Preferences for Captioning on Emulated Head Worn Displays While in Group Conversation

Gabriel Britain
Georgia Institute of Technology
Atlanta, Georgia, USA
gwbritain@gatech.edu

David Martin
Georgia Institute of Technology
Atlanta, Georgia, USA
dmartin305@gatech.edu

Tyler Kwok
Georgia Institute of Technology
Atlanta, Georgia, USA
tkwok7@gatech.edu

Adam Sumilong
Georgia Institute of Technology
Atlanta, Georgia, USA
adamsumi@gatech.edu

Thad Starner
Georgia Institute of Technology
Atlanta, Georgia, USA
thad@gatech.edu

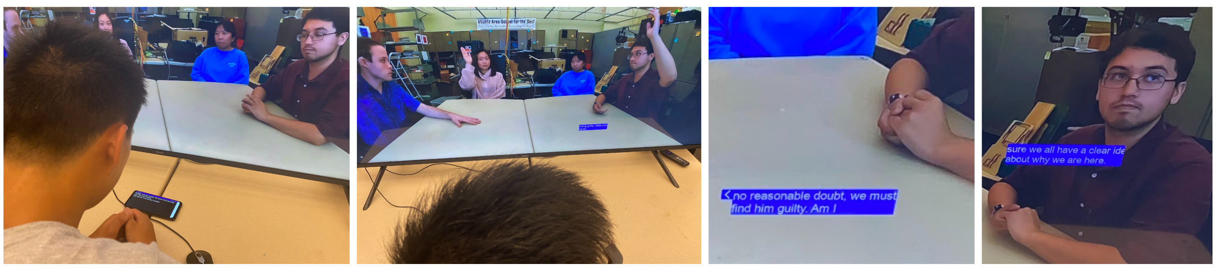

Figure 1: Phone, Non-registered (emulates a HWD), Indicators (adds arrow indicating sound direction), and Registered captions.

ABSTRACT

Head worn displays (HWDs) can provide a discreet method of captioning for people who are d/Deaf or hard of hearing (DHH); however, group conversations remain a difficult scenario as the wearer has difficulty in determining who is speaking and where to look. Using an HWD emulator during a group conversation, we compare eight DHH users' perceptions of four conditions: an 80 degree field-of-view (FOV) HWD that pins captioning text to each speaker (Registered), a HWD where the captioning remains in the same place in the user's visual field (Non-registered), Non-registered plus indicators as to which direction the current speaker is relative to the user's line of sight (Indicators), and a control of captions displayed on a Phone. Preference increased in order of Phone, Non-registered, Indicators, and Registered. While an 80 degree FOV HWD is not practical to create in a pair of normal looking eyeglasses, pilot testing with 12 hearing participants suggests a FOV between 20 and 30 degrees might be sufficient.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; *Accessibility design and evaluation methods.*

This work is licensed under a Creative Commons Attribution International 4.0 License.

ISWC '22, September 11–15, 2022, Cambridge, United Kingdom
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9424-6/22/09.
https://doi.org/10.1145/3544794.3558468

KEYWORDS

Deaf, Hard-of-Hearing, head-worn-display, captioning, accessibility

ACM Reference Format:
Gabriel Britain, David Martin, Tyler Kwok, Adam Sumilong, and Thad Starner. 2022. Preferences for Captioning on Emulated Head Worn Displays While in Group Conversation. In *The 2022 International Symposium on Wearable Computers (ISWC '22), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3544794.3558468

1 MOTIVATION AND RELATED WORK

Individuals who are d/Deaf or hard of hearing (DHH) report group conversations as being the most frequent context in which they encounter difficulties [13, 17]. Conspicuous accessibility strategies such as the use of a note-taker can lead to unwanted social attention [17, 23], and other strategies such as writing or e-mail summaries are often unsatisfactory [26]. For these reasons, people who are d/Deaf or hard-of-hearing often avoid participating in group conversations with hearing people entirely [7, 8, 13].

