# Regulating Dialogue with Gestures—Towards an Empirically Grounded Simulation with Conversational Agents

**Kirsten Bergmann**[1,2] **Hannes Rieser**[1] **Stefan Kopp**[1,2]

[1] Collaborative Research Center 673 „Alignment in Communication“, Bielefeld University
[2] Center of Excellence „Cognitive Interaction Technology“(CITEC), Bielefeld University

{kbergman,skopp}@TechFak.Uni-Bielefeld.DE
hannes.rieser@Uni-Bielefeld.DE

## Abstract

Although not very well investigated, a crucial aspect of gesture use in dialogues is to regulate the organisation of the interaction. People use gestures decisively, for example to indicate that they want someone to take the turn, to 'brush away' what someone else said, or to acknowledge others' contributions. We present first insights from a corpus-based investigation of how gestures are used to regulate dialogue, and we provide first results from an account to capture these phenomena in agent-based communication simulations. By advancing a model for autonomous gesture generation to also cover gesture interpretation, this account enables a full gesture turn exchange cycle of generation, understanding and acceptance/generation in virtual conversational agents.

## 1 Motivation

Research on gestures must combine empirical, theoretical and simulation methods to investigate form, content and function of gestures in relation to speech. Our work is based on a corpus of multimodal data, the Bielefeld Speech and Gesture Alignment corpus of route-description dialogues (SAGA corpus, Lücking et al. 2010). The point of departure of our research has been work on iconic and deictic gestures over many years. In this paper we focus on a not very well investigated function of gestures which we have repeatedly observed in this corpus, namely, the regulation of dialogue.

Most of current gesture research is oriented towards the semiotics of a Peircean tradition as can for instance be seen from McNeill's "Kendon's continuum" (McNeill 1992, p. 37). As a consequence of this Peircian orientation, gestures have been viewed as single signs interfacing with speech. Going beyond the integration of input/output modalities in *single* speech-gesture compositions (Johnston and Bangalore, 2005), little effort has been spent on the investigation of *sequences* of gestures and speech-gesture composition both within and across speakers (Hahn and Rieser 2010, Rieser 2010). Furthermore, research of gesture meaning was restricted to the contribution of gesture content to propositional content. An exception to this research line has been the work of Bavelas et al. (1992, 1995). It is characterised by two features, a functional perspective on gesture in opposition to purely classificatory and typological ones and an interest to systematically investigate the role of gesture in interaction. In particular, Bavelas et al. (1992) proposed a distinction between 'topic gestures' and 'interactive gestures': Topic gestures depict semantic information directly related to the topic of discourse, while interactive gestures refer to some aspect of the process of conversing with another person. Interactive gestures include *delivery* gestures (e.g. marking information status as new, shared, digression), *citing* gestures (acknowledging others' prior contributions), *seeking* gestures (seeking agreement, or help in finding a word), and *turn coordination* gestures (e.g. taking or giving the turn). Gill et al. (1999) noted similar functions of gesture use, adding body movements to the repertoire of pragmatic acts used in dialogue act theory (e.g. turn-taking, grounding, acknowledgements).

We aim to find out how gestures are related to and help regulate the structure of dialogue. We will call these gestures `discourse gestures´. Relevant research questions in this respect are the following: How can gesture support next speaker selection if this follows regular turn distribution mechanisms such as current speaker selects next? From the dialogues in SAGA we know that averting next speaker's self-selection is of similar importance as handing over the floor to the next speaker. So, how can averting self-selection of other be accomplished gesturally? A still different problem is how gesture is utilised to establish an epistemically transparent, reliable common ground, say a tight world of mutual belief. A precondition for that is how gesture can help to indicate a gesturer's stance to the information he provides. Natural language has words to indicate degrees of confidence in information such as *probably*, *seemingly*, *approximately*, *perhaps, believe, know, guess* etc. Can gestures acquire this function as well?

All these issues can be synopsised as follows: How can gestures—apart from their manifest contribution to propositional content—be used to push the dialogue machinery forward? In our research, gesture simulation and theory of speech-gesture integration are developed in tandem. Up to now, both have been tied to occurrences of single gestures and their embedding in dialogue acts. In this paper, we present first steps along both methodological strands to explore the use and function of gesture in dialogue. We start with an empirical perspective on discourse gestures in section 2. In section 3 we briefly describe our gesture simulation model which so far simulates gesture use employing the virtual agent MAX independent of discourse structures. Section 4 analyses a corpus example of a minimal discourse which is regulated mainly by gestures of the two interactants. This provides the basis for our proposed extension of the gesture generation approach to capture the discourse function of gestures as described in section 5. This extension will encompass a novel approach to employ the very generation model used for gesture production, and hence all the heuristic gesture knowledge it captures, also for gesture interpretation in dialogue. Section 6 discusses the difference between pure interactive gestures and discourse gestures and proposes further steps that need to be taken to elucidate how gestures are used as a vehicle for regulating dialogue.

