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.

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.
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


