Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 51 (2009) 630–639
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Studies on gesture expressivity for a virtual agent

Catherine Pelachaud *

IUT de Montreuil, Université de Paris 8, INRIA Paris-Rocquencourt, Paris

Received 5 February 2007; received in revised form 24 April 2008; accepted 25 April 2008

Abstract

Our aim is to create an affective embodied conversational agent (ECA); that is an ECA able to display communicative and emotional signals. Nonverbal communication is done through certain facial expressions, gesture shapes, gaze direction, etc. But it can also carry a qualitative aspect through behavior expressivity: how a facial expression, a gesture is executed. In this paper we describe some of the work we have conducted on behavior expressivity, more particularly on gesture expressivity. We have developed a model of behavior expressivity using a set of six parameters that act as modulation of behavior animation. Expressivity may act at different levels of the behavior: on a particular phase of the behavior, on the whole behavior and on a sequence of behaviors. When applied at these different levels, expressivity may convey different functions.

Keywords: Nonverbal behavior; Embodied conversational agents; Expressivity

1. Introduction

Nonverbal behaviors are a powerful means of communication. They help to produce speech, to formulate our thoughts, to communicate our feelings and so forth. Several studies have shown the tight link that exists between our cognitive and emotional states and our verbal and nonverbal behaviors. Some gestures and even facial expressions can replace words. They are commonly called emblems. A wink or fingers making a ring have precise meanings in given culture. Lexicons of these behaviors and even dictionaries are being gathered (Poggi, 2002; Posner and Serenari, 2003). Other behaviors have their meaning revealed when taking into account the context of their production. A lateral hand waving could be used to wipe dust away (gesture of action) or to convey the meaning of spreading butter on a piece of bread (iconic gesture).

Communicative behaviors are behaviors whose meaning arise from their interweaving with speech (Kendon, 2004). They cannot be interpreted without considering what is currently said. These behaviors are described by their shape (e.g., particular hand shape, eye direction) and place of occurrence in the discourse. They may be used to indicate a point in space (deictic gesture), to describe an action (e.g., the action of fishing) or the shape of an object (iconic gesture), to represent an abstract idea (metaphoric gesture) or mark the rhythm of speech (beat gesture) (McNeill, 1992). Nonverbal behaviors can also be classified by the type of information they provide on the mental state of the speaker. Isabella Poggi proposes a taxonomy based on the types of information that behaviors convey. They can provide information on the world (such as the deictic gestures and iconic eyes gestures), on the speaker's mind (raising eyebrows to show uncertainty) and on the speaker's identity (e.g., culture, gender and age) (Poggi, 2003).

Nonverbal behaviors are related to the quality of the mental state as the examples above show. But they also refer to quantity that can be linked to an intensity factor of the mental state (Wallbott, 1998). For example a raised eyebrow can be a signal of uncertainty (quality factor as noted by Wallbott (1998)). The eyebrows can be slightly or very much raised showing how uncertain the speaker

* Tel.: +33 148703702; fax: +33 148703467.
E-mail address: pelachaud@iut-univ.paris8.fr

0167-6393/$ - see front matter © 2008 Published by Elsevier B.V.
doi:10.1016/j.specom.2008.04.009

is (a little or a lot). The intensity of the muscular contraction is somehow linked to the intensity factor of uncertainty. Thus behaviors encode content information (the 'What' is communicating) and expressive information (the 'How' it is communicating).

In this paper we are concentrating on the expressivity factors of nonverbal behaviors. In previous work we have developed an expressivity model for an embodied conversational agent based on perceptual studies (Wallbott and Scherer, 1986; Wallbott, 1998). Six dimensions encompassing spatial and dynamic factors of a behavior have been defined and implemented (Hartmann et al., 2006). Evaluation studies have been conducted to validate the choice of the six dimensions as well as their adequacy to model behaviors with different expressivities (Buisine et al., 2004). In this paper, we report several studies we have conducted using our expressivity model. The aim of these studies is to understand at which level of animation the expressivity values act: do they act over a full sequence of gestures, on a single gesture or even on a gesture phase? Our exploratory studies are based on various data types: acted data, real data, 2D cartoon, and even literature.

