificial dialogues, as previously done in corpus analyses of linguistic alignment (Howes et al., 2010); for details see sect. 'Control Data'. Third, a characterizing feature for alignment in speech corpora is that the repetition probability is increased immediately after the prime and decreases toward the global mean with greater distances between prime and target (Reitter, 2008). A corpus analysis on extended dialogue allows to address this issue for gestural alignment, too. Fourth, we are able to investigate the *contingencies* involved in gestural alignment. Pickering and Garrod (2004) suggested to treat alignment not only as an *inter*-subjective phenomenon ('other-alignment'), but also *intra*-subjectively ('self-alignment'). Given the fact that the use of co-speech gestures is subject to major inter-individual differences (Hostetter & Alibali, 2007), the relationship between self- and other-alignment is important to assess the strength of inter-speaker gesture form convergence. Finally, the above mentioned studies have in common that they are limited to gestural alignment in repeated references to the same referent. Analyzing a large data sample allows to study the degree of gesture adaptation on a level of gesture form beyond the connection to specific referent objects, allowing to delineate grounding and mere motor resonances. In the following, we will briefly describe the corpus and explain how we framed the problem of detecting and measuring alignment between gestures occurring in dialogue.

Data Corpus The SaGA corpus consists of 25 dyads (21 female, 29 male participants) engaged in a spatial communication task combining direction-giving and sight description. This task required participants to convey the shape of objects and the spatial relations between them. The stimulus was an artificial town presented in a Virtual Reality environment, affording experimental control for the content of speaker messages. After taking a "bus ride" through the town, a router explained the route to an unknown and naïve follower. In total, the SaGA corpus consists of 280 minutes of video material containing 4449 iconic/deictic gestures. All dialogues are completely and systematically annotated based on an annotation grid, tested and refined using multi-coder agreement tests; for details see Lücking et al. (2010). Each gesture is demarcated by the beginning and end of the expressive, so-called 'stroke' phase. The gesture's form during the stroke phase, as far as relevant here, is annotated in terms of the following distinct form features: (1) HANDEDNESS: one-handed (either left- or right-handed), or two-handed gestures; (2) HANDSHAPE: ASL-based coding of hand configurations like ASL-B, ASL-C, etc. + modifiers (e.g. bent, loose); (3) PALM- AND FINGER ORIENTATION: up, down, sideways, towards body, away + combinations and sequences; and (4) WRIST MOVEMENT TYPE: static, linear, or curved + sequences. A further important and characterizing feature of an iconic gesture is the more general REPRESENTATION TECHNIQUE (e.g. Kendon (2004)). For the spatial domain of the SaGA dialogues, the following set of techniques proved to be adequate: indexing, placing (as if putting a virtual object somewhere), shaping (as if sculpting a 3D shape), drawing (as if sketching a 2D outline), and posturing (using the hand/arm as a model for something).

Prime-target Pairs From 4449 iconic/deictic gestures in the SaGA corpus, a total of 17130 prime-target pairs[1] were extracted. Figure 2 exemplifies the possible alignment-relevant influences between different prime-target pairs that can occur in dialogue. Each of these pairs is characterized by a distance DIST between prime and target gesture, taken to be the number of other gesture occurrences in-between plus 1. Each prime-target pair is further characterized by a CONTINGENCY TYPE: whether prime and target gestures are produced by the *same* speaker ('self-pair') or by *different* speakers ('other-pair'), respectively. In the SaGA corpus there are 17362 self-pairs and 3993 other-pairs.

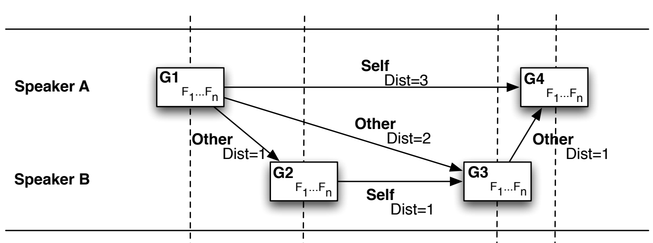

Figure 2: Possible alignment influences between gestures: Speaker A produces two gestures ($G1, G4$), while speaker B makes two gestures ($G2, G3$) in-between. Gestures are characterized by features ($F_1 \ldots F_n$) and can influence each other both within a speaker ('self') and across speakers ('other'), where the relation's distance (DIST) is determined by the occurrences of gestures in-between.

