CoCoMaps Project

CMLabs | IIIM

CoCoMaps Hannover Demo Report

Demonstration dates: 24-27 April 2017

Link to video on Youtube: https://youtu.be/HqtN0499Vq8

EXECUTIVE SUMMARY

This report describes the results of Demo-0 of the CoCoMaps Project which is a joint effort by CMLabs (UK) and the Icelandic Institute for Intelligent Machines (Iceland). The aim of Demo-0 is listed as Milestone 3 in the agreed proposal and shows off the following deliverables:

- T8.D1 Demo 0: Hannover Messe 2017

and was carried out at the Hannover Messe 2017.

The goal of Demo-0 is to demonstrate the initial advances made in the development of the Collaborative Cognitive Map architecture where two or more independent robots can work collaboratively by sharing and comparing live observations, negotiating the execution of tasks and finally make their own decisions on what to do next.

The CoCoMaps architecture is based on the Cognitive Map Architecture that was developed by CMLabs in collaboration with Honda Motors Research Institute (HRI) in California. The CoCoMaps project aims to extend the capabilities of the CMA, which worked for a single robot interacting with a single human, having the ultimate goal of having a new improved architecture that works for multiple robots and multiple humans, allowing them to communicate and collaborate. The project has four parts, the first being a demo of a simple virtual agent (Demo-0, this report), the second involving two robots collaborating on visually searching for humans in the scene (Demo-1, the next demo), the third adding basic human-robot communicative capabilities allowing the robots to extract information from humans using natural dialogue (Demo-2, to be done after Demo-1), and the fourth and final extending this by allowing robots and humans to talk about multiple pieces of information with dynamic feedback from the humans (Demo-3, to be done after Demo-2).

This report accompanies the video recorded of the system demonstrated at the Hannover Messe. It shows the initial development of integrating facial detection and tracking, dialogue planning and execution and Two-Party Turn-Taking. This is the first time that these technologies have been working side-by-side in the Psyclone system and shows the feasibility of the project as a whole.

INTRODUCTION

The overall goal of CoCoMaps is to demonstrate that our Cognitive Map Architecture (CMA) can be extended from single robot-human relatively simple interaction to multi-robot, multi-human more dynamic and social interaction. Getting to that final version of CoCoMaps in this project involves developing several sub-components which must be tested and demonstrated thoroughly to support continuing development.

Demo-0 was added as an extra demonstration as the project team was asked to participate with a technology demonstration at the ECHORD stand at the Hannover Messe. The system demonstrates the first integration of key technologies such as speech, vision, face recognition and social turn-taking, all running in real-time on the Psyclone 2 platform. In addition, it adds synchronised visual output to provide a more appealing and eye-catching interface for passers-by.

As the main purpose of this demonstration was to speed up development of software components that did not require testing on the final robot hardware, while some KPIs (Key Performance Indicators) for Demo-1, Demo-2 and Demo-3 are partially relevant, no specific KPIs had been created that were directly relevant to this activity.

The rest of the report is organised as follows: Following Experimental Setup we present the visual Results of Demo-0 which is offered as a video recording, followed by Discussion & Future Work.

EXPERIMENTAL SETUP

This section provides a description of, in the following order, physical space, computer hardware, computer software and experimental procedure / run.

PHYSICAL SPACE

The demonstration took place at the ECHORD stand at the Hannover Messe 2017. The lighting was provided using built-in overhead fluorescent lights.

Figure 1. Demonstration setup for CoCoMaps Demo-0. At the stand four projects were represented, one of which was the CoCoMaps interactive demonstration which ran on the laptop seen in the middle of the photo.

DEMO-0 COMPUTER HARDWARE

The laptop used in this work was a Windows-based Intel system. It has a built-in web camera which was used for vision and face detection, a full HD monitor for the visual output and an external speaker to provide better output audio quality and range.

DEMO-0 COMPUTER SOFTWARE & ARCHITECTURE

The laptop runs two programs. One is the Psyclone 2 system which controls every aspect of the input, processing and output and the other is a separate Windows application programmed in Qt providing the visual interface – connected to Psyclone via the local network.

The components running in the Psyclone system relevant for Demo-0 are listed in Table 1 below. Catalogs can be seen as containers and arbitrators of data and modules are the processors, detectors and decision makers.

Table 1.
Main software components used in Demo-1.

