ERICA and WikiTalk

Divesh Lala\(^1\), Graham Wilcock\(^1,2\), Kristiina Jokinen\(^3\) and Tatsuya Kawahara\(^1\)

\(^1\)Graduate School of Informatics, Kyoto University, Japan
\(^2\)CDM Interact, Helsinki, Finland
\(^3\)Artificial Intelligence Research Center, AIST Tokyo Waterfront, Japan

lala@sap.ist.i.kyoto-u.ac.jp, graham.wilcock@cdminteract.com, kristiina.jokinen@aist.go.jp, kawahara@i.kyoto-u.ac.jp

Abstract

The demo shows ERICA, a highly realistic female android robot, and WikiTalk, an application that helps robots to talk about thousands of topics using information from Wikipedia. The combination of ERICA and WikiTalk results in more natural and engaging human-robot conversations.

1 Introduction

The demo shows the results of combining a highly realistic android robot with an application that helps robots to talk fluently about thousands of topics using information from Wikipedia. The aim is to work towards a level of realism in spoken dialogue and associated non-verbal behaviours that enables more natural and more engaging conversations between humans and robots.

1.1 ERICA

ERICA is a female android that is being developed as a conversational robot with a variety of roles [Glas et al., 2016]. ERICA’s appearance is highly realistic so her conversational behaviors should also be human-like. Her text-to-speech synthesized voice is trained on a real voice actress, allowing her to generate realistic sounding backchannels, laughs and fillers. ERICA captures the user’s speech through a microphone array, so that the user is free to speak naturally without needing to hold a microphone. Speaker identification and human behavioral sensing is implemented through Kinect and a small camera.

We have already developed several functions for ERICA, including backchannels, statement response generation, engagement recognition and turn-taking models [Inoue et al., 2016; Lala et al., 2017; Lala et al., 2018; Inoue et al., 2018; Kawahara, 2018]. We have implemented ERICA in a number of roles such as an attentive listener and job interviewer. Using ERICA as an interface to Wikipedia is an additional role that we have implemented and will describe in Section 2.

1.2 WikiTalk

WikiTalk is an interactive spoken information access system that enables robots to talk fluently about thousands of topics using information from Wikipedia. The first implementation with the Pyro toolkit for AI and robotics [Blank et al., 2006] demonstrated the capability for talking about whatever topics an individual user finds interesting, and for making smooth topic shifts to related topics using links to related Wikipedia articles [Jokinen and Wilcock, 2012; Wilcock, 2012].

The implementation of WikiTalk on NAO humanoid robots added non-verbal interaction capabilities including head nods, gesturing, face-tracking, and interrupting the robot by touching its head [Csapo et al., 2012]. The robot uses beat gestures (short vertical hand movements) to mark mentioned items as potential topic shifts [Meena et al., 2012].

The NAO robot speech recognition component allows the speech recognition vocabulary to be changed dynamically. WikiTalk exploited this by predicting likely next topics as the dialogue proceeds and adding them to the vocabulary to support smooth topic shifts [Jokinen and Wilcock, 2014].

WikiTalk is multilingual and can switch between English, Finnish and Japanese [Wilcock et al., 2016; Laxström et al., 2017]. Ongoing work aims at developing long-term relations with human partners, and adding Wikipedia-based listening to Wikipedia-based talking [Wilcock and Jokinen, 2019].

2 ERICA and WikiTalk Implementation

When using ERICA as an interface to Wikipedia, our aim is to make the interaction as natural as possible, to be compatible with her realistic appearance. The user should be able...
to speak to ERICA using natural language as they would to a human, rather than a robotic command-based interface. ERICA should recognize when to start searching for a Wikipedia page without explicit notification from the user.

There are some differences in the way speech recognition works in the implementations on NAO and ERICA. As the NAO speech components allow the recognition vocabulary to be changed dynamically, WikiTalk on NAO continually updates the vocabulary by predicting likely next topics as the dialogue proceeds. This allows even very rare or very new topics to be recognized.

ERICA uses Julius as the automatic speech recognition (ASR) system with both English and Japanese recognition capability [Lee and Kawahara, 2009]. In this paper we assume the language is English. Julius currently uses a static but very large speech recognition vocabulary. Although it is not updated dynamically with changing topics, the large vocabulary allows the user to choose a very wide range of topics.

