Gesture Use – From Real to Virtual Humans and Back

Kirsten Bergmann
Bielefeld University
Faculty of Technology, CITEC
P.O. Box 100 131
33501 Bielefeld, Germany
kirsten.bergmann@uni-bielefeld.de

When we are face to face with others, we use not only speech, but also a multitude of nonverbal behaviors to communicate with each other. A head nod expresses accordance with what someone else said before. A facial expression like a frown indicates doubts or misgivings about what one is hearing or seeing. A pointing gesture is used to refer to something. More complex movements or configurations of the hands depict the shape or size of an object. Of all these nonverbal behaviors, *gestures*, the spontaneous and meaningful hand motions that accompany speech, stand out as they are very closely linked to the semantic content of the speech they accompany, in both form and timing. Speech and gesture together comprise an utterance and externalize thought; they are believed to emerge from the same underlying cognitive representation and to be governed, at least in part, by the same cognitive processes (Kendon, 2004; McNeill, 2005). Despite this important role of co-speech gestures in communication, little is known, however, about the mechanisms that underlie gesture production in human speakers (cf. Bavelas et al. (2008), de Ruiter (2007)) as well as the functions gesture use fulfills in communication and even beyond, e.g. in educational or therapeutic contexts. My talk at the symposium showed how building computational simulation models of natural, communicative behavior and employing these models in virtual humans allows to address these research issues.

1 A computational model for iconic gesture production

With the GNetIc approach (Generation Networks for Iconic Gestures; Bergmann & Kopp (2009)) we proposed a computational framework to automatically generate novel gesture forms to be realized with a virtual human. Based on extensive empirical data from human-human interaction (SaGA corpus; Lücking (2013)), GNetIc accounts for a number of factors, identified in empirical corpus analyses, which can roughly be divided into three kinds. First, since the meaning of iconic gestures is explained by similarity or resemblance to their referent, the way how meaning is mapped onto gesture form is decisive. This implies that iconic gesture use is dependent on an underlying imagistic representation and rises the question how this representation is transformed into gesture form constrained by the use of different gestural representation techniques (e.g., placing, drawing, or posturing; cf. Kendon (2004)). Second, a gesture's form is also influenced by specific discourse contextual constraints as well as its linguistic context. For instance, speakers tend to employ gestures rather to introduce new information into the discourse, than to refer to what is already known and acknowledged by the dialogue partners (McNeill, 1992). And third, inter-individual differences in gesture use are quite obvious, reflecting an individual, speaker-specific gesture style. Individual speakers differ obviously with respect to gesture frequency, handedness (one-handed vs. two-handed gestures), preference for particular handshapes etc.

The GNetIc approach takes all these factors into account and allows to derive gesture forms on the basis of characteristics extracted from the imagistic representation of a referent object. Going beyond a straightforward meaning-form mapping, contextual factors like the given communicative goal, information state, or previous gesture use are also taken into account. In particular, different gestural representation techniques are considered, mediating the meaning-form mapping. By combining rule-based and data-based models, GNetIc can simulate both systematic patterns shared among several speakers, as

K. Jokinen and M. Vels. 2015. Proceedings of The 2nd European and the 5th Nordic Symposium on Multimodal Communication.

well as idiosyncratic patterns specific to an individual. That is, GNetIc can produce novel gestures as if being a certain speaker. Further, building and comparing networks from different speakers allows to gain insights into how production processes might differ from individual to individual.

2 How do human observers judge virtual humans using gestures?

In an evaluation study of the GNetIc model, human observers were provided with an object description given by a virtual human (Bergmann et al., 2010). We manipulated the agent's gestural behavior and addressed in how far different GNetIc models (individualized ones and an 'average' one learned from the aggregated data of several speakers) as well as control conditions (no gestures; randomized gesture use) affect the perception of human observers. Spoken words remained the same across all conditions. Results showed that the two individual GNetIc conditions outperformed the other conditions in that gestures were perceived as more helpful, overall comprehension of the multimodal presentation was rated higher, and the agent's mental image was judged as being more vivid. Similarly, the two individual GNetIc conditions outperformed the control conditions regarding agent perception in terms of likeability, competence, and human-likeness. Moreover, the aggregated GNetIc condition was rated worse than the individual GNetIc conditions throughout. And finally, the no gesture condition was rated more positively than the random condition. That is, it seems even better to make no gestures than to randomly generate gestural behavior. Overall, this study provides evidence that building generative models of co-verbal iconic gesture use, instead of using pre-defined gestures from a lexicon or 'gesticon', can yield encouraging results with actual users.