## 2 Empirical Work on Discourse Gestures

In looking for discourse gestures we started from the rated annotation of 6000 gestures in the SAGA corpus. We managed to annotate and rate about 5000 of them according to traditional criteria using practices and fine-grained gesture morphology like hand-shape and wrist-movement. About 1000 gestures could not be easily subsumed under the traditional gesture types (iconics, deictics, metaphorics, beats). Furthermore, they were observed to correlate with discourse properties such as current speaker's producing his contribution or non-regular interruption by other speaker.

For purposes of the classification of the remaining 1000 gestures we established the following functional working definition: `Discourse gestures´ are gestures tied up with properties or functions of agents' contributions in dialogue such as successfully producing current turn, establishing coherence across different speakers' turns by gestural reference or indicating who will be next speaker.

What did we use for dialogue structure? Being familiar with dialogue models such as SDRT (Asher and Lascarides, 2003), PTT (Poesio and Traum, 1997), and KoS (Ginzburg, 2011) we soon found that these were too restricted to serve descriptive purposes. So we oriented our "classification of dialogue gesture enterprise" on the well known turn taking organisation model of Sacks et al. (1974) and Levinson's (1983) discussion of it. However, it soon turned out that even these approaches were too normative for the SAGA data: This is due to the fact that dialogue participants develop enormous creativity in establishing new rules of content production and of addressing violations of *prima facie* rules.

Rules of turn-taking, for example, are not hard and fast rules, they can be skirted if the need arises, albeit there is a convention that this has to be acknowledged and negotiated. A very clear example of an allowed interruption of an on-going production is a quickly inserted clarification request serving the communicative goals of current speaker and the aims of the dialogue in general. Another problem with the Sacks et al. model consists in the following fact: Since its origination many dialogue regularities have been discovered which cannot be easily founded on a phenomenological or observational stratum which is essentially semantics-free. This can for example be seen from the develop-

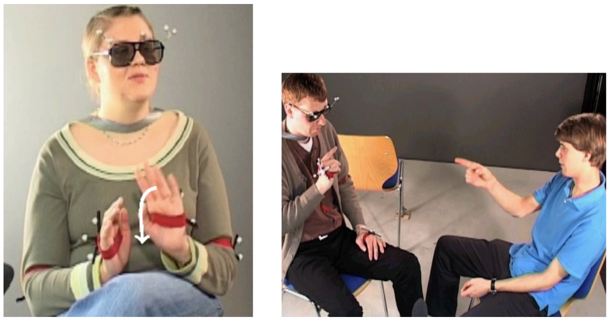

**Figure 1:** Examples of discourse gestures: the brush-away gesture (left) and situated pointing to the upper part of the interlocutor's torso (right) used for next speaker selection in a "Gricean" sense (see text for explanation).

ment of the notion of grounding and common ground as originally discussed by Stalnaker (1978), Clark (1996) and others. Nevertheless, grounding (roughly, coming to agree on the meaning of what has been said (see e.g. Traum, 1999; Roque and Traum, 2008; Ginzburg 2011, ch. 4.2 for the options available) generates verbal structure and verbal structure interfaces with gesture. Other examples in this class are acknowledgements or accepts discussed in more detail below.

How did we decide on which distinctions of gesture annotation have to be used for characterising discourse gestures? In other words, how did we conceive of the map between gestures of a certain sort and discourse structures? First of all we observed that two types of discourse gestures emerge from the SAGA data. Some of them come with their own global shape and are close to emblems, (i.e. conveyors of stable meaning like the victory sign). This is true for example of the "brush aside or brush away" gesture shown in Figure 1 (left), indicating a gesturer's assessment of the downrated relevance of information, actions or situations.

Discourse gestures of the second class exploit the means of, for instance, referring gestures or iconic gestures. An example of an iconic gesture in this role will be discussed to some extent in section 4. Its simulation will be described in sections 3 and 5. Here we explain the phenomenon with respect to referring pointing gestures which are easier to figure out (see Figure 1 (right)). Their usage as under focus here is not tied to the information under discussion but to objects in the immediate discourse situation, preferably to the participants of the dialogue. These uses have a Gricean flavour in the following way: Only considerations of relevance and co-occurrence with a turn transition relevance place together indicate that *prima facie* not general reference is at stake but indication of next speaker role. It wouldn't make sense to point to the other person singling her or him out by indexing, because her or his identity is clear and well established through the on-going interaction. Thus we see that a gestural device associated with established morphological features, pointing, acquires a new function, namely indicating the role of next-speaker.

