An Approach to Audio-Only Editing for Visually Impaired Seniors

Robin N. Brewer[1], Mark Cartwright[2], Aaron Karp[2], Bryan Pardo[2], Anne Marie Piper[1]
[1]Inclusive Technology Lab, Northwestern University, 2240 Campus Drive, Evanston, IL 60208
[2]Interactive Audio Lab, Northwestern University, 2133 Sheridan Road, Evanston, IL 60208
{rnbrewer, mcartwright, aaronkarp2017}@u.northwestern.edu
{pardo, ampiper} @northwestern.edu

ABSTRACT
Older adults and people with vision impairments are increasingly using phones to receive audio-based information and want to publish content online but must use complex audio recording/editing tools that often rely on inaccessible graphical interfaces. This poster describes the design of an accessible audio-based interface for post-processing audio content created by visually impaired seniors. We conducted a diary study with five older adults with vision impairments to understand how to design a system that would allow them to edit content they record using an audio-only interface. Our findings can help inform the development of accessible audio-editing interfaces for people with vision impairments more broadly.

CCS Concepts
• Human-centered computing → Accessibility systems and tools;

Keywords
Older adults; vision impairments; audio interface; editing

1. INTRODUCTION AND RELATED WORK
Seniors (60+) are increasingly relying on phones and audio content to access information (e.g. receiving weather updates by phone) and want to publish their own content [1]. We are interested in helping seniors with vision impairments create and edit audio content to be shared with others, (e.g. for podcasts). Despite being a task that is not inherently visual, existing audio editing software such as Avid's Pro Tools relies heavily on graphical interfaces. Yet, such interfaces present many challenges for with vision impairments. Screen readers are a common assistive technology that help visually impaired people use graphical interfaces, yet are difficult to learn and maintain (e.g. install software updates) [2]. It also can be difficult to access screen readers and to navigate them, due to complex navigational structures that are difficult to learn and to use [3]. Another approach is to use audio-only interfaces. However, graphical tools have not been developed or deployed for audio-only editing. Prior work has shown how traditional phone interfaces on landline and non-smart cell phones can be used to access online content for seniors and suggests this modality can be useful for visually impaired seniors [1]. However, it is not obvious how best to translate audio-only interfaces for editing voice content due to the lack of visual input and feedback.

Here, we present an in-depth diary study with seniors with vision impairments to learn about their post-processing needs for audio content. We use this diary study to design a system that would allow visually impaired seniors to edit content they record using an audio-only interface (e.g. phone). Our findings can help inform developers on how to create simple, accessible audio-editing interfaces for people with vision impairments. Our findings may also extend to other populations without easy access to mainstream graphical interfaces such as people in developing countries who may rely on phones and IVR systems.

ASSETS '16, October 23-26, 2016, Reno, NV, USA
ACM 978-1-4503-4124-0/16/10.
http://dx.doi.org/10.1145/2982142.2982196

2. METHOD
We conducted a two-week diary study with older adults (over the age of 60) with vision impairments to understand how they would create and potentially want to edit audio stories. We conducted pre-interviews to learn about their current technology use and provided them with an accessible audio recorder (Wilson Digital Voice Recorder, v5) to record their stories. Participants were instructed to record on any topic as often as they wanted for two weeks. In the post-interviews, we asked participants if and how they would want to edit their stories.

Participants were recruited through audio and printed flyers at a local residential community for people with vision impairments and at organizations that have support groups for visually impaired seniors. Five people participated in the diary study (average age=72.2 years old, min=60, max=96, male=3). Two participants are blind and three reported having low vision.

3. FINDINGS AND RECOMMENDATIONS
We learned that participants employed different preferred recording strategies. P2 wanted to record his diary entry in its entirety without stopping, but noted that he would repeat a phrase until he was satisfied before moving on. The "flow" of the recording was very important to P4 and P5. P4 preferred not to edit his recording at all and would rather rerecord his whole diary entry. P5 also recorded his entry in full to maintain the flow of the story, but he would go back to edit individual phrases, rather than rerecord a whole entry.

Also, participants wanted the quality of their recordings to resemble that of audio content with which they were familiar, where quality was both a reflection the quality of the audio production and performance. For example, P5 said, "*Make me sound like a professional"*. This participant was a musician. He enjoyed uploading music online and listening to music recorded by other artists on sites such as SoundCloud. Therefore, he wanted his recordings to mimic a similar level of audio-production quality. Also, P1 described how she wanted her recordings to be similar to the quality of the audio books she listens to weekly. Participants explained that removing filler words like "ahs", "ums", long pauses, or repeated words would increase the quality of performance and production. However, P5 noted that the occasional "um" is okay, a reflection of his desire to maintain a natural flow to his recordings. P2 and P4 said they would want their recordings to have limited background noise or to be able to remove any background noises such as thumps or deep breaths. P4 tried to achieve this same goal by creating a relaxing and quiet environment for recording.

