On How Deaf People Might Use Speech to Control Devices

Jeffrey P. Bigham[1], Raja Kushalnagar[2], Ting-Hao Kenneth Huang[1],
Juan Pablo Flores[3,4], Saiph Savage[4]

[1] Carnegie Mellon University, <jbigham,tinghaoh>@cs.cmu.edu
[2] Gallaudet University, raja.kushalnagar@gallaudet.edu
[3] Universidad Nacional Autonoma de Mexico, jp@juanpabloflores.com
[4] West Virginia University, saiph.savage@mail.wvu.edu

ABSTRACT

Smart devices connected to the Internet are proliferating. To reduce costs of devices that have traditionally been inexpensive (toasters, microwaves, printers, etc), many of these devices have chosen to use a speech interface rather than a visual one. This transition has been hastened by the increasing capabilities of speech interfaces, exemplified by products like Amazon Echo and Apple's Siri. A consequence of these products moving to voice control is that people who are deaf and hard of hearing (DHH) may be unable to use them. In this paper, we briefly introduce two technical approaches we are pursuing for enabling DHH people to provide input to these devices: *(i)* human computation workflows for understanding "deaf speech," and *(ii)* mobile interfaces that can be instructed to speak on the user's behalf.

CCS Concepts

•**Human-centered computing** → *Accessibility theory, concepts and paradigms; Empirical studies in accessibility;*

1. INTRODUCTION

More and more devices are becoming speech-controlled, now that speech recognition is becoming better, and more devices are becoming "smarter" and thus need an interface for users (Figure 1). Speech is often preferred because devices using speech can avoid display screens that are expensive and bulky, and difficult to operate in a hands-free way. There are speech-controlled devices that are quite general (*e.g.*, Amazon's Alexa, Google Home, and Apple's HomePod), and those that are quite specific (*e.g.*, microwaves. Unfortunately, speech input/output is inaccessible for people who are deaf, and that is the only way to access much of the functionality of today's smart devices.

In this poster, we present initial work on enabling deaf people to access speech-controlled devices. We first overview the problem, and demonstrate that deaf speech is not recognized by current speech recognition technology. We then present initial work exploring whether human computation might be a more effective way to recognize deaf speech, and finally present evidence that using an accompanying mobile application that produces speech on behalf of the user may be a more fruitful path forward. We conclude with challenges for future work in this area.

Figure 1: Voice-activated devices have become common-place. Users can use voice to control their smart watches, mobile phones, personal assistant devices such as Amazon's Echo or Apple's HomePod, voice-activated vacuum robots, even smart cars. However, most voice-activated devices are not accessible to deaf people.

2. RECOGNIZING DEAF SPEECH

Deaf speech is the speech produced by deaf individuals. Because the individual cannot hear their own speech, deaf speech generally sounds different from the speech of hearing people (the so-called deaf accent). Unlike a true accent, deaf speech is characterized by large variation in pronunciation both between individuals and within a single individual. This variation can lead to difficulty in understanding, even among people familiar with deaf speech or even those familiar with the specific speech of an individual [5]. As a result, speech recognition, including that on today's proliferating smart devices, does not work for most deaf people [2].

To begin exploring the problem of recognizing deaf speech, we collected a corpus from five deaf individuals, each of whom contributed 10 common commands [4] used to control the Amazon Echo device. Example commands included: "Alexa, what is the weather today?", "Alexa, stop.", "Alexa, tell me about the movie Ghost in the Shell." We chose to use Echo as a platform of inquiry because of its popularity, because of its state-of-the-art speech recognition capabilities, and because, despite its popularity, much of its functionality, cannot be accessed without speech (even though there is an accompanying mobile application).

We first explored how well state-of-the-art general speech recognition works in understanding the commands that deaf people pro-

ASSETS '17 October 29-November 1, 2017, Baltimore, MD, USA

ACM ISBN 978-1-4503-4926-0/17/10.

DOI: https://doi.org/10.1145/3132525.3134821

duce for controlling an Amazon Echo device. For this purpose, we took all of the commands that deaf individuals produced, inserted them into Google's Speech Recognition system, and calculated the word error rate (WER). The word error rate helps us asses the performance of a speech recognition system as it measures the number of words that the system correctly identifies. We found that Google's Speech Recognition system had an average word error rate of over 40 % and a standard deviation of .4. In other words, over 40 % of the words that a deaf person pronounced were on average incorrectly classified. Notice that this error is much larger than Google's standard word error rate, which is less than 5 % [6]. Given that state of the art speech recognition did not work especially well on deaf speech, we next explored how human computation might work for understanding this speech. We posted to Amazon Mechanical Turk our 50 clips of deaf people stating different commands. Workers were pretty good at understanding the speech of one of our participants (who became deaf later in life, and whose speech was most understandable). However, this method overall worked much poorer than via automated speech recognition systems. With human computation we had an average word error rate of over 90 %, and a standard deviation of .62.