[1] We employ the term 'prime-target pair' in lack of a better one. This is not to imply that alignment is due to *priming*.

Control Data A common problem in studies on behavioral coordination is to lay down a baseline of how much coordination can occur simply by chance, regardless of any contingencies between primes and targets. Adopting the approach of Howes et al. (2010) in their corpus analyses of lexical and syntactic alignment in speech, we created 'fake' dialogues by re-combining the gestures of two speakers from originally different dialogues. This is done in an interleaved fashion, i.e., the whole sequence of gestures produced by one particular direction-giver is kept, but merged with the complete gesture sequence produced by a different direction-follower. This way we created 25 control dialogues with randomly chosen participants while respectively maintaining the participants' role (direction giver vs. direction follower). As a matter of course, although the total number of gestures remains the same, this results in a different number of prime-target pairs in the control data set with regard to CONTINGENCY TYPE: 16523 self-pairs and 2407 other-pairs.

Metric Considering gestural alignment necessitates to define a metric estimating the similarity between prime and target gesture. Since we want to be able to assess alignment even at the level of single features of a gesture, we define a metric for each particular gesture feature. To make results comparable with each other, we employ a binary metric for all variables: it scores 1 if prime and target gesture are identical in a particular gesture feature, and 0 otherwise. For some features this definition can be applied straightforwardly (e.g. HANDEDNESS: one-handed vs. two-handed), for others it is reasonable to allow some minor variation between prime and target gesture. Palm and finger orientation, for instance, are coded as combinations of five basic values (up, down, sideways, towards, away). That is, a palm orientation of 'down' and an orientation of 'down/away' would count as a mismatch although the actual difference in palm orientation is 45° which can be regarded a slight deviation given the natural fuzziness of human gesture use. Accordingly, the binary metric is applied to the gesture features as follows, whereby for features which allow sequential coding the final segment of the prime's value and the first segment of the target's value are considered:

- REPRESENTATION TECHNIQUE and HANDEDNESS: A score of 1 is given only if the values for prime and target gesture are identical, 0 otherwise.
- HANDSHAPE: Any modifiers of ASL handshapes like 'spread' or 'loose' are omitted, i.e. 'ASL-B-spread' and 'ASL-B-loose-spread' both fall into the basic category 'ASL-B'. A score of 1 is given only if prime and target are identical in this basic category for both hands, 0 otherwise.
- PALM AND FINGER ORIENTATION: A score of 1 is given if prime and target match in at least one part of the annotation value for both hands, 0 otherwise.

- WRIST MOVEMENT TYPE: A score of 1 is given if prime and target are identical or – in case of a two-handed gesture with different movement types – if the value for one hand is identical with the other gesture's value, 0 otherwise.

Results

Is there gestural alignment in dialogue? This analysis aims to show whether gesture use in real dialogues shows reliably more other-alignment than would occur by chance. To this end, we compare similarity scores in real vs. control dialogues for each gesture feature with a one-way analysis of variance for each of the gesture features. We only consider prime-target pairs with DIST=1 here, since it is more likely that alignment occurs in consecutive gestures than in more distant pairs. With regard to Figure 2 this means that we take prime-target pairs like (G1,G2) or (G3,G4) into account (N=950 pairs; 579 from the original data, and 371 from control dialogues). Exact means and standard deviations are given in Table 1.

For REPRESENTATION TECHNIQUE ($F_{(1,948)} = 24.61$, $p < .001$), HANDSHAPE ($F_{(1,948)} = 17.92$, $p < .001$), and PALM ORIENTATION ($F_{(1,948)} = 6.65$, $p = .01$), there is a reliable difference between the two groups such that the mean similarity in control dialogues is significantly lower than in real dialogues. For HANDEDNESS ($F_{(1,948)} = 3.47$, $p = .063$) the analysis marginally fails to reach significance, but by trend the mean similarity in control dialogues is significantly lower than in real dialogues. For FINGER ORIENTATION ($F_{(1,948)} = .16$, $p = .69$) and WRIST MOVEMENT TYPE ($F_{(1,948)} = .06$, $p = .94$) the analysis shows no significant main effect between real and control data.

