Android-based Speech Processing for Eldercare Robotics

Tatiana Alexenko
University of Missouri
Columbia, MO
ta7cf@mail.missouri.edu

Megan Biondo
Lindenwood University
St. Charles, MO

Deya Banisakher
Bethune-Cookman University
Daytona Beach, FL

Marjorie Skubic
University of Missouri
Columbia, MO
SkubicM@missouri.edu

ABSTRACT
A growing elderly population has created a need for innovative eldercare technologies. The use of a home robot to assist with daily tasks is one such example. In this paper we describe an interface for human-robot interaction, which uses built-in speech recognition in Android phones to control a mobile robot. We discuss benefits of using a smartphone for speech-based robot control and present speech recognition accuracy results for younger and older adults obtained with an Android smartphone.

Author Keywords
Speech recognition; user study; older adults; eldercare; robotics; Android; smartphones; human-robot interaction.

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces - Voice I/O, Natural Language.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Recent studies have shown that one of the top five tasks noted by seniors for assistive robots is help with fetching objects, for example, retrieving missing eyeglasses [3], and the preferred form of communication with the robot is speech [7]. Older voices pose a challenge to Automatic Speech Recognition (ASR) systems because aging affects many acoustic parameters such as frequency, jitter and harmonic-noise ratios, and aging voices also have increased breathiness and slower speaking rates; however, ASR is still possible with specialized acoustic models [5,9].

We investigated the built-in speech recognition in Android phones for a robot fetch task. We created an Android application and implemented the underlying network and process communication system to support its use. A human subject experiment was then conducted to evaluate ASR accuracy. Voice recognition transcriptions were collected from older and younger adults. Word and whole-sentence accuracy of ASR for older and younger adults, as well as males and females, was measured and compared.

IUI'13 Companion, March 19–22, 2013, Santa Monica, CA, USA.
ACM 978-1-4503-1966-9/13/03.

OVERVIEW
This paper is part of an ongoing collaborative project to develop a robot system for the fetch task. Carlson et. al. investigated spatial language by collecting speech samples from older and younger adults [4] and created a robot capable of recognizing furniture and processing textual spatial descriptions [8]. However, there was a need for accurate ASR which is addressed in this paper.

SPEECH RECOGNITION SYSTEM
Google's ASR was chosen because it is freely available in Android-based devices which are being activated at a rate of 1 million per day worldwide [2]. Google's approach to ASR is also unique because it continuously integrates speech samples from users in addition to existing acoustic models [6]. We created an Android application that handles the audio data and sends the transcription to the robot for language processing over a wireless network. We also integrated a TCP server into the robot code. See Figure 1.

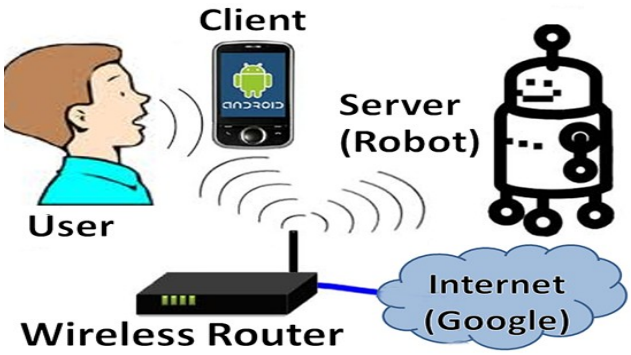

Figure 1.View of the overall system of communication.

The use of a smartphone for the voice control of a home robot has several benefits. Smartphones have built-in microphones. An ASR mobile app allows the user to decide when they want to communicate with the robot, which prevents the robot from reacting to speech directed to other people and allows the user to see and cancel incorrect transcriptions. Android ASR API also has the option of letting the user choose the best transcription from a list of suggestions; we used this feature, increased the amount of time the speaker is silent before the command is complete to accommodate older adult speech and used the Free-Text language model instead of the Web-Search model to accommodate the spatial language descriptions. Figure 2 shows our Android App at various steps; in (d) the user selects the best transcription from a list.

