

Representing Audio Narrations of Photos

by Amanda Legge

Committee Members:
James Hollan, Primary Advisor
Nadir Weibel, Secondary Advisor
Anne Marie Piper

Table of Contents

1. Introduction	page 3
1.1 Inspiration: Case Study of Photo Sharing for Social Therapy	page 3-5
1.2 Research Questions	page 5
1.3 Design Process	page 6
2. Background and Related Work	
2.1 Story-telling Devices	page 6-7
2.2 Digital Pen and Paper Technology	page 7
2.3 Write-N-Speak Application	page 7-8
3. Methods	
3.1 Introduction	page 8
3.2 Specific Aims	page 8
3.3 Procedure	page 8
3.4 Early Prototype Exploration	page 8-12
3.5 Photo-based Narrative	page 12
3.6 Usability Testing	page 12-14
4. Results	
4.1 Free Exploration	page 14-15
4.2 Accessing Audio	page 15-16
4.3 Survey	page 16-20
4.4 Paper Prototype	page 20-21
4.5 Analysis/Expected Results	page 21
5. Conclusion	page 21-22
6. Future Work	page 22
7. References	page 22-23
Appendix	
1. Survey Results	page 24-25

1. Introduction

Pictures are often a central feature of storytelling and how we share experiences with others. This can be observed in everyday life interactions ranging from social networks like Facebook to household items such as photo albums. Furthermore, technological advancements in cell-phones enable users to take and send photos through text or email. The prevalence of these smart phones further promotes this photo sharing activity. Because photos provide real meaning and insight into people's lives, photo sharing is a common endeavor.

Audio is another powerful tool that further enhances story-telling by adding an interaction component. Accompanying photos with audio explanations is frequently overlooked because this photo sharing activity occurs from a distance, through phones with cameras or over the internet. The combination of visual and audio data provides a more complete understanding of unique images. In addition, the extra sense of audio adds dimensionality in a distinct way. The flexibility of audio information also encourages an interaction component. Luan et al claim, "Enhancing images with visual and audio annotations can provide added value to the images" [10].

Nowadays photos are often shared digitally, but printing photos on paper is still a common practice. The inability to use traditional paper methods in combination with sound is one shortcoming of audio information. This problem can be solved, however, as "with the advent of ubiquitous digital technologies, high-fidelity sound samples have become increasingly easy and inexpensive to produce and implement" [12]. With access to digital pens and paper, this holds especially true. Because technology enabling users to attach audio to paper is easily accessible, the versatility of audio information demands further exploration.

1.1 Inspiration: Case Study of Photo Sharing for Social Therapy

Aphasia, commonly caused by stroke, is "an acquired communication disorder [...] that impairs a person's ability to understand, produce, and use language" [6]. This frustrating condition often leads to isolation and depression in the individual. In an effort to improve quality of life and communication, social approaches to treating aphasia have been employed. The most noteworthy group that practices such treatment methods where the underlying goal is "improved feelings/attitudes, connections with others, and participation in daily activities" is the Life Participation Approach to Aphasia (LPAA) [6].

As already mentioned, depression is not uncommon in people affected by aphasia. Furthermore, "a body of research, from various fields, indicates that social connection is highly correlated with longevity" [9]. Because social interactions play such an important role in emotional well-being, it is important to encourage communication (in one form or another) and participation in activities. Maintaining and strengthening relationships is often the most valuable target objective in aphasia therapy [9].

In an effort to follow this advice and mimic these approach methods, we conducted weekly observations of a 92 year old woman who was diagnosed with aphasia, hereto referred to as AT¹, for three months. The first day meeting AT (and consequently the first observation session,) she immediately initiated conversation, asking about family members, origins, and childhood life. We proceeded to tell stories about loved ones. Besides her mumbling, her speech was audible and understandable. She did, however, confuse peoples' relation to her. During this same visit, she called her actual mother, her daughter, and her caretaker "mother".

¹ All involved participants have given consent to use photos, quotes, etc.

