Mubin, O., Henderson, J., & Bartneck, C. (2013). Talk ROILA to your Robot.
Proceedings of the International Conference on Multimodal Interaction, Sydney pp. 317-318. | DOI: 10.1145/2522848.2531752

Talk ROILA to your Robot

Omar Mubin
School of Computing, Engineering and Mathematics
University of Western Sydney
Australia
o.mubin@uws.edu.au

Joshua Henderson
School of Computing, Engineering and Mathematics
University of Western Sydney
Australia
joshhendo@outlook.com

Christoph Bartneck
HIT Lab
University of Canterbury
New Zealand
christoph.bartneck@canterbury.ac.nz

ABSTRACT

In our research we present a speech recognition friendly artificial language that is specially designed and implemented for humans to talk to robots. We call this language Robot Interaction Language (ROILA). In this paper, we describe our current work with ROILA that utilizes the Nao humanoid robot. Our current demo implementation will allow users to interact with the Nao robot without the usage of any external laptops or microphones. Therefore the purpose of our demo is two-fold: 1) to demonstrate "live" that ROILA has improved recognition accuracy over English and 2) to demonstrate that users can interact with the Nao robot in ROILA without the use of any external devices.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

General Terms

Design, Languages

Keywords

Human Robot Interaction; Speech Recognition.

1. INTRODUCTION

ROILA is a speech recognition friendly artificial language that was designed to be easy for machines/robots to understand (recognize) and also easy for humans to learn. ROILA was created on the backdrop of the idea that natural languages are difficult for machines to understand and therefore speech recognition for natural languages has not advanced to an extent that users can use it extensively in their daily life [10, 3]. This is primarily due to how natural language has evolved, for e.g. words can sound the same but have different meaning (i.e. homophones). We have also witnessed progress in the development of algorithms for the recognition of natural language but we are yet to see an as evident usage of speech interaction in our daily lives [1]. Problems with speech recognition are a cause for concern for most researchers in Human Robot Interaction (HRI) [5]. Prasad et al [8] go as far as describing Speech Interaction with robots as the Holy Grail. Therefore in some case the robot's behavior is not autonomous but rather is controlled by the researchers (wizard of oz setups) [9] while the robots are interacting with humans. ROILA was designed on the premise that natural language speech recognition is not reliable and efficient and therefore instead of improving recognition technology perhaps optimizing the input would be a better trade off. A detailed description of ROILA is available in [7], here we summarise its main design principles. ROILA was designed on the basis of two key attributes, i.e. the proposed language should be easy for robots to recognize and at the same time easy for humans to learn. In order to provide ease in pronunciation, the phonetics of the language were built from phonemes comprised in the most widely spoken languages. In addition, word structure was based on what would be easiest to pronounce (only consonant-vowel units were included). The grammar rules were regular and inflections were not allowed (thereby reducing the number of rules speakers had to remember). To support ease of recognition by a robot, a genetic algorithm generated the vocabulary such that words would be acoustically unique from each other. Given are some sample ROILA sentences (see Table 1).

Table 1: Sample ROILA sentences

English Sentence	ROILA Translation	Literal Translation
I am walking to the house	Pito fosit bubas	I walk house
I walked to the house	Pito fosit jifi bubas	I walk <word marker> house.
I will walk to the house	Pito fosit jifo bubas	I walk <word marker> house.

We conducted an evaluation of ROILA with high school children [7], who learnt the language for 3 weeks and then took part in a controlled experiment where they used ROILA to interact with a LEGO Mindstorms robot. ROILA was shown to outperform English by 18.9% using the open source Sphinx-4 speech recognizer [6] and Festival Speech synthesis system [2]. We used the North American Acoustic Model from Sphinx-4 for the recognition of ROILA because of a) All phonemes of ROILA exist in North American English and are meant to be pronounced in the same way and b) we do not have any native speakers of ROILA hence we cannot derive data to create our own acoustic model.

We acknowledged that the results of our initial evaluation were derived under certain constrained conditions. Firstly, speech recognition was being processed on an external ma-

ICMI'13, December 9–13, 2013, Sydney, Australia
ACM 978-1-4503-2129-7/13/12.
http://dx.doi.org/10.1145/2522848.2531752.

chine and not the LEGO Mindstorms robot. Secondly, LEGO Mindstorms robots do not have a native or in-built microphone, hence only an external (desktop) based microphone was used. To fully acknowledge and determine the value of ROILA and to consequently provide an innovative and novel research platform for Human Robot Interaction ROILA needs to be easily and seamlessly portable to other platforms, robots and systems.