Now both classes of gestures, "brush away" used to indicate informational or other non-relevance and pointing, indicating the role of being next speaker exploit the motor equipment of the hands. For this reason, annotation of discourse gestures can safely be based on the classification schemas we have developed for practices like indexing, shaping or modelling and for the fine-grained motor behaviour of the hands as exhibited by palm orientation, back-of-hand trajectory etc. In work by Hahn & Rieser (2009-2011) the following broad classes of discourse gestures were established. We briefly comment upon these classes of gestures found in the SAGA corpus relevant for dialogue structure and interaction:

- **Managing of own turn**: A speaker may indicate how successful he is in editing out his current production.
- Mechanisms of **next-speaker selection** as proposed in classical CA research, for instance, pointing to the other's torso is often used as a means to indicate next speaker.
- In **grounding acts and feed-back** especially iconic gestures are used to convey propositional content.
- **Clarification requests** to work on contributions: An addressee may indicate the need for a quick interruption using a pointing to demand a clarification. In contrast, a current speaker can ward off the addressee's incipient interruption using a palm-up gesture directed against the intruder thus setting up a "fence".
- **Evidentials for establishing a confidence leve:** There are fairly characteristic gestures indicating the confidence a speaker has in the information he is able to convey.
- **Handling of non-canonical moves by discourse participants**: Interaction sequences consisting of attempts by other speaker to interrupt and to thwart this intention by current speaker or to give way to it show how

discourse participants handle non-canonical moves.

- **Assessment of relevance by discourse participants**: Speakers provide an assessment of which information is central and which one they want to consider as subsidiary.
- An **indication of topical information with respect to time, place or objects** is frequently given by pointing or by "placing objects" into the gesture space.

We know that this list is open and could, moreover, depend on the corpus. In this paper the focus will be on grounding acts and feedback (see sections 3-5). The reason is that this way we can provide an extension of existing work on the simulation of gesture production in a fairly direct manner.

## 3 Simulating Gesture Use: The Generation Perspective

Our starting point to simulate gestural behavior in dialogue is a gesture generation system which is able to simulate speaker-specific use of iconic gestures given (1) a communicative intention, (2) discourse contextual information, and (3) an imagistic representation of the object to be described. Our approach is based on empirical evidence that iconic gesture production in humans is influenced by several factors. Apparently, iconic gestures communicate through iconicity, that is their physical form depicts object features such as shape or spatial properties. Recent findings indicate that a gesture's form is also influenced by a number of contextual constraints such as information structure (see for instance Cassell and Prevost, 1996), or the use of more general gestural representation techniques such as shaping or drawing is decisive. In addition, inter-subjective differences in gesturing are pertinent. There is, for example, wide variability in how much individuals gesture when they speak. Similarly, inter-subjective differences are found in preferences for particular representation techniques or low-level morphological features such as handshape or handedness (Bergmann & Kopp, 2009).

To meet the challenge of considering general and individual patterns in gesture use, we have proposed GNetIc, a gesture net specialised for iconic gestures (Bergmann & Kopp, 2009a), in which we

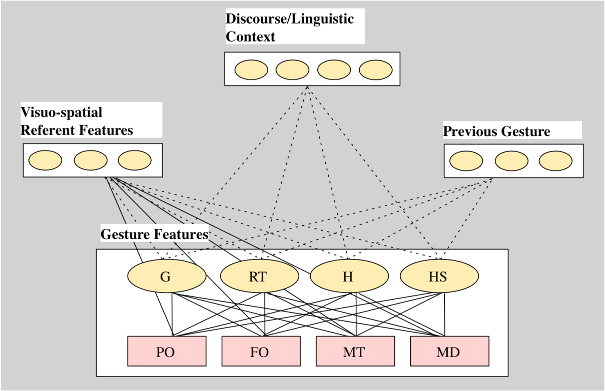

**Figure 2:** Schema of a gesture generation network in which gesture production choices are considered either probabilistically (chance nodes drawn as ovals) or rule-based (decision nodes drawn as rectangles). Each choice is depending on a number of contextual variables. The links are either learned from speaker-specific corpus data (dotted lines) or defined in a set of if-then rules (solid lines).

