Loneliness in Blind and Low-Vision Individuals and the Use of AI Voice Generation to Facilitate Social Connection

Jordyn Anthony, University of Virginia

***Abstract*—The portion of the population experiencing moderate to severe vision loss has been increasing since 1990, and this trend is expected to continue at least through 2050. Yet daily activities seem to revolve increasingly around sight as digitalization finds its place in every facet of modern life. This shift compounds with the long-studied correlation between vision loss and negative emotional health, notably social isolation and loneliness. I designed a messaging application for iPhone devices that uses artificial intelligence voice cloning technology to read aloud messages in the voices of their senders. This innovative design improves upon the customizability of current screen readers and seeks to increase humanity in digital interactions for blind and low-vision users. In the future, a usability study should use this design to test the effectiveness of cloned, familiar voices in a digital messaging application on loneliness in the blind and low-vision community.**

I. Introduction

The International Agency for the Prevention of Blindness (IAPB) reported that 1.1 billion people, or 14% of the global population, were living with vision loss in 2020. Among those, almost 340 million, or 4.3% of the global population, had moderate to severe vision loss or blindness [1]. In 1990, the IAPB reports that roughly 580 million people, or 10.9% of the global population, were affected by vision loss, with over 182 million, or 3.44% of the global population, experiencing moderate to severe vision loss or blindness [1]. Across both metrics, there is a clear increase in vision loss cases proportional to population [1]. Unfortunately, this trend is not expected to cease; the IAPB expects vision loss to rise to over 18% in 2050, a 55% increase from the 2020 figures [2]. Despite the enormity of these statistics, societies remain ill-equipped for visual impairments. Blind or visually impaired (BVI) individuals continue to have inequitable experiences with technology, despite its enormous presence in modern daily life. This paper will use the term BVI to strictly refer to individuals who are experiencing moderate to total vision loss, in which individuals' lives and routines are significantly affected by their visual impairment. I first examine the negative emotional effects that BVI individuals face as a result of their impairment (Motivation). Then, I investigate relevant common vision accessibility aids (Related Work). I continue with a proposal for a mobile application, Talk To Me, that aims to facilitate social connection by BVI individuals with their friends and family members (Design). In this Design section, I also describe and present images of a horizontal, high-fidelity prototype of this mobile application. Finally, I conclude with the significance of my work and propose a future usability study of my tool with BVI participants.

II. Motivation

There is a large body of literature studying the emotional effects of visual impairments on individuals. Among the most commonly reported conditions are anxiety and depression [3]. These conditions often stem from concern of continued vision loss, interruptions to daily routines, and exclusion from social and career pursuits [3]. According to the Centers for Disease Control and Prevention (CDC), social isolation and loneliness can have serious emotional and physical effects: "heart disease and stroke, Type 2 diabetes, depression and anxiety, addiction, suicidality and self-harm, dementia, earlier death" [4]. These conditions underscore the danger of allowing BVI individuals to continue to be excluded from society. Especially as the BVI population grows, taking active steps to reduce the loneliness felt by this community is critical. The National Center for Chronic Disease Prevention and Health Promotion, a branch of the CDC, defines social isolation as "the lack of relationships with others and little to no social support or contact" and loneliness as "feeling alone or disconnected from others" [3]. These definitions highlight the low barriers to which any person can start to feel the harmful effects listed above, but the situation is yet more precarious for BVI individuals. In 2022, researchers proposed that BVI individuals may have difficulty "effectively engag[ing] in social interactions, due to a loss of visual information" [3, pp. 3]. It is possible that negative social experiences may cause BVI individuals to become hesitant or averse to future social interactions.

Loneliness among BVI individuals is a studied phenomenon in the research community. In 2019, a team of researchers interviewed 736 BVI Norwegian adults and recorded their rates of loneliness and general life satisfaction [5]. They compared these rates to those of a sample of almost 15,000 non-BVI Norwegian adults [5]. The team found that 28.7% of the interviewed participants were experiencing moderate to severe loneliness, and the researchers correlated loneliness with lower life satisfaction [5]. Though these findings cannot be generalized for a global population, they nonetheless underscore the necessity of improving the social inclusion of BVI individuals. Brunes' team's study further draws a relationship between loneliness and "younger age, blindness, having other impairments, unemployment, and a history of bullying or abuse" [5, pp. 1]. It is this first association, of loneliness and younger age, that I believe yields the most actionable area for solutions.