During an observation session two weeks later, we brought old pictures of AT and her family provided by AT's daughter. When we showed her the pictures, she enthusiastically took them with a huge smile on her face [See Figure 1]. Pointing to people in one of the pictures, she labeled them "Monday," "Tuesday," and "Wednesday" but was unaware of her mistake. While she did label her children correctly in the photos, she had trouble recognizing other people. With some prompting, however, (i.e. "And this is your granddaughter, right?") she spoke and nodded in agreement. Her enthusiasm for telling stories with photos was evident, but she often had trouble explaining the setting of photos, labeling people in them, and explaining the relation of these people to her.

Taking this into account, we decided to design and implement an interactive audio photo album for AT [See Figure 2]. The photo album consisted of pictures of AT and her family, all printed on paper with a special dot pattern that lets a digital pen decode user interactions. When tapped with the digital pen, these photos play voice recordings. AT's daughter was asked to record anecdotes as well as labels for each person in the photos so that the voice on the recordings would be familiar and friendly for AT. Using this interactive photo album, the pen could supplement AT's own audio. The hope for this interactive photo album was that the assistance of details from the recordings would encourage AT to share her pictures and stories with others.



*Figure 1. AT's big smile upon looking at pictures of her family.
(Photo taken with Microsoft SenseCam to reduce intrusiveness.)*



Figure 2. Constructing AT's interactive audio photo album.

This early field study inspired the idea for a more general audio photo-sharing project. The larger project has two components: (1) A paper-digital version (like that used in AT's photo album) to be used in combination with a digital pen and (2) A digital version to be accessed from a computer. The end goal is to combine both versions so that changes to either can be synchronized. Amanda Lazar, an Electrical Engineering major at UCSD, will be driving the development and programming of the digital version of the photo album. Other members of the team include Anne Marie Piper whose dissertation focuses on assistive devices for therapy, and Nadir Weibel, a current postdoc in the Human Computer Interaction lab at UCSD. While the goal of this larger project covers multiples aspects of photo sharing, my honors project focuses on one aspect: representing audio narrations on a printed photo.

1.2 Research Questions

This research project is motivated by the following two questions:

- (1) How might audio narrations improve the paper-based photo viewing experience?
- (2) How should audio narrations be visualized on photos to maximize ease-of-use and enjoyment?

The novelty of this study derives from the fact that very little research on the topic of representing audio on digital-paper exists. In addition to identifying effective ways of visualizing audio narrations, this project will also result in the development of a multimodal photo album toolkit accessible from a Web-based server as well as a pen-based interface. Audio through this photo medium aims to provide a more enjoyable way to share stories. Further contributions of the audio photo album include the potential to strengthen methods of maintaining contact and socialization between family and friends.

1.3 Design Process

The figure below [Figure 3] illustrates the design process of this project, including the specific steps toward rapid iterative prototyping [16]. More detail about these steps will be discussed in the Method section later on.

