Spontaneous spoken dialogues with the Furhat human-like robot head

Samer Al Moubayed, Jonas Beskow, Gabriel Skantze
KTH Royal Institute of Technology
Department of Speech, Music and Hearing
Lindstedtsvägen 24, 10044 Stockholm, Sweden
{sameram, beskow, skantze}@kth.se

ABSTRACT

Furhat [1] is a robot head that deploys a back-projected animated face that is realistic and human-like in anatomy. Furhat relies on a state-of-the-art facial animation architecture allowing accurate synchronized lip movements with speech, and the control and generation of non-verbal gestures, eye movements and facial expressions.

Furhat is built to study, implement and validate patterns and models of human-human and human-machine situated and multiparty multimodal communication, a study that demands the co-presence of the talking head in the interaction environment, something that cannot be achieved using virtual avatars displayed on flat screens [2,3]. In Furhat, the animated face is back-projected on a translucent mask that is a printout of the animated model. The mask is then rigid on a 2DOF neck to allow for the control of head movements. Figure 1 shows a snapshot of Furhat in interaction.

We will show in this demonstrator an advanced multimodal and multiparty spoken conversational system using Furhat, a robot head based on projected facial animation. Furhat is an anthropomorphic robot head that utilizes facial animation for physical robot heads using back-projection. In the system, multimodality is enabled using speech and rich visual input signals such as multi-person real-time face tracking and microphone tracking. The demonstrator will showcase a system that is able to carry out social dialogue with multiple interlocutors simultaneously with rich output signals such as eye and head coordination, lips synchronized speech synthesis, and non-verbal facial gestures used to regulate fluent and expressive multiparty conversations. The dialogue design is performed using the IrisTK [4] dialogue authoring toolkit developed at KTH. The system will also be able to perform a moderator in a quiz-game showing different strategies for regulating spoken situated interactions.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – Natural Language; D.2.2 [Software Engineering] Design Tools and Techniques – State diagrams; D.2.11 [Software Engineering] Software Architectures – Languages

Keywords

Human-Robot Interaction, Multiparty interaction, Gaze, Gesture, Speech, Spoken dialog, Facial animation, Robot head, Furhat.

General Terms

Design, Algorithms, Human-Factors.

ACKNOWLEDGMENTS

This work has been partly funded by the EU project IURO (Interactive Urban Robot) No. 248314, and the SAVIR project (Situated Audio-Visual Interaction with Robots) funded by the Swedish Government (strategic research areas), and the KTH ICT Strategic Research Areas Embodied Multimodal Communication.

REFERENCES


1 For more info on Furhat, see http://www.speech.kth.se/furhat
2 For a video of the demonstration, please check: www.youtube.com/watch?v=v84e6HMFbyc