Animating a Conversational Agent
with User Expressivity

M. K. Rajagopal[1]   P. Horain[1]   C. Pelachaud[2]

[1] Institut Telecom, Telecom SudParis, 9 rue Charles Fourier, 91011 Évry Cedex, France
[2] CNRS Telecom ParisTech, 37-39 rue Dareau, 75014 Paris, France

[1]{Manoj_kumar.Rajagopal, Patrick.Horain} @Telecom-sudparis.eu

[2]Catherine.Pelachaud@telecom-paristech.fr

Abstract. Our objective is to animate an embodied conversational agent (ECA) with communicative gestures rendered with the expressivity of a real human user it represents. We describe an approach to estimate a subset of expressivity parameters defined in the literature (namely spatial and temporal extent) from captured motion trajectories. We first validate this estimation against synthesis motion and then show results with real human motion. The estimated expressivity is then sent to the animation engine of an ECA that becomes a personalized autonomous representative of that user.

1   Introduction

In day-to-day life, people express gestures with their own natural variations. These variations are consistent with individuals across gesture, so we call them "expressivity". We considered the expressivity parameters defined by Hartmann *et al.* [1] that are based on the wrist movement in 3D space, irrespective of joint angles (shoulder, elbow, etc.) information. In this work, we estimate spatial extent (SPC) and temporal extent (TMP) of the Hartmann's *et al.* [1] expressivity parameters from captured user motion and then animate an ECA to render the captured expressivity.

2   Estimating Expressivity Parameters and its Validation

Generally a gesture is formed from a set of key poses of wrist positions p having SPC and TMP as 0.0 called "basic gesture". From the wrist positions of human communicative gestures, we compute the 3D distance between the wrists and the sacroiliac vertebra. This distance should be mapped to SPC from -1 to +1. The wrist position p' of captured communicative gestures of a person is varied by a factor of SPC from basic gesture wrist positions p. (i.e.) $p' = [SPC] \cdot p$. For zero SPC value, the wrist position p' is same as p. When value of p is known, the SPC is determined by back substituting p in the above equation.

TMP is measured using wrist speed. Wrist speed in whole motion is determined from the instant speed of the wrist among each poses. Distance covered between consecutive poses defines instant speed of the motion. From the example trajectories we considered, instant speed among each pose is sorted with descending order and the range of 5 to 5.5% of upper quantile gives 99% correlation with TMP values. The 99% correlated upper quantile value is mapped to TMP ranging from -1.0 to +1.0 through linear regressions.

We tested the estimated method against synthesized motion with known SPC and TMP values for validating of our process. The absolute mean error of estimated SPC with respect to ground truth SPC is 0.13. Similarly absolute mean error for estimated TMP with respect to actual TMP value is 0.15. This error value shows our method is working well for estimating expressivity parameters.

3   Experiments

We used motion data captured by computer vision from two video lectures (hereafter named V1 and V2) using software developed by Gómez Jáuregui *et al.*, [2] and it outputs the upper body joint angles. 3D wrist positions are obtained from upper body joint angles using forward kinematics [3]. Experiment yields SPC as 0.8 for V1 and 0.6 for V2 and TMP as -1.0 for V1 and -0.7 for V2. These estimated values are given as input to the Greta [4] and the conversational agent is animated.

4   Conclusion

By estimating the SPC and TMP from motion by a real user, we can then animate the conversational agent to render the expressivity captured from a real user. This animation can be played virtually when the user is not available to control his avatar. Rendering the expressivity parameters allows generating personalized animations, so that the viewer can have the feeling of interacting with an expressive virtual human.

5   References

1. Hartmann, B., Mancini, M., Pelachaud, C.: Implementing Expressive Gesture Synthesis for Embodied Conversational Agents. In : Gesture Workshop,LNAI 3881. Springer (2006) 188-199

2. Gómez Jáuregui, D., Horain, P., Rajagopal, M., Karri, S.: Real-Time Particle Filtering with Heuristics for 3D Motion Capture by Monocular Vision. In : Multimedia Signal Processing, Saint-Malo, France., pp.139-144 (2010)

3. Craig, J.: Forward Kinematics. In : Introduction to robotics : mechanics and control 3rd edn. Prentice hall (1986)

4. Pelachaud, C.: GRETA. http://perso.telecom-paristech.fr/~pelachau/Greta/