model the process of gesture formulation with Bayesian decision networks (BDNs) that supplement standard Bayesian networks by decision nodes. This formalism provides a representation of a finite sequential decision problem, combining probabilistic and rule-based decision-making. Each decision to be made in the formation of an iconic gesture (e.g., whether or not to gesture at all or which representation technique to use) is represented in the network either as a decision node (rule-based) or as a chance node with a specific probability distribution. Factors which contribute to these choices (e.g., visuo-spatial referent features) are taken as input to the model (see Figure 2) The structure of the network as well as local conditional probability tables are learned from the SAGA corpus by means of automated machine learning techniques and supplemented with rule-based decision making. Individual as well as general networks are learned from the SAGA corpus by means of automated machine learning techniques and supplemented with rule-based decision making. So far, three different factors have been incorporated into this model: discourse context, the previously performed gesture, and features of the referent. The latter are extracted from a hierarchical representation called Imagistic Description Trees (IDT), which is designed to cover all decisive visuo-spatial features of objects one finds in iconic gestures (Sowa & Wachsmuth, 2009). Each node in an IDT contains an imagistic description which holds a schema representing the shape of an object or object part. Features extracted from this representation in order to capture the main characteristics of a gesture's referent are whether an object can be decomposed into detailed subparts (whole-part relations), whether it has any symmetrical axes, its main axis, its position in the VR stimulus, and its shape properties extracted on the basis of so called multimodal concepts (see Bergmann & Kopp, 2008).

Analyzing the GNetIc modelling results enabled us to gain novel insights into the production process of iconic gestures: the resulting networks for individual speakers differ in their structure and in their conditional probability distributions, revealing that individual differences are not only present in the overt gestures, but also in the production process they originate from.

The GNetIc model has been extensively evaluated. First, in a prediction-based evaluation, the automatically generated gestures were compared against their empirically observed counterparts, which yielded very promising results (Bergmann & Kopp, 2010). Second, we evaluated the GNetIc models in a perception-based evaluation study with human addressees. Results showed that GNetIc-generated gestures actually helped to increase the perceived quality of object descriptions given by MAX. Moreover, gesturing behaviour generated with individual speaker networks was rated more positively in terms of likeability, competence and human-likeness (Bergmann, Kopp & Eyssel, 2010).

GNetIc gesture formulation has been embedded in a larger production architecture for speech and gesture production. This architecture comprises modules that carry out content planning, formulation, and realisation for speech and gesture separately, but in close and systematic coordination (Bergmann & Kopp, 2009). To illustrate gesture generation on the basis of GNetIc models, consider the following example starting upon the arrival of a message which specifies the communicative intent to describe the landmark townhall with respect to its characteristic properties:

```
lmDescrProperty (townhall-1).
```

Based on this communicative intention, the imagistic description of the involved object gets activated and the agent adopts a spatial perspective towards it from which the object is to be described (see Figure 3). The representation is analyzed for referent features required by the GNetIc model: position, main axis, symmetry, number of subparts, and shape properties. Regarding the latter, a unification of the imagistic townhall-1 representation and a set of underspecified shape property representations (e.g. for „longish“, „round“ etc.) reveals „U-shaped“ as the most salient property to be depicted. All evidence available (referent features, discourse context, previous gesture and linguistic context) is propagated through the network (learned from the data of one particular speaker) resulting in a posterior distribution of probabilities for the values in each chance node.

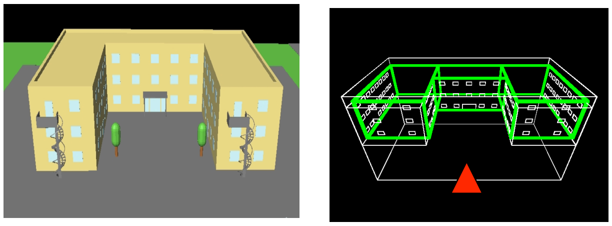

**Figure 3**: The townhall in the virtual world (left) and schematic of the corresponding IDT content (right); activated parts are marked.

This way, it is first decided to generate a gesture in the current discourse situation at all, the representation technique is decided to be „drawing“, to be realized with both hands and the pointing handshape ASL-G. Next, the model's decision nodes are employed to decide on the palm and back of hand (BoH) orientation as well as movement type and direction: as typical in drawing gestures, the palm is oriented downwards and the BoH away from the speaker's body. These gesture features are combined with a linear movement consisting of two segments per hand (to the right and backwards with the right hand; accordingly mirror-symmetrical with the left hand) to depict the shape of the townhall.

Accompanying speech is generated from selected propositional facts using an NLG engine. Synchrony between speech and gesture follows co-expressivity and is set to hold between the gesture stroke (depicting the U-shape property) and corresponding linguistic element. These values are used to fill the slots of a gesture feature matrix which is transformed into an XML representation to be realized with the virtual agent MAX (see Figure 4).

