2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust

Social and emotional turn taking for embodied conversational agents

Merijn Bruijnes
Human Media Interaction, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
m.bruijnes@utwente.nl

***Abstract*—In this doctoral consortium paper I describe the theme of my research; the model-based generation of consistent emotional turn taking behavior in virtual human conversations and the evaluation of this behavior. My goal is to investigate and *generate* convincing social behavior in embodied conversational agents.**

***Keywords*-Social Interaction; Conversation; Emotion; Turn Taking; Virtual Humans; Human-Computer Interaction**

INTRODUCTION

Natural interaction with an embodied conversational agent (ECA) is something that has proven to be difficult to achieve. In particular, turn taking in a conversation between an ECA and a human is often very unnatural. Agent systems that immediately halt their speech if the user makes a sound are abundant. The same goes for systems that only pay attention to the user at a system-determined time. Albeit workable, this is not a natural way of interacting for humans. Recent ECAs are capable of more natural turn taking. However, their turn taking strategies are often based on statistical features of human-human turn taking. This does not take into account the complex underlying reasons humans can have for taking a turn in a conversation. An emotion can be one of the underlying reasons for (not) taking a turn. Most ECAs are capable of displaying emotions, like facial expressions. Very few display emotions in relation to turn taking[5]. An ECA that has a model for the relation between emotion and turn taking will be able to exhibit more natural turn taking behavior[4].

For many professions conversations are central, often emotionally laden, and difficult (e.g. police interrogation). A realistic conversational system might help train some of the needed skills for such professions. The Dutch National Police (KLPD), a partner in the natural interaction project (COMMIT), has interest in such training solutions. Therefore, my research and the ECA developments will be in the context of (but not limited to) police interviews. Eventually, our goal is to develop a serious game / training exercise. Here the ECA can 'play' the police officer or suspect in a role-playing exercise (as explained later).

The theme of my research is: the model-based generation of consistent emotional turn taking behavior in virtual human conversations and the evaluation of this behavior. The goal is to investigate and *generate* convincing social behavior. It is explicitly not the goal to create systems and models for the detection of social behavior, nor achieving state of the art 3D models and animations, nor make an ECA that is able to predict the effect of its actions. I want to achieve generation of sufficiently rich social behavior in conversations in an ECA, that is perceived as convincing and natural. The ECA should have a character, a presence, and it should *'feel alive'*.

We are developing a conceptual and computational model for emotional turn taking. In this emotional model, social emotions such as the stances from Leary's rose [3] play an important role. The position an ECA has on Leary's rose provides the interaction style that the agent displays, e.g. a dominant position might mean the ECA interrupts more. Further, turn taking is implemented in a finite state machine (similar to [2]). Currently, we are implementing how the position on Leary's rose modulates the turn taking of the finite state machine.

In parallel to this, human-human behavior in emotional conversations are being recorded. This is in the context of a role-playing exercise where an actor plays the suspect and police trainees interrogate them. These conversations will be recreated in ECA-ECA behavior, resulting in natural human-human behavior played out by two ECAs. The naturalness of the human-human and ECA-ECA conversations will be compared, by asking people to rate the naturalness of these clips. This will give insight into the effect of translating human behavior to ECA behavior and establishes a baseline to which model generated behavior can be compared.

The human-human recordings will also be analyzed on turn taking behavior and social emotions (the styles from Leary's rose). We hope to find prototypical pieces of behavior: things people tend to say and do in similar situations. Placing such pieces in a different order constitutes to generating new behavior, based on the original. This means we can create different scenarios or conversations by selecting an appropriate piece of behavior for the new situation. The naturalness of these new scenarios (ECA-ECA interactions) will be compared to the baseline, showing the feasibility of manipulating natural behavior in this manner. The prototypical pieces of behavior will also be related to the styles in Leary's rose, as we annotate the style of the behavior. When the emotional turn taking model for an ECA 'moves over' Leary's rose, it can select the appropriate behavior for the ECA to display. Hopefully, this leads to

978-0-7695-4848-7/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialCom-PASSAT.2012.29

model generated ECA behavior that will 'feel' natural.

