Human Robot Interaction Through an Extensible Avatar

Enas Tarawneh and Michael Jenkin
Lassonde School of Engineering, York University, Toronto, Canada
(enas, jenkin)@eecs.yorku.ca

Introduction

• A key problem in the development of interactive robotic systems is the lack of common techniques and tools.
• The Extensible Avatar Project (EAP) seeks to address this failing by providing an easy to use ROS toolkit for understanding and generating natural language input/output capabilities.
• The toolkit consists of two software components:
  • Speech-to-text module.
  • Text-to-animated avatar module.

A Formal Utterance Markup Language

• The Avatar Utterance Markup Language (AUML) is a formal language for avatar uttering.
• This is an XML representation, thus it defines a set of rules for encoding a desired output using a textual data format.
• An utterance includes an avatar's detailed description, language, spoken words, expression and general mood.

Overview of the AUML to Utterance Process

Load system
Query for available avatars, expressions, mood, languages etc.
Using given AUML expression, select utterance transitional state
Implement using Blender
Lip sync spoken words

Text to Animated Avatar

• Goal is to provide spoken word-based control of the robot.
• Operates as a ROS node that converts spoken text into actions or an appropriate response text.
• The Google Engine continuously listens for input
• The engine reduces the noise, supports many languages, supports context aware recognition
• If the engine recognizes what is being said it generates a sentence, if it does not the system raises an exception.
• The WolframAlpha engine is used to parse the sentences and return the result of the query using the WolframAlpha engine or the commanded robot action as specified by the context-specific patterns.

Lip-syncing the Spoken Words

Preprocessing:
Duration of words in dictionary calculated using same text-to-speech engine
Get words
Get word duration
Select first frame based on expression first frame
Plot word markers according to duration
Use Google text-to-speech engine to get audio
Plot audio at first frame
Blend and Render
Plot lip-syncing of word sounds using Blender's QuickTalk plugin

Building a Realistic Utterance State Transition

- NORMAL
- SAD
- HAPPY
- EXCITED
- When not generating utterances, the avatar animates through one of a number of ‘waiting’ states to simulate a non-engaged speaker.
- Different ‘waiting’ states are supported to suggest different moods of the avatar.
- Common states within each ‘waiting’ animation and utterance animation allow seamless transitions between them.

The Avatars

• Available avatars are sorted based on the values given to the attributes of the AUML tag <AVATAR>.
• Maria and Jim are currently implemented.
• They were created using a free and open source software to create realistic 3d humans avatars called Make Human. Mhx2 was used to export the avatar to blender.

Ongoing and Future Work

• We are working on having our avatar gaze in the direction of the speaker through motion identification using a front view camera.
• Documentation and release of standard ROS packages so as to encourage their use in other robotic systems.
• Development of a full example system based on a ‘store greeter’ scenario.

Acknowledgment

The financial support of the NCFRN is gratefully acknowledged.

An example Avatar Utterance Markup Language.

[XML code for AUML example]

A display of our current avatars and a simplistic one that is not.

A) Meet Maria
B) Meet Jim
C) Not one of ours

[Image of avatars]