COMPONENT	ROLE
CoCoMapsGUI	This is the visual interface showing the animated face and the textual information about the speech interaction of the demonstration. It also shows the running log of Psyclone messages down the right hand side of the interface.
SpeechRecogniser	Analyses audio frames detecting speech events and utterances
PitchTrackerInput	Detects pitch, pitch derivative and slope, silences and hums in a continuous speech.
OverlapAnalyzer	Alerts to temporal overlaps between two or more voices
YTTM	Turn-taking module which controls all reactive (low-level) turn decisions.
InteractionManager	Manages all low-level events with regards to the communication
TaskDialogManager	Manages all high-level decision wrt dialog and task execution.
FaceRecognition	This module receives the video stream from the USB camera on the robot and analyses it for faces.
NuanceTTS	Takes textual utterances and turns them into audio frames
Others	Numerous other system components have been developed that are fundamental (navigation, motor control, etc.) and not detailed here for brevity sake or because they are not essential for Demo-0.

A screenshot of the visual interface during a conversation can be seen in Figure 2.

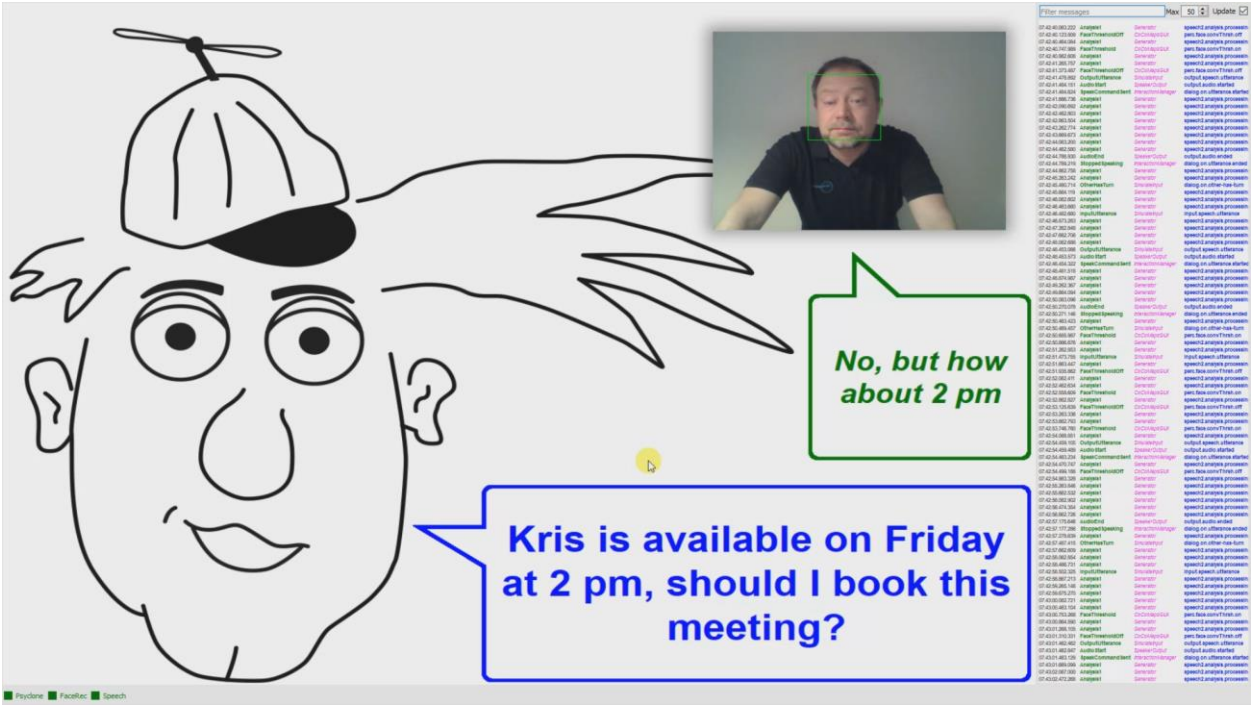

Figure 2. Screenshot of the system running on the demonstration laptop

The face displays relevant social interaction expressions and speech bubbles to highlight the audible conversation. The face of the person interacting with the system is displayed to the right and to the far right a stream of Psyclone messages are shown which describe the data communication happening in real-time between the components inside the Psyclone system.

