An Extensible Avatar (EA) Toolkit for Human Robot Interaction

Enas Al Tarawneh and Michael Jenkin
Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Canada
{enas, jenkin}@eecs.yorku.ca

ABSTRACT

A key problem in the development of interactive robotic systems is the lack of common tools and tool chains to support critical aspects of the interaction between the robot and a human. The Extensive Avatar (EA) Human-robot interaction toolkit seeks to address this failing. The EA Toolkit consists of two core software components: a generic speech-to-text module that converts utterances from a speaker in proximity to the robot to a standard ROS (Robot Operating System) text message, and a generic text-to-utterance module that presents a realistic 3D avatar to the user that utters natural language speech while presenting an animated avatar that is synchronized with this utterance. Each of these modules is designed to be ROS-based and to be designed to be easily integrated into general robot systems. Details of the two core modules are described below.

The speech-to-text module utilizes cloud-based software to perform generic speech-to-text mapping. This provides for continuous and active listening that detects speech in the environment, reduces the surrounding noise, and obtains the spoken words as text, simulating human listening. In addition to performing general speech to text translation, the speech to text module can be tuned to expected queries/commands from human operators thus enhancing the expected accuracy of the process and ensuring that the resulting text maps to pre-determined commands for the robot itself.

The text-to-speech module combines a standard text-to-speech generation system with a 3D avatar (puppet) whose facial animation is tied to the utterance being generated. Text messages presented to the text-to-speech module are embedded within an XML structure that allows the user to tune the nature of the puppet animation so that different emotional states of the puppet can be simulated.

The combination of these two modules enables the avatar representing the robot to appear as if they listen and recognize vocal commands given to it. The robot can answer and respond to questions given to it and can be programmed to answer customized questions, such as “take me to the manager”.

The system described here relies on a number of state of the art software modules. In particular it relies on a cloud-based speech to text recognition, a knowledge engine, a text to speech engine, a 3D character designing program, a 3D animation program, a lip syncing plugin for the animation program that extracts the sounds in words, maps them to mouth shapes and plots them according to duration and occurrence in the text in real time. An expression package controls the animated character’s mood and facial expressions. An “idle loop” process animates the avatar puppet between utterances so that the character being rendered is never still but rather interacts with external users even when not being spoken to directly.