This means that there exists other-alignment in gesture use, but not all gesture features are subject to this effect. Only for the features REPRESENTATION TECHNIQUE, HANDSHAPE, PALM ORIENTATION, and HANDEDNESS the mean similarity of prime and target is higher as to be expected by chance. We continue with these features to a finer analysis.

Table 1: Mean similarity of gesture features for real and control dialogues (standard deviations in parentheses).

	Real	Control
Representation Technique	.31 (.37)	.16 (.46)
Handedness	.68 (.47)	.62 (.49)
Handshape	.37 (.48)	.24 (.43)
Palm Orientation	.49 (.50)	.41 (.50)
Finger Orientation	.61 (.49)	.60 (.49)
Wrist Movement Type	.40 (.49)	.40 (.49)

Self- vs. Other-Alignment? To compare the effects of self- and other-alignment, we now investigate the difference between prime-target pairs in the same speaker (CONTINGENCY TYPE = 'self') vs. prime-target pairs with different speakers (CONTINGENCY TYPE = 'other'). We only consider adjacent prime-target pairs with DIST=1 (for instance, in Figure 2 self-pairs like (G2,G3) with other-pairs like (G1,G2). The total number of pairs amounts to N=4317 (3738 self-pairs, and 579 other-pairs). Again we employ a one-way analysis of variance, exact means and standard deviations are given in Table 2.

For all variables under consideration the analysis reveals significant main effects: REPRESENTATION TECHNIQUE ($F_{(1,4315)} = 25.05$, $p < .001$), HANDSHAPE ($F_{(1,4315)} = 51.86$, $p < .001$), HANDEDNESS ($F_{(1,4315)} = 39.38$, $p < .001$), PALM ORIENTATION ($F_{(1,4315)} = 67.95$, $p < .001$). These effects are due to the fact that mean similarity of prime and target are higher for self-pairs than in other-pairs. That is, the alignment between gestures is reliably stronger *within* speakers than it is *across* speakers.

Table 2: Mean similarity of gesture features for self- and other-speaker pairs (standard deviations in parentheses).

	Self	Other
Representation Technique	.41 (.49)	.31 (.46)
Handedness	.79 (.40)	.68 (.46)
Handshape	.53 (.50)	.37 (.48)
Palm Orientation	.67 (.47)	.49 (.50)

Effect of temporal distance on other-alignment? To elucidate how gestural alignment is affected by temporal distance, we analyze how the similarity score depends on the distance between prime and target gestures. For this analysis we consider other-pairs of distance 1-4. In Figure 2 examples of other-pairs with DIST=1 would be (G1,G2) or (G3,G4), examples of an other-pair with DIST=2 are (1,3) or (2,4). We employ a one-way analysis of variance for the dependent variable SIMILARITY and the independent variable DIST. A total of 3081 primed-target pairs is analyzed (N(DIST=1)=579, N(DIST=2)=758, N(DIST=3)=843, N(DIST=4)=901).

For REPRESENTATION TECHNIQUE there is a main effect of DIST and similarity score $F_{(3,3077)} = 6.22$, $p < .001$): the similarity score is smaller the greater the distance between prime and target. This is due to significant differences between prime-target pairs with DIST=1 and others (DIST=2: $t_{(1335)} = 1.96$, $p = .05$; DIST=3: $t_{(1420)} = 2.51$, $p = .012$; DIST=4: $t_{(1478)} = 4.30$, $p < .001$), as well as between distances 2 and 4 ($t_{(1657)} = 2.40$, $p = .017$). Likewise, for HANDSHAPE the similarity scores decrease significantly with increasing distance between prime and target DIST ($F_{(3,3077)} = 7.10$, $p < .001$). This is due to the fact that the similarity of prime-target pairs with DIST=1 is higher than for prime-target pairs with higher distances (DIST=2: $t_{(1335)} = 2.63$, $p = .09$; DIST=3: $t_{(1420)} = 3.37$, $p = .001$; DIST=4: $t_{(1478)} = 4.51$, $p < .001$). In contrast, for HANDEDNESS ($F_{(3,3077)} = .045$, $p = .99$), and PALM ORIENTATION ($F_{(3,3077)} = .41$, $p = .75$) there is no main effect of distance between prime and target gesture.