2. DEMO IMPLEMENTATION

Our current implementations are focusing on using the Nao humanoid robot for implementing ROILA. This is primarily because firstly, the Nao robot is widely used by HRI researchers and secondly it is Linux based therefore our system could be extended to other products. We have accomplished our implementation by firstly implementing a ROILA installation script. The script runs on any Linux machine and installs the required libraries to run the speech recognizer and speech synthesis system. The script also provides an API that can be used to write ROILA applications. Therefore, before using the Nao for interaction in ROILA we run the script on the robot. The primary limitations of our earlier research on ROILA with the LEGO Mindstorm robots was that speech recognition could not run on the robot (due to limited processing and resources) and therefore an external microphone and computer had to be used. We are now able to have users interact in ROILA with the Nao robot using the existing microphone of the Nao robot (see Figure 1). This provides a much more user-friendly and seamless interaction process. Our current setup also provides the user with feedback on how much confidence the robot has in recognizing what was said (67% as in Figure 1).

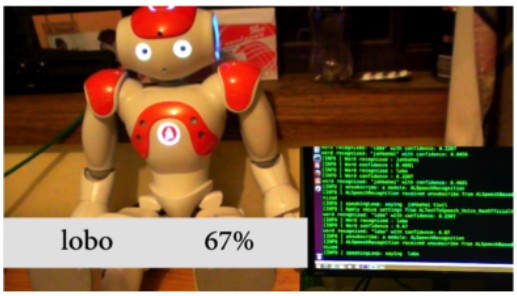

Figure 1: Running ROILA on the Nao robot

3. FUTURE RESEARCH IMPLICATIONS

Our future research road map also includes evaluating our new and improved ROILA system by conducting an empirical study. We will compare the recognition accuracy of ROILA against English of native Australian English speakers across two input conditions: headset microphone and native microphone of the Nao robot. Prior research [4] has investigated similar comparisons on the Nao robot, where a ceiling microphone was also employed. However, speech recognition was conducted offline using recordings and hence there was limited interaction between the robot and the user. Therefore it is doubtful is such a setup can be used to create HRI applications. Eventually, successful completion of our research will accomplish the following:

- Users will be provided with an "off the shelf" platform to use ROILA on not only the Nao robot but other Linux based products.
- Provide empirical proof/validation of the recognition advantage that ROILA has over English across different conditions.
- Provide researchers in Human Robot Interaction with a novel and innovative research platform.

In conclusion, we would like to acknowledge the trade-off of employing ROILA in HRI scenarios. A definite learning curve is involved and our earlier research [7] has shown that the benefits of learning ROILA lie in the long term, i.e. like learning to type with ten fingers. Users would need to interact sufficiently long enough in ROILA to gain time because of efficient recognition thereby overcoming the time spent to learn ROILA.

4. ACKNOWLEDGMENTS

We would like to acknowledge Chris Stanton who assisted in technical aspects related to the Nao robot.

5. REFERENCES

[1] J. Aron. How innovative is apple's new voice assistant, siri? *New Scientist*, 212(2836):24, 2011.

[2] A. Black, P. Taylor, R. Caley, R. Clark, K. Richmond, S. King, V. Strom, and H. Zen. The festival speech synthesis system, version 1.4. 2. *Unpublished document available via http://www. cstr. ed. ac. uk/projects/festival. html*, 2001.

[3] M. Forsberg. Why is speech recognition difficult. *Chalmers University of Technology*, 2003.

[4] S. Heinrich and S. Wermter. Towards robust speech recognition for human-robot interaction. In *Proceedings of the IEEE RSJ International Conference on Intelligent Robots and Systems*, pages 468–473, 2011.

[5] V. A. Kulyukin. On natural language dialogue with assistive robots. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 164–171. ACM, 2006.

[6] P. Lamere, P. Kwok, W. Walker, E. B. Gouvêa, R. Singh, B. Raj, and P. Wolf. Design of the cmu sphinx-4 decoder. In *INTERSPEECH*, 2003.

[7] O. Mubin, C. Bartneck, L. Feijs, H. Hooft van Huysduynen, J. Hu, and J. Muelver. Improving speech recognition with the robot interaction language. *Disruptive Science and Technology*, 1(2):79–88, 2012.

[8] R. Prasad, H. Saruwatari, and K. Shikano. Robots that can hear, understand and talk. *Advanced Robotics*, 18(5):533–564, 2004.

[9] M. Saerbeck, T. Schut, C. Bartneck, and M. D. Janse. Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1613–1622. ACM, 2010.

[10] B. Shneiderman. The limits of speech recognition. *Communications of the ACM*, 43(9):63–65, 2000.