Deaf Individuals' Views on Speaking Behaviors of Hearing Peers when Using an Automatic Captioning App

Matthew Seita, Matt Huenerfauth
Rochester Institute of Technology
Rochester, NY 14623, USA
mss4296@rit.edu, matt.huenerfauth@rit.edu

Abstract
As automatic speech recognition (ASR) becomes more accurate, many deaf and hard-of-hearing (DHH) individuals are interested in ASR-based mobile applications to facilitate in-person communication with hearing peers. We investigate DHH users' preferences regarding the behaviors of the hearing person in this context. Using an ASR-based captioning app, eight Deaf/deaf participants held short conversations, with a hearing actor who exhibited certain behaviors, e.g. speaking quietly/loudly or slowly/quickly. Participants indicated some of the hearing individual's behaviors were more influential as to their subjective impression of the communication efficacy. We also found that these behaviors differed in how noticeable they were to the Deaf participants. This study provides guidance, from a Deaf perspective, about the types of behaviors hearing users should ideally exhibit in this context, motivating a focus on such behaviors in future design or evaluation of ASR-based communication apps.

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3383083

Author Keywords
Deaf and Hard of Hearing; Accessibility; Automatic Speech Recognition; Speaking Behavior

CSS Concepts
Human-centered computing → Accessibility design and evaluation methods; *Empirical studies in accessibility; Accessibility technologies*;

Introduction
Approximately 20% of the United States population report some level of hearing loss [5]. The vast majority of DHH individuals interact with hearing people regularly on a daily basis, e.g. family, friends, and co-workers. Depending on an individual's residual hearing, communication preferences, and other factors, it may be difficult to communicate using spoken English. While accessibility accommodations such as professional

captioning services or sign-language interpretation are available in some contexts, there are many settings in day-to-day life when these services are not legally required or provided. The result is that DHH individuals may miss out on information during conversations with hearing colleagues. This communication barrier can lead to feelings of isolation, frustration, and, especially in the workplace, reduced effectiveness and productivity [7]. Automatic speech recognition (ASR) technologies are an emerging solution for supporting in-person communication in spoken English for DHH individuals. These technologies automatically transcribe spoken English words to English text on a device screen (e.g. a phone or tablet); mobile apps, e.g. Google's Live Transcribe [4], can provide users with a convenient method of supporting impromptu conversations, for which it would be difficult to anticipate or schedule professional accessibility accommodation services. While ASR technologies are rapidly improving in accuracy, they are still imperfect, especially in some noisy environments or for some speaker's voices [1]. More research is needed to determine if these tools are sufficiently useful and reliable for DHH users.

While ASR researchers are working to improve recognition accuracy, HCI researchers may also be able to design such applications to support DHH individuals' understanding of conversations. In prior work, we have found that hearing people tend to change their speaking behaviors when talking to a DHH person using ASR technology [10]. This finding motivates several questions: Does this change in behavior of hearing participants make them easier to understand, whether to the technology or the DHH person themselves? And are these induced behavioral changes in hearing individuals looked upon favorably by DHH people? Understanding the viewpoint of DHH users is important because even if certain behavior changes are beneficial for increasing ASR accuracy, they could still detract from the experience of DHH people who are attempting to communicate in spoken English. The contribution of this study is **empirical**: We conduct in-person interviews among Deaf/deaf participants, after they have short ASR-app-supported conversations with a hearing actor who exhibits a variety of behaviors. We report which speaking and communication behaviors of hearing speakers were noticed by the participants and which behaviors they subjectively preferred. These findings motivate additional research on influencing these behaviors among hearing users in this context and suggests monitoring these behaviors during evaluation studies of such applications.

Related Work

Prior work has investigated DHH individuals' subjective opinions regarding ASR technologies as an accessibility tool, e.g. identifying shortcomings or areas for improvement [6]. Elliot et al. found that the DHH community was interested in using ASR technologies and was satisfied with a prototype app they used [3]. Researchers have also interviewed DHH participants to gather opinions on these technologies for use in small group meetings [2] or in the workplace [8].