After understanding what types of edits they would make, we asked participants how would they make such edits by voice. They described two primary models of navigation which we call \ *standard* and *bookmarking*. P1, P2, and P3 said they wanted to use physical buttons for standard audio navigation similar to how one may navigate an audio cassette---*stop*, *play*, *rewind*, and *fast forward*. However, in contrast to a cassette player in which playback speed and pitch are dependent on each other, participants noted they preferred "fast playback" (e.g. 1.5x, 2x, 4x speed), in which just the playback speed (and not the pitch) is affected---this is similar to the fast playback functionality found in text-to-speech screen readers and audio books which the participants are used to. P4 elaborated further and explained that he would want to perform such navigation using a combination of keypad and voice input. For example, he may use buttons to fast-forward to the approximate subsection of the recording and dictate by voice to delete "*a sentence before that.*" P5 noted that this relative navigation could work well with absolute navigation where users can also say "*fast forward to one minute*" to find the appropriate place to edit, especially for longer recordings. These known editing locations are similar to what P1, P4, and P5 describe as navigation using bookmarks. People would be able to set bookmarks either while recording or during playback to quickly navigate recordings.

Therefore, we recommend that an accessible audio editing system provide multiple methods for navigation: *Standard*--- *stop*, *play*, *increase speed*; *Bookmarks*---jump to points in time saved previously the user (e.g. by user-provided or default labels) or defined by the system (e.g. filler words identified by the system); *Variable time interval*---jump forward/backward in time by "audio chunks", which are segmented by significant regions of silence; and *Fixed time interval*---jump forward/backward in time specified by minutes/seconds. To perform local edits, users would specify the boundaries of regions to be altered either by predefined locations *current location*, beginning, and *end*, or by bookmark locations.

4. PROTOTYPE DESIGN

Based on our findings above, we developed an initial voice interface with fixed and variable navigation functionality, the ability to delete and play segments, and a global and local silence reducer. To segment the audio and split it into sensible chunks for editing, we calculated the average power over a given frame length of the audio signal using the root-mean-square of the amplitude (RMS) for every frame of the given signal. Using these values, we determined a cutoff point to differentiate between periods of relative silence and periods of relative sonic activity. This resulted in "audio chunks" that primarily consisted of individual words or short phrases spoken in an elided manner. This "separation by silence" technique allows for variable time interval navigation. Navigation was implemented through two commands---next chunk, and previous chunk. Users can delete segments of audio by navigating to the chunk and pressing the corresponding "delete" button. The prototype's chunking system gives users the ability to remove individual words or phrases that are unnecessary. Given the importance placed on the awkward feeling of extended pauses in the post-interviews, the system also allows users to reduce the length of silences both locally and globally. Locally, this technique reduces the currently selected silence's length by a given scaling factor. Globally, all silence segments above a certain length are reduced by a given scaling factor, which is useful for longer recordings, as navigating linearly through the recording to find silences would be time-consuming for the user. We chose to focus on segmenting the audio because this is crucial for the functionality of the fixed and variable time navigation mentioned above and automatically creates bookmarks for users to more easily traverse through their recordings.

5. CONCLUSIONS AND FUTURE WORK

In this paper we investigated how older adults with vision impairments would edit voice input to explore how they could edit audio off of a computer. Our findings show how navigation and strategic audio deletion are preferred to produce high quality recordings. In the future we will further develop navigation functionality and test the prototype with older adults. This research contributes to designing voice interfaces that are accessible, flexible, and easy-to-use.

6. REFERENCES

1. Robin N. Brewer and Anne Marie Piper. 2016. "Tell It Like It Really Is": A Case of Online Content Creation and Sharing Among Older Adult Bloggers. *Proceedings of CHI 16*. http://doi.org/10.1145/2858036.2858379

2. Shaun K Kane, Chandrika Jayant, Jacob O Wobbrock, and Richard E. Ladner. 2009. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, ACM, 115–122. Retrieved March 12, 2012 from http://dl.acm.org/citation.cfm?id=1639663

3. Anne Marie Piper, Robin N. Brewer, and Raymundo Cornejo. 2016. Technology learning and use among older adults with late-life vision impairments. *Universal Access in the Information Society*.