Multilingual WikiTalk:
Wikipedia-based talking robots that switch languages

Graham Wilcock
CDM Interact and
University of Helsinki, Finland
graham.wilcock@cdminteract.com

Kristiina Jokinen
University of Tartu, Estonia and
University of Helsinki, Finland
kjokinen@ut.ee

Abstract

At SIGDIAL-2013 our talking robot demonstrated Wikipedia-based spoken information access in English. Our new demo shows a robot speaking different languages, getting content from different language Wikipedias, and switching languages to meet the linguistic capabilities of different dialogue partners.

1 Introduction

In the digital world, information services need to be multilingual. While there has been much progress in some areas such as on-line translation, it is less clear in other areas such as interactive applications. For many people, the most effective form of communication is face-to-face, and it is important to be able to use one's mother tongue when dealing with interactive services.

Our previous demo at SIGDIAL-2013 (Jokinen and Wilcock, 2013) showed spoken information access dialogues in English with a monolingual humanoid robot. Our new demo shows a robot speaking different languages, getting information from different language Wikipedias, and switching languages to meet the linguistic capabilities of different dialogue partners.

Section 2 gives a summary of our spoken information access system, which has been described in more detail in previous papers, and Section 3 outlines the development of multilingual versions of the system. A description of the language-switching demo is given in Section 4.

2 Outline of WikiTalk

WikiTalk (Wilcock, 2012) is a spoken dialogue system for Wikipedia-based information access. On humanoid robots WikiTalk uses face-tracking, nodding and gesturing to support interaction management and the presentation of new information (Jokinen and Wilcock, 2014).

The dialogue model uses a finite state machine but the states function at a dialogue management meta-level dealing primarily with topic initiation, topic continuation, and topic switching (Wilcock, 2012; Jokinen, 2015).

An important feature is the ability to make smooth topic shifts by following hyperlinks in Wikipedia when the user repeats the name of a link. For example if the robot is talking about Japan and mentions "kanji" when explaining the Japanese name for Japan, the user can say "kanji?" and the system will smoothly switch topics and start talking about kanji after getting information from Wikipedia about this new topic.

To jump to an unrelated topic, an awkward topic shift can be made by saying "Alphabet!" and spelling the first few letters of the new topic using a spelling alphabet (Alpha, Bravo, Charlie, etc.).

The user can interrupt the robot at any time by touching the top of the robot's head. The robot stops talking, says "Oh sorry!" and waits. The user can tell the robot to continue, go back, skip to another chunk, or switch to a new topic.

The robot can take the initiative by suggesting new topics, using the "Did you know ...?" sections from Wikipedia that are new every day.

The interaction history is stored by the dialogue manager. Using heuristics, the robot avoids giving the same instructions to the user in the same way. At first the robot gives simple instructions so the user can learn the basic functionalities. Later, it suggests new options that the user may not know.

3 Multilingual WikiTalk

The first version of WikiTalk was developed with the Pyro robotics simulator (Wilcock and Jokinen, 2011; Jokinen and Wilcock, 2012). This version was monolingual and used English Wikipedia and English speech components.

A humanoid robot version of WikiTalk was implemented at 8th International Summer Work-
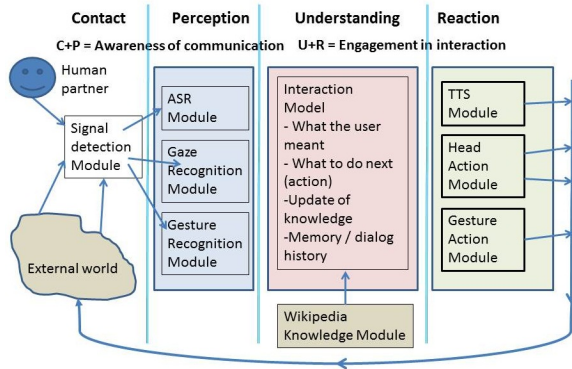
Figure 1: Architecture of WikiTalk for humanoid robots, from (Jokinen and Wilcock, 2014).

shop on Multimodal Interfaces (Csapo et al., 2012; Meena et al., 2012; Han et al., 2012). This version was also monolingual English. The system architecture is shown in Figure 1. An annotated video of the first demo can be seen at `https://drive.google.com/open?id=0B-D1kVqPMlKdOEcyS25nMWpjUG8`.

WikiTalk is very suitable for making multilingual versions. The essential requirements are the availablity of a Wikipedia in a given language and suitable speech components (recognition and synthesis) for the language. Advanced NLP tools such as syntactic parsers can also be useful but WikiTalk does not depend on them.

In order to prepare for making different language versions of WikiTalk for humanoid robots, an internationalized version of the software was developed (Laxström et al., 2014). The first two localizations were for English and Finnish. Each localized version is based on the internationalized system. Each version uses its own Wikipedia and its own speech components (i.e. English WikiTalk uses English Wikipedia and English speech components, Finnish WikiTalk uses Finnish Wikipedia and Finnish speech components).

Finnish WikiTalk was first demonstrated at EU Robotics Week 2014 in Helsinki. A video report by Iltalehti newspaper titled "This robot speaks Finnish and can tell you what is a robot" can be seen at `www.iltalehti.fi/iltvdigi/201411290140927_v4.shtml`.

