Grounding the Simulation of Iconic Gestures in Gesture Typology

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

[1] Collaborative Research Center 673, "Alignment in Communication", Bielefeld University
[2] Center of Excellence in "Cognitive Interaction Technology" (CITEC), Bielefeld University
{kbergman,skopp}@techfak.uni-bielefeld.de
hannes.rieser@uni-bielefeld.de

Keywords: Iconic Gesture, Gesture Typology, Gesture Simulation, Virtual Agents

How are co-speech iconic gestures used to convey visuo-spatial information? We investigate this question, which is still relatively unexplored [1], with an interdisciplinary methodology combining the empirical study of speech and gesture use, the elaboration of theoretical reconstructions and the formulation of generation models that enable the simulation of such communicative behaviour with virtual humans. In our talk we will focus on two topics, first, how gesture simulation is grounded in empirical gesture typology and second, how gesture simulation can be used methodologically, looping back to the empirical data on which both, simulation and theoretical modelling are based. Regarding current gesture research, the second topic is entirely new. Even traditionally, gesture research and gesture typology are intimately intertwined as one can see from [4], [6], or [8]. Recently, new options for gesture typology have arisen due to systematically collected and annotated data such as the Bielefeld Speech And Gesture Alignment corpus ([7, SaGA]). Corpus-based empirical methods proceed from rated annotations to classification of recurrent structures and ultimately to an investigation of its generalizability supported by statistical investigations [5]. Computational simulation opens up new possibilities enriching this set of methods in many ways. Obviously, gesture simulation has its independent goals in endowing virtual agents with human-like expressiveness. In addition, we use it as a methodological device, more specifically for the post-hoc evaluation of decisions made at various levels of the theory construction process, in other words, as a method of Popperian falsification. As an illustration of every aspect of our methodology, we will discuss a church-window-example from the SaGA corpus shown in Figure 1 throughout the talk (restricted to the top of the window).

Empirical Study and Theoretical Perspective The empirical study slot in Figure 1 shows under (1) the pointed gothic church-window presented in a VR-video film as perceived from an agent called Router. Empirical study slot (2) shows the gesture of the Router reporting his ride through a VR-town to a Follower, especially his drawing of a pointed top. The Router's gesture modifies his words "these typical uhm church-windows". Empirical study slot (3) presents

information coming from the rated annotation: The gesture is located in the centre of the Router's torso (CC). The trajectory described by his right hand is an ARC followed by another ARC. The movement of the right hand's wrist is a sequence of "move right- move up" (MR/MU) and "move right-move down" (MR/MD). The Router's palm points to the ground (PDN). The hand-shape is ASL-G oriented away from his body (BAB).

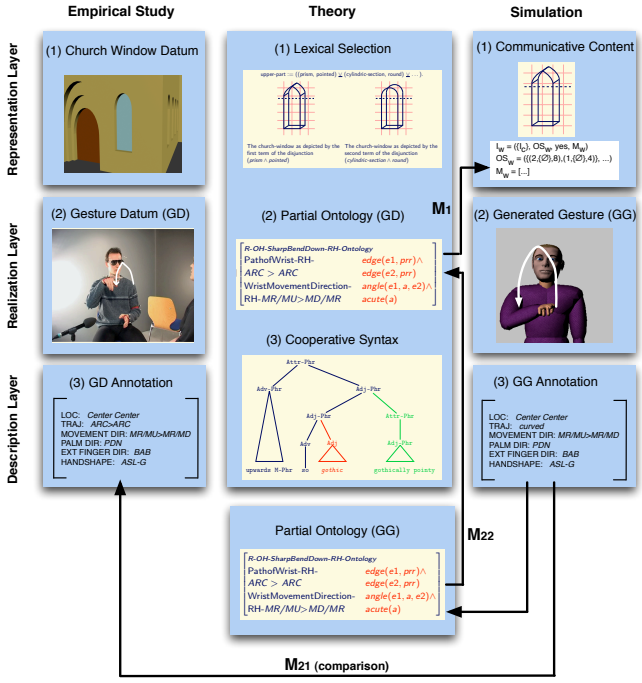

Fig. 1. Our methodology to study iconic gesture combines empirical study, theoretical modeling and computational simulation across three layers: (1) representation layer (stimulus, attributed partial ontology, conceptual representation); (2) realization layer (speaker's iconic gesture, virtual agents gesture); (3) description layer (annotation of the stimulus, annotation of the generated gesture). This Figure also shows the matching the generated gesture with the originally annotated datum (GD-Annotation). The match uses mapping M_{21} from the annotation of the Generated Gesture (GG-Annotation) to the GD-Annotation and the mapping M_{22} from the GG-Partial Ontology to the original GD-Partial Ontology.

Information as contained in the Empirical Study slot is used for establishing a gesture typology. The logical information defined is shown under theory (2). It provides a matrix with the annotation predicates applied in the gesture morphology composed with their respective values as affixes. These newly formed attributes like PathofWrist-RH-ARC>ARC are in turn mapped onto a

parameterised quantifier-free PL1-expression which encodes the ontological information. This ontological information singles out the class of upright standing pointed solids. More in general: the gesture typology set up in [10] rests on gesture form features like hand-shape, palm-direction or wrist-movement extracted from systematic annotation. Clusters of features then provide entities of different dimensions such as lines, regions, partial objects and composites of these. These are provided with a partial ontology interfacing with verbal semantics.