**RH (mirSym)**

| | |
|---|---|
| MOVEMENT TYPE: | *linear* |
| MOVEMENT DIR: | *right > back* |
| PALM ORIENT: | *down* |
| BoH ORIENT: | *away* |
| HANDSHAPE: | *ASL-G* |
| LOC*:* | *LocUpperChest*<br>*LocCCenter*<br>*LocNorm* |

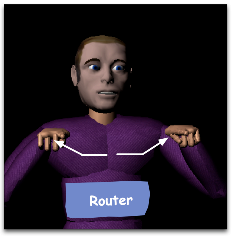

**Figure 4**: Specification (left) and realization (right) of an autonomously generated drawing gesture which depicts the U-shaped townhall.

## 4 Example of a Minimal Discourse

To start with the analysis of how gestures are not only employed to carry referential content but also to regulate dialogue and discourse, we first present a datum from the SAGA corpus showing how the Follower's gesture aligns with the Router's gesture to indicate acknowledgement or accept. The situation is as follows: the Router describes to the Follower that he would approach the town-hall and how it looks to him. A transcription of the initial dialogue passage by the Router and the subsequent crucial speech-gesture annotation, including the Follower, in ELAN looks as displayed in Figure 5 (*placing*, *drawing*, and *shaping* are names of annotated gestural representation techniques).

A short comment on the data might be in order: When introducing the townhall as a U-shaped building, the Router draws the boundary of it, namely a "U". He then goes on to describe how the on-looker apprehends the building. This is accompanied by a forward-oriented direction gesture with both hands, mimicking *into it*. In principle, all the information necessary to identify the townhall from a front perspective is given by then. There is a short pause and we also have a turn transition relevance place here. However, there is no feedback by the Follower at this point. Therefore the Router selects a typical pattern for self-repairs or continuations in German, a *that is* construction in the guise of a propositional apposition. Overlapping the production of *kind*, he produces a three-dimensional partial U-shaped object maintaining the same perspective as in his first drawing of the U-shaped border.

Observe that the Follower already gives feedback after *front*. The most decisive contribution is the Follower's acknowledgement, however. She imitates the Router's gesture but from her perspective as a potential observer. Also, at the level of single form features, she performs the gesture differently. (different movement direction, different symmetry) The imitating gesture overlaps with her nod and her contribution *OK*. It is important to see that her gesture provides more than a repetition of the word *townhall* could possibly give. It refers at the same time to the town-hall (standing for a discourse referent) and provides the information of a U-shape indicating property, in other words, it expresses the propositional information "This building being U-shaped" with *this building* acting as a definite anaphora to the occurrence of *a building* in the first part of the Router's contribution. Hence, assessed from a dialogue perspective the following happens: The grounding process triggered by the Follower's acknowledgement amounts to mutual belief among Router and Follower that the town hall is U-shaped

Router: Das ist dann das Rathaus [placing].
*This is then the townhall [placing].*
Das ist ein u-förmiges Gebäude [drawing].
*That is a U-shaped building [drawing].*
Du blickst praktisch da rein [shaping].
*You look practically there into it [shaping].*
Das heisst, es hat vorne so zwei Buchtungen.
*That is, it has to the front kind of two bulges.*
und geht hinten zusammen dann.
*and closes in the rear then.*

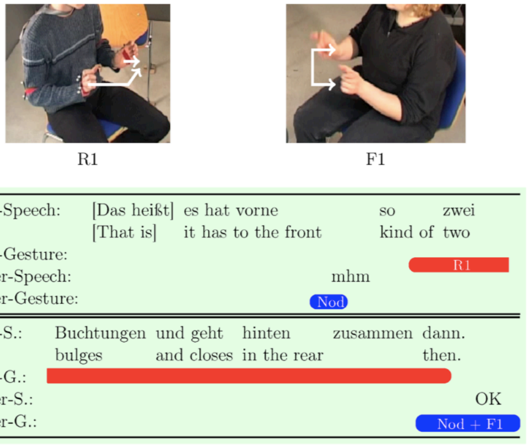
R1
F1
Router-Speech: [Das heißt] es hat vorne so zwei
[That is] it has to the front kind of two
Router-Gesture: R1
Follower-Speech: mhm
Follower-Gesture: Nod
Router-S.: Buchtungen und geht hinten zusammen dann.
bulges and closes in the rear then.
Router-G.:
Follower-S.: OK
Follower-G.: Nod + F1

**Figure 5:** Example showing the Router's and the Follower's gestures and their crucial exchange in terms of the Router's assertion and the Follower's acknowledgement.