The social isolation felt by BVI individuals is not exclusive to adults. A 20-year study followed 161 Dutch adolescents with visual impairments as they grew into adulthood [6]. The researchers found that, while parental support did not correlate with loneliness in adulthood, peer support was critical to decreasing future loneliness [6]. The study reported that teenagers without much peer support at the onset of their impairment experienced some of the most loneliness in adulthood [6]. Thus, not only does support in adulthood matter, but the systems of support for a BVI individual's entire life also matter. I decided to tackle this issue of loneliness at this point of a user's lifetime: teenage to early adulthood. Thus, I needed to investigate the technologies that would meet the accessibility needs of a BVI population while also improving the social connection for this particular age group.

Advances in technology have made instant, virtual social connection ubiquitous. Among individuals born after 2000, there is even a societal expectation to be perpetually digitally connected [7]. While intense digital engagement certainly has its downsides, younger generations rely on it to communicate. Accordingly, a tool that facilitates social connection among BVI teenagers or young adults needed to be digital. I looked to the place where many social interactions for this age demographic occur: instant text messaging. Texting has become a standard part of most daily lives, with over 2 trillion SMS text messages sent in the United States in 2021 [8]. This figure excludes the extremely popular instant messaging services, such as iMessage, Google Messages, and social media direct messaging. Thus, my goal became to create an innovative technology that meets the accessibility needs of the BVI community while utilizing text messaging to facilitate greater social interaction among a younger demographic.

III. Related Work

A. Background

There are two categories of what are colloquially referred to as "screen readers" [9]. The first are true screen readers, which read aloud content on the screen and change the host machine's default shortcuts and input gestures. These systems give users high control and high degrees of freedom over their devices. The second category consists of text-to-speech (TTS) systems. TTS software is much more lightweight, and it can only read aloud certain, recognized text. These systems do not change interaction patterns on the host device, and as such, they generally allow for less user control. Both screen readers and TTS systems allow users to adjust the rate of speech. Henceforth, I shall make the distinction between the two categories by referring to the first as screen readers and the second as TTS systems.

TTS systems have become ubiquitous in the past decade. Usually, highlighting text on a device prompts a menu that includes a "Speak text" function. However, for individuals who depend on a majority of their screen to be read aloud, simple TTS systems often slow down their workflow. The specialized gestures of screen readers allow them to reclaim reading efficiency, though there is an initial learning period while users adjust to the new interaction patterns. Hence, the following discussion will center on screen reader systems as they are best suited to the situations of BVI users.

Screen reader software differ in their degree of user control and host platform. The most popular screen reader is JAWS, a paid software available on Windows systems [10]. NonVisual Desktop Access (NVDA) is also very popular and designed for Windows systems, but it is free [10]. For MacOS users, the most widely-used option is VoiceOver, Apple's native screen reader [10]. VoiceOver also exists on Apple's mobile devices, iPhones and iPads. For mobile devices running the Android operating system, Google offers a native screen reader system called TalkBack. In general, it appears that BVI users prefer a combination of a Windows computer and an iPhone [11], [12], [9]. Though social pressures about owning certain devices surely also come into play, this preference likely stems from the superior functionality of JAWS for Windows devices and VoiceOver for iPhone. Unlike with a computer, users are not able to change the screen reader on their mobile device, which forces BVI individuals into certain screen reader systems. If the user is experiencing difficulties with the native screen reader, they have no choice but to adapt to the system or buy a new phone running a different operating system. However, switching between Apple and Android ecosystems can be difficult due to data compatibility issues.

