GNetIc – Using Bayesian Decision Networks for Iconic Gesture Generation

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

Abstract. Expressing spatial information with iconic gestures is abundant in human communication and requires to transform a referent representation into resembling gestural form. This task is challenging as the mapping is determined by the visuo-spatial features of the referent, the overall discourse context as well as concomitant speech, and its outcome varies considerably across different speakers. We present a framework, GNetIc, that combines data-driven with model-based techniques to model the generation of iconic gestures with Bayesian decision networks. Drawing on extensive empirical data, we discuss how this method allows for simulating speaker-specific vs. speaker-independent gesture production. Modeling results from a prototype implementation are presented and evaluated.

Keywords: Nonverbal Behavior, Gesture Generation, Inter-subjective Differences, Bayesian Decision Networks.

1 Introduction

The use of speech-accompanying iconic gestures is a ubiquitous characteristic of human-human communication, especially when spatial information is expressed. It is therefore desirable to endow virtual agents with similar gestural expressiveness and flexibility to improve the interaction between humans and machines. This is an ambitious objective, as de Ruiter [4, p. 30] recently put it: "The problem of generating an overt gesture from an abstract [...] representation is one of the great puzzles of human gesture, and has received little attention in the literature". The intricacy is due to the fact that iconic gestures, in contrast to language or other gesture types such as emblems, have no conventionalized form-meaning mapping. Apparently, iconic gestures communicate through iconicity, i.e., their physical form corresponds with object features such as shape or spatial properties. Empirical studies, however, reveal that similarity with the referent cannot fully account for all occurrences of iconic gesture use [10]. Recent findings actually indicate that a gesture's form is also influenced by specific contextual constraints and the use of more general gestural representation techniques such as shaping or drawing [9,2] .

In addition, human beings are all unique and inter-subjective differences in gesturing are quite obvious (cf. [6]). Consider for instance gesture frequency: While some people rarely make use of their hands while speaking, others gesture almost without

Zs. Ruttkay et al. (Eds.): IVA 2009, LNAI 5773, pp. 76–89, 2009.

interruption. Similarly, individual variation becomes apparent in preferences for particular representation techniques or the low-level choices of morphological features such as handshape [2]. See Figure 1 for some examples of how people perform different gestures that refer to the same entity, a u-shaped building. The speakers differ, first, in their use of representation techniques. While some speakers perform drawing gestures (the hands trace the outline the referent), others perform shaping gestures (the referent's shape is sculpted in the air). Second, gestures vary in their morphological features even when speakers use the same representation technique: Drawing gestures are performed either with both hands (P1 and P5) or with one hand (P8), while the shaping gestures are performed with differing handshapes (P7 and P15).

Fig. 1. Example gestures from different speakers, each referring to the same u-shaped building

Taken together, iconic gesture generation on the one hand generalizes across individuals to a certain degree, while on the other hand, inter-subjective differences also must be taken into consideration by an account of why people gesture the way they actually do. In previous work we developed an overall production architecture for multimodal utterances incorporating model-based techniques of generation as well as data-driven methods (Bayesian networks) [2]. In this paper we now present a complete model for the generation of iconic gestures combining both methods. We present GNetIc, a gesture net specialized for iconic gestures, which is a framework to support decision-making in the generation of iconic gestures. Individual as well as general networks are learned from annotated corpora and supplemented with rule-based decision making. Employed in an architecture for integrated speech and gesture generation, the system allows for a speaker-specific gesture production which is not only driven by iconicity, but also by the overall discourse context. In the following, we survey existing approaches to model gesture generation (Section 2) and present our integrated approach of iconic gesture generation (Section 3). We discuss how it accounts for both, general characteristics across speakers and idiosyncratic patterns of the individual speaker. In Section 4 we describe modeling results from a prototype implementation, and present an evaluation of the gesturing behavior generated with GNetIc in Section 5.

2 Related Work

Work on the generation of speech-accompanying iconic gestures and its simulation with virtual agents is still relatively sparse. The first systems investigating this challenge were lexicon-based approaches [3]. Relying on empirical results, these systems focus on the context-dependent coordination of gestures with concurrent speech, whereby gestures are drawn from a lexicon. Flexibility and generative power of gestures to