Systematicity and Idiosyncrasy in Iconic Gestures—A Probabilistic Model of Gesture Use

Kirsten Bergmann and Stefan Kopp

Sociable Agents Group, CITEC, Bielefeld University
P.O. Box 100 131, D-33615 Bielefeld, Germany
{kbergman,skopp}@techfak.uni-bielefeld.de

Key words: gesture generation, idiosyncrasy, probabilistic model

Introduction Current literature on gesture research states that the question "why different gestures take the particular physical form they do is one of the most important yet largely unaddressed questions in gesture research" [1, p. 499]. This holds especially for iconic gestures, for which information has to be mapped from some mental image into (at least partly) resembling gestural form. This transformation is neither direct nor straightforward but involves a number of issues like the composability of a suitable linguistic context, the choice of gestural representation technique (e.g., placing, drawing etc.), or the low-level choices of morphological features such as handshape or movement trajectory. Current research does not offer much concrete findings, that could inform the meaning-form mapping. We present an empirical study to elucidate some of these puzzles and a probabilistic model to simulate human gesture production.

Empirical Basis A study on spontaneous speech and gesture use in direction-giving was conducted (25 dialogs, ∼5000 gestures). All coverbal gestures have been segmented and coded for their representation technique (e.g., drawing, placing) and their gesture morphology in terms of handshape, hand position, palm and finger orientation, and movement features (e.g. trajectory, direction). In addition, we transcribed the spoken words and coded further information about the dialog context. Findings indicate that a gesture's form is not only determined by characteristic features of its referent. Rather, these choices are made under specific contextual constraints, e.g., we found the communicative intention underlying an utterance to be decisive for the representation technique chosen [**?**]. In addition, our corpus analysis shows that individuals differ significantly in the way they gesture about the same thing, and these differences concern all abovementioned decision levels. In other words, people differ in the number of gestures they employ, in their choice of representation techniques, in their handedness and also in their choice of morphological gesture features. This suggests that, besides systematic *inter*-personal factors, the role of idiosyncrasy, taken to be a form of *intra*-personal systematics, must be taken into consideration by an account of why people gesture the way they actually do.

Modeling Approach Our goal is a model that, given a particular discourse context and a given communicative intention such as "describe object X", allows virtual agents to automatically select the content and derive the form of coordinated language and iconic gestures. The fact that the form of a gesture is influenced by both, systematic and idiosyncratic factors, has severe implications for the design of such a computational model. We therefore present a simulation account going beyond previous systems that either applied systematic meaning-form mappings in the formation of iconic gestures [3], or model the individual gesturing patterns of specific speakers in generating animations for new texts [4]. Based on data extracted from our annotated corpus we employ machine learning algorithms to create Bayesian networks which allow to make use of both, general characteristics across speakers and idiosyncratic patterns within the current individual speaker.

This probabilistic approach is advantageous for a number of reasons. First, a network can be easily extended by introducing further variables, either annotated in the corpus, or inferred from that data. Further, the same network can be used both for simulating gestural behavior in an agent and, given a particular gesture, for inferring the states of any causal variables. Finally, the approach provides the possibility to detect clusters of speakers who share particular interrelations between causal nodes and observable behavior, i.e. it enables us to distinguish between different degrees of inter- and intrapersonal systematics in the production of iconic gestures.

A first prototype of the computational model has been realized for which we trained Bayesian networks from the empirically observed behavior of different subjects. The system succeeds in reproducing similar gestural expressions in corresponding contexts for embodied agents like the virtual human Max.

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. J. Bavelas, J. Gerwing, C. Sutton, and D. Prevost. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58:495–520, 2008.
2. K. Bergmann and S. Kopp. Increasing expressiveness for virtual agents–autonomous generation of speech and gesture. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '09)*, 2009.
3. S. Kopp, P. Tepper, K. Ferriman, K. Striegnitz, and J. Cassell. Trading spaces: How humans and humanoids use speech and gesture to give directions. In T. Nishida, editor, *Conversational Informatics*, chapter 8, pages 133–160. John Wiley, 2007.
4. M. Neff, M. Kipp, I. Albrecht, and H.-P. Seidel. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions on Graphics*, 27/1:1–24, 2008.