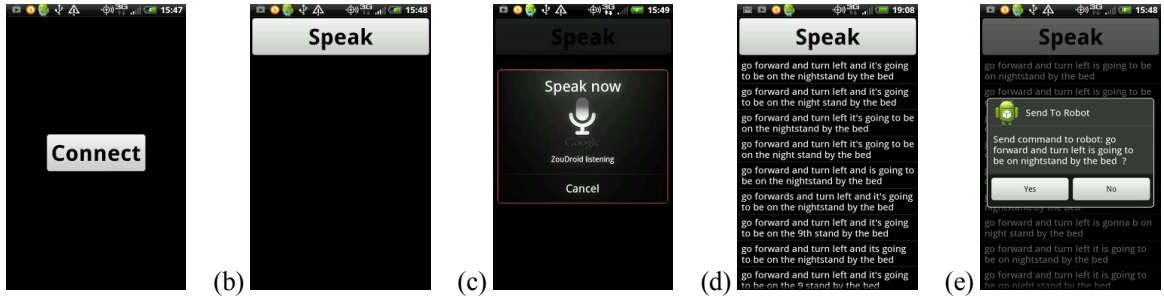

Figure 2: Android speech processing user interface at various stages: (a) Connect to robot; (b) Speak button; (c) Microphone active after "Speak" initiated; (d) List of possible transcriptions; (e) User has option of sending to robot or cancelling.

EXPERIMENTAL RESULTS

A human subject experiment was conducted to test the accuracy of Android ASR. Older (>60 years old) and younger adults were asked to read previously collected spatial descriptions [4] out loud into an Android phone with our Android app installed. The experimenter chose the most accurate transcription from a list of possibilities and this was logged. Each experimental subject read two descriptions randomly chosen from a list of sixteen. Two metrics were then calculated: per description word accuracy (number of correctly transcribed words divided by the total number of words in the description) and a binary metric that represents whether the entire description was transcribed correctly. These results are reported in Table 1 along with the total number of descriptions read by each group.

	# Descr.	Word Acc.	Binary Acc.
Older Adults			
Men	22	**79.25% (15.86)**	**2 (1.35)**
Women	31	84.66% (16.96)	10 (2.60)
All	53	82.41% (16.58)	12 (3.05)
Younger Adults			
Men	28	**94.25% (9.69)**	**17 (2.58)**
Women	20	90.18% (14.67)	8 (2.19)
All	48	92.55% (12.05)	25 (3.46)

Table 1. Experimental results. First column: total # descriptions transcribed. Second column: word accuracy rate (std). Third column: # whole descriptions correct (std).

CONCLUSION

We developed an Android application and underlying network and software system to allow remote voice control of a home robot. The human subject study has shown a statistically significant difference (p=0.01) of higher per description word accuracy rates for younger vs. older adults. Older women had 5% higher per description word accuracy than older men; while not statistically significant in this study, a similar difference was reported in [1]. The difference in binary accuracy between genders within each age group was statistically significant.

Overall, Android's ASR proved to be successful for the spatial descriptions and sample population of older and younger adults, especially considering that no modifications of acoustic models were made to accommodate older adult voices. However a higher accuracy could potentially be achieved if separate acoustic models are used for older voices and different genders.

ACKNOWLEDGMENTS

This work was supported by the NSF under grants CNS-1004606 and IIS-1017097.

REFERENCES

1. Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B., Hudson, R. Recognition of elderly speech and voice-driven document retrieval. In *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, (1999), 1:145-148.
2. Android, the world's most popular platform. 2012. http://developer.android.com/about/index.html
3. Beer, J.M., Smarr, C., Chen, T.L., Prakash, A., Mitzner, T.L., Kemp, C.C. & Rogers, W.A. The domesticated robot: design guidelines for assisting older adults to age in place. In *Proc. ACM/IEEE Intl. Conf. on HRI*, (2012), 335-342.
4. Carlson, L., Skubic M., Miller J., Huo Z., Li, X.O. A Corpus of Spatial Descriptions for the Development of Human-Driven Spatial Language Algorithms. *52nd Meeting of the Psychonomics Soc.* (2011).
5. Hooper, C.R., Cralidis, A. Normal Changes in the Speech of Older Adults: You've still got what it takes; it just takes a little longer! *Perspectives on Gerontology*, (2009), 14:47-56.
6. Now You're Talking! http://www.slate.com/articles/technology/technology/2011/04/now_youre_talking.html
7. Scopelliti, M., Giuliani, M., and Fornara, F. Robots in a domestic setting: a psychological approach. *Universal Access in the Information Soc.*, (2005), 4(2): 146-155.
8. Skubic, M., Alexenko, T., Huo, Z., Carlson, L., Miller, J. Investigating spatial language for robot fetch commands, *AAAI Technical Report*, WS-12-07, 39-45.
9. Vipperla, R., Wolters, M., Georgila, K., Renals, S. Speech Input from Older Users in Smart Environments: Challenges and Perspectives. In *Proc. Intl. Conf. on Univ. Access in HCI. Part II*. (2009), 117-126.