Laughing with a Virtual Agent

(Extended Abstract)

Florian Pecune
CNRS - LTCI, Telecom ParisTech, Paris (France)
pecune@enst.fr

Maurizio Mancini
DIBRIS - Università di Genova, Genova (Italia)
maurizio.mancini@unige.it

Beatrice Biancardi
CNRS - LTCI, Telecom ParisTech, Paris (France)
biancardi@enst.fr

Giovanna Varni
DIBRIS - Università di Genova, Genova (Italia)
giovanna.varni@unige.it

Yu Ding
CNRS - LTCI, Telecom ParisTech, Paris (France)
ding@enst.fr

Catherine Pelachaud
CNRS - LTCI, Telecom ParisTech, Paris (France)
pelachaud@enst.fr

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Artificial, augmented, and virtual realities

Keywords

Virtual agents, Expressivity, Copying, Movement analysis, Laughter, Music, Evaluation

1. INTRODUCTION

Laughter is an important social signal which may have various communicative functions [3]. It can show pleasure, mask embarrassment or even act as an indicator of in-group belonging. A laughing agent can modify the perception of funniness a user has of an event, compared to a smiling agent [4]. Endowing machines with laughter capabilities is still a crucial challenge to develop virtual agents and robots able to act as companions, coaches, or supporters in a more natural manner. In this work, we follow the paradigm introduced by previous scholars on copying behaviors ([2]) and we extend it to allow for copying expressivity on the fly, that is copying the expressivity dynamically as it evolves in human's performance. Our aim is to study how a virtual agent able to copy and to adapt its laughing and expressive behaviors on the fly participates in enhancing user's experience in the interaction. We present an experiment in which a human participant listens to audio stimuli in the presence of a laughing agent performing scripted Vs. copying expressivity. More precisely, in our setup: the agent always laughs at pre-defined moments which are linked to the audio stimulus playback; the agent's quality of movement (direction and amplitude of laughter movements) can be modulated on-the-fly depending on user's quality of movement (copying condition) or not (scripted condition).

2. GENERAL ARCHITECTURE

The architecture of laughing agent (see Figure 1) is the result of the link between two platforms: the *Detection Platform* implemented via EyesWeb XMI platform [6], and the *Virtual Agent* designed using the Greta agent platform [7]. These two platforms communicate via ActiveMQ messages.

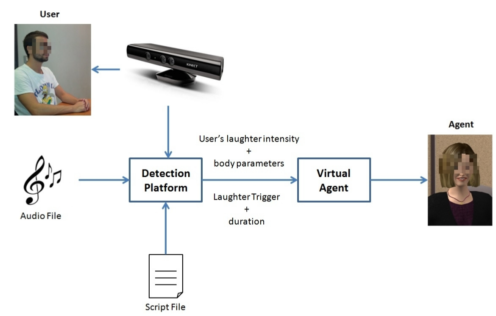

Figure 1: the architecture of laughing agent.

2.1 Detection Platform

The Detection Platform plays audio files, sends triggers for laughter, and analyzes the user's quality of movement. Since we could not automatically compute laughter triggers from audio features, we use script file containing: *laughter timings* and *laughter intensities*. Instead, user's quality of movement (direction and Quantity of Movement) is extracted in real-time and continuously transmitted to the Virtual Agent. Body features extraction is computed from the depth map video of a Kinect [1] sensor. Head's and shoulder's positions are determined by recursively segmenting the user's silhouette in EyesWeb XMI. We split the user's silhouette in two halves separated by the horizontal line passing through the user's silhouette barycenter. For both halves, we compute the mean depth of the pixels belonging to each area. The front/back leaning is the ratio between these 2 values.

2.2 Virtual Agent

During laughter the whole body moves. Our synthesized animation model computes facial expressions, head and torso motions. It relies on our previous data-driven works [4]. Our laughter animation generator takes as input laughter audio signals, such as its transcription into pseudo-phonemes

[1]http://www.microsoft.com/en-us/kinectforwindows

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.*

(in reference to speech phoneme) including their duration and their intensity, as well as two speech prosodic features, namely pitch and energy. Our method allows us to generate facial expression including lip shape of laughter as well as 3 types of motions: (1) modeling shaking-like movement (for shoulder and torso movements); (2) emphasizing prosodic features influence on the synthesized animation; (3) taking into account the dependencies between the movements across modalities (head and torso).

The Greta agent is able to adapt its laughter according to the user's behavior. While there is a direct mapping between the user's and the agent's leaning, the user's laughing intensity has an overall influence on the laughter animation. If the user laughs with large movements, the amplitude of the agent's movements is increased. On the other hand, if the user does not laugh at all or perform small motions, then the amplitude of the agent's laughter motions is decreased; it could decrease up to extremely low amplitude making the agent appears to stay still.