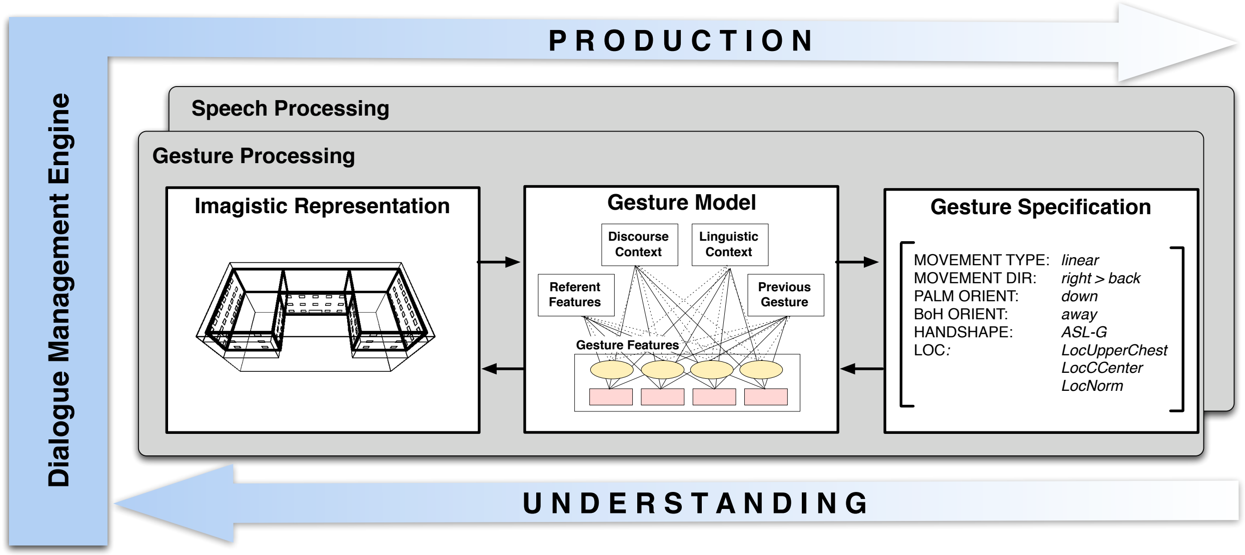

**Figure 6**: Overview of the production and understanding cycle in the simulation model.

and the approaching on-looker on the route perceives it from the open side of the U.

## 5 Extending the Simulation: The Understanding-Acceptance/Generation Cycle

How can we go beyond the simulation of isolated speaker-specific gestures towards the generation of gestures in dialogues? We build on our findings in the corpus study, briefly taken up here again (see list in section 2 and the respective comments): Gesture helps in structuring the dialogue supporting next speaker selection or indicating non-regular contributions of other speaker. It enables assessment of the current speaker's (Router's or Follower's) communicative intentions by the addressee, for example of whether the Router wants to keep the turn but indicates current memory and recapitulation problems thus appealing to the addressee's cooperation. In addition, appraisal of the reliability of the information given by the Router can be read off from some of the Router's gestures. Finally, as shown in section 4, gestures complementing or even replacing verbal information is used in acknowledgements.

Building on these observations, our goal is to simulate such dialogic interaction with two virtual agents (Router and Follower), each of whom provided with a speaker-specific GNetIc model. In the minimal discourse example Router and Follower use similar gestures which, notably, differ with respect to some details (e.g. speaker's perspective). In the simulation we essentially capture the Router's contribution in Figure 5 (R1) and the subsequent acknowledgement by the Follower (F1). In order to vary the Router's gesturing behavior we use the representation technique of drawing instead of shaping in the simulation.

What we need to extend the model with is an analysis of the Follower's understanding of the Router's gesture. Psychologically plausible but beyond commonly specialised technical approaches, we want to employ the same model of an agent's „gesture knowledge" for both generating and understanding gestures. For an overview of the production and understanding cycle see Figure 6.

Here we can make use of the fact that the BDN formalism allows for two different types of inference, causal inferences that follow the causal inter actions from cause to effect, and diagnostic inferences that allow for introducing evidence for effects and infer the most likely causes of these effects. This bi-directional use of BDNs could be complementary to approaches of plan/intention recognition such as in Geib and Goldman (2003).

To model a use of gestures for regulation as observed with the Follower F1, the Router agent's gestural activity is set as evidence for the output nodes of the Follower's BDN. A diagnostic inference then yields the most likely causes, that is, the most likely referent properties and values of discourse contextual variables. In other words, we employ the same speaker-specific GNetIc model for generation and for understanding. That is, information about the physical appearance of the Router's gesture (as specified in Figure 4) is provided as evidence for the Follower's GNetIc model revealing—correctly—that the gesture's representation technique is "drawing" and the shape property is "U-shaped".