Leary's rose

The rose (see figure 1) is defined by two axes: a dominance axis (vertical: above-below), which tells whether the speaker is acting dominant or submissive; and an affect axis (horizontal: together-opposed), which says something about the speaker's willingness to co-operate. The axes divide the rose into four quadrants, and each quadrant can again be divided into two octants, each stands for an interaction style. Further, Leary's theory states two rules: above-below are complementary and opposed-together are symmetric. This means that, for example, opposed behavior invokes opposed behavior and above behavior invokes below behavior.

Literature study

We conducted a literature survey on social and emotional turn taking in conversations and related computational models [4]. In this review, we made the point that emotion has an important influence on turn taking behavior. However, we found no turn taking models that consider emotion, nor any emotion models that describe turn taking. In addition, there are many computational models on emotion and on turn taking, however, few that combine the two (e.g. [5]).

Demonstrator Scenario

The demonstrator will be a serious game where the goal is to train social behaviors in conversations. The KLPD, as the end user, has a strong say in the scenario. They provide us with expertise on credible social scenarios and recordings of actual social interactions between police (trainees) and suspects (actors). In the demonstrator, a police trainee will interact with an ECA through the MultiLis setup (see *Setup*). The goal of this demonstrator is to let the trainee practice his conversation style, according to Leary's rose.

Consider the example scenario: juveniles have been loitering around a shop and have been insulting passersby. Shop owners have reported this to the police many times and, yet again, the police arrives to confront the teenagers. Our demonstrator starts the moment a police officer and a juvenile engage in conversation. The ECA might play the role of officer or of juvenile and engage in conversation with the trainee.

Setup

The recording and evaluation of human-human behavior will ideally take place in a video mediated setting. The setup used was created for the MultiLis project [1]. In this setup, each participant faces a cubicle with a one-way mirror set at an angle. A camera is placed behind this mirror to record the participant and the video feed of an other participant is projected onto this mirror. The MultiLis setup creates the illusion of direct eye contact in a mediated setting.

Another advantage of using the MultiLis setup is that it can easily be used for human-computer interactions. An

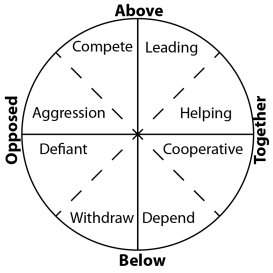

Figure 1. Leary's rose.

ECA can be projected onto the mirror as well. In this setup, the conversation with an ECA is an immersive experience, as the user has 'eye-contact' with the ECA. Also important, the recorded data of the human-human conversation is in the same shape and format as the behavior of the ECA when it is generated. This means that when the ECA acts out an earlier human conversation recording, the recording does not need to be translated; the recorded human data is compatible with the generated ECA data.

We will add a Kinect and a face reader to the setup. The Kinect will record head and body movements and the face reader can automatically extract facial expressions and gaze direction. For the recreation of human-human behavior in ECAs, a detailed annotation of the human behavior is needed. It is necessary to annotate body and head movement, gaze, and facial expression and how large they are. This is a very time consuming process when done by hand. Using the Kinect and face reader, we will try to annotate some of the human behavior automatically.

ACKNOWLEDGMENT

This publication was supported by the Dutch national program COMMIT

REFERENCES

[1] I. de Kok and D. Heylen. The MultiLis Corpus Dealing with Individual Differences in Nonverbal Listening Behavior. In *Proceedings of COST 2102*, pages 362–375, 2011.

[2] F. Kronlid. Turn taking for artificial conversational agents. In *Cooperative Information Agents X*, volume 4149 of *LNCS*, pages 81–95. Springer Berlin / Heidelberg, 2006.

[3] T. Leary. *Interpersonal Diagnosis of Personality: Functional Theory and Methodology for Personality Evaluation*. Ronald Press, New York, 1957.

[4] R. op den Akker and M. Bruijnes. Computational models of social and emotional turn-taking for embodied conversational agents : a review. Technical report, CTIT - University of Twente, 2012.

[5] B. Steunebrink, N. Vergunst, C. Mol, F. P. M. Dignum, M. Dastani, and J. Meyer. A generic architecture for a companion robot. In *Proceedings of ICINCO08*, 2008.