
Better Captions With Glass: Building Apps for Captions and Lyrics

Eric Yurko

Stanford University
PO Box 12657
Stanford, CA 94309
eyurko@cs.stanford.edu

Brie Bunge

Stanford University
353 Serra Mall
Stanford, California 94305
brie@cs.stanford.edu

Alex Hsu

Stanford University
353 Serra Mall
Stanford, California 94305
kzm@stanford.edu

Zheng Shen

Stanford University
353 Serra Mall
Stanford, California 94305
zhengs@stanford.edu

Abstract

This project looked at the interaction of movies and subtitles (and music and lyrics) and tried to develop a novel interface to present and augment that interaction on the new Google Glass platform. Though not much Glass-specific work has been made public, we found many documents that explained how best to present text on a small screen, informing our design. By presenting results of our user studies with Glass, we hope to demonstrate our belief that this idea has potential to make a very positive impact in the lives of moviegoers and music lovers everywhere.

Introduction

Imagine you're in a movie theater and are having a hard time understanding what's being said. In the United States, you may be part of 13% of the

population for which English is not their first language, or the 3% of the population that has some sort of hearing impairment. Or, less severely, you're listening to music with your friends and they're all singing along, but you don't know any of the words. Within these use cases is an underlying pattern, that there is text that can enhance your listening or speaking activities. This text may be known to you (though you don't know it explicitly), but it's not necessarily synced up with what is happening. To that end, we decided to build out an app on the Google Glass platform that would bridge that gap, allowing users to better understand the media that they're experiencing.

Previous and Related Work

Currently, no development work has been done in this area specifically related to the Google Glass platform, though some work has been done on how best to present different designs on Glass via user studies [1]. The findings influenced the present research in that we followed its suggestions to align text left (rather than centered), not to use muted colors, not to use colors unless for visual separation, and to be sensitive that the use of color is a divisive issue among users.

Current systems for movie theater captioning are rear-window captioning systems that project subtitles at the

back of the theater which are visible to the user via a reflective pair of glasses. This solution is less than ideal due to the expense of installation [2]. Other options include Sony's Entertainment Access Glasses which are see-through glasses available at select movie theaters [3] [4]. Google has proposed a similar idea via their Live Subtitles for Google Glass video [5], but it is little more than a concept video at this point. This work is encouraging because the need for such technology is clearly demonstrated. Our contribution to these offerings is that it leverages a technology that the user may already own and be familiar with, so the additional cost for the movie theater would be dramatically reduced to just running a subtitle broadcasting system.

Design

We spent a fair amount of time just wearing and experiencing the Glass platform, as our entire team was new to it. We wanted to see how it performed in various contexts and what modes of presentation were the most useful and the least obtrusive. To that end, we determined that microinteractions were the best form of interactions with the Glass, as this reduced the amount of time the user had to awkwardly focus on the Glass as well as stopped the Glass from heating up. Outside of that, we primarily focused on readability across both apps, and our specific design and implementation choices are outlined below.

Caption This

The movie system consists of a private server and Glass app, and uses Firebase (www.firebase.com) (an API for syncing data in realtime) to communicate between them. A webpage on the server can play movies, and when one is played, it broadcasts the time and the name of the movie to Firebase at an interval of

one second. When the Glass application starts up, it queries Firebase for the current movie information, fetches the subtitles from an endpoint on the server, and begins playing the subtitles from the current time in the movie. If the movie skips ahead or the subtitles get out of sync, the user can tap the Glass to resync. The user also has the option of swiping through the subtitles with their finger to navigate the movie's subtitles—this helps for refining the subtitle playback in the case that it is slightly off.

The source broadcasts what is playing and the time stamp in order to synchronize reliably. In an earlier iteration, the Glass app read the current time from Firebase every second, which caused the Glass to heat up rather quickly. We minimized the number of network requests by changing the model to one in which the Glass app runs its own internal clock and scheduler, and only resets this when a user taps.