3 A virtual human as a tutor in second language learning

Beyond beneficial effects of virtual humans' gesture use regarding presentation quality and agent perception, virtual tutors also have potential to support learners' performance, e.g., in learning linguistic materials. In an empirical study we addressed whether the memory-supporting effect of iconic gesture imitation in vocabulary learning, so far shown for real human tutors (e.g. Macedonia et al. (2011)), is also valid for virtual tutors. The study employed a within-subject design manipulating the type of training in terms of (1) gesture-based training with human stimuli, (2) gesture-based training with agent stimuli, and (3) a control condition without any gestures. A total of 32 participants learned 45 vocabulary items on three consecutive days. Training materials comprised 45 nouns in German and 'Vimmi' (an artificial corpus created for experimental purposes to avoid associations and to control for different factors that might constrain the memorization of particular vocabulary items; see Macedonia et al. (2011)).

Short-term learning performance was measured the next day prior to the training session, respectively. The long-term effect of information decay was measured additionally four weeks after training was finished. In every test session, participants conducted a free and thereafter a cued recall test. Results showed, for both types of long-term measures (free and cued recall), better memory performance for items learned in the virtual human condition over items learned in the control condition. The same effect was present for short-term measures of free recall. Notably, in all tests performed, there was a trend for the virtual agent leading to better memory performance than training with a human (for details see Bergmann & Macedonia (2013)). These findings doubtlessly need follow-up studies to elucidate which factors and constraints constitute the beneficial effect of the virtual character. Nevertheless, they clearly substantiate the view that virtual pedagogical agents might play a crucial role in future language learning.

Acknowledgements

This research is supported by the Collaborative Research Centre SFB 673 "Alignment in Communication" and the Cluster of Excellence in "Cognitive Interaction Technology" (CITEC), both funded by the German Research Foundation (DFG).

References

J. Bavelas, J. Gerwing, C. Sutton, and D. Prevost. 2008. Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58:495–520.

K. Bergmann and S. Kopp. 2009. GNetIc—Using Bayesian decision networks for iconic gesture generation. In Z. Ruttkay, M. Kipp, A. Nijholt, and H. Vilhjalmsson, editors, *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, pages 76–89. Springer, Berlin/Heidelberg.

K. Bergmann and M. Macedonia. 2013. A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents*, pages 139–148, Berlin/Heidelberg. Springer.

Kirsten Bergmann, Stefan Kopp, and Friederike Eyssel. 2010. Individualized gesturing outperforms average gesturing–evaluating gesture production in virtual humans. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, editors, *Proceedings of the 10th Conference on Intelligent Virtual Agents*, pages 104–117, Berlin/Heidelberg. Springer.

J.P. de Ruiter. 2007. Some multimodal signals in humans. In *Proceedings of the 1st Workshop on Multimodal Output Generation*, pages 141–148. CTIT.

A. Kendon. 2004. *Gesture—Visible Action as Utterance*. Cambridge University Press.

A. Lücking, K. Bergmann, F. Hahn, S. Kopp, and H. Rieser. 2013. Data-based analysis of speech and gesture: the bielefeld speech and gesture alignment corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces*, 7(1-2):5–18.

M. Macedonia, K. Müller, and A.D. Friederici. 2011. The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, 32:982—998.

D. McNeill. 1992. *Hand and Mind—What Gestures Reveal about Thought*. University of Chicago Press, Chicago.

D. McNeill. 2005. *Gesture and Thought*. University of Chicago Press, Chicago, IL.