In recent years, the available voices for VoiceOver and TalkBack have received a quality boost, sounding more human and less robotic. However, users cannot customize the voices for these native screen readers beyond choosing from the pre-provided lists, so there is little flexibility of choice. In 2022, a BVI user reported that Google released new high-quality voices, but the update arrived only on Google Pixel phones and not all phones running Android [13]. This user also recounted that these higher fidelity voices could not be used at high rates of speech [13]. This limitation excludes power users of screen readers from enjoying the improved voices, but some such users claim that they prefer the robotic voice for its clarity.

B. Implications for Design

Despite using screen readers, many BVI individuals continue to struggle with understanding on-screen content. This can be attributed, in part, to common content organization schemes that do not consider non-visual access [14]. In 2023, the Nielsen Norman Group (NN/g) conducted usability research into how BVI individuals use screen readers on their mobile devices [14]. Screen reader users discover content in two ways: dragging a finger over on-screen elements and listening to the text or alt text (textual descriptions of non-text elements); and automatically having the entire page read aloud. These interaction patterns make it difficult for screen reader users to navigate pages that contain two-dimensional or non-sequential layouts [14]. In other words, designs that are appealing to someone with sight can be very difficult to navigate for someone with little to no sight [14]. A 2022 study investigated the swiping patterns with which screen reader users explored a mobile phone screen [15]. The researchers found that participants tended to start in the upper left corner of the screen and most often proceed with a zigzagging motion [15]. The design implication from this finding is that content should be laid out along the vertical axis with only similar content being grouped within the same row. Joh's team also noted that the BVI participants tended to start their finger nearer to the target after becoming familiar with the target locations [15]. This highlights the UI/UX guideline of maintaining standards across applications and emphasizes the necessity of doing so especially when designing for accessibility groups.

The NN/g researchers also found that certain features were difficult for BVI participants to understand because their on-screen placement provided visual context [14]. Though a research team in 2012 investigated whether spatializing screen reader output improved content understandability, they could not establish a significant difference between spatialized audio and mono channel audio [16]. These findings emphasize the need for sequential information hierarchies to accommodate the behavior of screen readers.

IV. Design: Talk To Me

I propose a messaging application, Talk To Me, for Apple smartphones that reads text messages in the voices of the messages' senders. The creation of these personalized voices would be achieved using artificial intelligence (AI) voice cloning technology accessible through a web API. This design differs from existing technology in that it reads different content in different voices, rather than all content in the same voice. My hypothesis is that hearing messages in simulated familiar voices will increase the perceived human connection felt by teenage and young adult BVI individuals, especially those who are losing or have recently lost their vision. Though screen reader power users may prefer the efficiency and clarity of more robotic-sounding voices, BVI individuals who are transitioning to a new vision state may benefit from the aural tangibility of their loved ones' messages. This assumption is predicated on the finding that more face-to-face interactions decreased loneliness as compared to more phone calling [17]; simulated familiar voices will add more humanity to text message conversations, thus hopefully decreasing a BVI individual's loneliness. I created a high-fidelity, horizontal prototype of the Talk To Me application, which I will reference below.

A. Implementation

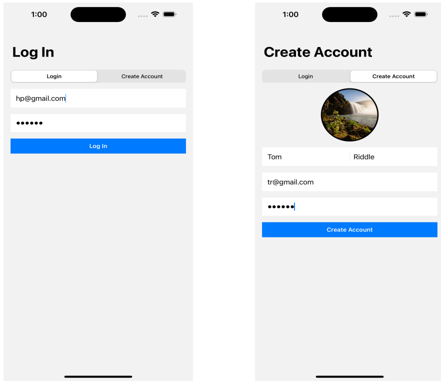
As discussed in Motivation, a text messaging application best fits the methods with which a teenage and young adult demographic communicate. And following the discussion in Implications for Design, I designed Talk To Me for Apple phones to accommodate BVI individuals' device preference. The application was designed for the iPhone 15 Pro running iOS 17.0. Even though the iPhone 15 Pro is the high-end model of the newest iPhone release, the screen size is similar to the previous recent generations. Also, coding in Apple's SwiftUI framework allows for high compatibility between phone models. iOS 17 is the newest iOS generation as of May 2024, and its September 2023 release date ensures that many iPhones are already running it. I decided to develop for iOS 17 to delay function deprecation.