Real-time captioning applications using automatic speech recognition (ASR) on a smartphone may help people who are DHH participate in everyday situations. While these applications can be employed in group contexts, they struggle to accommodate multiple speakers [17]. Furthermore, presenting captions on a mobile device out of the user's direct line of sight creates a visual dispersion (a visual distance between caption and subject) that can cause a loss of information [18] and detract from the perceived quality of the conversation [1, 19]. The presence of smartphones running ASR technologies can decrease bystander social comfort [1, 19] and cause hearing individuals to change their behavior in ways that can make people who are DHH feel awkward [25, 27].

Head-worn displays (HWDs) may enable a more discreet and usable method of captioning [3, 9–11, 18, 21, 24]. Captioning on HWDs may reduce visual dispersion [18] and increase environmental awareness [9, 11], with users of HWD captions reporting improved visual contact with speakers [9]. However, prior investigations frequently report wearer concerns with the conspicuousness of bulky, distinctive-looking HWDs [6, 9, 11, 18, 20, 21], which they fear attract unwanted attention to their disability and could lead to discrimination or exploitation [4, 23].

In particular, HWDs that can render captions affixed to objects and people [21] via object tracking and spatial mapping require heavier, hotter and larger devices, unsuitable for extended use [6, 9, 11, 21]. Similarly, large field-of-view (FOV - the visual angle subtended by the display) optics are too bulky to be mistaken for normal eyeglasses. Lighter, smaller HWDs prioritize comfort and extended use and often display non-registered text which sits in a relatively small FOV and in a consistent spot in the user's visual field. However, Jain et al's work highlights the need to help people who are DHH determine who is speaking in group conversations as sound localization is an everyday struggle [6, 10, 14, 15, 22, 29].

This paper compares captions on a smartphone to those on a HWD and explores if unregistered captions with arrows pointing to the direction of the current speaker might be an alternative to registered captions (and their bulky head wear). In addition, it performs a preliminary experiment to determine what the smallest FOV (which corresponds to the lightest and least noticeable optics) might be needed for captioning.

2 EXPERIMENT

Because group conversations are a complex social environment that are difficult to reproduce consistently, we chose to film a recording. This approach enabled us to simulate "perfect" captioning, avoiding shortcomings in speech recognition and speaker segmentation technology by pre-defining the caption timings. We recorded four actors seated around a table, each playing a role from MacIntyre et al.'s *Four Angry Men* [16]. To maintain consistency, simultaneous speech was not allowed. We employed a captioning timing scheme that divided consequent utterances by their duration and their number of words so that each word lasts a uniform duration. The main captions file was then split into four files, each file containing the captions of one of the actors. We divided the video into four 2.5 minute sections. Once captioned and partitioned, we presented the video on a 165cm x 93cm flat screen TV, 98cm from the participant (approximately 120° max horizontal FOV). The video was captured and aligned on the television to give the illusion that the viewer is seeing through the TV to the physical desk at which they are seated (see Figure 1 Non-registered). The actors were arrayed to span a visual arc of 90 degrees such that the participants would naturally rotate their head when attending a given actor in the video (as they would if the actor was live).

Emulating the HWD eliminates many confounds found in other studies, such as discomfort due to head and nose weight, different focal depths of the HWD graphics and real world objects (the actors in this case), lag, resolution, contrast, brightness, and novelty effects. For consistency across participants (who had various levels of residual hearing), no sound was played. We used Google Glass Enterprise Edition 2 (GGEE2) to track the participant's head orientation using the onboard gyroscope, magnetometer, and accelerometer.