With both NAO and ERICA, the user can ask about a Wikipedia topic they wish to hear about by simply saying for example I’d like to hear about [topic name]. NAO currently uses simple word-spotting with dynamically updated vocabularies, but ERICA uses a more flexible approach for extracting the Wikipedia topic, by applying natural language processing techniques to extract the desired topic from the input utterance as follows.

Firstly, since the ASR system outputs the result in lowercase, we use truecasing1 to estimate the correct capitalization. Then we find all the named entities and nouns in the utterance using the Python module spaCy2. Since there can be both named entities and nouns (both compound and simple) in the utterance, we use a heuristic to prioritize each candidate. Named entities are given preference over compound nouns, which are prioritized over single nouns. If there are multiple named entities or noun types, we choose the one that was said last in the utterance. For example, if the user says Well, you’re a smart robot, so I want to know about philosophy the latest noun (philosophy) would be chosen by ERICA. With NAO, the word-spotting would depend on the dynamic vocabulary, so either robot or philosophy (or neither) would be recognized depending on the context.

Within Wikipedia there are many terms which could be related to multiple different topics, so disambiguation pages are required. For example, John Smith is the name of multiple famous people. Our system recognizes this disambiguation page and ERICA presents the list of candidate topics to the user. When the user hears the topic they wish to know about, they can say yes to confirm.

With NAO speech components, turn-taking is very basic. Either the robot is listening or talking, but not both at the same time. To grab the turn while NAO is talking, WikiTalk supports interrupting by touching a sensor on the robot’s head.

For ERICA, we use a more natural form of interaction. The user only has to start to talk and ERICA will capture their speech. The end of the user’s turn is recognized after the user is silent for a fixed period of time. ERICA signals that she has heard the user and is searching Wikipedia. When ERICA is speaking and the user wishes to hear about a related topic, the user can simply say that topic and if it is available, ERICA notifies the user and switches to that topic. If the user wishes to interrupt ERICA during her explanation, they can barge in and say Stop at any time. ERICA will stop speaking and apologize, allowing the user to choose a different topic.

On NAO, beat gestures are used to mark potential topic shifts, but synchronizing speech and gestures is difficult [Meena et al., 2012]. On ERICA, we tried marking potential topic shifts by changing the voice slightly, and this seemed to work quite well. This feature is optional and can be switched on or off. It is not used in the demo video.

3 Description of the Demo Video

The video (8.5 min) can be seen at https://www.youtube.com/watch?v=Aq4Rfwkrtr0.

In the first part of the video (6 min), ERICA talks about classical languages and robots, and then talks more extensively about artificial intelligence. In the later part (2 min), ERICA talks to another person about android robots.

The video shows two ways that ERICA can change topics. In smooth topic shifts ERICA moves to a related topic. At the start of the video, she is talking about classical languages and mentions Greek and Latin. When the person asks her to talk about Greek, she makes a smooth topic shift to ancient Greek. There are many other Wikipedia articles with Greek in the title, but WikiTalk finds a link from classical languages to ancient Greek so that is the most relevant continuation.

To change to a completely unrelated topic, the person first asks ERICA to stop talking about Greek. She stops, shows surprise, and asks him to suggest another topic. He asks about robots, an unrelated topic. ERICA checks Wikipedia and starts talking about robots. When he asks to hear more about autonomous robots, she makes a smooth topic shift from robots to autonomous robots, and when he asks about artificial intelligence, she makes another smooth topic shift and starts talking about AI.

ERICA continues talking about the current topic if the person continues listening. In the video she keeps talking about AI for more than two minutes, until he asks her to switch to cognitive intelligence. As there is no Wikipedia article with that title, ERICA talks about intelligence.

In the later part, another person asks ERICA about android robots. She recognizes Android but not android robots. There are multiple articles about Android, so ERICA asks him to disambiguate from a list. As the first item is Android (robot), there is no need to go through the other alternatives. While talking about android robots, ERICA says that a robot with a female appearance can also be referred to as a gynoid.

Acknowledgments

This work was supported by JST ERATO Ishiguro Symbiotic Human-Robot Interaction program (Grant Number JPM-JER1401), Japan. The third author acknowledges the support of the New Energy and Industrial Technology Development Organisation (NEDO), Japan.
References


