24th European Conference on Artificial Intelligence - ECAI 2020
Santiago de Compostela, Spain

WikiTalk and WikiListen: Towards listening robots that can join in conversations with topically relevant contributions

Graham Wilcock[1] and **Kristiina Jokinen**[2]

Abstract. The WikiTalk open-domain dialogue system performs Wikipedia-based talking. This paper highlights the potential for a new *Wikipedia-based listening* capability. The aim is to enable social robots to follow the changing topics in human-human conversations, opening the possibility for robots to join in human conversations by making topically relevant contributions at appropriate points.

1 INTRODUCTION

A Japanese project for symbiotic interaction between humans and socially-aware robots has developed ERICA [1], a female android robot with advanced conversational skills. ERICA's appearance is highly realistic (Figure 1), and to make her speech more human-like, her synthesized voice is trained on a real voice actress, allowing her to generate realistic sounding backchannels, laughs and fillers.

ERICA has several social roles [3], including *attentive listening* which encourages older people to talk and to keep talking over longer periods in order to maintain mental health. To allow people to keep talking smoothly, attentive listening focuses on backchanneling and filler generation, especially on timing, prosody and synchrony. People are encouraged to make longer, more informative contributions by simple responsive questions and flexible turn-taking [5].

A high level of attentive listening has been achieved for ERICA without requiring any kind of world knowledge, but her conversational capabilities can be extended further by using information from Wikipedia. A new *Wikipedia-based talking* role for ERICA [6] was demonstrated at IJCAI 2019 and is described in Section 2.

We highlight the potential for a further *Wikipedia-based listening* capability in Sections 3 and 4. The aim is to enable social robots like ERICA to silently follow the changing topics in human-human conversations, opening the possibility for robots to join in human conversations by making topically relevant contributions.

2 WIKITALK: WIKIPEDIA-BASED TALKING

The WikiTalk robot dialogue system [2] enables NAO robots to talk fluently and at length about thousands of different topics using information from Wikipedia. WikiTalk also follows the user's changing interests by making smooth topic shifts to related Wikipedia topics. To switch to a related topic, the user just says the name of the new topic. For example, if a NAO robot is talking about Shakespeare and says *Shakespeare was born in Stratford-upon-Avon*, the human can say *Stratford-upon-Avon?* and the robot will smoothly switch topics and start talking about Stratford-upon-Avon.

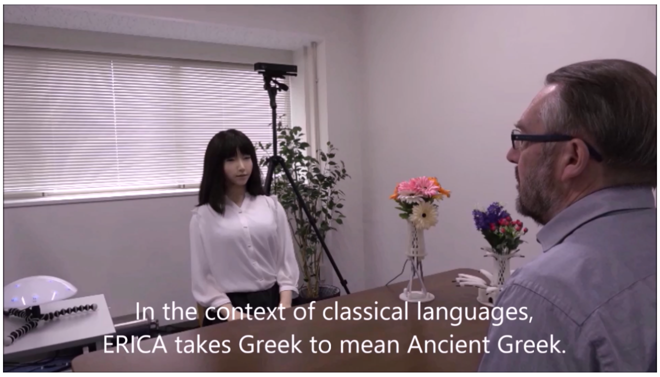
In the context of classical languages,
ERICA takes Greek to mean Ancient Greek.

Figure 1. Screenshot from a video showing ERICA and WikiTalk.

WikiTalk was used to implement a Wikipedia-based talking role for ERICA [6]. In the video at `https://www.youtube.com/watch?v=Aq4Rfwrktr0` ERICA talks about classical languages and robots (Figure 1), and more extensively about artificial intelligence. Later, ERICA talks to another person about android robots.

There are differences in speech recognition on NAO and ERICA. WikiTalk on NAO [2] continually updates the recognition vocabulary by predicting likely next topics as the dialogue proceeds, allowing even new Wikipedia topics to be recognized. ERICA uses Julius [7] for speech recognition with a static but very large vocabulary, which allows the user to choose a very wide range of topics.

With WikiTalk on NAO, turn-taking is very basic: either NAO is listening or talking, but not both at once. ERICA can support more natural interaction. Whenever the user starts to talk, Julius captures their speech. Even when ERICA is speaking, the user can simply say a topic they wish to hear about, and if it is available, ERICA switches to that topic. If the user wishes to interrupt ERICA, they can barge in at any time and say *Stop*. ERICA will stop speaking and apologize, allowing the user to choose a different topic.

3 TOWARDS WIKIFICATION OF SPEECH

For written texts, named entity recognition is a standard NLP task that recognises entities and classifies them as Location, Date, Person, etc. *Wikification* [8, 10] is a more fine-grained classification that links named entities to related Wikipedia articles. Open-source tools [11] and web services are available for wikification of written texts.

[1] CDM Interact, Helsinki, Finland, email: graham.wilcock@cdminteract.com
[2] Artificial Intelligence Research Center, AIST Tokyo Waterfront, Japan, email: kristiina.jokinen@aist.go.jp

Kim et al [4] propose wikification of concept mentions in spoken dialogues using domain constraints from Wikipedia. They identify particular differences between written texts and spoken dialogues: (1) at least two speakers are engaged in dialogues, while texts are mostly written by a single author, (2) references in spoken dialogues depend not only on the explicit context but also on speakers' background knowledge, (3) spoken utterances are more informal and noisy than written sentences, with more ambiguous and variable expressions. To solve these issues, they propose a three step approach for wikification of spoken dialogue: (1) use classifiers to analyze the dialogue-specific aspects of a given mention, (2) determine criteria for selecting concept candidates, (3) rank the filtered candidates to identify the concept most relevant to the mention.

