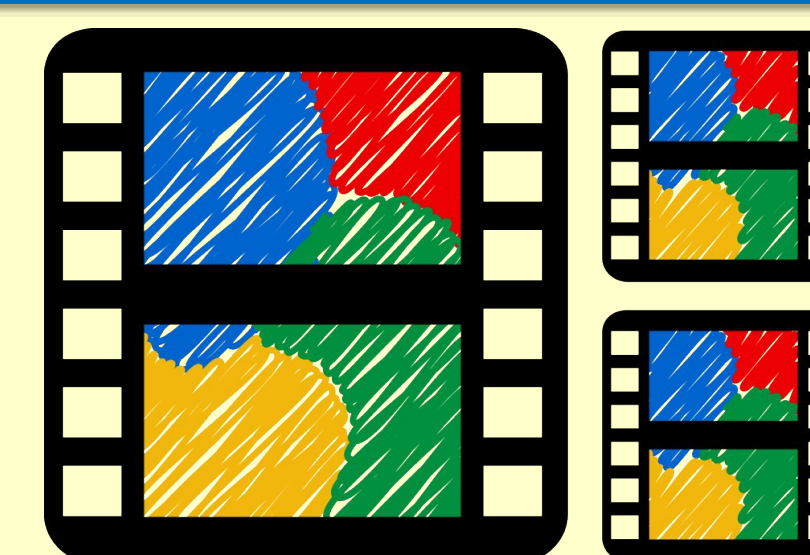


STT Studying the Impact of Captioning Delays on User Understanding

REU Accessible Information and Communication Technologies Site, Gallaudet University

Carrie Bergstedt, University of North Carolina at Asheville
Chloe Keilers, Stanford University

Mentors: Dr. Christian Vogler
Dr. Linda Kozma-Spytek



GALLAUDET
UNIVERSITY



Background

Internet Protocol Captioned Telephone Services (IP CTS) primarily relies on live human captioning which causes a delay between the audio of the conversation and the following captions on the screen. The delay between audio and captions causes problems - for example, the delay could disrupt the flow of a conversation or the delay could impact understanding where the users are listening to audio and following the delayed captions. The STT team focused on the second example.

Research Question

What is the cognitive impact of varying caption latency relative to audio for Deaf or Hard-of-Hearing (D/HH) IP CTS users?

Hypothesis

The smaller delay will have a more positive impact - i.e. better understanding of the audio and more comfortable and usable captions.

Literature Review

Existing literature on caption latencies is in short supply. What exists generally fell into two categories:

1. Studies that focused on the technology itself
2. Studies that focused on the people using the technology.

Additionally, there is one study from 1998 that had a similar methodology to what the STT team used but lacked participants (Burnham). Overall, literature review indicates that this study would be covering new territory.

Methodology

Interested potential participants filled out a basic demographics survey which determined if they were eligible for the study. Participants that met the requirements (had used a captioned telephone before, identified as Deaf/Hard of Hearing/had hearing loss) then came in for the in-person study.

There were 9 participants total. See demographics for more.

Procedure

1. Online demographic survey
2. Vision screening
3. Pilot audio clip for baseline (no captions)
4. Six experimental clips in random order
 - i. Test each delay twice with different audio.
 - ii. Comprehension summary to test the participants understanding of the clips
 - iii. Rate on a scale of 0-10: confidence of understanding, how comfortable were the delayed captions, how usable were the delayed captions, and how much they think they understood.
 - iv. What they think of the clip, how they felt, and how long they think the delay was.
5. Exit interview (which clips seemed best, worst, what are their thoughts on this experiment).

Results

We use a one-way repeated measures ANOVA to separately analyze ratings on four dependent variables and determine if there was a significant effect of the independent variable. The independent variable is caption delay of either 0, 2, or 4 seconds. The table below shows the results of the ANOVA. Red numbers indicate a significant effect. Black numbers indicate no significance.