Slot (1) under Theory indicates the function of the Router's gesture. It selects from a lexicon providing a disjunction for pointed-or-round-church-windows the pointed-church-window-reading. This way, gestural meaning specifies lexical meaning. Theory slot (3) gives an indication of the syntactic problems involved. The attributive phrase Attr-Phr depicted is produced by Router and Follower together, forming a so-called split utterance or completion. Green parts represent Router's and red ones Follower's contributions. The Router's gesture is produced right after the split point "so". Upon it, "gothic" is contributed by the Follower. Subsequently, we have a repair by the Router extending the Follower's completion. How verbal semantics and gestural semantics interface in the end is not shown in [9].

The Generation Perspective Generation aims to produce a context-dependent dialogue act starting from an initial communicative intention and based on an imagistic representation of content (see simulation slot (1) in Figure 1). The virtual agent MAX in his role as the Router selects the "right" lexical entry from the start. This is the natural perspective, given the assumption that MAX is provided with information about the object to be depicted.

To simulate the use of iconic gestures, a situated gesture formulation model has been realized as a Bayesian decision network combining machine learning (data-based) and model-based techniques ([3, GNetIc]). Besides providing an analysis tool for the dependencies and strategies involved in cognitive gesture generation processes, this method allows for simulating speaker-specic as well as speaker-independent gesture production with virtual humans. Further, this model allows to formulate iconic gestures not only based on principles of iconicity (focusing resemblance) but it also takes into account empirically determined factors like discourse context, own previous gestures, or syntactic decisions in the language system. Gesture production is based on a hierarchical representation of imagistic knowledge which allows to extract features of the object(s) to be described, e.g., position, shape properties, symmetry information etc. The combination of gestures with speech is done using a realization engine that turns the behavior specification into synthetic speech and synchronized gesture animations under consideration of body kinematics. The resulting gesture realization with the virtual agent Max is given in Figure 1, simulation slot (2) and (3).

Mapping Generated Gestures onto Theoretical Gesture Types Theory set up and generation are carried out independently on exactly the same set of SaGA-data. Independence is essential to avoid viscous circles. Founding the simulation on gesture typology is accomplished as follows: We define a mapping

M_1 from the partial ontology of abstract gesture description (GD-Partial Ontology) to the semantic representation that determines the generated gesture's morphology, i.e. its Communicative Content which is a specialization of the partial ontology. Two additional mappings M_{21} and M_{22} are established between the generated gesture and the gesture in the original datum. Now the methodological step, mentioned at the beginning of the paper, is arrived at: The mappings M_{21} and M_{22} are defined, if we take the simulated gesture as a datum. Its annotation (GG-Annotation) is provided with a partial ontology (GG-Partial Ontology) and compared with the originally annotated and interpreted real-world datum (GD-Annotation, respectively GD-Partial Ontology). For an evaluation of the GNetIc generation model with respect to the corpus data see, e.g., [2]. to evaluate the goodness of fit of the simulation and its explanatory power in theoretical terms.

Acknowledgements

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

1. Bavelas, J., Gerwing, J., Sutton, C., and Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58:495–520.
2. Bergmann, K., and Kopp, S. (2010). Modelling the Production of Co-Verbal Iconic Gestures by Learning Bayesian Decision Networks. In *Applied Artificial Intelligence* 24(6):530–551.
3. Bergmann, K., and Kopp, S. (2009). GNetIc—Using Bayesian Decision Networks for Iconic Gesture Generation. In Z. Ruttkay et al. (Eds.), *Proceedings of the 9th Conference on Intelligent Virtual Agents* (pp. 76–89). Berlin: Springer.
4. Ekman, P. and Friesen, W. (1969) The repertoire of nonverbal behaviour: categories, origins, usage and coding. In *Semiotica*, 1, pp. 49–98.
5. Hahn, F. and Rieser, H.: 2010, Explaining Speech Gesture Alignment in MM Dialogue Using Gesture Typology. In P. Lupowski and M. Purver (Eds.), *Proceedings of SemDial 2010*. Polish Society for Cognitive Science. Poznan 2010, 99–111.
6. Kendon, A.: 2004, Gesture: Visible Action as Utterance. Cambridge: Cambridge University Press.
7. Lücking, A., Bergmann, K., Hahn, F., Kopp, S., and Rieser, H. (2010). The Bielefeld Speech and Gesture Alignment Corpus (SaGA). In M. Kipp, J.-C. Martin, P. Paggio and D. Heylen (Eds.), LREC 2010 Workshop: Multimodal Corpora—Advances in Capturing, Coding and Analyzing Multimodality.
8. McNeill, D. (1992). *Hand and Mind. What Gestures Reveal about Thought.* Chicago: University of Chicago Press.
9. Rieser, Hannes and Poesio, Massimo, 2009. Interactive Gesture in Dialogue: a PTT Model. In P. Healey et al. (Eds.) *Proceedings of the SIGDIAL 2009 Conference London*, UK: ACL, pp. 87–96.
10. Rieser, H. (2010). On Factoring Out a Gesture Typology from the Bielefeld Speech-and-Gesture-Alignment Corpus (SAGA). In: Kopp S. and Wachsmuth, I. (Eds.), *Gesture in Embodied Communication and Human-Computer Interaction* (pp. 47–60). Berlin: Springer.