To modify the generated animation of the agent on-the-fly, we designed a graphical tool allowing us to manipulate different inputs and to blend them to obtain a new animation as output. Here, we blend two types of inputs: (1) animation parameters generated by our laughter animation model and (2) user's parameters extracted by the Detection Platform, that is, body leaning and laughter intensity. The user's body leaning is directly mapped to the agent's body leaning: if the user leans forward, the agent leans forward as well. User's laughter intensity has a global influence on the agent's body movements. A high intensity augments the amplitude of the agent's movements, whereas a small intensity reduces this amplitude.

3. METHODOLOGY AND EVALUATION

Participant is instructed to consider the agent as a colleague that works sitting on a desk in the same room; he/she will hear some music pieces, coming either from the agent's laptop or from the external environment. The participant's attention is focused on a "fake" research goal: we tell the participant that we aim to enable it to recognize laughter from humans, and to express and to exploit laughter to improve its interaction with humans.We ask participant to teach the agent "how and when to laugh" in a natural, effective, and convincing way. The choice of using audio stimuli as laughter-eliciting signals has been made in order to eliminate any visual factors. In such a way, participants are expected to spend more time looking at the agent and are more attentive to its expressive nuances. Studies by D. Huron [5] identified a set of *Humor-Evoking Devices* in classical music pieces. We exploit some of them to design and synthesize the audio stimuli for our experiment. The performance of the laughing virtual agent and its impact on the user's experience were evaluated by using questionnaires built from well-established ones described in [8, 1], in order to verify the impact of Greta and its behaviour on 3 dependent variables (DVs): user's perception of music funniness; user's mood; agent's social presence, spatial presence and believability.

The main part of the study included the participants' listening of the same piece of music twice; first alone, and then with the presence of Greta. At the end of each listening, the questionnaires relative to each of the DVs were proposed to the participants. two independent variables (IVs) were manipulated: *Agent* (within subject factor), i.e. the presence or not of Greta during the listening, and *Copying* (between subject factor), concerning the agent's behaviour during the listening. This last factor had the two following levels: *no_Copying*, when Greta laughed at prefixed times with prefixed intensity, and *yes_Copying*, when Greta laughed at prefixed times, and copied user's laughter intensity and trunk movements.

We collected data from 2 groups, each one randomly assigned to one of the *Copying* conditions (n1 = n2 = 16; 13 females, 19 males; average age range 26-35). We performed t-tests in order to check an effect of the different IVs on the rating of every questionnaire (paired t-test for the factor *Agent*, Welch t-test for factor *Copying*). The tests were carried out first considering each questionnaire as a whole, then by splitting them into constructs and finally considering each single question. Results show a strong effect of Greta's behaviour: the group which listened to the music with the agent in *yes_Copying* condition rated its social presence with higher scores than the group in *no_Copying* condition (p=0.01).

4. ADDITIONAL AUTHORS

Additional authors: Antonio Camurri and Gualtiero Volpe (DIBRIS, Università degli Studi di Genova email: `antonio.camurri@unige.it gualtiero.volpe@unige.it`) [2].

5. REFERENCES

[1] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[2] G. Castellano, M. Mancini, C. Peters, and P. McOwan. Expressive copying behavior for social agents: A perceptual analysis. *IEEE Transactions on Systems Man and Cybernetics-Part A-Systems and Humans*, 42(3):776, 2012.

[3] A. Chapman. Humor and laughter in social interaction and some implications for humor research. In P. McGhee and J. Goldstein, editors, *Handbook of humor research, Vol. 1*, pages 135–157, 1983.

[4] Y. Ding, K. Prepin, J. Huang, C. Pelachaud, and T. Artières. Laughter animation synthesis. In *AAMAS'14*, pages 773–780, 2014.

[5] D. Huron. Music-engendered laughter: An analysis of humor devices in PDQ Bach. In *Proc. of Int. Conf. on Music Perception and Cognition. Evanton, Illinois*, pages 700–704, 2004.

[6] M. Mancini, G. Varni, R. Niewiadomski, G. Volpe, and A. Camurri. How is your laugh today? In *CHI'14*, pages 1855–1860, 2014.

[7] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud. Greta: an interactive expressive ECA system. In *AAMAS'09*, pages 1399–1400, 2009.

[8] R. Prada and A. Paiva. Teaming up humans with autonomous synthetic characters. *Artificial Intelligence*, 173(1):80–103, 2009.

[2]Acknowledgements: The research leading to these results has received funding from the from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 270780 and from the ANR project MoCA