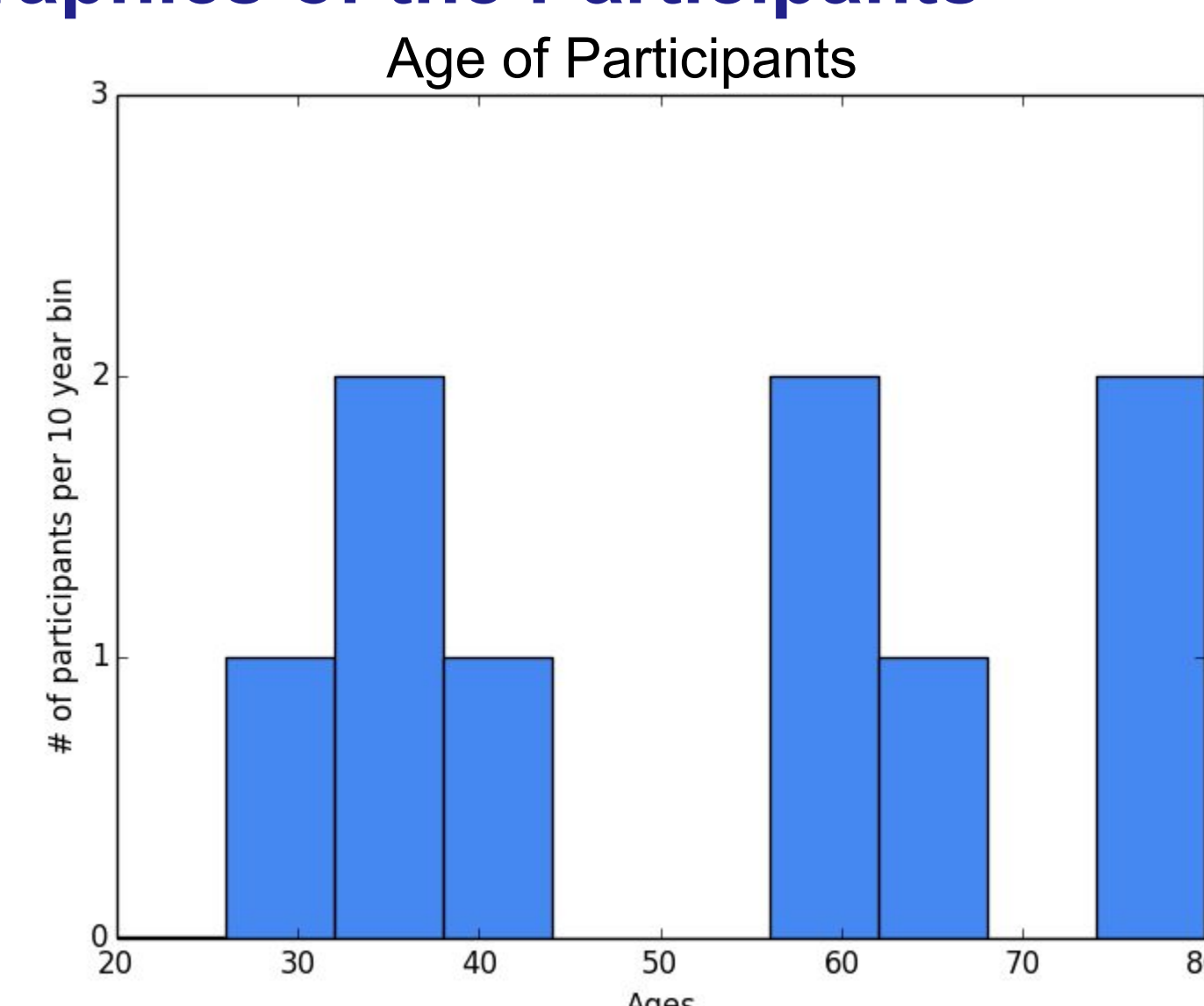
To the right is a graph showing the means of each of the ratings with standard error bars.