After presenting existing studies on behavior expressivity, we present our expressivity model then we turn our attention towards the various studies we have worked on. They are very much diverse. Section 4 describes works where the expressivity model acts over the whole animation. Two studies will be described: EmoTV that uses video corpus from real data and a system that analyzes automatically the expressivity behavior of an actor and that reproduces it on an ECA. Section 5 presents a study that starts from a manual annotation of behavior shape and expressivity from a corpora of acted data. The last section, Section 6, reports on a study where the expressivity dimensions may act on a particular phase of the gesture.

2. State of the art

Embodied conversational agents, ECAs, are software entities capable of communicating with users through verbal and nonverbal means. Most of the time, ECAs take humanoid aspects. As such they are endowed with the capability to display human-like nonverbal behaviors to convey information on their mental and emotional states (Poggi, 2003). ECAs can smile, nod and even show iconic gesture (Cassell et al., 1999; Kopp and Wachsmuth, 2004; Stone et al., 2004; Lundeberg and Beskow, 1999; Pelachaud, 2005). These ECA models have been mainly concentrated on modelling and representing the quality factors of nonverbal behaviors. When emphasizing a given word they can show a raised eyebrow (Lundeberg and Beskow, 1999; Stone and DeCarlo, 2003; Pelachaud, 2005; Krahmer and Swerts, 2004) or a beat gesture (Cassell et al., 2001). ECAs can complement speech by providing additional information: the speaker may mimic the surface quality of a building while describing it (Cassell et al., 2007), or the size of an object (Kopp and Wachsmuth, 2004). ECAs can show basic emotional expressions (Ruttkay et al., 2003; Becker and Wachsmuth, 2006) or complex ones (Niewiadomski and Pelachaud, 2007; Bui et al., 2004).

On the quantity side, fewer models have been proposed for expressive communicative gestures. Ruttkay and her colleagues proposed a behavior representation to encompass styles (Ruttkay et al., 2003). An ECA is described over a large set of dimensions ranging from its culture and profession to its emotional and physical state. All these dimensions affect the way an ECA moves and gesticulates. EMOTE (Chi et al., 2000) implements the Laban annotation scheme for dance to change, through a set of parameters, the way a gesture looks depending on values such as the strength of the gesture and its tempo. EMOTE works as a post-filter after a gesture animation has been computed and adds expressivity to its final animation. Other works are based on motion capture data and modeled gesture movements such as walking, but not necessarily communicative gestures (Ménardais et al., 2004; Liu and Hertzmann, 2005; Neff and Fiume, 2004).

Our model differs from previous studies as it is designed for communicative gestures (Hartmann et al., 2006). It acts on the gesture phases as defined by McNeill (1992) and gesture description parameters that are based on Prillwitz et al. (1989). Lately we have expanded our model to work on facial expressions. Contrarily to Ruttkay et al.'s approach, gesture expressivity does not affect gesture selection (which is done through modalities preference (Mancini and Pelachaud, 2007)). Our model modifies a gesture by modulating its spatial and temporal properties. Our work differs from Emote as, in our approach, the gestures are first modified and then the animation is computed, while Emote acts as a filter of a pre-calculated animation.

3. Our expressivity model

We have developed a model of behavior expressivity based on perceptual studies conducted by Wallbott (1985), Wallbott (1998), Wallbott and Scherer (1986) and Gallaher (1992). These studies describe expressivity along several dimensions.

In his study, Wallbott (1998) asked actors to portray fourteen different emotions for a given scenario. His aim was to see if emotions could be characterized by specific body movement and posture. A coding schema for body movement and posture was designed. Body movement for each anatomy part (hand, arm, head, shoulder, and upper body) was encoded as well as movement quality. For this last item, three dimensions were annotated: 'movement activity' (overall quantity of movements), 'expansiveness/spatial extension' (of body parts), and 'movement dynamics/energy/power' (of body parts). In total there were twenty-six categories in the annotation schema. From the analysis of the annotation it was apparent that most of the categories (seventeen in total) served to differentiate emotions. In particular the three movement quality dimensions showed significant differences for the fourteen emo-