As a modification of this, we posted the speech samples again, along with the 10 phrases, and asked workers to simply choose which of the 10 phrases was most likely to be the one said. Workers were much better at this. Workers were incorrect in only 33 % of cases, *i.e.*, 67 % of the commands were correctly classified through this approach. Overall, we expect recognizing deaf speech to remain challenging with both automatic and human-powered approaches, although for some individuals it may be possible to develop hybrid recognition approaches that work.

3. INTERFACES TO PRODUCE SPEECH

In addition to improving the accuracy of automatic recognition of deaf speech, one potential direction is to empower deaf people to generate natural speech easily. If smartphones can convert text (or deaf speech) to speech that the devices such as Amazon's Echo can understand, smartphones can act as convenient and powerful tools to intermediate deaf users and voice-controlled devices. While the technology of speech synthesis (also known as Text-To-Speech, TTS) have been developed for decades~[1], literatures had little to say about its capability of generating speech commands for personal assistant devices. To understand the feasibility of using smartphones as a TTS device for communicating with voice-controlled devices, we manually typed the text of ten Alexa commands that were used in our previous speech recognition study into a TTS application[1] powered by Google Speech Recognition via an LG's Nexus 5X Android phone, and then had the phone display the synthesized speech to an Amazon's Echo device. We found that Echo was able to understand speech produced from text-to-speech and respond with valid answers in all ten trials. However, we found that some parameter tuning allowed for the best recognition. In particular, we found that the phone volume needed to be set fairly loud for it to work, and the phone needed to still be fairly close to the device. It may be difficult for a person who is unable to hear their device to do this efficiently, and so likely one challenge going forward will be to give users feedback about good placement of the phone and the audio it is outputting.

4. FUTURE WORK

Our on-going work focuses on two main activities, as represented by the preliminary work discussed here. First, we are attempting to create speech recognition that is likely to work for deaf speech. We are working with speech recognition researchers to build models that adapt to deaf speech. This is difficult because deaf speech is variable from person to person, and even within the same person across different times.

We are also pursuing more complex human computation workflows that may allows crowd workers to better recognize deaf speech. We are borrowing the iterative model from Turkit that may allow a sequence of workers to do better than a single worker [3]. We are also implementing human interface equivalents to the language models used by speech recognition systems. For example, if a particular device can only recognize certain templated commands, we are exploring ways to have workers interact within those constraints. This may make them more efficient and accurate.

Finally, we are building a mobile application that can produce speech on behalf of the deaf user. The design of this application will necessarily go beyond simply producing speech, by also providing feedback to the user about how well the produced speech was understood and recognizing speech and other audio output from the device as well.

5. CONCLUSION

In this paper, we have presented an initial investigation of the accessibility challenges presented by speech-controlled devices for deaf users. Using a sample of inputs to these devices using deaf speech, we first demonstrated that off-the-shelf speech recognition does not work well for this user group. We then presented two alternative approaches for enabling deaf people to interact with these devices: *(i)* using human computation to recognize the speech, and *(ii)* using text-to-speech functionality on a mobile application to produce the speech.

6. REFERENCES

[1] Sadaoki Furui. 2000. *Digital speech processing: synthesis, and recognition*. CRC Press.

[2] Linda~G Gottermeier, Carol~L De~Filippo, R~Aja Kushalnagar, and Bonnie~L Bastian. 2016. User Evaluation Of Automatic Speech Recognition Systems For Deaf-hearing Interactions At School And Work. *Audiology Today* 28, 2 (2016), 20–34.

[3] Greg Little, Lydia~B Chilton, Max Goldman, and Robert~C Miller. 2009. Turkit: tools for iterative tasks on mechanical turk. In *Proceedings of the ACM SIGKDD workshop on human computation*. ACM, 29–30.

[4] Taylor Martin and David Priest. 2017. The complete list of Alexa commands so far. (April 2017). https://www.cnet.com/how-to/amazon-echo-the-complete-list-of-alexa-commands/

[5] Nancy~S. McGarr. 1983. The Intelligibility of Deaf Speech to Experienced and Inexperienced Listeners. *Journal of Speech, Language, and Hearing Research* 26, 3 (1983), 451–458. DOI: http://dx.doi.org/10.1044/jshr.2603.451

[6] John Shinal. 2017. Making sense of Google CEO Sundar Pichai's plan to move every direction at once. (May 2017). http://www.cnbc.com/2017/05/18/google-ceo-sundar-pichai-machine-learning-big-data.html

[1] https://codepen.io/SteveJRobertson/pen/emGWaR