ToozKit: System for Experimenting with Captions on a Head-worn Display

Peter Feng
pfeng32@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

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

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

Figure 1: Left: Captions displayed on Android application on mobile phone and on screen of head-worn display. Center: Person wearing tooz DevKit head-worn display. Right: Person conversing using captioning on tooz DevKit head-worn display.

ABSTRACT

The advent of Automatic Speech Recognition (ASR) has made real-time captioning for the Deaf and Hard-of-Hearing (DHH) community possible, and integration of ASR into Head-worn Displays (HWD) is gaining momentum. We propose a demonstration of an open source, Android-based, captioning toolkit intended to help researchers and early adopters more easily develop interfaces and test usability. Attendees will briefly learn about the the technical architecture, use-cases and features of the toolkit as well as have the opportunity to experience using the captioning glasses on the tooz HWD while engaging in conversation with the demonstrators.

CCS CONCEPTS

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

KEYWORDS

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

ACM Reference Format:
Peter Feng, David Martin, and Thad Starner. 2023. ToozKit: System for Experimenting with Captions on a Head-worn Display. In *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), October 08–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3594739.3610715

UbiComp/ISWC '23 Adjunct, October 08–12, 2023, Cancun, Quintana Roo, Mexico

ACM ISBN 979-8-4007-0200-6/23/10.
https://doi.org/10.1145/3594739.3610715

1 INTRODUCTION

Head-worn Display (HWD) captioning systems have caught the eye of researchers in recent years due to the increased capabilities of both HWDs and Automatic Speech Recognition (ASR). The advances seen in ASR with tools such as Google's Live Transcribe [4] have demonstrated its capabilities to provide real-time captioning for the Deaf and Hard-of-Hearing (DHH) in a variety of settings such as conversations, lectures, meetings, and other forms of communication. However, ASR integration into HWDs that are inconspicuous remain under-researched.

In this paper we present an open-source HWD-based captioning system that employs a smartphone, utilizing its microphone as well as its computation. This system design lowers the financial

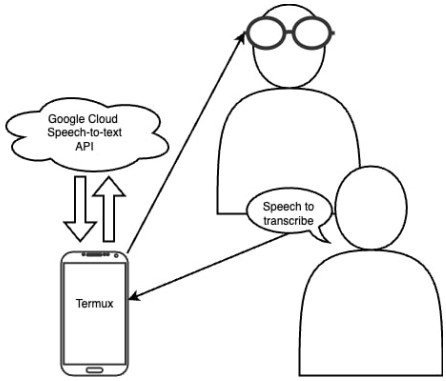

Figure 2: Technical Schematic of the ToozKit

Figure 3: Tooz DevKit Head-worn Display

barrier of entry, making it accessible to researchers, as well as DHH individuals.

Our system provides an user-friendly interface which allows for users to alter text size, text color, and text placement. This interface allows researchers to further investigate HWD-based captioning systems. Furthermore, it also allows for DHH earlier adopters to begin leveraging this system, improving their access to human-to-human communication.

2 RELATED WORKS

2.1 Social Acceptability

With the development of Assistive Technologies (AT), researchers' consideration for the social acceptability of such ATs is crucial [3, 12, 14]. Some important factors to consider when designing wearable AT is its size and weight, noticability due to fear of discrimination, and its utility to ensure that users both benefit and adopt the wearable AT created [3, 14].

2.2 Head-worn Display Captioning

As ASR's capabilities have advanced, HWD captioning has been an area of interest for many researchers [1, 9, 11, 13, 15]. Researchers have investigated a variety of uses cases ranging from lectures [10], group conversations [1], conversations while walking [9], and all-day use [11]. Furthermore, there has been investigation into HWD specific topics such as the placement of the text on the display [15], field-of-view [1, 13], and presentation of the captions [1, 8].

While investigations into HWD captioning have revealed valuable insights, many use off-the-shelf HWDs that have bulky form factors, high headset weights, and are prohibitively expensive such as the Microsoft Hololens [9, 13]. In addition, with the advances in HWD technologies and ASR utility, bringing captioning out of the laboratory and into the field in the hands of its user can provide researchers and HWD manufacturers with feedback that is much needed to create an off-the-shelf HWD that is socially acceptable, inconspicuous, and ready for every-day by DHH individuals to enhance their access to communication.

3 TOOZKIT

3.1 Technical Implementation

The ToozKit captioning system[1] integrates three separate components: the Google Cloud speech-to-text service [7], a custom Android application, and a physical HWD.

The Speech-to-Text service provides real time text transcription using a machine learning model trained and hosted by Google on the cloud.