I use a combination of the Swift programming language and the SwiftUI framework to develop Talk To Me. Since most of the project's programs are client-facing, they are implemented as SwiftUI files. SwiftUI allows flexibility in creating and modifying adaptive containers and UI elements while retaining the functional capabilities of Swift. When designing the application's views, I implement View structs to comply with framework requirements and contain the aesthetic elements. Each View struct instantiates an ObservedObject View Model class. The View Model classes are implemented as ObservableObjects, a conformance that allows the Views to use instances of the customized View Model classes. The View Models wrap the core functionalities of the application, such as retrieving or storing information in Google Firebase, that are unable to be implemented in the Views due to nonconformity. Some structures with specific properties, such as messages and user objects, are shared across view files, so these structs are implemented in distinct Swift files.

I use Google Firebase to handle back-end cloud computations such as user authentication functionality, user information storage, and message storage. The online Firebase Console allows the developer high control over enabled cloud services and application information, which facilitated testing of my prototype's features. When a user first creates an account, their email and password are registered with Firebase Authentication, and the authentication service automatically creates a unique user identification number (UID). Concurrently, their name, email, and profile picture are saved to the Firestore Database. The profile picture is uploaded to Firebase Storage and given a unique access URL. The Firestore Database accesses the picture via URL rather than storing the image file to maximize storage efficiency. The log-in and account creation prototype pages are shown below (Fig. 1a & 1b). Once the user logs in, the home page (Fig. 2a) displays all of their conversations by pulling message and user information from Firestore Database. Similarly, when the user presses the New Message button in the upper right corner of home page, the list of Talk To Me users (Fig. 2b) is pulled from a master user folder in Firestore Database. When viewing a conversation (Fig. 3a), the application locates the messages in Firestore Database using the UID of the logged-in user, the UID of the other user, the message content, and the message timestamp.

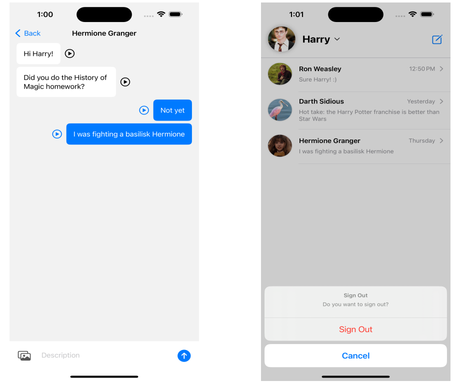
B. Application Features

Talk To Me was designed to closely mimic the look and functionality of iMessage, Apple's native messaging app, and Instagram, one of the most downloaded social media networks. This was a deliberate choice to reduce the requisite learnability time of Talk To Me, following the discussion of Joh et al.'s study in the previous section. Talk To Me was built with as many standard Apple library functions and symbols as possible to maximize compatibility with Apple VoiceOver and maintain standards seen elsewhere on the iPhone.



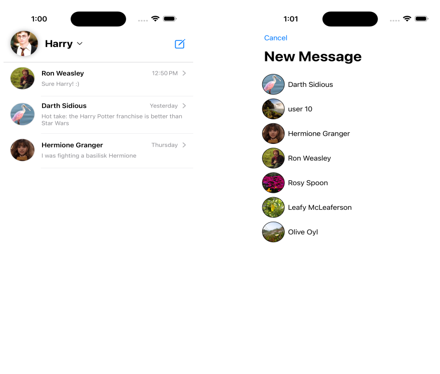
(a) Log In (b) Create Account

Fig. 1: Log In and Create Account pages



(a) Conversation (b) Sign Out

Fig. 3: Conversation and Sign Out pages



(a) Home (b) New Message

Fig. 2: Home and New Message pages

The conversation log on the home page (Fig. 2b) imitates that of iMessage. The horizontal design allows only one contact's information per row, separated by a standard Divider symbol. This accommodates the sequential discovery nature of screen reader software. Clicking anywhere within the contact's row will open the conversation with that contact, decreasing error probability by increasing target area according to Fitts' Law [18]. Joh's team found that BVI users benefited from larger targets in proportion to screen size, emphasizing the necessity of large targets in vision accessibility designs [15].