	Overall p-value	Pairwise comparisons		
		0 ⇒ 2 sec	0 ⇒ 4 sec	2 ⇒ 4 sec
Confidence	0.0093	0.032	0.005	0.575
Understanding	0.0224	0.118	0.006	0.266
Comfort of Captions	0.0025	0.111	0.009	0.022
Caption Usability	0.0266	0.071	0.012	0.525
How Much Understood				

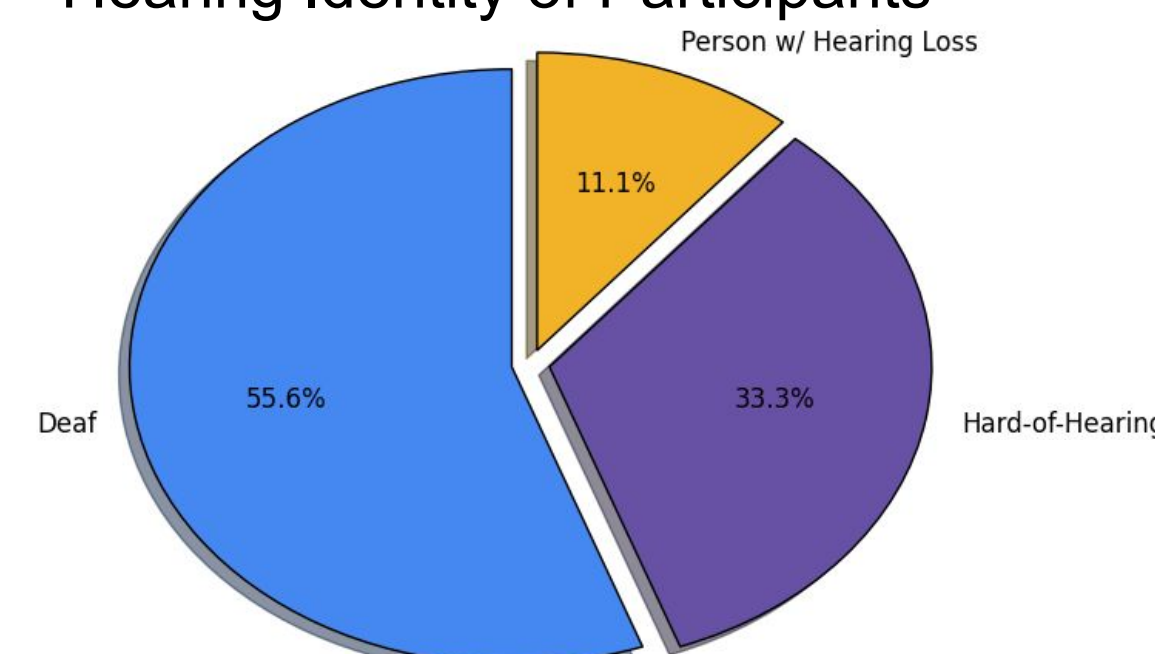
Qualitative Results

- Audio alone not comfortable
- 4-second delay incredibly frustrating
 - Captions help, but too delayed to help much
- Highly subjective whether or not slight delay helps
 - Most prefer captions to be in sync with audio
 - Some listen first, then refer back to the captions
- Issues with variation in voice pitch or intonation
 - Related to the audio quality of podcasts
- Most have issues pinning down precise delays in timing *but* usually judge ratios of delays fairly well

