

THE DESIGN AND IMPLEMENTATION OF AN IOS MESSAGING APPLICATION
EQUIPPED WITH DEEPFAKE VOICE TECHNOLOGY

(Technical Paper)

AN INVESTIGATION OF THE PRIVACY CONCERNS AND EMOTIONAL
IMPLICATIONS OF DEEPFAKE VOICE TECHNOLOGY

(STS Paper)

A **Thesis Prospectus** submitted to the
Faculty of the School of Engineering and Applied Science
University of Virginia | Charlottesville, Virginia

In partial fulfillment of the requirements of the degree
Bachelor of Science, School of Engineering

Jordyn Anthony

On my honor as a University Student, I have neither given nor received unauthorized
aid on this assignment as defined by the Honor Guidelines for Thesis-Related
Assignments.

Signature _____ Date _____

Jordyn Anthony

Signature _____ Date _____

Richard Jacques, Department of Engineering and Society

Signature _____ Date _____

Panagiotis Apostolellis, Department of Computer Science

Introduction

In 2017, the CDC estimated that over 7 million Americans experience vision loss (20/40 vision or worse) or blindness (20/200 or worse) (CDC, 2023). The world has always been built for those with standard vision, and this development bias is becoming increasingly significant as increasingly of daily life is transferred from the physical world to the digital one. The percentage of the population accessing the internet through a smartphone was 79.64% in 2019, and that number is projected to rise to 88.75% in 2028 (Statista, 2023). In fact, about 15% of American adults in 2021 used their smartphones for primary internet access (Perrin, 2021). Clearly, the country is experiencing a profound shift to full smartphone integration, and this pattern is not unique to America. In 2021, mobile phones generated about 52% of global internet traffic (Techjury, 2023). Fortunately, there has been a push in recent years for improving the accessibility of smartphones. For example, smartphones now arrive standard with a feature that allows users to enlarge text across the entire device. This improves the ease with which those with low vision can use distinct functions of their smartphone, but depending on the degree of visual impairment, large font can slow reading speed significantly.

To address this drop in efficiency, some elect to use screen readers, of which there are two informal categories (S. Humphreys, personal communication, October 23, 2023). The first group of screen readers is the “true” screen reader: an installed application that changes the machine’s controls and allows the user to selectively read-aloud parts of the screen (S. Humphreys, personal communication, October 23, 2023). The second group of screen readers is more appropriately referred to as text-to-speech. These applications simply read aloud text that the user has selected; unlike the “true” screen readers, they do not have the authority to change the device’s controls, and they

do not help the user find text on the screen (S. Humphreys, personal communication, October 23, 2023). The following discussion pertains only to text-to-speech applications. These services are often referred to simply as screen readers, so I will follow this naming convention for the remainder of the paper.

As part of the previously mentioned push for accessibility, smartphone manufacturers have built screen reader functionalities into their phones. The inclusion of this service certainly improves the experiences of millions of users, but mobile users are confined to the operating system's native screen reader. As of the end of 2023, there do not exist reasonable methods for modifying or replacing a smartphone's screen reader software. This situation forces the mobile screen reader user to choose one of various predetermined robotic voices. My technical topic seeks to address this issue by changing the text-to-speech voice of text message conversations from a preset robot to the cloned voices of the user's contacts. My STS topic will explore the psychological and emotional effects of such a change.

Technical Discussion

The goal of my project is to develop a messaging application for the Apple iPhone. This messaging application will read a user's conversation aloud using the cloned voices of their contacts. First, I will discuss the voice cloning aspect.

"Deepfaking" is a recently coined term. It refers to items of media that have been manufactured by artificial intelligence. For example, deepfake technology has been used to take an image of a celebrity and replace that celebrity's face with the face of another person. This may seem insignificant; Adobe Photoshop has existed for years. However, deepfake software smooths the inconsistencies that the normal person would notice,

and the result is a very realistic picture that is only discernible to forensic experts. In June 2023, as part of his presidential campaign, Florida governor Ron DeSantis released a smear video of former President Donald Trump containing deepfaked images (Nehamas, 2023). This pitch was a success for the DeSantis campaign as many people believed that the images were true. This story is not unique; indeed, deepfake software has been misused countless times in the past few years since its pop culture debut. As artificial intelligence continues to improve and deepfake technology becomes more widely available, it is highly likely that these misuses of technology will continue to occur. However, deepfaking can do more than harm. For example, deepfake technology could improve the efficiency of visual effect generation or be used to develop images of crime scenes from multiple cell phone videos (Knowledgenile, 2023). Or, in my case, this technology can be used to improve the daily experience for screen reader users.