The New Message button in the upper right corner of the home screen has the same icon and placement as in iMessage. The icon is referenced by the SF Symbol name "square.and.pencil". SF Symbols is Apple's standard icon library, so VoiceOver will easily be able to recognize and describe this symbol to users. Figure 2b shows the overlay that appears when this button is clicked.

On the home page, the current user is displayed as the header with a large picture and large text. Next to the user's bolded name is a downward facing arrow, an icon which is also pulled from SF Symbols. This line of text and arrow is designed to mimic the header on the profile page on Instagram, though the functionality differs from that of Instagram. In Talk To Me, when the user clicks on the name or the arrow, a standard SwiftUI structure, Action Sheet, is opened. Action Sheet temporarily disables the rest of the screen, so the screen reader user is brought directly to the Action Sheet menu's buttons (Fig. 3b). The name and arrow combine to form one large target for the Action Sheet to reduce the likelihood of error.

C. Human Connection Features

Users must upload a sample audio file of their speech to their profile for the voice cloning software to replicate. This file is stored in Firebase Storage and linked to the user's Firestore Database entry. Depending on the API used, the cloned voice's metadata may also be stored, or the application would run a new API call on the voice sample each time a message needs to be read in the cloned voice. The BVI user would also need to upload their voice sample to listen to fully-cloned-audio conversations. This feature contributes the most to increasing human connection, following the discussions and assumptions in the previous sections.

On the home page, unread messages are differentiated by a blue circle and darker, bold text, which maintains the convention of displaying unread messages seen on most messaging services. When the user taps on a conversation with an unread message, the application enters that conversation and simultaneously reads aloud the new message in the correct cloned voice. Users can also select any message in the conversation thread to have it read aloud, a function communicated by the play buttons located next to every message. To increase the target area, clicking either the button or the message will read the message content aloud. There would also be some gesture pattern or additional button that would allow users to start listening to the conversation at any given point in the thread's history, and there would be a corresponding gesture to stop the automatic reading.

D. Limitations

Due to time and financial constraints, the app prototype did not incorporate the voice cloning function. Also, users cannot send images or emojis, nor can they use dictation to enter text in any text fields. The prototype was created as a proof of concept, so demonstrating the potential for various accessibility features was more important than full implementation of such features.

V. Conclusion

To reduce loneliness of teenage and young adult BVI individuals, I proposed a design for an iOS texting application that reads aloud messages in the voices of their senders by employing AI voice cloning technology. This innovative design adapts an emerging technology, AI voice generation, to a unique use case of improving human connection among BVI individuals. In recent years, AI voice generation has been marketed as a method for podcasters or content creators to automate advertisement reads, but my design gives this new technology a more meaningful application. Furthermore, I propose a future usability study to investigate the effectiveness of Talk To Me on the emotional states of BVI individuals. The BVI participants would use the application for an extended period of time as their primary messaging service, and the study would evaluate the application on various metrics, such as perceived social connection and satisfaction.

Acknowledgment

I would like to thank Dr. Panagiotis Apostolellis for supporting my interest in UI/UX during my time at the Univeristy of Virginia, as well as for guiding my research this semester. I would also like to thank Sarah Humphreys, Assistant Director of University of Virginia's Student Disability Access Center, for taking the time to share her observations and experience with me early on in my project. Finally, I would like to thank Brian Voong from the Lets Build That App channel on YouTube for producing a thorough SwiftUI tutorial that proved an incredible help in developing the Talk To Me prototype.