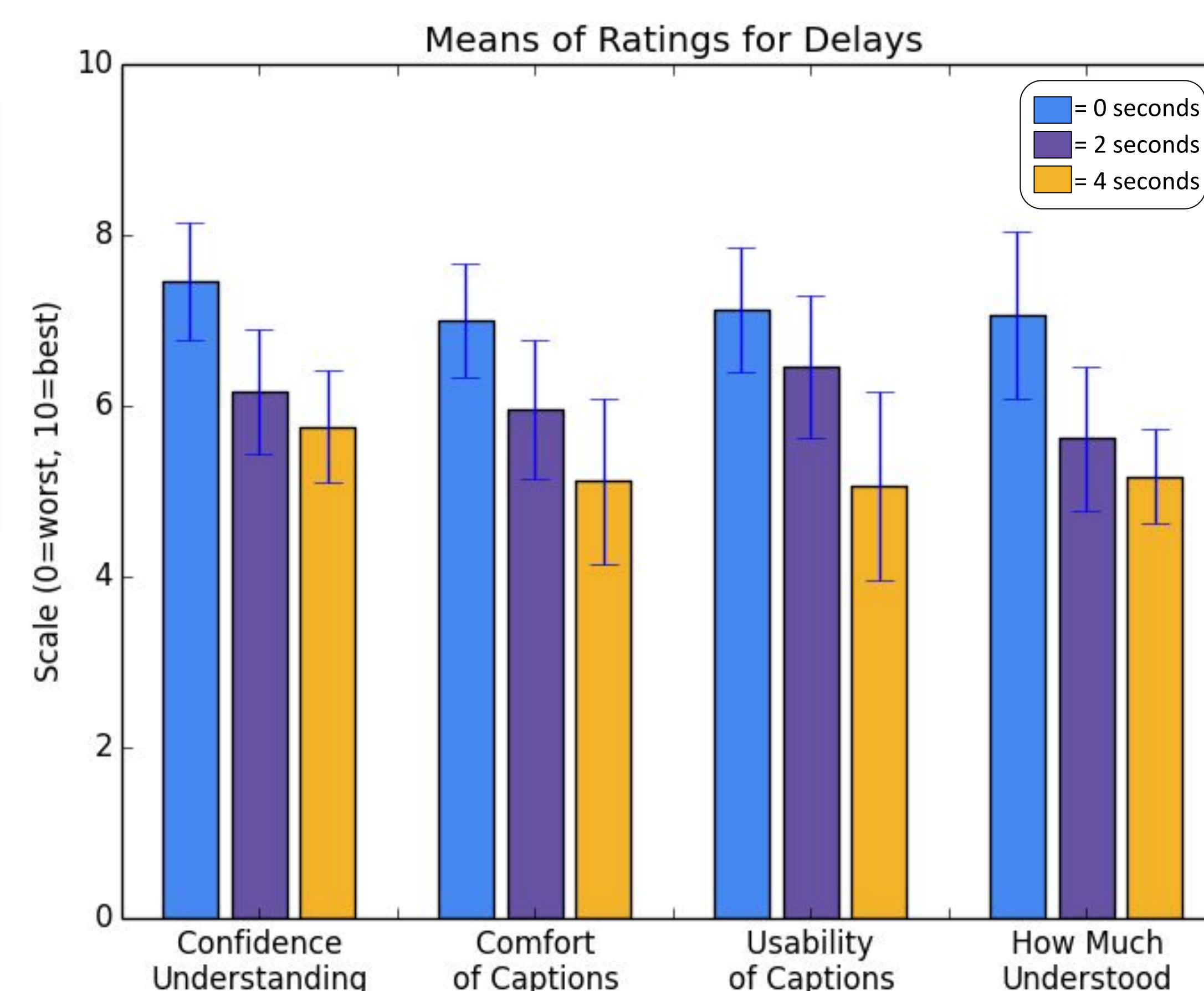
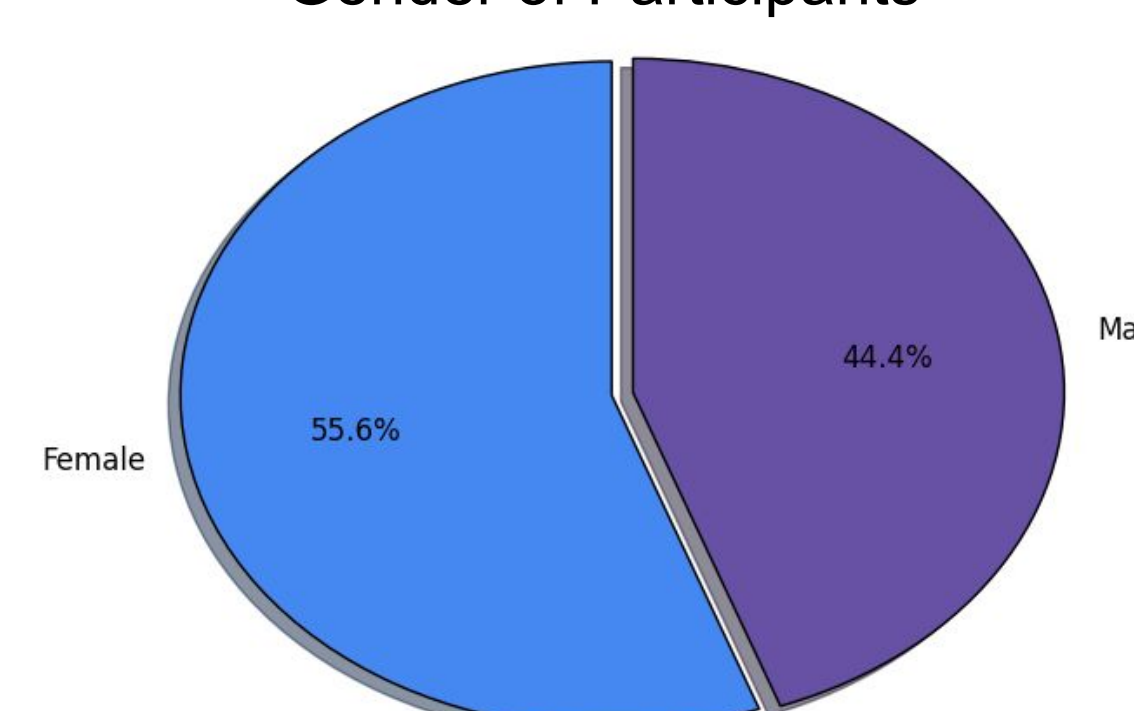
Demographics of the Participants