We compare four methods of presenting captions to participants: Registered, Nonregistered, Indicators, and Phone (which acts as a control condition). In all conditions, words appear one at a time, simulating the pace of speech recognition systems. In HWD conditions, captions were presented in a manner consistent with Debernardis et al.'s design suggestions [2]: white text (RGBA(255, 255, 255, 128)) on a blue (RGBA(0, 0, 255, 128)) background. The text was rendered in Arial font at 30pt size, resulting in a widest single-character horizontal visual angle of 1.06°, and a widest single-character vertical visual angle of 1.19°. The widest caption had a horizontal visual angle of 13°, and the overall field of regard for the group conversation was 82°. The Registered condition was intended to emulate the behavior of more-powerful HWDs rendering captions fixed "in space" to the person speaking [12, 21]. Depending on where the participant is looking, captions rendered outside of the emulated HWD's field of view (defined in software as 80°) are clipped (the rightmost image in Figure 1 shows an example where the word "idea" is clipped). 80° horizontal FOV is considered sufficient for situation awareness in highly mobile conditions [5], and while the optics for such a HWD would be prohibitively heavy, they are commercially available.

The Non-registered condition is representative of current capabilities of HWDs that might be mistaken as normal eyeglasses [20]: captions are rendered at a fixed location in the participant's visual field and follow their line of sight as their head turns (Figure 1 Non-registered). For the sake of simplicity, we track only the changes in the participant's head's azimuth, so captions move horizontally across the display, instead of in two dimensions. The Indicators condition adds a directional indicator to the Non-registered condition, with arrows on the left or right of the text in the direction of the current speaker if the participant's head is pointed elsewhere. The choice to use arrows as indicators was inspired by Jain et al. [10]. Our goal for this presentation method was to identify whether coarse directional feedback could provide speaker identification information without the need for registered graphics that involve intensive computation (and power-hungry cameras).

To compare our HWD conditions against a smartphone's performance, we developed a simple Android application that rendered the captions onscreen, simulating popular transcription applications such as Google's LiveTranscribe or Microsoft Translator using white text on a black background. The text size was 32 scaled pixels, and the font face used was Roboto. The Android application had a maximum line height of 10 lines.

We recruited eight DHH participants, with ages ranging from 18 to over 65, mild to profound hearing loss, and predominantly ASL/English users. The participants were recruited through personal networks, a Facebook group for DHH individuals in the local area, and faculty in our department. Participants were given a digital survey to complete over the course of the study. The participants were then presented with Google Glass EE2 and asked to position the display in a way that was comfortable for them while seated (note the display itself is not used in the experiment). We then