Notably, just as the gesture generation process has to make choices between similarly probable alternatives, not all diagnostic inferences which are drawn by employing the Follower agent's GNetIc model are necessarily in line with the evidence from which the Router agent's gesture was originally generated. For instance, the communicative goal as inferred by the Follower agent is "lmDescrPosition" (with a likelihood of .65) instead of "lmDescrProperty". Nevertheless, the inferred knowledge reveals an underspecified representation of the referent (see Figure 7) as well as the most likely specification of the discourse context. That way, the Follower agent develops his own hypothesis of the Router agent's communicative goal and the content being depicted gesturally.

This hypothesis is forwarded to the follower agent‘s dialogue manager, which responds to such declaratives by the Router with an acknowledgement grounding act. Now the very same generation process as described in section 3 sets in. The Follower agent's feedback is generated by employing his GNetIc model for causal inference. The resulting gesture is, notably, different from the Router agent's gesture: it is a two-handed shaping gesture with handshape ASL-C. Movement type and movement features are the same as in the Router agent's drawing gesture. Palm and BoH orientation are different due to representation technique specific patterns which are implemented in the decision nodes (see Figure 7). This case of using iconic gesture for regulating dialogue has been successfully implemented using GNetIc and the overall production architecture.

## 6 Discussion and further research agenda

In this paper we addressed the dialogue-regulating function of gestures. Based on empirical observations of interactional patterns from the SAGA corpus, the starting points for the simulation of these gestures were non-interactional propositional ones such as iconics used to describe routes or landmarks. We achieved to simulate such iconic gestures used in their function as acknowledgements shown in section 3 which clearly transcends their mere representational task.

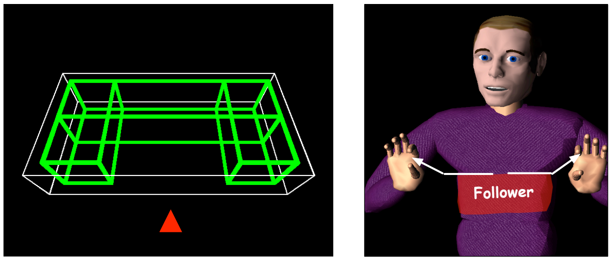

**Figure 7**: Imagistic representation of what the Follower understood from the Router's gestural depiction of the townhall (left) and the simulation of the Follower's autonomously generated shaping gesture used as an acknowledgement.

We first note that we draw a distinction between gestures relevant for dialogue structure such as next speaker selection or acknowledgement and those which focus on influencing the social climate among the dialogue participants. We did not have many of the latter in SAGA but observed some which we classified as "calming down" and "don't bother". In certain communication cultures also touching the other's body is accepted.

As for a research agenda to elucidate further the functions of gestures in dialogue, we do not go too deeply into matters of dialogue theory here. We already have shown that gestures accompanying base-line information, being part of the Router's report or the Follower's uptake can be modelled in PTT (Poesio and Rieser 2009, Rieser and Poesio 2009), if one assumes a unified representation for verbal and gestural meaning. Here we concentrate on how the simulation work can be pushed forward based on theoretical analyses of empirical data.

Note that on the list of discourse gestures given in section 2 the following items are tied to Router's behaviour and can be generated in an autonomous fashion:

- managing of own turn
- evidentials for establishing a confidence level
- assessment of relevance by discourse participants
- indication of topicality with respect to time, place or objects.

Observe, however, that these will also have an impact on the mental state of the Follower as is e.g., obvious for evidentials or the "brush away gesture" (Figure 1). Relevant for the sequencing of multimodal contributions are clearly the following:

- mechanisms of next-speaker selection as proposed in classical CA research
- grounding acts and feedback
- handling of non-canonical moves by discourse participants
- clarification requests to work on contributions.

These are intrinsically involved in the production of adjacency pairs, having a current and a next contribution and it is on these that simulation will focus on in future work. In combination with an information state-based multimodal discourse record (Traum & Larsson, 2003), the implementated cycle of generation, understanding and acceptance/ generation provides the basis for modeling this kind of gesture-based discourse regulation.

## Acknowledgments

This research is partially supported by the DFG in the CRC 673 "Alignment in Communication" and the Center of Excellence "Cognitive Interaction Technology".

## References

Asher, N. and Lascarides, A. (2003). *The Logic of Conversation*. Cambridge University Press

Bavelas, J., Chovil, N., Lawrie, D., and Wade, A. (1992). Interactive gestures. *Discourse Processes*, 15(4):469–491.