That is, the more gestures occur between prime and target, the smaller is their similarity with respect to REPRESENTATION TECHNIQUE and HANDSHAPE, which is corroborated when checking the actual temporal distances in milliseconds.

By contrast, the similarity score remains more or less constant for the features HANDEDNESS and PALM ORIENTATION.

Table 3: Mean similarity for varying distances between prime and target gesture (standard deviations in parentheses).

	DIST=1	DIST=2	DIST=3	DIST=4
Representation Technique	.31 (.46)	.26 (.44)	.25 (.43)	.21 (.41)
Handedness	.68 (.47)	.67 (.47)	.68 (.47)	.67 (.47)
Handshape	.37 (.48)	.30 (.46)	.29 (.45)	.26 (.44)
Palm Orientation	.49 (.50)	.49 (.50)	.48 (.50)	.47 (.50)

Together with results from the first analysis that revealed highest *F*-scores for the former two features in comparison with control data, this provides the following picture: For REPRESENTATION TECHNIQUE and HANDSHAPE there is a strong other-alignment effect which decreases with greater distances from the prime gesture. For HANDEDNESS and PALM ORIENTATION there is a rather weak difference when comparing original and control data and the effect is more or less constant. In other words, there seems to be a general (weak) tendency to produce gestures with a certain amount of similarity in these two features, but this is not biased by the other's directly preceding gesture use.

Discussion

In this paper we reported results from the first fine-grained and systematic analysis of alignment in co-speech (iconic) gesturing in natural direction-giving dyads. What did we find? First, there is significant gestural alignment in dialogue. That is, a speaker's use of co-speech gestures is affected by the other's gestures in the dialogue. Remarkably, not all gesture features seem to be equally sensitive, with WRIST MOVEMENT and FINGER ORIENTATION being most resistant. Second, alignment effects are significantly stronger *within* speakers than *across* speakers. That is, a speaker's gestures influence each other more than the gestures the interlocutor performs, albeit the effectiveness of other-alignment. Third, regarding the relation between the strength of other-alignment and the prime-target distance, a multi-faceted picture emerges: alignment in handshape or representation technique becomes weaker with greater distances, while alignment in handedness and palm orientation remain constant.

These findings can shed new light on iconic co-speech gestures, as well as the cognitive processes underlying their production in dialogue. To start with, how can we make sense of the heterogeneity of feature-based gestural alignment? A closer look at the role of gestural representation techniques might be informative. Each of these techniques is characterized by a specific pattern of how meaning is depicted. For example, in drawing gestures as in Figure 1 it is the *wrist trajectory* that conveys most of the intended meaning, while in indexing or placing gestures the *position* of the hands is of major importance to convey meaning. That is, some features can be considered more communicatively significant than others. Indeed, variation within the less significant features has also been reported to reflect individual gesturing style (Bergmann, 2012). Our results here suggest that those features, like handshape, are also more amenable to inter-personal coordination, while the communicatively more significant features tend to be more resistant. This suggests a notion of gestural alignment as adaptation within the degrees of freedom available under given communicative constraints.

Existing cognitive models of speech and gesture production lack an account of other-alignment. However, our findings suggest a row of implications for such models. At first it is important to note that our analysis of gestural alignment at the level of different features supports a view that a gesture is not produced as a whole, but in different steps that can exert influences over the constitution of different features of a gesture. We thus hypothesize that different gesture form features are determined at different points in time with other-alignment arising potentially from high-level mechanisms in terms of full grounding, i.e. signaling established links between form and meaning, as well as low-level mechanisms of priming or motor resonance (Montgomery, Isenberg, & Haxby, 2007). That is, such a model provides (at least) two routes than could mediate alignment. However, our finding that communicatively significant features are less affected by alignment may be a consequence of a hypothesized more "ego-centric" nature of early, high-level stages of the production process, which may be more concerned with the necessary, communicatively intended functions and less with corrective or audience-design functions (Keysar & Henly, 2002). Lower-level processes, on the other hand, involve connected sensory and motor processes which have been shown to be effective also in gesture (Montgomery et al., 2007) and have often been assumed to mediate interpersonal coordination. Our empirical findings suggest that the sensorimotor route is particularly effective. To further distinguish between the two routes, we are concerned with measuring the degree of form similarity in relation to referent similarity in ongoing work.

