Android Speech Interface to a Home Robot

Deya Banisakher[1], Megan Biondo[2], Tatiana Alexenko[3] , Zhiyu Huo[3]
Marjorie Skubic[3], Ph.D., Wenjun Zeng[3],Ph.D.
[1]Bethune-Cookman University, [2]Lindenwood University, [3]University of Missouri - Columbia

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 (Beer et al., 2012), and the preferred form of communication with the robot is a speech interface (Scopelliti et al., 2005). We investigated the use of the built-in speech recognition in Android phones for use in this scenario. We created an Android application and implemented the underlying network and process communication system to support its use. We also collected voice recognition transcriptions from old and young people; they spoke into an android device that had a testing application installed which we have developed. We also compared the accuracy of speech recognition on the Android phone for older and younger adults, as well as male and female ones.

Previous Work

Skubic et al. have studied spatial language in older and younger populations. In collaboration with Carlson et al. at Notre Dame Dept. of Psychology, they collected speech samples of older and younger adults giving spatial descriptions (Carlson et al, in review). They also created a robot capable of recognizing furniture and processing textual spatial descriptions, in addition to the common robot capabilities such as obstacle avoidance. The robot was made to listen to commands coming from the user through a computer's keyboard that is wired to the robot itself. Since it is impractical to type the spatial descriptions, there is a need for an accurate speech recognition which we addressed in our research.

Overview

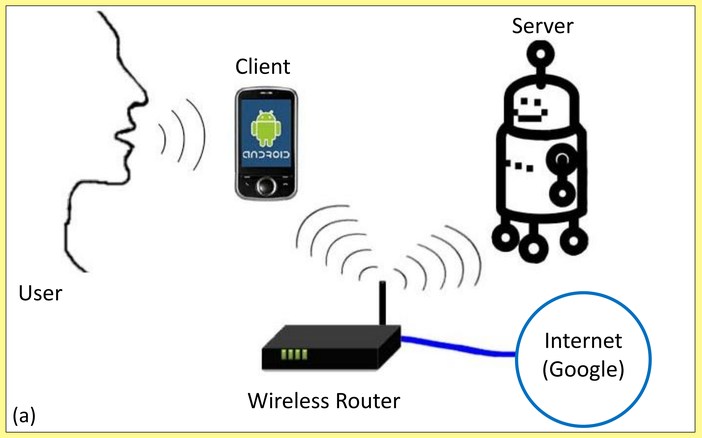
We decided to test Android's speech interface, created by Google, because it is known for high accuracy and is freely available in Android-based devices which are being activated at a rate of 1 million devices per day worldwide (Android, 2012). Google's approach to speech recognition is also unique because it relies on crowd-sourcing in addition to integration of existing acoustic models.

We created an Android application that handles the audio data and sends the transcription to the robot for processing.

The use of Android devices for this purpose also has technical benefits:

- The audio processing and transcription is handled by Google's servers.
- Android application is easy to install on any Android device.
- Android devices and the operating systems support a wide range of accessibility features for helping the elderly use the different applications installed.
- Android devices have built-in microphones, eliminating the need for the user to purchase a headset or other microphone.
- A speech recognition application 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.

Technical Overview



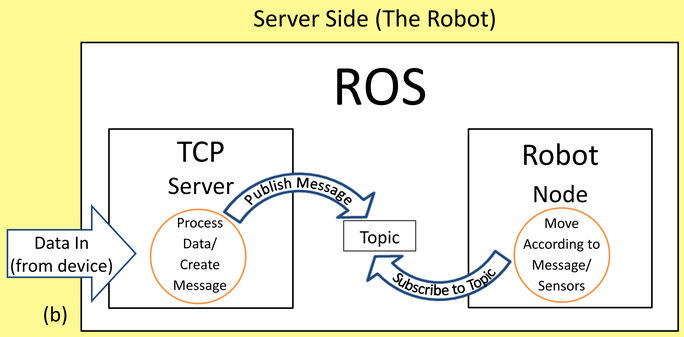


Figure 1. (a) View of overall system communication. (b) Server communication within ROS .