Prior work has found that hearing people tend to adjust their speaking patterns when speaking to DHH people or speaking to speech recognition software. Prior work has revealed that when hearing speakers spoke to an ASR system that had errors [9, 12], or to a non-native English speaker [11], they spoke more slowly, with more pauses, and with increased articulation. In prior work, we found that when hearing individuals are speaking to a DHH person while using ASR technologies for automatic captioning, hearing speakers spoke more loudly and with a higher harmonics-to-noise ratio (an indicator of voice quality) [10].

In summary, there has been prior research among DHH users to investigate their interest in using ASR tools, their subjective impression of such tools, and the effectiveness of the tools. We also know that many hearing people adjust their speech patterns when speaking to a DHH person with or without ASR technology. **There is a gap in the literature here:** None of these prior studies specifically asked DHH participants: if they notice these behavioral changes in the speech of their hearing conversation partner or whether they view these behaviors favorably. In our study, we examine many of the speech behavioral changes identified in these prior works, e.g. hyper-articulation, speech rate, and voice intensity.

Research Questions

Our research questions are as follows:

1. Which speaking/conversational behaviors of hearing people do Deaf/deaf individuals notice when using an ASR-based live captioning app?
2. Which of these behaviors do Deaf/deaf individuals subjectively view more favorably when using an ASR-based app in this context?

Methodology

Our goal was to investigate DHH participants' preferences regarding the behaviors of hearing people during live in-person conversation in spoken English, while using an ASR-based automatic captioning mobile app. As a key use-case for such apps is to support impromptu conversations in the workplace among individuals who are not able to perceive spoken English through the use of assistive listening devices or speech-reading, our study focused on individuals who identify as (culturally) Deaf or deaf, and we focus on individuals with university-level education who will enter the workforce. Our IRB-approved, in-person, experimental and interview study was conducted at our lab at Rochester Institute of Technology (RIT). A native-ASL-signer researcher and a hearing researcher (who did not know any ASL) met with each participant separately in a private room. The native-ASL-signer researcher asked all questions in ASL to the participants. Our 8 participants included 5 females, 2 males, and 1 non-binary. The median age was 25.5 years (ranged from 22 to 32). Seven of our participants identified as Deaf, one as deaf, and none as hard-of-hearing. All participants lost their hearing and learned ASL at very young ages (while they were 0-3 years old) and all said that ASL was their primary *and* preferred language of communication. All participants self-identified as having strong reading and writing English skills (either 5 or 6 on a 6-point Likert scale, with 6 indicating having very strong English skills). Six had a bachelor's degree, one had a master's degree, and one had an M.D. degree.

After answering demographic questions, participants were given a 10-inch tablet with Google Live Transcribe [4] installed. Our researchers took a few minutes to explain to each participant how the application works. As shown in Figure 1, the upper region of the screen displays text produced by ASR of speech captured by the device's microphone from the hearing person. The deaf individual can respond by clicking a keyboard icon on the bottom of the screen and using an onscreen

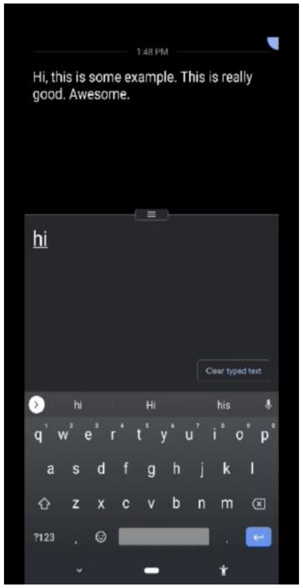

Figure 1: A screenshot of the Google Live Transcribe app, obtained from [4]

keyboard to type their response. After this familiarization period, the main portion of the study consisted of structured back-and-forth conversation between the hearing researcher and the DHH participant; during these brief conversational interactions, the hearing researcher adjusted their behavior during each interaction according to a predetermined schedule, to exhibit six categories of speech behaviors:

1. **Speech Rate** – How quickly the hearing person speaks (*fast*, *slow*, or *normal*)
2. **Voice Intensity** – How loudly the hearing person speaks (*loud*, *quiet*, or *normal*)
3. **Over-Enunciating** – Exaggerating the lip movements while speaking to emphasize each sound (*over-enunciating* or *speaking normally*)
4. **Eye Contact** – How often the hearing person maintains eye contact (*lots of eye contact*, or *mostly looking at the tablet screen*)
5. **Gesturing** – Using arm and body movements while speaking (*with gesturing*, or *without*)
6. **Intermittent Pausing** – While speaking, making sure to pause about one second after each sentence (*with pausing*, or *without*)

Our goal was to determine *to which extent a hearing person should strive to exhibit a certain behavior category* – e.g. for Speech Rate, to determine whether the hearing person should speak quickly, slowly, or normally. Each behavior category was tested following these steps (for this example, assume the behavior category being tested is **Speech Rate**):

1. The DHH participant was instructed to ask the hearing participants a specific question. The participants were given a list of seven questions beforehand (one to be used for each of the behavior categories listed earlier) and e.g. they were told "Ask the hearing participant question number one," using the app. The list of questions is provided later in this section.
2. The hearing actor responded with a specific level of the behavior, e.g. if **Speech Rate**–*fast* then the person spoke quicker than normal.
3. The DHH participants were instructed to repeat the *same* question again to the hearing person.
4. The hearing researcher would provide the *same* response, exhibiting the second level of this behavior (e.g. they would exhibit **Speech Rate**–*slow* by speaking slower than normal).
5. The DHH participants were instructed to repeat the *same* question.
6. The hearing researcher would provide the *same* response, exhibiting the last level of this behavior (e.g. if **Speech Rate**–*normal* then speak at a normal pace). Some behaviors only had two levels; so, steps 5 and 6 were omitted.
7. The participant was asked if they noticed any differences in how the hearing person responded across the two or three times they answered the question. After obtaining a yes/no answer, the researcher revealed what had been different about the speech behavior (if the participant had not noticed any change).
8. Steps above repeat for each behavior category.
9. At the end of the study, the participant was asked to give a priority score (from 1 to 10, with 10 indicating highest priority) to each of The list of seven conversational-prompt questions provided to DHH participants were as follows:

Ethical concerns for future researchers

For future researchers who are interested in replicating this study, it is important to note that care must be taken when conducting a study like this so that the participants are not exposed to communication behaviors that are rude or offensive. Our IRB-approved study was conducted under the supervision of a Deaf researcher, and the behaviors exhibited by the hearing actor were within the typical range of socially-acceptable conversational behaviors: For instance, while the actor spoke more loudly than their natural speaking voice at times, they did not yell. While the actor over-enunciated at times, they did not do so to an exaggerated degree that would appear rude, in the judgement of the DHH researcher who supervised the study.

a) What is your favorite cuisine and why?
b) What do you like about Rochester?
c) What do you NOT like about Rochester?
d) Where do you want to travel to next and why?
e) What are your plans for the winter holidays?
f) Do you play any sports, and which one do you like the most?
g) How many family members do you have and who are they?

The order of prompts was rotated across participants, but the order in which the behaviors were tested (as listed above, with Speech Rate first and Intermittent Pausing last) remained consistent. The prompts were designed to have short answers so that each question-answer exchange would not exceed 30 seconds.

Results

When our 8 participants were asked if they had noticed each of the speech behaviors: 5 noticed speech rate, 2 noticed voice intensity, 6 noticed over-enunciating, 6 noticed eye-contact, 8 noticed gesturing, and 5 noticed intermittent pausing. A Fisher's exact test did not reveal any statistically significant differences across groups, but the small sample size may have limited the power of our analysis. To determine whether there were significant differences in DHH participants' priority scores for each of the behavior categories (collected on a 1 to 10 scale at the end of the study), a Friedman rank sum test was performed on the priority scores and a significant main effect was found. Pairwise post-hoc Wilcoxon signed rank tests were performed, and there was one pairwise significant difference: DHH participants significantly prioritized **Speech Rate** over **Voice Intensity** ($p=0.001 < 0.05$).

