— Draft of the extended abstract for MLMI2005 —

Analysis and Synthesis of Head Motion for Lifelike Conversational Agents

Hiroshi Shimodaira[1], Keisuke Uematsu[2], Shin'ichi Kawamoto[3], Gregor Hofer[1], and Mitsuru Nakai[2]

[1] CSTR, University of Edinburgh (http://www.cstr.ed.ac.uk/)
[2] Japan Advanced Institute of Science and Technology
[3] ATR Spoken Language Communication Research Laboratories

1 Introduction

This study aims to investigate which and what motions of lifelike conversational agents play essential role to make the agents natural. Some preliminary experimental results and future plan are shown.

Embodying conversational agents is intended to imitate paralinguistic channel or back channel which plays important role in human-human conversation. Although there have been not a few lifelike conversational agents developed, appearance and behaviour of the agents are far from being natural, and merits of embodying have not been realized that much. One of the main reasons is the lack of natural motions of the agents, including the motion of head, mouth, eyes and eye brows. To improve the quality of these motions, several pioneering works have been reported [1–3].

To tackle the problem, the present study was started conducting a subjective test of head motion and analysis of head motion using signal processing techniques.

2 Subjective Evaluation of Agent Motions

We carried out an experiment to investigate the effect of three major parts of motion, eye gaze, blink and head motion.

A customizable software agent toolkit 'Galatea' [4, 5] was used for this purpose. Galatea provides a photo-realistic talking head whose facial animation is synthesized using a single snapshot of a real human.

To realize natural motions, agent motions were manually controlled as natural as possible according to the motions of a real human. Mouth motion was controlled to synchronize with recorded speech of a real human. 13 male subjects were used, and a paired comparison analysis and the Thurstone's analysis were employed to evaluate the effect of the three types of motions.

The result is shown in Fig. 1, where we can see that head motion is more effective than the combination of eye gaze and blink, and the motion of either eye gaze or blink alone is less effective.

3 Dynamic Characteristics of Human Head

Unlike the previous studies for controlling head motion of talking heads, where direct mapping between head motion and speech features [1, 2] is used, the present study tries to develop hierarchical model that takes into account physical characteristics of human head. As a first attempt, we have employed a rather simple model, a linear model, where dynamic characteristic of head is modeled in terms of an impulse response.

We measured human head motions using a motion capture system, and estimated an impulse response assuming an IIR (all-pole) filter. The estimated impulse response was evaluated carrying out an subjective test where head motion of an agent was synthesized using the estimated impulse response when a step pulse was given. Fig. 2 shows naturalness of head motion with respect to filter order M. It can be found from the figure that the filter with $M = 4$ achieved the best naturalness.

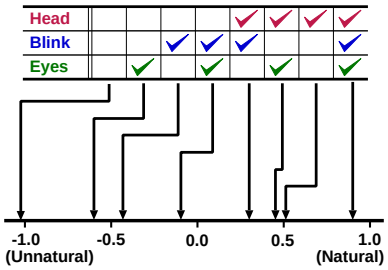

Fig. 1. Naturalness for different type of motions

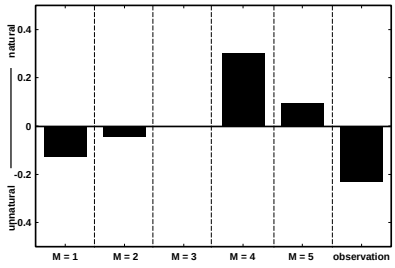

Fig. 2. Evaluation of different impulse-response filters

4 Discussions

We will discuss a difference in the manner of head rotation between real humans and the agent synthesized by Galatea. We also discuss an approach for automatic head motion synthesis based on trainable stochastic models.

References

1. Yehia, H.C., Kuratsake, T., Vatikiotis-Bateson, E.: Linking facial animation, head motion and speech acoustics. Journal of Phonetics **30** (2002) 555–568
2. Costa, M., Chen, T., Lavagetto, F.: Visual Prosody Analysis for Realistic Motion Synthesis of 3D Head Models. In: Proc. of ICAV3D'01 - International Conference on Augmented, Virtual Environments and 3D Imaging, Mykonos, Greece. (2001)
3. Lee, S.P., Badler, J., Badler, N.: Eyes Alive. ACM Trans. Graphics (SIGGRAPH2002) **21** (2002) 637–644
4. Kawamoto, S., Shimodaira, H., et al.: Galatea: Open-source Software for Developing Anthropomorphic Spoken Dialog Agents. In: Life-Like Characters – Tools, Affective Functions and Applications –. Springer (2003) 187–212
5. : (Galatea Toolkit) http://hil.t.u-tokyo.ac.jp/˜galatea/.