The robot uses ROS (Robot Operating System) which is based around publish-subscribe pattern. The server process inside of ROS publishes the textual transcriptions it receives from the Android device while other processes in the robot (primarily language processing) subscribe to the server's feed .

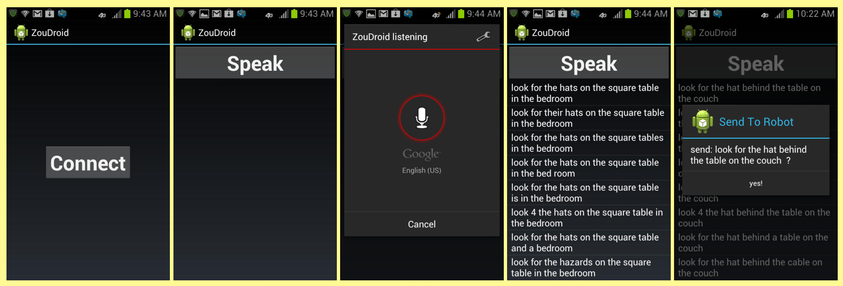
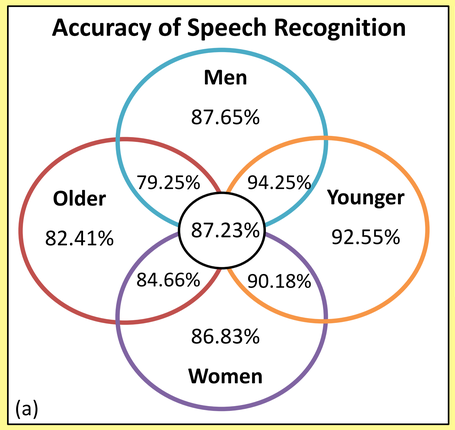


Figure 2. (a) User connects to robot. (b) User chooses to speak into phone. (c) User speaks into phone. (d) Phone displays the possible transcriptions to user. (e) Phone prompts user to send transcription selected to the robot.

Results

We tested the accuracy of Android speech recognition for older and younger adults. Accuracy is one way to measure effectiveness of Speech-to-Text. It is calculated by taking the number of correctly transcribed words and dividing by the total number of words spoken.



Older Adult Voices

	# Trans.	Average	# Perfect
Men	22	79.25%	2
Women	31	84.66%	10
All	53	82.41%	12

(b)

Younger Adult Voices

	# Trans.	Average	# Perfect
Men	28	94.25%	17
Women	20	90.18%	8
All	48	92.55%	25

(c)

Figure 3. (a) Accuracy of speech recognition comparing older adults vs. younger adults as well as male and female. (b) Accuracy results of older adult voices. (c) Accuracy results of younger adult voices.

Conclusions

- The developed Android Application has proved to be effective in sending transcriptions to the server.
- There has been a significant difference of 10% between older and younger adults' word accuracy rates with the younger voices leading.
- The binary comparison between older and younger adults' transcriptions has also shown that younger voices get transcribed better than older voices.
- However, Android's speech recognition proved to be very successful for the overall sample population of older and younger voices.

Acknowledgements

This project was made possible by the National Science Foundation under the award CNS1004606 as well as IIS-1017097.

References

1. Android, 2012. Android, the world's most popular mobile platform. http://developer.android.com/about/index.html
2. Robot Operating System (ROS), http://www.ros.org/wiki/
3. Beer, J.M., Smarr, C., Chen, T.L., Prakash, A., Mitzner, T.L., Kemp, C.C. & Rogers, W.A. 2012. The domesticated robot: design guidelines for assisting older adults to age in place. In Proc., ACM/IEEE Intl. Conf. on Human-Robot Interaction, 335-342, March, 2012, Boston, MA
4. Scopelliti, M., Giuliani, M., and Fornara, F. 2005. Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, 4(2): 146-155.
5. Carlson, L., Skubic, M., Miller, J., Huo, Z., and Alexenko, T. In Review. Investigating Spatial Language Usage in a Robot Fetch Task to Guide Development and Implement of Robot algorithms for Natural Human-Robot Interaction. *Topics in Cognitive Science*.

See also http://people.cs.missouri.edu/~reu/REU12/