Our observation of strong self-alignment effects may be explained by a strong role of internal priming or caching in the cognitive speech-gesture production process. This conforms hypotheses of self-routinization or *expert performance* effects, which state that over repeated encounters with a particular problem, memory traces build up that directly map a problem stimulus to a solution (e.g., Logan (1988)). It is important to note, however, that gesture use also seems to be subject to adaptations taking place in extended dialogue with repeated references. Such processes have been found, e.g., to lead to considerable *reduction* of the complexity of speech and gestures (Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2011). That is, gestures can become simpler or less precise over repeated uses when referring to the same entity. Further research is needed to elucidate how different mechanisms and driving forces (e.g., alignment through repetition vs. simplification through reduction) compete and interact with each other.

With all these raised questions, we think that computational simulation can provide a valuable tool to test whether different kinds of cognitive mechanisms result in the effects we observe empirically. We have developed a computational model for speech and gesture production in previous work (Kopp, Bergmann, & Wachsmuth, 2008). Opening this to the effects of dialogical interaction, e.g. endowing it with perceptive abilities, will enable us to complement this empirical work with computational studies.

Acknowledgments

This research is partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center 673 "Alignment in Communication" and the Center of Excellence in "Cognitive Interaction Technology" (CITEC).

References

Bergmann, K. (2012). *The production of co-speech iconic gestures: Empirical study and computational simulation with virtual agents.* PhD Thesis: Bielefeld University.

Branigan, H., Pickering, M., Pearson, J., & McLean, J. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics*, *42*, 2355—2368.

Brennan, S., & Clark, H. (1996). Lexical choice and conceptual pacts in conversation. *Journal of Experimental Psychology*, *22*(6), 1482—1493.

Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*, 893–910.

Chartrand, T., Maddux, W., & Lakin, J. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of unconscious mimicry. In R. Hassin, J. Uleman, & J. Bargh (Eds.), *The new unconscious* (pp. 334—361). New York: Oxford University Press.

Dijksterhuis, A., & Bargh, J. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, *33*, 1–40.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2011). Greebles greeble greeb: On reduction in speech and gesture in repeated references. In *Proc. of the 33rd annual conference of the cognitive society.*

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

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.

Hostetter, A., & Alibali, M. (2007). Raise your hand if you're spatial—relations between verbal and spatial skills and gesture production. *Gesture*, *7*, 73–95.

Howes, C., Healey, P., & Purver, M. (2010). Tracking lexical and syntactic alignment in conversation. In *Proc. of the 32nd annual conference of the cognitive science society.*

Kendon, A. (2004). *Gesture—visible action as utterance*. Cambridge University Press.

Keysar, B., & Henly, A. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, *13*, 207–212.

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

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

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Kopp, S. (2010). Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, *52*, 587–597.

Kopp, S., Bergmann, K., & Wachsmuth, I. (2008). Multimodal communication from multimodal thinking—towards an integrated model of speech and gesture production. *Semantic Computing*, *2*(1), 115–136.

Kopp, S., Tepper, P., Ferriman, K., Striegnitz, K., & Cassell, J. (2007). Trading spaces: How humans and humanoids use speech and gesture to give directions. In *Conversational informatics* (pp. 133–160). New York: John Wiley.

Lakin, J., & Chartrand, T. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*, *14*, 334—339.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010). The Bielefeld Speech and Gesture Alignment corpus (SaGA). In *Proceedings of the LREC 2010 workshop on multimodal corpora.*

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

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

Montgomery, K., Isenberg, N., & Haxby, J. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Social Cognitive and Effective Neuroscience*, *2*(2), 114-122.

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

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–226.

Reitter, D. (2008). *Context effects in language production: Models of syntactic priming in dialogue corpora.* PhD Thesis: University of Edinburgh.