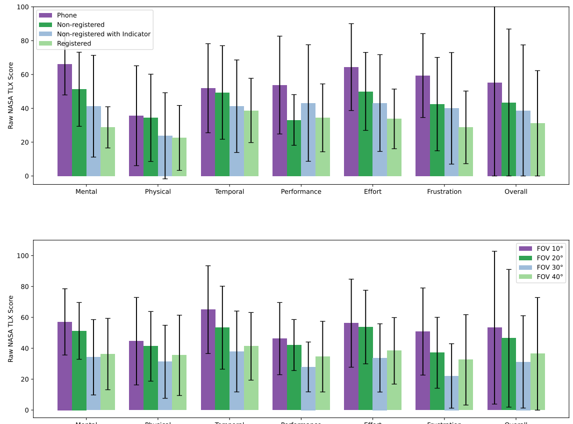

Figure 2: NASA-TLX scores for (a) caption and (b) FOV study

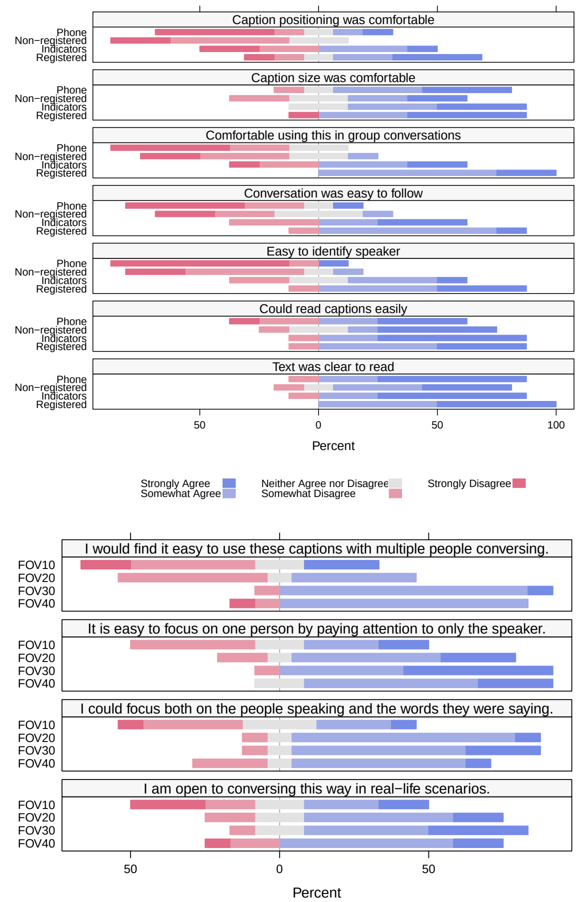

Figure 3: Subjective results for (a) caption and (b) FOV study

played the four *Four Angry Men* video sections in sequence, changing the order of conditions according to a balanced 4x4 Latin square design to avoid order effects.

Participants were asked a series of Likert-scale questions intended to evaluate the participants' sentiments towards the caption presentation method, as well as the overall experience. Before continuing on to the next section, participants were asked to complete a raw NASA-TLX form and to rank the conditions in order of preference (1 being best). Lastly, we conducted a brief semi-structured interview with the participant, to further understand their experiences and identify motivations that our written questions would not catch. This process was repeated for each captioning condition.

Note that the Registered condition is not realistic. Repeatedly, the DHH community has expressed that any captioning glasses must not call attention to themselves [20, 28]. Yet, performing registration often implies an onboard camera, whose aperature in a pair of eyeglasses may cause unwelcome attention for the wearer. Even without this constraint, smartglasses that look like normal eyeglasses have small FOVs, often around 10 degrees [20]. Would DHH users prefer the Registered condition if the display's FOV requires them to move their head back and forth in order to scan the group to find where the current captions are displayed? What FOV is sufficient for DHH users to feel comfortable using a Registered captioning system? To begin exploring this issue, we ran a pilot experiment with 12 hearing participants to determine what FOVs might be adequate (before attempting to recruit more DHH participants, who are harder to acquire). We repeated the Registered condition four times, counterbalanced with FOVs of 10, 20, 30, and 40 degrees, using the NASA-TLX and a Likert questionnaire.

3 RESULTS

There was a significant linear trend in the overall workload scores (Figure 2): F(1,7)=13.430, p = 0.008. A Friedman's test revealed that the ranked preferences varied significantly (χ^2(3, N=8) = 11.250, p = 0.010) across the four conditions: Registered (1.50), Indicators (2.26), Non-registered (2.63), and Phone (3.63). The series of Likert-scale questions, which was intended to evaluate participants' sentiment towards the conditions and towards the overall experience of participating in the group conversation, shows a familiar trend as responses tended to improve from control (Phone) to Non-registered to Indicators to Registered (Figure 3).

In the semi-structured interviews, of the three conditions intended to emulate a HWD, participants felt most positively about Registered Captions. This result aligns with prior literature, in which co-design sessions with DHH participants resulted in a captioning system that appeared similar. Participants mentioned the presence of captions fixed in space underneath a person associated the words with the person speaking (n = 3). **P6**: *When the captions started going under the person who was speaking, I knew okay, this person is speaking or about to speak.*

However, when participants could not see a caption because it appeared outside their field of view, identifying the current speaker required sequentially scanning the faces of the actors (n = 2), reinforcing the notion that participants trust registered captions as indicators of the current speaker and override the participants' instinctive method of identifying the speaker in a group conversation by looking for mouth movement or body language changes.