TURN-TAKING IN DEMO-0

The turn-taking modules in the system perform the following functions:

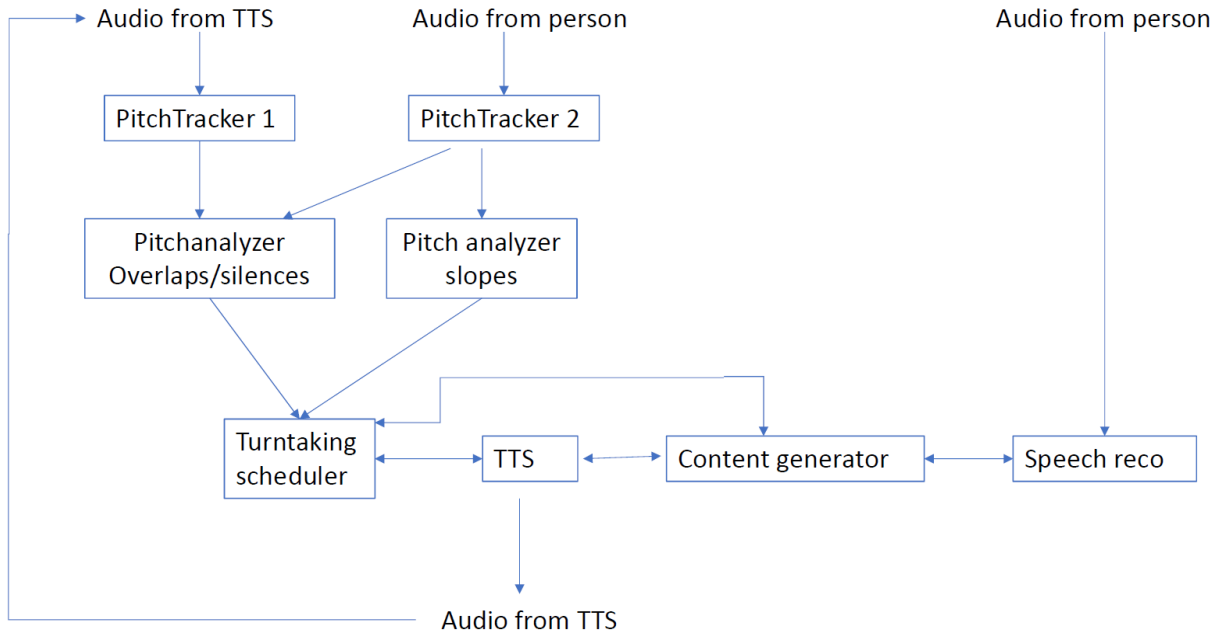

Figure 3. Turn-taking diagram as used in the Demo-0 system

In 2-party conversation each participant must keep track not only of what the other is saying and doing, but also whether they are paying attention, whether they want the turn and other indicators of importance to the dialogue state. Our model is based on the Ymir Turn-Taking Model, for which we have implemented from the ground up a new structure that will serve as the basis for both the 2-party interaction (Demo-2) and multi-party interaction (Demo-3). The model is based on a finite state machine approach with time-sensitive rules for transitions. Feeding that model is a network of modules (see Figure 3) which tracks the robots' self-speech, the human's speech, and allows future additions of information sources for determining how the dialogue state is controlled.

EXPERIMENTAL PROCEDURE & EXECUTION

The demonstration starts up in idle mode where the character is attempting to attract attention from people passing by the stand. Once a human stops it encourages the person to step closer and once close enough the actual demonstration starts.

The demonstration shows a conversation between the human and the character where they attempt to agree on a timeslot for a meeting with a third party. The reason we chose to pipe in pre-recorded audio instead of using audio from the microphone is that the noise level in the exhibition hall made speech recognition unreliable. Beyond the audio input every other part of the system was fully working as can be seen in the stream of real-time messages visible in the right hand side of the screen.

RESULTS

The Demo-0 system shows that the system works as a whole reliably with audio and visual input, turn-taking and speech and visual output, running for several hours at a time in a busy exhibition hall.

DISCUSSION & FUTURE WORK

The system development and implementation has reached a milestone for reliability and functionality, including integration of key components for future work (Demo-1, Demo-2 and Demo-3).

Next task is to integrate the system on the actual robots, making use of the onboard sensors and available processing resources and devices to perform the speech interaction.