The Android application is an extension of the Termux software, which is an open source Android terminal emulator and Linux environment app [16]. It also makes use of the Live Transcribe Speech Engine library [5] for communicating with Google Cloud's Speech-to-Text service. It serves as the main control point of the system, communicating with both the Cloud service and physical HWD as seen in Figure 2.

3.2 Physical System

For the purposes of this demo the HWD chosen is the tooz DevKit smart glasses [17] by tooz technologies as seen in Figure 3. The DevKit is a pair of smart glasses running its own proprietary OS, which can be communicated to and from using Bluetooth to display images. It maintains an array of sensors including a gyroscope, accelerometer, brightness sensor, and more. It has a total weight of 67 grams and a 20 degree monocular (right eye) field of view, with a screen that is 400px x 640 px. One of the major appealing factors of the DevKit is that it is relatively unobtrusive and subtle, making it a more feasible display for captioning in real-world scenarios.

The Android application cannot directly run on the DevKit. Instead, it maintains a Bluetooth connection with the DevKit, mirroring the contents of the terminal emulator on its screen to the display on the HWD. A major limitation of the DevKit is that the display can only process about one full frame per second. This problem is remedied in the Android application using a series of workarounds including sending snapshots of only characters that have changed,

[1]Source code can be found at https://github.com/Peter-Feng-32/Terminal-For-Smart-Glasses

rate-limiting the numbers of frames per second sent to the display, and formatting the captions to update the minimum number of characters possible on the terminal emulator. Furthermore, the Android application is able to run captioning as a background service, allowing the mobile phone running the Android application to be put away while not interrupting the captioning. This feature provides a better hands-free experience for the user.

3.3 Toolkit Features

Upon opening the app, users are presented with a terminal emulator display. Users are able to access a control menu by swiping inward from the right of the screen as in Figure 4. The control menu allows the user to input their API key used to connect to their Google Cloud account, providing access to the Google Cloud captioning service. In the menu are fields for customizing font size and font color for the captions which can be seen in Figure 5. The X and Y fields can be used to position the captions on the screen, and the width and height fields are used to control the maximum size of the captions. Once the user has adjusted the settings as desired, they are able to enable and pause captioning sessions from this menu. Finally, there are options for displaying phone notifications on the

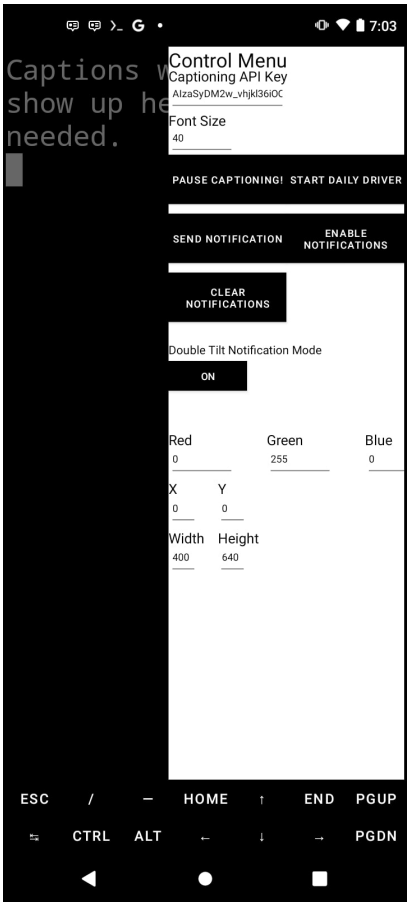

Figure 4: Control menu on Android application for captions

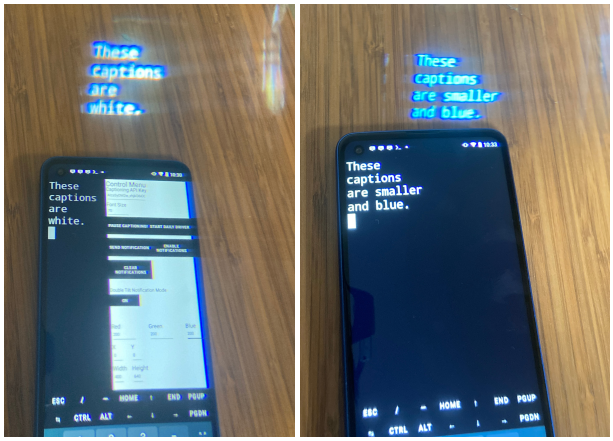

Figure 5: Comparison of different font size and color captions looking into the display of the DevKit

HWD, which is a feature currently only compatible with the tooz DevKit, but is under development for other HWDs.