Participants expressed frustrations with the Non-registered condition. Half of the participants did not feel that the conversation was easy to follow or comfortable using the technology in a group conversation, and 66% felt that it was difficult to identify the speaker. In our semi-structured interviews, participants were frustrated with the lack of indication as to who was speaking ($n = 5$) at a given time, and the lack of indication when the speaker changed ($n = 4$). **P5**: *Compared to* [Registered Captions]*, it was harder to figure out who was actually speaking, because it only depended on the speaker: the only clue was that their lips were moving.*

In contrast to the Non-registered condition, participants expressed excitement about the Indicators condition. The Likert scale responses were much more positive: participants found it much easier to identify which person was speaking and follow the conversation. Of particular note is the significant shift in sentiment on the "I would feel comfortable using this in group conversations" question. In the Non-registered condition, only 13% of participants agreed with the statement, while >50% responded negatively to some degree. In contrast, simply adding binary direction indicators created a remarkable shift in sentiment, with 63% of participants stating they would feel comfortable to some degree using the technology in group conversations. Participants expressed that the appearance of an indicator reduced the number of people they had to scan to determine who was speaking ($n = 3$), which made the conversation easier to follow. **P6**: *I mean, the arrows help, right? It's not like I'm scanning all four actors to see who's speaking, and [I] know in which direction [I] need to look, but that's only crossed off two people.* While there was still some level of effort involved in identifying the current speaker, participants stated that the process was easier. **P1**: *It still takes a little to figure out who's speaking and it's harder to focus on the speaker and the captions. Other than that I thought it was great, the arrows make a huge difference.* **P6**: *I mean, because [the indicators] showed me which...direction I need to look at whoever speaking it's coming from, but then it didn't exactly show me who. And I still felt like I had to work a little bit.* That said, some participants found the mechanics of the indicators to be distracting from the conversation ($n = 1$). Some participants expressed a desire to see captions anchored closer to the speaker's face ($n = 2$), which would further minimize the visual dispersion between the speaker and their words.

Participants expressed the most frustration with the Phone condition, which was designed to emulate the state of current transcription applications like Google Live Transcribe [1] or Microsoft Translator. These kinds of technologies are limited to mobile devices, require the user to look down at the captions and back up at the speakers, and occupy the user's hands. Based on the Likert scale results, participants overwhelmingly found it difficult to identify the speaker and to follow the conversation. The positioning of the captions on the phone seemed especially frustrating to participants, which was corroborated in our semi-structured interviews.

Like the Non-registered condition, participants found it difficult to identify who was speaking at a given time ($n = 4$), as well as changes in speaker ($n = 5$). **P2**: *...there came a point where it was hard... to tell who was saying which line.*

[1]https://www.android.com/accessibility/live-transcribe/

Alarmingly, during this condition participants mentioned feeling like an outsider to the conversation ($n = 3$), comparing the experience to reading a script or a book. **P1**: *It's like I'm just reading a script...I also felt much more removed from the conversation because without positioning...I'm having to just look at [the phone] because there's no captions near [the speakers].* **P3**: *It felt like I had note pad open on my phone and the audio [muted].*

Figures 2b and 3b show the NASA-TLX and Likert results, respectively, for the FOV pilot. As expected for the FOV experiment, both workload and Likert scores improve between 10 to 30 degrees. However, above 30 degrees, increasing the FOV seems to have diminishing returns. Assuming a sigmoidal psychometric function between FOV and workload/perceived benefit, a good compromise FOV might be around 25 degrees. Future work will recruit DHH participants and test FOVs of 15, 20, 25, and 30 degrees to try to better establish this parameter. Font size might affect preferences. With our current fonts and captioning line widths, the 10 degree FOV could barely fit the entire caption area, causing the user difficulty in keeping a given speaker's captions on the screen. Smaller fonts might allow a smaller FOV but would be less visible. We hope to investigate this trade-off as another parameter in determining the minimum comfortable FOV for DHH participants for group captioning.