Milde et al. [9] demonstrated *Ambient Search* at COLING 2016. This system for wikification of speech was tested by the first author, speaking into a microphone and mentioning numerous topics one after another. He was impressed by the speed and accuracy with which the relevant Wikipedia articles appeared on the demo screen. Although the demo is impressive (see the video at `https://raw.githubusercontent.com/bmilde/ambientsearch/master/demo_video_august_2016.mp4`), the system used only the smaller Simple English Wikipedia. In free testing we found that the ranking of linked articles can be puzzling, but this work shows that wikification of speech is now becoming feasible.

4 TOWARDS WIKIPEDIA-BASED LISTENING

This paper highlights the potential for Wikipedia-based listening to enable robots to silently follow the changing topics in human-human conversations using wikification of speech. This will open up new possibilities for social robots to join in the conversations by making topically relevant contributions at socially appropriate points.

The WikiTalk system already performs Wikipedia-based listening but only with one dialogue partner in a dyadic conversation. The new challenge for Wikipedia-based listening (*WikiListen* [12]) is for the robot to listen to two or more humans talking to each other (not to the robot), and to identify the changing topics that the humans are talking about by linking the topics to Wikipedia articles.

However, topic-tracking is more difficult in overheard speech than in face-to-face dialogue. With NAO robots, the robot's face-tracking (turning its head towards the partner) also optimizes the orientation of its built-in microphones to pick up speech from that partner, giving higher speech recognition accuracy in one-to-one interaction. With ERICA, the standard setup includes an external microphone array, which is more suitable for supporting multi-party interaction.

The current topic is always *grounded* in dyadic dialogues because the human partner would clarify (*No, I said elephant, not telephone!*) if the robot mis-hears and starts talking about the wrong topic. Knowing the current topic, WikiTalk predicts next topics using topic links in Wikipedia, and these predictions help ongoing speech recognition. But when humans talk to each other, topic grounding is between the humans and the robot is out of the loop. If it mis-hears a topic it will not be corrected, and its next topic predictions will also be wrong. It must then try to recover the topic thread by further listening.

5 CONCLUSION

WikiTalk [2] demonstrated Wikipedia-based talking, enabling robots to talk fluently and at length about thousands of topics. We now highlight the challenge of Wikipedia-based listening, to enable robots to listen to open-domain human conversations and recognize topics that have further information in Wikipedia. This information can then be used to generate topically relevant dialogue contributions.

A capability for social robots to first listen and then join in human conversations in an appropriate manner will be a major step forward towards symbiotic human-robot interaction. It will raise further challenges, including ethical, legal and social issues: When should robots listen and when not listen? When should they join in and not join in? When should they forget what they have overheard?

ACKNOWLEDGEMENTS

The first author thanks Prof. Tatsuya Kawahara for the opportunity to participate in the ERICA robot project at Kyoto University. The second author acknowledges the support of the New Energy and Industrial Technology Development Organisation (NEDO) in Japan.

REFERENCES

[1] Dylan F. Glas, Takashi Minato, Carlos T. Ishi, Tatsuya Kawahara, and Hiroshi Ishiguro, 'ERICA: The ERATO Intelligent Conversational Android', in *25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 22–29, New York, USA, (2016).

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

[3] Tatsuya Kawahara, 'Spoken dialogue system for a human-like conversational robot ERICA', in *Ninth International Workshop on Spoken Dialog Systems (IWSDS 2018)*, Singapore, (2018).

[4] Seokhwan Kim, Rafael E. Banchs, and Haizhou Li, 'Wikification of concept mentions within spoken dialogues using domain constraints from Wikipedia', in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2225–2229, Lisbon, Portugal, (2015). Association for Computational Linguistics.

[5] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara, 'Attentive listening system with backchanneling, response generation and flexible turn-taking', in *Proceedings of the SIGDIAL 2017 Conference*, pp. 127–136, Saarbrücken, Germany, (2017).

[6] Divesh Lala, Graham Wilcock, Kristiina Jokinen, and Tatsuya Kawahara, 'ERICA and WikiTalk', in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 6533–6535. International Joint Conferences on Artificial Intelligence Organization, (2019).

[7] A. Lee and T. Kawahara, 'Recent Development of Open-Source Speech Recognition Engine Julius', in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, (2009).

[8] Rada Mihalcea and Andras Csomai, 'Wikify! Linking documents to encyclopedic knowledge', in *Proceedings of the 16th ACM Conference on Information and Knowledge Management (CIKM 2007)*, pp. 233–242, Lisbon, (2007).

[9] B. Milde, J. Wacker, S. Radomski, M. Muhlhuser, and C. Biemann, 'Ambient Search: A Document Retrieval System for Speech Streams', in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, Osaka, (2016).

[10] David Milne and Ian Witten, 'Learning to link with Wikipedia', in *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM 2008)*, pp. 509–518, Napa Valley, (2008).

[11] David Milne and Ian Witten, 'An open-source toolkit for mining Wikipedia', *Artificial Intelligence*, **194**, 222–239, (2013).

[12] Graham Wilcock and Kristiina Jokinen, 'Towards increasing naturalness and flexibility in human-robot dialogue systems', in *Proceedings of Tenth International Workshop on Spoken Dialogue Systems*, Siracusa, Italy, (2019).