Starting a captioning session first edits the terminal emulator to match the desired settings by calculating the number of rows and columns needed to fit the desired screen and text size. It then opens a connection with Google Cloud for the Speech-to-Text service. Audio data is continuously recorded and transmitted from the Android application to the cloud service, and a callback function runs whenever the cloud service provides an updated transcript of captions. In the callback, the transcript is formatted so that no word appears in two separate rows on the terminal emulator. This formatting is performed by inserting spaces in between each word so that words which occupy the end of one row and the beginning of another are "pushed" to the next row. Finally, the terminal emulator screen is cleared and the new formatted transcript is written to the screen, lasting until changes occur or it times out.

As it is controlled and interfaced by an Android application, this captioning system is designed to work out-of-the-box with any HWD that runs an Android or Android compatible operating system, such as the Google Glass Enterprise Edition 2 [6] or the EPSON Moverio BT-300 [2], by directly installing and running the captioning app on the HWD. In these cases, the Bluetooth communication aspect of the system would be unused, as it merely mirrors the captions from the app's interface to the DevKit, and the app's interface would already be displayed on the Android based HWD. This compatibility with various HWDs allows the system to be versatile and fit the price points and needs of different users, whether it be for research or personal use.

4 DEMO OVERVIEW

Our demo will provide attendees an opportunity to interact with, understand, customize, and use the captioning system for conversation. It will also provide a unique opportunity to interact with some of the extended features implemented for the tooz DevKit specifically. The demo will begin with a short introduction to the system and the motivation for developing a toolkit, which will last under five minutes. This introduction will cover the major pieces

of the architecture, including the Android application design and the HWD hardware. Attendees will then be handed a mobile phone running the Android application, which they can use to customize their captioning settings. Then they will be able to wear the tooz DevKit glasses, start the captioning session, and view the current conversation as captions on the DevKit. At this point, they will be able to ask any questions they might have about the system as well as make general conversation to test the system. They will also be able to change the captioning settings as they desire in order to experiment with how different settings affect their captioning experience. Finally, the demo will conclude with an opportunity to test the DevKit specific notification functionality with motion controls, using sample text messages sent to the phone.

5 FUTURE WORK

In this paper we present ToozKit, a HWD captioning system, which has various customizable features. We plan to implement more customization and accessibility features going forward. Some features include customizable font families for the captions, allowing users to use different colored opaque backgrounds, and using a light sensor to perform automatic brightness adjustment and reduce eye strain. We will also explore ways to receive phone notifications on the HWD, mitigating potential redundancy in needing to use a mobile device for notifications while already wearing a HWD. Currently HWD notifications are only available in conjunction with the tooz DevKit. The functionality uses the DevKit's accelerometer to allow the wearer's head nods to display and dismiss the notifications. We would like to investigate different ways of receiving notifications and expand this feature to other HWDs. Finally, there is work to be done in terms of collecting user feedback to improve the captioning system. We intend to use and test this system both in carefully controlled experiment settings and rigorous, real-world use cases, allowing us to both study various aspects of HWD captioning and refine the system further.

REFERENCES

[1] Gabriel Wade Britain, David Martin, Tyler Kwok, Adam Sumilong, and Thad Starner. 2022. Preferences for Captioning on Emulated Head Worn Displays While in Group Conversation. In *Proceedings of the 2022 ACM International Symposium on Wearable Computers* (Cambridge, United Kingdom) *(ISWC '22)*. Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/3544794.3558468

[2] EPSON. 2023. Moverio Basic Function SDK | Developer's Guide for Android & Windows. https://tech.moverio.epson.com/en/basic_function_sdk/developers_guide/introduction.html.

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

[4] Google. 2019. Live Transcribe | Speech to Text App | Android. https://www.android.com/accessibility/live-transcribe/.

[5] Google. 2019. Live Transcribe Speech Engine. https://github.com/google/live-transcribe-speech-engine/.

[6] Google. 2023. Get Started | Glass Enterprise Edition 2. https://developers.google.com/glass-enterprise/guides/get-started.

[7] Google. 2023. Google Cloud Speech to Text. https://cloud.google.com/speech-to-text/.

[8] 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/2702123.2702393

[9] 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* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[10] 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* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 1909–1915. https://doi.org/10.1145/3027063.3053117

[11] 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* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 1108–1120. https://doi.org/10.1145/3379337.3415817

[12] Phil Parette and Marcia Scherer. 2004. Assistive technology use and stigma. *Education and training in developmental disabilities* 39, 3 (2004), 217–226.

[13] 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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[14] 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* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4884–4895. https://doi.org/10.1145/2858036.2858130

[15] 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* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 365–366. https://doi.org/10.1145/3132525.3134786

[16] termux. 2023. Termux. https://termux.dev/.

[17] Tooz. 2023. Tooz DevKit. https://tooz.com/devkit/.