In open-ended feedback comments collected after each behavior group, participants indicated that they were especially sensitive to several of these aspects of speech behavior among hearing individuals, e.g.:

- **Speech Rate:** Several participants indicated the importance of this aspect of speech behavior, e.g. "I don't feel it matters a lot to me personally because the device will translate everything, but it could make more mistakes if speak too fast and if she speaks fast it makes it hard to understand her and if she speaks slowly it makes me feel dumb." (Participant 8)
- **Voice Intensity:** Some participants noted differences in voice intensity, especially among individuals who use assistive listening devices, such as hearing aids or cochlear implants, e.g. "I wear a cochlear implant. Don't speak loudly, I would feel awkward like don't yell at me I know I'm deaf but that doesn't help. Loudness doesn't help clarity. Quiet voices are also not clear. Normal is best." (Participant 7)
- **Over-enunciating:** Some participants who use speech reading indicated sensitivity to this issue, saying, e.g. "I would get annoyed because the over exaggeration may actually make it harder to read lips. If there's a little exaggeration it is fine but a lot makes it harder for me to process it and piece together what they're saying. And in deaf culture, it's rude and unnecessary." (Participant 2)
- **Eye Contact:** Participants felt more connected to their conversational partner when there was greater eye contact, e.g. "When someone gives me eye contact, I feel validated. If they don't, sometimes I feel something is wrong. Also, if

Acknowledgements
We thank our participants, especially Nayeri Jacobo who played the role of the hearing participant throughout this study. This material is based upon work supported by the National Science Foundation under Award No. 1746056, by the Department of Health and Human Services under Award No. 90DPCP0002-01-00, and by a Microsoft AI for Accessibility (AI4A) Award.

there's no eye contact it gives me the feeling they are sad or need help." (Participant 1)

- **Gesturing:** Some participants indicated that while gesturing can be helpful, large or frequent gestures can be distracting when using the app or lipreading with a hearing individual, e.g. "Gesturing can be helpful in clarifying what they say. Must be accurate gesturing though; bad gesturing can be detrimental. It should be used for key points only. Gesturing all the time makes it hard for me to lipread, look at their hands and keep my eyes moving everywhere." (Participant 3)
- **Intermittent Pausing:** Participants indicated that they preferred occasional pauses (but not for *too* long) in the conversation to enable them to keep pace with the conversation, e.g. "I feel like some pauses were too long it's important to pause but the length of the pause is important as well. The pauses do help me follow along." (Participant 4)

Discussion, Limitations, and Future Work

Our study investigated the subjective preferences of DHH participants regarding specific behaviors of the hearing participant. Specifically, our quantitative analysis revealed that users significantly prioritized their hearing partner having appropriate **Speech Rate,** as compared to their prioritization of **Voice Intensity**. In addition, participants indicated whether they had noticed various speech behaviors and provided open-ended comments about their impression of these properties. These findings suggest a set of desirable behaviors among hearing people when they speak to a DHH person using ASR technologies, as well as help future researchers design ASR technologies that might persuade hearing people to behave in ways that are appealing to DHH individuals and aid their understanding of conversations. This study was just a small-scale initial study, and more work is needed to confirm these findings.

A clear **limitation** of this study is that the sample size (n=8) is small, and **future work** is needed to replicate this study with additional participants. Another **limitation** is that this study was not conducted in a multi-factorial manner: We tested each of the six behavior categories individually, with the hearing speaker encouraged to behave naturally regarding the other five categories while they focused on a specific behavior. We also did not randomize the order of behaviors presented to the participants. **Future work** could also include testing multiple behaviors at the same time, such as Speech Rate and Voice Intensity simultaneously. Another **limitation** is that while we did observe that the majority of DHH participants were using both the onscreen text and lipreading, we did not do a rigorous analysis of the behaviors of the DHH participants. A final **limitation** was that the nature of the interaction in this study was very brief and scripted. The question prompts were predetermined so that the hearing researcher could prepare responses beforehand, reducing the need for improvisation. This simplification in our study enabled the hearing actor to devote their mental energy toward displaying the specific behaviors adequately, but **future work** is needed to determine whether our findings generalize to more natural conversational interactions, including longer and more natural unscripted conversations. Such studies could employ a more naturalistic observational methodology or make use of hearing actors with additional training and improvisational skills.