4 DISCUSSION

Based on our qualitative and quantitative results, registered captions on a simulated HWD significantly outperform captions displayed on a handheld Android smartphone, which is how many in the DHH community approach spoken group conversations. In fact, all HWD conditions tend to be rated better than the Phone condition. Participants prefer registered captions because they provide an authoritative indicator of the current speaker, minimize the visual dispersion between the speaker and their words, and make the participant feel more of a part of the conversation.

4.1 Limitations and Future Work

These results should be tested against simultaneous speaker situations, which are a common occurrence in group conversations. Future work should also endeavor to measure the cognitive load of both registered and non-registered captions in group conversations. Our field of view for Registered captions condition was 80° horizontal, which is wider than a HWD with a traditional glasses form factor could support. A direct comparison of the Indicators condition with a more reasonable 25 degree FOV Registered condition, perhaps also equipped with directional indicators, is warranted. In practice, both the Registered and Indicators conditions would have tracking errors, which should be emulated in future testing as they may significantly affect user preference. While there is significant work on typefaces for captioning movies, experiments are needed for verifying that such typefaces are also optimal for captioning on HWDs (i.e., font family, size, color, etc.).

5 CONCLUSION

In this paper, we emulated idealized HWDs using head-tracking and a large display to test different methods of captioning group conversations for people who are deaf or hard-of-hearing. Our findings indicate that DHH users find HWD captions a more pleasant experience than smartphone captions, perhaps as it minimizes the visual dispersion between speaker and caption. Finally, we highlight the potential of combining coarse indications of speaker direction with non-registered captions as an intermediate step for making spoken group conversations more accessible.

REFERENCES

[1] Matthias Böhmer, T. Scott Saponas, and Jaime Teevan. 2013. Smartphone use does not have to be rude: making phones a collaborative presence in meetings. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (MobileHCI '13)*. Association for Computing Machinery, New York, NY, USA, 342–351. https://doi.org/10.1145/2493190.2493237

[2] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (Jan. 2014), 125–139. https://doi.org/10.1109/TVCG.2013.86 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[3] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300276

[4] Gyeong-Nam Gimhae. 2013. Six human factors to acceptability of wearable computers. *International Journal of Multimedia and Ubiquitous Engineering* 8, 3 (2013), 103–114. Publisher: Citeseer.

[5] Mackenzie Glaholt. 2016. *Field of view requirements for night vision devices: A review of empirical research.* Technical Report.

[6] Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-Mounted Display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3373625.3418031 event-place: Virtual Event, Greece.

[7] Lillemor R-M. Hallberg and Sven G. Carlsson. 1993. A qualitative study of situations turning a hearing disability into a handicap. *Disability, Handicap & Society* 8, 1 (Jan. 1993), 71–86. https://doi.org/10.1080/02674649366780051 Publisher: Routledge _eprint: https://doi.org/10.1080/02674649366780051.

[8] Judith Harkins and Paula Tucker. 2007. An Internet Survey of Individuals With Hearing Loss Regarding Assistive Listening Devices. *Trends in Amplification* 11, 2 (June 2007), 91–100. https://doi.org/10.1177/1084713807301322 Publisher: SAGE Publications.

[9] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems (DIS '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 7–11. https://doi.org/10.1145/3197391.3205404

[10] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/2702123.2702393

[11] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People Who Are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362 event-place: Galway, Ireland.

[12] Dhruv Jain, Angela Lin, Rose Guttman, Marcus Amalachandran, Aileen Zeng, Leah Findlater, and Jon Froehlich. 2019. Exploring Sound Awareness in the Home for People Who Are Deaf or Hard of Hearing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300324 event-place: Glasgow, Scotland Uk.

[13] Adam Jaworski and Dafydd Stephens. 1998. Self-reports on silence as a face-saving strategy by people with hearing impairment. *International Journal of Applied Linguistics* 8, 1 (1998), 61–80. https://doi.org/10.1111/j.1473-4192.1998.tb00121.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1473-4192.1998.tb00121.x.

[14] Sophia E. Kramer, Theo S. Kapteyn, Joost M. Festen, and Hilde Tobi. 1995. Factors in Subjective Hearing Disability. *Audiology* 34, 6 (Jan. 1995), 311–320. https://doi.org/10.3109/00206099509071921 Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.3109/00206099509071921.