I plan to use deepfake voice software to clone the voices of users' contacts for the conversational read-aloud portion of my app. In my preliminary research, I tried Speechify, a popular text-to-speech application that recently introduced a voice cloning function. In its free demo, the software only requires a one-minute audio clip of the user's voice to generate a very realistic clone. While the voice still sounds slightly inhuman, the difference between a traditional text-to-speech voice and a voice cloned through Speechify is immense. I plan to use Speechify's API to implement the personalized voice cloning in my messaging app. API stands for application programming interface, and it essentially allows a programmer to easily incorporate a different application's functionality into their project.

The remainder of this technical discussion pertains to the second aspect of my project: to develop the application for Apple iPhone. Apple iPhones are widely

recognized as some of the best smartphones in terms of accessibility (Callen, 2021). In 2021, a survey of 1568 screen reader users was conducted, and the study found that about 70% of participants used Apple iPhone, iPad, or iPod Touch (WebAIM, 2023). Consequently, I decided to develop this project for iOS, the operating system of Apple iPhone and iPod Touch. iPad uses a separate operating system, iPadOS, which I am excluding from my scope.

Apple has a closed ecosystem, which means that Apple Inc. has full control over the device, the device's operating system, and installable apps (LeasePilot, 2023). For example, only Apple devices can run Apple operating systems; it is impossible to install iOS on any device that is not an iPhone. Conversely, the Android platform is an open ecosystem, which means that Android OS can be installed on essentially any device, including Samsung, Google, and Motorola phones. Another manifestation of Apple's closed ecosystem is that a user must receive their text messages through Apple Messages. Users can download other messaging apps, such as GroupMe or Facebook Messenger, which send instant messages over the internet, but messages sent to their phone number will always appear through Apple Messages. However, on Android devices, the user can change their default messaging app. For example, a Samsung Galaxy phone will arrive with Samsung Messages installed, but the user could elect to change the default messaging app to Messages by Google. In this situation, all messages sent to the user's phone number would appear in Messages by Google.

The implication for my project of Apple's ethos is that I must design a new messaging app that functions well enough to entice users to not use Apple Messages. By the limitations of the operating system, I can neither design an app that automatically syncs with Apple Messages, nor can I design an app that can be set to the default

messenger. The app that I create must be reasonably comparable to Apple Messages in terms of existing functionality.

STS Discussion

Those with visual impairments tend to experience higher rates of social isolation and loneliness (Biomedcentral, 2023). Considering how much of daily life depends on vision, whether the task be driving or seeing a movie with friends, this statistic is not surprising. This perceived isolation is worsened by the permeation of technology, especially mobile technology, into daily life. Increasingly, text messages are used to replace phone calls or even face-to-face interactions. Users who rely on screen readers then experience much of modern social interaction with, essentially, a robot, due to the robotic nature of the screen reader's voice. Addressing this disparity is one of my primary goals for this project. My hypothesis is that with cloned voices, much more familiar though not entirely realistic, users who rely on screen readers will experience a decrease in feelings of social isolation and loneliness. My app aims to simulate in-person conversations by using cloned voices, which I hypothesize will increase the perceived human connection between the user and their contact.

There is an ethical concern with this deepfake technology being used to create false soundbites. For example, two malefactors may contrive a fake conversation in which a politician "says" uncharacteristic things. However, deepfake software to perform this action already exists on the internet, and my project's application of deepfaked audio would not create any new avenue for false information. Also, the voice clones will be linked to the users' Apple accounts and automatically read by the app.

This prevents a malefactor from directly accessing another's voice clone and using it to say untrue things.