A localized Japanese version of WikiTalk was developed in 2015 (Okonogi et al., 2015). This version uses Japanese Wikipedia and Japanese speech components.

We also intend to develop localized versions of WikiTalk for smaller languages such as Northern Sami which is spoken by a few thousand people in Lapland. For the revitalization of under-resourced languages in the digital world it is important for speakers of such languages to see that their language is part of the future as well as part of the past. This view may be strengthened by hearing robots speaking their language.

Currently the robot does not perform automatic language recognition, it switches language only when this is explicitly requested by the user. For example, the user says "Nihongo" to switch to Japanese, "Suomi" to switch to Finnish, "English" to switch to English. Robot-initiated language-switching raises interesting issues which will be addressed in future work.

4   The language-switching demo

The demo starts in English. The robot identifies a human face and makes eye-contact. It explains that it can talk about any topic in Wikipedia, and suggests some favourites such as Shakespeare and Manchester United. When the human moves, the robot does face-tracking to maintain eye contact.

The user selects a suggested topic, Shakespeare, so the robot downloads information about this topic directly from Wikipedia via a wifi network. The robot begins talking about Shakespeare and continues talking about this topic for a while as the human does not interrupt. After a paragraph, the robot stops and asks explicitly whether to continue or not.

After the user has listened to another paragraph about the same topic, the robot explains "You can change to other topics related to Shakespeare simply by saying them". The user then asks about Shakespeare's son Hamnet so the robot makes a smooth topic shift and talks about Hamnet Shakespeare.

The robot mentions Shakespeare's play Julius Caesar and the human says "Julius Caesar", so the robot starts talking about Julius Caesar (the play). While talking about the play, the robot mentions the historical person Julius Caesar and the human again says simply "Julius Caesar". This time the robot starts talking about the person Julius Caesar, not the play.

When the English-speaking user says "Enough" and moves away, a Japanese-speaking person approaches the robot and says "Nihongo". The robot makes eye-contact with the new person,

and switches to Japanese speech. It explains in Japanese that it can talk about any topic in Wikipedia, suggesting some favourite topics. The Japanese user also selects Shakespeare, and the robot gets information about Shakespeare, but this time from Japanese Wikipedia.

While talking about Shakespeare in Japanese, the robot also explains the Japanese versions of some basic commands and interactions. After a while the Japanese-speaking user decides to stop. The English-speaker returns. He says "English" and the robot switches back to English speech.

An annotated video (Figure 2) of the English-Japanese language-switching demo can be seen at `https://drive.google.com/open?id=0B-D1kVqPMlKdRDlkVHh4Z2tUTG8`.

Figure 2: Annotated video of an English-Japanese language-switching robot.

Acknowledgements

References

Adam Csapo, Emer Gilmartin, Jonathan Grizou, Jing-Guang Han, Raveesh Meena, Dimitra Anastasiou, Kristiina Jokinen, and Graham Wilcock. 2012. Multimodal conversational interaction with a humanoid robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 667–672, Kosice.

JingGuang Han, Nick Campbell, Kristiina Jokinen, and Graham Wilcock. 2012. Investigating the use of non-verbal cues in human-robot interaction with a Nao robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 679–683, Kosice.

Kristiina Jokinen and Graham Wilcock. 2012. Constructive interaction for talking about interesting topics. In *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul.

Kristiina Jokinen and Graham Wilcock. 2013. Open-domain information access with talking robots. In *14th Annual SIGdial Meeting on Discourse and Dialogue: Proceedings of the SIGDIAL 2013 Conference*, pages 360–362, Metz.

Kristiina Jokinen and Graham Wilcock. 2014. Multimodal open-domain conversations with the Nao robot. In Joseph Mariani, Sophie Rosset, Martine Garnier-Rizet, and Laurence Devillers, editors, *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice)*, pages 213–224. Springer.

Kristiina Jokinen. 2015. Bridging gaps between planning and open-domain spoken dialogues. In Núria Gala, Reinhard Rapp, and Gemma Bel-Enguix, editors, *Language Production, Cognition, and the Lexicon*, pages 347–360. Springer.

Niklas Laxström, Kristiina Jokinen, and Graham Wilcock. 2014. Situated interaction in a multilingual spoken information access framework. In *Proceedings of the Fifth International Workshop on Spoken Dialog Systems (IWSDS 2014)*, Napa, California.

Raveesh Meena, Kristiina Jokinen, and Graham Wilcock. 2012. Integration of gestures and speech in human-robot interaction. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, pages 673–678, Kosice.

Kenichi Okonogi, Graham Wilcock, and Seiichi Yamamoto. 2015. Nihongo WikiTalk no kaihatsu (Development of Japanese WikiTalk). In *Forum on Information Technology (FIT 2015)*, Matsuyama, Japan. (in Japanese).

Graham Wilcock and Kristiina Jokinen. 2011. Adding speech to a robotics simulator. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 371–376, Granada.

Graham Wilcock. 2012. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. In *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*, pages 57–69, Mumbai.