The Bielefeld Speech and Gesture Alignment Corpus

Kirsten Bergmann

Sociable Agents Group, CITEC, Bielefeld University
SFB 673 'Alignment in Communication', Bielefeld University
P.O. Box 100 131, D-33615 Bielefeld, Germany
kbergman@techfak.uni-bielefeld.de

The *Speech and Gesture Alignment* (SaGA) corpus is built from 25 dyads of spatial communication combining direction-giving and sight description. The SaGA corpus consists of 280 minutes of video material containing 4961 iconic/deictic gestures, approximately 1000 discourse gestures and 39,435 words. This one of the largest and most comprehensive collection of naturalistic, yet controlled, and systematically annotated speech-gesture data currently available.

The data has been completely and systematically annotated based on an annotation grid developed according to theoretical considerations, tested and refined using multi-coder agreement tests. The grid comprises gesture segmentation and classification (iconics, deictics, beats), gestural representation techniques (e.g., drawing, placing), morphological gesture features (e.g., handshape, hand position, palm orientation, movement features), transcription of the spoken words and dialogue context information (DAMSL dialogue acts, information focus, thematization, elemental actions of direction giving). In addition, a subpart of the corpus (290 gestures, 5 dyads) has been coded for the gestures' referent objects and their spatio-geometrical properties (dimensionality, extents, symmetries, profiles, etc.).

Since manual annotations of gestures are extremely time-consuming and come along with the difficulty of estimating position and orientation values from perspectively distorted video images, we developed data analysis tools to automatically annotate morphological gesture features such as hand position, movement extent, or palm orientation from motion tracking data. A comparative analysis of manually and automatically annotated data allows to appraise the quality as well as pros and cons of both kinds of coding.

The SaGA corpus is used to shed light onto the alignment of speech and gesture from different perspectives. In addition to statistical analyses we make use of the database in theorectical linguistic analyses and reconstructions (cf. Rieser (in press)), as well as in the formulation of generation models that enable the simulation of multimodal behaviour with virtual agents (Bergmann and Kopp (2009); Kopp et al. (2008)). To be realized with any virtual agent, the results of behavior planning are described in a representation language such as MURML (Multimodal Utterance Representation Markup Language; Kranstedt et al. (2002)) or BML (Behavior Markup Language, Vilhjalmsson et al. (2007)). Such representation formats precisely specify multimodal behavior and are in turn a promising candidate for a standardization of multimodal databases.

Bibliography

Bergmann, K. and Kopp, S. (2009). Increasing expressiveness for virtual agents–Autonomous generation of speech and gesture in spatail description tasks. In Decker, K., Sichman, J., Sierra, C., and Castelfranchi, C., editors, *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, pages 361–368, Budapest, Hungary. IFAAMAS.

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

Kranstedt, A., Kopp, S., and Wachsmuth, I. (2002). MURML: A multimodal utterance representation markup language for conversational agents. In *AAMAS'02 Workshop Embodied conversational agents–let's specify and evaluate them!*

Rieser, H. (in press). On factoring out a gesture typology from the bielefeld on factoring out a gesture typology from the bielefeld speech-and-gesture-alignment corpus (saga). In Kopp, S. and Wachsmuth, I., editors, *Gesture in Embodied Communication and Human-Computer Interaction*. Springer, Berlin/Heidelberg.

Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thorisson, K., van Welbergen, H., and van der Werf, R. (2007). The behavior markup language: Recent developments and challenges. In Pelachaud, C., Martin, J.-C., Andre, E., Chollet, G., Karpouzis, K., and Pelé, D., editors, *Proceedings of the 7th International Conference on Intelligent Virtual Agents*, pages 99–111, Berlin/Heidelberg. Springer.