There exists another ethical concern in regard to the storage of cloned voices since they may be considered sensitive data. I will need to investigate safe data storage and understand the laws concerning confidential information. This concern would only be relevant if a hacker were to enter the system and steal voice clone metadata to use for nefarious purposes, but it is still an issue that I would like to address.

This project's app also has the potential to help other portions of the population. While the idea was originally conceived with the visually impaired in mind, this technology also has the potential to help those with learning disabilities, people with autism, and language learners. For example, dyslexics experience difficulty reading, but being able to listen to text conversations in facsimiles to the real-world parallel may improve their texting experience. Another example is a non-native English speaker whose listening and speaking skills are much stronger than their reading skills. This hypothetical user could understand the conversation more quickly by listening to it, perhaps even with a familiar accent, and could use dictation, rather than typing, to formulate their response. This described potential touches on universal design. Universal design is a concept in which a product that is designed to be accessible for a particular sector of the population has the potential to benefit all sectors of the population. This is, essentially, my goal for this project. While I am striving to address a particular need felt by those with low vision, I am also aiming to create an app that everyone enjoys and in which they can find significant use.

Research Question and Methods

This project's research question is whether the usage of cloned, familiar voices of a user's contacts when using text-to-speech improves the satisfaction of the user. I will use satisfaction surveys which compare key aspects of the new application to Apple Messages to gauge the users' satisfaction. Such testing requires full development of my app and deployment to the Apple App Store to allow test participants to download it. I also plan to conduct further research on strategies that have been utilized in the past to decrease social isolation both for those with visual impairments and average people. I will conduct this research by reading scholarly articles and journals. My project and research will be overseen by my technical advisor, Panagiotis Apostollelis, Assistant Professor in the Computer Science department.

Conclusion

By the end of my project, I will have a technical deliverable of a unique messaging app for iPhone and an STS deliverable of an analysis of user satisfaction data after using the app. Through my use of cloned voices, I hope to contribute to a reduction in the perceived social isolation felt by those with low vision. Simultaneously, I hope that the features of my messaging app foster greater human connection between all users and their contacts.

Word count: 1974

References

- 45+ Mobile Internet Usage Stats You Need To Know [2023]. (n.d.). Techjury. Retrieved November 22, 2023, from <https://techjury.net/blog/mobile-internet-usage/>
- Applications of Deepfake Technology: Its Benefits and Threats.* (n.d.). Retrieved November 22, 2023, from <https://www.knowledgenile.com/blogs/applications-of-deepfake-technology-positives-and-dangers>
- Callen, K. (2021, November 18). Accessible Cell Phones: Which Model is Best for You? *Wisconsin Council of the Blind & Visually Impaired.*
<https://wcblind.org/2021/11/accessible-cell-phones-which-is-best-for-you/>
- Loneliness among adults with visual impairment: Prevalence, associated factors, and relationship to life satisfaction | Health and Quality of Life Outcomes | Full Text.* (n.d.). Retrieved November 22, 2023, from <https://hqlo.biomedcentral.com/articles/10.1186/s12955-019-1096-y>
- Nehamas, N. (2023, June 8). DeSantis Campaign Uses Apparently Fake Images to Attack Trump on Twitter. *The New York Times.*
<https://www.nytimes.com/2023/06/08/us/politics/desantis-deepfakes-trump-fauci.html>
- “Open” vs. “Closed” Software Ecosystems: A Primer—LeasePilot.* (n.d.). Retrieved November 22, 2023, from <https://leasepilot.co/blog/open-vs-closed-software-ecosystems-a-primer/>
- Perrin, A. (2021, June 3). Mobile Technology and Home Broadband 2021. *Pew Research Center: Internet, Science & Tech.*
<https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/>
- Prevalence Estimates Vision Loss and Blindness | Vision and Eye Health Surveillance System |*

CDC. (2023, July 19). <https://www.cdc.gov/visionhealth/vehss/estimates/vision-loss-prevalence.html>

U.S.: Mobile internet usage reach 2019-2028. (n.d.). Statista. Retrieved November 22, 2023, from <https://www.statista.com/statistics/275587/mobile-phone-internet-user-penetration-us/>

WebAIM: Screen Reader User Survey #9 Results. (n.d.). Retrieved November 22, 2023, from <https://webaim.org/projects/screenreadersurvey9/>