References

[1] R. Bourne *et al.*, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study," *The Lancet Global Health*, vol. 9, no. 2, pp. e130–e143, Feb. 2021, publisher: Elsevier. [Online]. Available: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30425-3/fulltext

[2] I. A. for the Prevention of Blindness, "Projected Change in Vision Loss 2020 to 2050." [Online]. Available: https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/projected-change/

[3] S. Klauke, C. Sondocie, and I. Fine, "The impact of low vision on social function: The potential importance of lost visual social cues," *Journal of Optometry*, vol. 16, no. 1, pp. 3–11, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9811370/

[4] CDC, "Health Risks of Social Isolation and Loneliness," May 2023. [Online]. Available: https://www.cdc.gov/emotional-wellbeing/social-connectedness/loneliness.htm

[5] A. Brunes, M. B. Hansen, and T. Heir, "Loneliness among adults with visual impairment: prevalence, associated factors, and relationship to life satisfaction," *Health and Quality of Life Outcomes*, vol. 17, no. 1, p. 24, Feb. 2019. [Online]. Available: https://doi.org/10.1186/s12955-019-1096-y

[6] E. C. M. Heppe *et al.*, "Loneliness in young adults with a visual impairment: Links with perceived social support in a twenty-year longitudinal study," *Research in Developmental Disabilities*, vol. 101, p. 103634, Jun. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0891422220300640

[7] X. Xu, W. Han, and Q. Liu, "Peer pressure and adolescent mobile social media addiction: Moderation analysis of self-esteem and self-concept clarity," *Frontiers in Public Health*, vol. 11, p. 1115661, Apr. 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1115661/full

[8] M. Armstrong, "Texting Is Alive and Well at 30," Dec. 2022. [Online]. Available: https://www.statista.com/chart/12109/sms-volume-in-the-united-states

[9] S. Humphreys, Oct. 2023.

[10] G. Thompson, "Screenreader Comparisons." [Online]. Available: https://www.perkins.org/resource/screenreader-comparisons/

[11] itisisidneyfeldman, "Blind professionals, what computer do you use and/or recommend?" Mar. 2020. [Online]. Available: www.reddit.com/r/Blind/comments/fjmmxv/blind_professionals_what_computer_do_you_use/

[12] WebAIM, "WebAIM: Screen Reader User Survey #9 Results," Jun. 2021. [Online]. Available: https://webaim.org/projects/screenreadersurvey9/

[13] Carrie on Accessibility, "Talkback now has New High Quality Voices? | Google Speech Services Update #accessibility," Nov. 2022. [Online]. Available: https://www.youtube.com/watch?v=Yy2_Iz9nvA8

[14] T. Kohler, "Challenges for Screen-Reader Users on Mobile," Apr. 2023. [Online]. Available: https://www.nngroup.com/articles/screen-reader-users-on-mobile/

[15] H. Joh, Y. J. Lee, and U. Oh, "Understanding the touchscreen-based nonvisual target acquisition task performance of screen reader users," in *Proceedings of the 19th International Web for All Conference*, ser. W4A '22. New York, NY, USA: Association for Computing Machinery, Apr. 2022, pp. 1–10. [Online]. Available: https://dl.acm.org/doi/10.1145/3493612.3520454

[16] J. Cofino, A. Barreto, and M. Adjouadi, "Sonically spatialized screen reading: Aiming to restore spatial information for blind and low-vision users," in *2012 Proceedings of IEEE Southeastcon*, Mar. 2012, pp. 1–6, iSSN: 1558-058X. [Online]. Available: https://ieeexplore.ieee.org/document/6196889

[17] B. Jin and N. Park, "Mobile voice communication and loneliness: Cell phone use and the social skills deficit hypothesis," *NEW MEDIA & SOCIETY*, vol. 15, no. 7, pp. 1094–1111, Nov. 2013, num Pages: 18 Place: London Publisher: SAGE Publications Ltd Web of Science ID: WOS:000329434400006. [Online]. Available: https://www.webofscience.com/api/gateway?GWVersion=2&SrcAuth=DOISource&SrcApp=WOS&KeyAID=10.1177%2F1461444812466715&DestApp=DOI&SrcAppSID=USW2EC0E314GXo2AbXEczXGYbd3It&SrcJTitle=NEW+MEDIA+%26+SOCIETY&DestDOIRegistrantName=SAGE+Publications

[18] P. M. Fitts, "The Information Capacity of the Human Motor System in Controlling The Amplitude of Movement," *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, Jun. 1954.