Beat Gestures in Direction Giving

Chris J. Brandhorst and Mariët Theune

University of Twente, Human Media Interaction group
PO Box 217, 7500 AE Enschede, The Netherlands
`chris.brandhorst@gmail.com, m.theune@ewi.utwente.nl`

1 Introduction

When humans speak, they use gestures that "are not random but convey to listeners information that can complement or even supplement the information relayed in speech" [1]. To illustrate this, consider three of the main gesture types from [2]. *Deictic gestures* are pointing movements indicating the location of items being referred to. *Iconic gestures* depict a physical aspect of what is spoken about, such as the shape of an object or the trajectory of a movement. Finally, *beat gestures* are simple flicks of the hand that do not convey any semantic content, but reflect discourse structure by marking important words and phrases.

In this paper, we investigate the use of beat gestures in direction giving. This research is motivated by our work on the Virtual Guide, an animated character that can give directions to visitors in a 3D environment, using speech and gesture [3]. Currently, the Virtual Guide uses deictic or iconic gestures when mentioning directions (left, right) or objects with a certain shape. It is not yet clear when beat gestures should be used. Other work on direction giving such as [4] also neglects the generation of beat gestures. Therefore, we have analysed a corpus of direction giving monologues to find out when beat gestures are used, and to create a beat usage model based on the results.

2 Corpus analysis

We analysed 16 short video clips (average duration about 45 seconds) of Dutch speakers giving directions in a building on our university campus. The video corpus contains 133 gestures, which we scored using McNeill's 'beat filter' [2]: a system to determine the probability of a gesture being a beat gesture, based on the shape of its movement. We classified each gesture as a certain type based on its beat score. Gestures that could not be classified based on shape alone were disambiguated based on the accompanying speech. For instance, if a beat-like gesture to the left or right was accompanied by a mention of *left, right* or *here*, it was classified as deictic, based on the assumption that gestures reflecting the content of the speech (deictic or iconic) would be preferred over beats, cf. [5].

To determine which gesture types are used in which speech contexts, we identified speech classes such as spatial expressions, landmarks (i.e., salient objects found along a route), directions, etc. For all occurrences of these classes in

the corpus we annotated what type of gesture accompanied them (if any). The results show that beats most often mark words expressing locations or movements in space (e.g., *through, in, across*), with 11% of these expressions being augmented with a beat gesture. Next were hesitations (9.5%), descriptions of the duration or timing of actions (8.1%) and references to landmarks (4.5%). For the first three classes, beat gestures were used more often than any other gesture type. For landmarks, however, in most cases another gesture type was preferred, in particular at first mention. Similarly, when mentioning a direction, our speakers used pointing gestures much more often than beat gestures.

3 A simple model of beat gesture use

A well-known framework for the automatic generation of gestures by animated characters is the BEAT toolkit [5]. Despite its name, BEAT only generates beats when no other gesture is available. However, our findings show that this approach is too simple: beats are also used when other gesture types are appropriate. Therefore, we propose to give beat gestures the same basic priority as other gesture types. Given the results of statistical corpus analysis (preferably using a much larger corpus than used for this initial study), the probability that a beat is generated in a certain context can be given by $P(B|u) = P(B|c_u)m_s$, where B is the generation of a beat, u is the word or phrase to be uttered, c_u is the class of the utterance, and $P(B|c_u)$ is the corpus frequency of beat gestures accompanying this class. Finally, m_s is a multiplier for a specific speaker, reflecting that beat usage also depends on personal style. This probability function can be easily applied in the gesture generation model of the Virtual Guide [3].

Although the proposed model is an improvement over the way beat gestures are currently handled in the Virtual Guide and other systems, we believe that ultimately a purely probabilistic model will be too limited. Further research is needed to determine what factors influence a speaker's choice to use a beat, another gesture, or no gesture at all when expressing certain types of information.

References

1. Iverson, J., Goldin-Meadow, S.: Why people gesture when they speak. Nature **396** (1998) 228
2. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press, Chicago, IL/London, UK (1992)
3. Theune, M., Hofs, D., van Kessel, M.: The Virtual Guide: A direction giving embodied conversational agent. In: Proceedings of Interspeech 2007. (2007) 2197–2200
4. Kopp, S., Tepper, P., Striegnitz, K., Ferriman, K., Cassell, J.: How humans and humanoids use speech and gesture to give directions. In Nishida, T., ed.: Engineering Approaches to Conversational Informatics. John Wiley & Sons (2007)
5. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of SIGGRAPH '01, Los Angeles, CA (2001)