References

[1] Jon P. Barker, Ricard Marxer, Emmanuel Vincent, Shinji Watanabe. 2017. The CHiME challenges: Robust speech recognition in everyday environments. In: Watanabe S., Delcroix M., Metze F., Hershey J. (eds.), New Era for Robust Speech Recognition. Springer, Cham, 327-344. https://doi.org/10.1007/978-3-319-64680-0_14

[2] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17). ACM, NY, NY, USA, 155–164. https://doi.org/10.1145/3132525.3132541

[3] Lisa B. Elliot, Michael Stinson, Syed Ahmed, and Donna Easton. 2017. User Experiences When Testing a Messaging App for Communication Between Individuals Who Are Hearing and Deaf or Hard of Hearing. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17). ACM, NY, NY, USA, 405–406. https://doi.org/10.1145/3132525.3134798

[4] Google Live Transcribe. 2019. Introducing Live Transcribe. Retrieved January 6, 2019 from https://www.android.com/accessibility/live-transcribe/

[5] Hearing Loss Association of America. 2017. Basic Facts About Hearing Loss. Retrieved December 17, 2017 from http://www.hearingloss.org/ content/basic-facts-about-hearing-loss

[6] Saba Kawas, George Karalis, Tzu Wen, and Richard E. Ladner. 2016. Improving Real-Time Captioning Experiences for Deaf and Hard of Hearing Students. In Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '16). ACM, NY, NY, USA, 15–23. https://doi.org/10.1145/2982142.2982164

[7] Ronald R. Kelly. 2015. The Employment and Career Growth of Deaf and Hard of Hearing Individuals. Raising and Educating Deaf Children: Foundations for Policy, Practice, and Outcomes. Retrieved from http://www.raisingandeducatingdeafchildren.org/2015 /01/12/the-employment-and-career-growth-of-deafand-hard-of-hearing-individuals/

[8] James R. Mallory, Michael Stinson, Lisa Elliot, and Donna Easton. 2017. Personal Perspectives on Using Automatic Speech Recognition to Facilitate Communication Between Deaf Students and Hearing Customers. In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17). ACM, NY, NY, USA, 419–421. https://doi.org/10.1145/3132525.3134779

[9] Sharon Oviatt, Gina-Anne Levow, Elliott Moreton, and Margaret MacEachern. 1998. Modeling Global and Focal Hyperarticulation during Human–Computer Error Resolution. J. Acoust. Soc. Amer. 104, 3080– 3098. https://doi.org/10.1121/1.423888

[10] Matthew Seita, Khaled Albusays, Sushant Kafle, Michael Stinson, and Matt Huenerfauth. 2018. Behavioral Changes in Speakers who are Automatically Captioned in Meetings with Deaf or Hard-of-Hearing Peers. In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18). Association for Computing Machinery, New York, NY, USA, 68–80. DOI:https://doi-org.ezproxy.rit.edu/10.1145/3234695.3236355

[11] Rein Ove Sikveland. 2006. How do We Speak to Foreigners? — Phonetic Analyses of Speech Communication between L1 and L2 Speakers of Norwegian. Working Papers 52, 109–112. Centre

for Language and Literature, Lund University, Sweden.

[12] Amanda J. Stent, Marie K. Huffman, and Susan E. Brennan. 2008. Adapting Speaking After Evidence of Misrecognition: Local and Global Hyperarticulation. Speech Commun. 50, 3 (March 2008), 163–178. https://doi.org/10.1016/j.specom.2007.07.005