[15] S. Köbler and U. Rosenhall. 2002. Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification: Localización horizontal y discriminación del lenguaje con adaptación unilateral y bilateral de auxiliares auditivos. *International Journal of Audiology* 41, 7 (Jan. 2002), 395–400. https://doi.org/10.3109/14992020209090416 Publisher: Taylor & Francis _eprint: https://doi.org/10.3109/14992020209090416.

[16] Blair MacIntyre, Jay David Bolter, Jeannie Vaughan, Brendan Hannigan, Emmanuel Moreno, Markus Haas, and Maribeth Gandy. 2002. Three angry men: dramatizing point-of-view using augmented reality. In *ACM SIGGRAPH 2002 conference abstracts and applications (SIGGRAPH '02)*. Association for Computing Machinery, New York, NY, USA, 268. https://doi.org/10.1145/1242073.1242281

[17] Emma J. McDonnell, Ping Liu, Steven M. Goodman, Raja Kushalnagar, Jon E. Froehlich, and Leah Findlater. 2021. Social, Environmental, and Technical: Factors at Play in the Current Use and Future Design of Small-Group Captioning. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 434:1–434:25. https://doi.org/10.1145/3479578

[18] Ashley Miller, Joan Malasig, Brenda Castro, Vicki L. Hanson, Hugo Nicolau, and Alessandra Brandão. 2017. The Use of Smart Glasses for Lecture Comprehension by Deaf and Hard of Hearing Students. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1909–1915. https://doi.org/10.1145/3027063.3053117

[19] Shalini Misra, Lulu Cheng, Jamie Genevie, and Miao Yuan. 2016. The iPhone Effect: The Quality of In-Person Social Interactions in the Presence of Mobile Devices. *Environment and Behavior* 48, 2 (Feb. 2016), 275–298. https://doi.org/10.1177/0013916514539755 Publisher: SAGE Publications Inc.

[20] Alex Olwal, Kevin Balke, Dmitrii Votintcev, Thad Starner, Paula Conn, Bonnie Chinh, and Benoit Corda. 2020. Wearable Subtitles: Augmenting Spoken Communication with Lightweight Eyewear for All-day Captioning. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1108–1120. https://doi.org/10.1145/3379337.3415817

[21] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[22] Erin M. Picou, Elizabeth Aspell, and Todd A. Ricketts. 2014. Potential Benefits and Limitations of Three Types of Directional Processing in Hearing Aids. *Ear and Hearing* 35, 3 (June 2014), 339–352. https://doi.org/10.1097/AUD.0000000000000004

[23] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4884–4895. https://doi.org/10.1145/2858036.2858130

[24] Chris Schipper and Bo Brinkman. 2017. Caption Placement on an Augmented Reality Head Worn Device. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 365–366. https://doi.org/10.1145/3132525.3134786

[25] 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. https://doi.org/10.1145/3234695.3236355

[26] Matthew Seita, Sarah Andrew, and Matt Huenerfauth. 2021. Deaf and hard-of-hearing users' preferences for hearing speakers' behavior during technology-mediated in-person and remote conversations. In *Proceedings of the 18th International Web for All Conference (W4A '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3430263.3452430

[27] Matthew Seita and Matt Huenerfauth. 2020. Deaf Individuals' Views on Speaking Behaviors of Hearing Peers when Using an Automatic Captioning App. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383083

[28] Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 705–714. https://doi.org/10.1145/

1978942.1979044
[29] Tim Van den Bogaert, Thomas J. Klasen, Marc Moonen, Lieselot Van Deun, and Jan Wouters. 2006. Horizontal localization with bilateral hearing aids: Without is better than with. *The Journal of the Acoustical Society of America* 119, 1 (Jan. 2006), 515–526. https://doi.org/10.1121/1.2139653 Publisher: Acoustical Society of America.