Bavelas, J., Chovil N., Coated, L., Roe, L. (1995). Gestures Specialised for Dialogue. *Personality and Social Psychology Bulletin*, 21(4):394–405

Bergmann, K., & Kopp, S. (2010). Modelling the Production of Co-Verbal Iconic Gestures by Learning Bayesian Decision Networks. *Applied Artificial Intelligence,* 24(6):530–551.

Bergmann, K. & Kopp, S. (2009). Increasing expressiveness for virtual agents–Autonomous generation of speech and gesture in spatial description tasks. In *Proceedings of AAMAS 2009*, pages 361–368.

Bergmann, K. & Kopp, S. (2009a). GNetIc–Using Bayesian Decision Networks for iconic gesture generation. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 76–89.

Bergmann, K., Kopp, S., and Eyssel, F. (2010). Individualized gesturing outperforms average gesturing–Evaluating gesture production in virtual humans. In *Proceedings of IVA 2010*, pages 104–117, Berlin/Heidelberg. Springer.

Cassell, J. and S. Prevost (1996). Distribution of Semantic Features Across Speech and Gesture by Humans and Computers. Proceedings of the Workshop on the Integration of Gesture in Language and Speech.

Clark, H.H. (1996). *Using Language*. CUP

Geib, C., Goldman, R.,(2003). Recognizing Plan/Goal Abandonment. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1515–1517.

Gill, S. P., Kawamori, M., Katagiri, Y., and Shimojima, A. (1999). Pragmatics of body moves. In *Proceedings of the 3rd International Cognitive Technology Conference*, pages 345–358.

Ginzburg, J. (2011). The Interactive Stance. Meaning for Conversation. Oxford University Press (in press).

Hahn, F. and Rieser, H. (2009-2011): Dialogue Structure Gestures and Interactive Gestures. Manual, 1st version. CRC 673 Working Paper. Bielefeld University

Hahn, F. and Rieser, H. (2010): Explaining Speech-Gesture Alignment in MM Dialogue Using Gesture Typology. In P. Lupowski and M. Purver (Eds.), *Aspects of Semantics and Pragmatics of Dialogue*. SemDial 2010, pp. 99–111.

Levinson, St. C. (1983). *Pragmatics*. Cambridge University Press.

Lücking, A., Bergmann, K., Hahn, F., Kopp, S., & Rieser, H. (2010): The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In M. Kipp et al. (Eds.), *LREC 2010 Workshop: Multimodal Corpora*.

McNeill, D. (1992). *Hand and Mind*. Chicago University Press.

Poesio, M. & Rieser, H. (2009). Anaphora and Direct Reference: Empirical Evidence from Pointing. In J. Edlund et al. (Eds.), *Proceedings of the 13th Workshop on the Semantics and Pragmatics of Dialogue (DiaHolmia)* (pp. 35–43). Stockholm, Sweden.

Rieser, H. (2010). On Factoring out a Gesture Typology from the Bielefeld Speech-And-Gesture-Alignment Corpus (SAGA). In Kopp and Wachsmuth (Eds.), *Proceedings of GW 2009*. Springer, pp. 47–61.

Rieser, H. & Poesio, M. (2009). Interactive Gesture in Dialogue: a PTT Model. In P. Healey et al. (Eds.), *Proceedings of the SIGDIAL 2009 Conference* (pp. 87–96). London, UK: ACL.

Poesio, M. and Rieser, H. (2010). Completions, coordination and alignment in dialogue. *Dialogue and Discourse* 1(1), 1–89

Poesio, M. and Traum, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3): 309–347

Roque, A. and Traum, D. (2008). Degrees of Grounding Based on Evidence of Understanding. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pp. 54–63

Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50: 696–735

Stalnaker, R. (1978): Assertion. In Cole, P. (Ed.) *Syntax and Semantics 9: Pragmatics,* pp. 315–322.

Sowa, T. and Wachsmuth, I. (2009). A computational model for the representation an processing of shape in coverbal iconic gestures. In K. Coventry et al. (Eds.), *Spatial Language and Dialogue*, pages 132–146. Oxford University Press.

Traum, D. (1999). Computational models of grounding in collaborative systems. In *Working Notes of AAAI Fall Symposium on Psychological Models of Communication*, pp. 124–131.

Traum, D., & Larsson, S. (2003). The information state approach to dialogue management. In R.W. Smith and J.C.J. van Kuppevelt (Eds.), *Current and New Directions in Discourse & Dialogue* (pp. 325–353). Kluwer Academic Publishers.