Hearing Identity of Participants



Gender of Participants



Conclusions

- Tested the effects of captioning delays on users
- Met a wide variety of enthusiastic people who participated in our study
- Participants generally preferred the smaller delays over the larger delays
- Results shows there is a significant difference between 0 and 4 second delays, but smaller differences between 0 and 2 second delays and 2 and 4 second delays.

Future Work

- Look for better audio sample - i.e. from audiobooks
- Generate captions more representative of a captioned telephone
- Could have asked participants whether they leaned on captions or audio more (for each clip)
- Test more delays with more data - i.e. +1 second or -1 second delay

References

1. Burnham, D., Rober-Ribes, J., & Ellison, R. (Dec 4, 1998). Why captions have to be on time. Auditory-Visual Speech Processing, AVSP '98. Sydney, Australia: ISCA.
2. E. Sohoglu, J. E. Peelle, R. P. Carlyon, and M. H. Davis. "Top-down influences of written text on perceived clarity of degraded speech," Journal of Experimental Psychology: Human Perception and Performance, vol. 40, no. 1, pp. 186-199, 2014.
3. Kwas, S., Karalis, G., Wen, T., & Ladner, R. E. (2016). Improving real-time captioning experiences for deaf and hard of hearing students. Paper presented at the ASSETS 2016 - Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, 15-23. doi:10.1145/2982142.2982164 Retrieved from www.scopus.com
4. Kozma-Spytek, L. (2013, November/December). IP Captioned Telephone Service What Consumers Told Us and What the FCC Ruled. Hearing Loss Magazine, 20-23.

Acknowledgements

Sincerest thank you to our mentors and Kirsten Fargas who guided us through this project.

The contents of this poster were developed under a grant from the National Science Foundation, grant #1757836 (REU AICT).

Contacts:

Carrie Bergstedt, carrie.bergstedt@gallaudet.edu
Chloe Keilers, chloe.keilers@gallaudet.edu