Show Me Lyrics

The lyrics display, we reasoned, would be much like a karaoke app. To that end, we designed an interface inspired by the text highlighting found in most karaoke. Currently, the architecture is the same as Caption This, but it uses songs and .lrc files rather than movies and .srt files. Like the subtitles, we use .lrc's because it makes the syncing more reliable.

As a recent update, the user can find lyrics by listening to a song, like Shazam or Soundhound. The user plays a song from their phone and Glass will recognize it and show appropriate lyrics for that point in the song. Currently, we have it working for two test songs, and hope to have more songs working in the future.

As a second app, users can also say the lyrics to find the song, and the lyrics start from the match in the song. This is great for users who are listening to the song and want to continue to sing along from that point, as they can just pause their song, check with Glass, and then resume from there. In another case, the user may only know the chorus or part of a song ("Hey, what's that song that goes, 'blah blah blah'", for instance) and wants to be able to sing the entire song.

Discussion

From testing the captions app, we found that the overwhelming majority of our users (~93% of 15 users tested) agreed in some way that the Glass's displayed text was easy to read, validating our design choices. However, the major issue most had was that they had some difficulty switching focus between the Glass and the screen. In either case, only ~27% of the users agreed in any way that the app was comfortable to use, citing these concerns. Even then, perhaps the novelty of the app alone was convincing, as 73% of the users agreed in some way that the tool was fun, and 87% of the users agreed in some way that the tool was easy to use as well as cool.

Overall, it seems like the user response has been overwhelmingly positive. We believe, as said earlier, that that is in part due to the novelty of the platform. After testing with a student with hearing impairments, she was very excited about the potential applications of the app (albeit with some caveats that we then worked to correct), and testing with nonnative English speakers was interesting as well. [Zheng, could you briefly talk about that here?] We hope that in demonstrating the potentially socially-conscious uses of Glass, we can also start to remove some of the stigma that has plagued

the platform as of late, and our users seem inclined to agree. While they did not necessarily love the Glass, they found the idea of the app and the experience compelling, and we would claim that makes our app a success.

Future Work

Across both apps, we would have liked a more robust listening interaction (for movie and song recognition) Perhaps subsequent iterations (or a custom wearable) would have a more robust microphone attachment for this use case. We'd also have been interested in adding different tracks for trivia or commentary tracks. Also, some suggested that we branch into different areas, such as flash mob choreography instructions or even Dance Dance Revolution-inspired apps. These all seem like fun new directions to take this idea.

Conclusions

Overall, we think that our subtitles app and lyrics display app accomplished our goal, which was to provide a novel interface for users to utilize text to enhance their listening or speaking abilities. In developing that app, we referenced previous work in subtitle and text presentation to try and design an optimal experience, though we noted that not much work has been done with Glass. After building the systems, we were pleased to find that our user tests came back mostly positive, as many people saw the app for what it could be and looked past some of the technical issues they had with the Glass platform. While there is much work still to be done, we believe this is an excellent first step in this field, and we are excited to see what future developments take place in this space.

References

- [1] Dima, E. Determining UI design principles for Google Glass and other over-eye interactive device applications.
- [2] Meddaugh, J. Audio description in theaters: Making theaters more accessible.
- [3] Mukawa, H., Akutsu, K., Matsumura, I., Nakano, S., Yoshida, T., Kuwahara, M., Aiki, K., and Ogawa, M. 8.4: Distinguished paper: A full color eyewear display using holographic planar waveguides. In SID

Symposium Digest of Technical Papers (2008), vol. 39, Wiley Online Library, pp. 89–92.

- [4] Sony's entertainment access glasses can open your doors to a whole new audience.
<http://pro.sony.com/bbsc/ssr/mkt-digitalcinema/resource/latest.bbseccms-assets-mkt-digicinema-latest-EntertainmentAccessGlasses.shtml>.
- [5] Neto, A., and V. Amaro. Live Subtitles for Google Glass.
https://www.youtube.com/watch?v=4s_DbSp9PFw