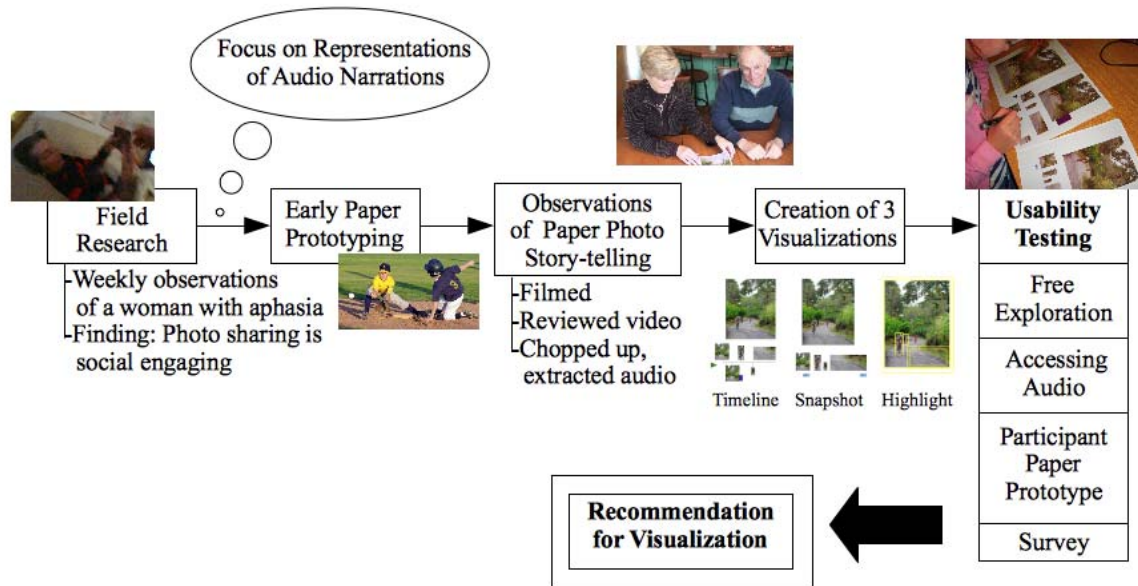


Figure 3. Flowchart of design process.

2. Background and Related Work

2.1 Story-telling Devices

Some story-telling devices that combine pictures with audio already exist. One such tool that takes advantage of these multimodal capabilities, including the ability to draw annotations, is VoiceThread [See Figure 4]. Allowing people to converse from anywhere, VoiceThread is “a collaborative multimedia slide show that holds images, documents, and videos, and allows people to navigate slides”² and leave five different types of comments: voice, typed text, and pre-recorded audio or video files. It even allows users to make doodle annotations on slides that automatically sync to his/her comments. This toolkit functions as an excellent way to stay in touch with family and friends remotely.

² <http://voicethread.com>



Figure 4. VoiceThread, a multimodal story-telling device (<http://voicethread.com>).

One of the limitations of the VoiceThread system, however, is the inability to use it in combination with traditional paper methods, as paper is both familiar and mobile [7].

2.2 Digital Pen and Paper Technology

Instead of “replacing real-world objects like paper and pen, we can successfully augment them, combining the advantages of the physical world with the capabilities of digital technology” [18]. Digital pen and paper technology make this augmentation process a reality. The Livescribe pen³ [See Figure 5] is a smart pen with an infrared camera, integrated speaker, and microphone. When used with the Livescribe dot pattern paper, which is normal paper imprinted with the Anoto pattern⁴, the pen can log handwritten notes as well as record, pause, and play audio.

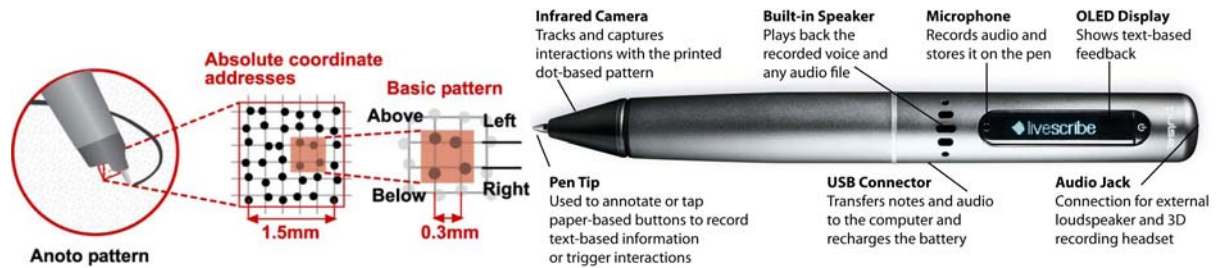


Figure 5. Anoto pattern and Livescribe Pulse Smart pen with labeled parts (source: N. Weibel).

2.3 Write-N-Speak Application

The Write-N-Speak application [See Figure 6] works in conjunction with the Livescribe pen [14]. Write-N-Speak allows users to create and customize materials easily without any previous programming experience. Included in the software are detailed instructions describing how content can be added. The interface consists of many features, including recognizing writing, recording audio, and attaching audio to specific regions on Livescribe paper.

³ <http://livescribe.com>

⁴ <http://www.anoto.com>

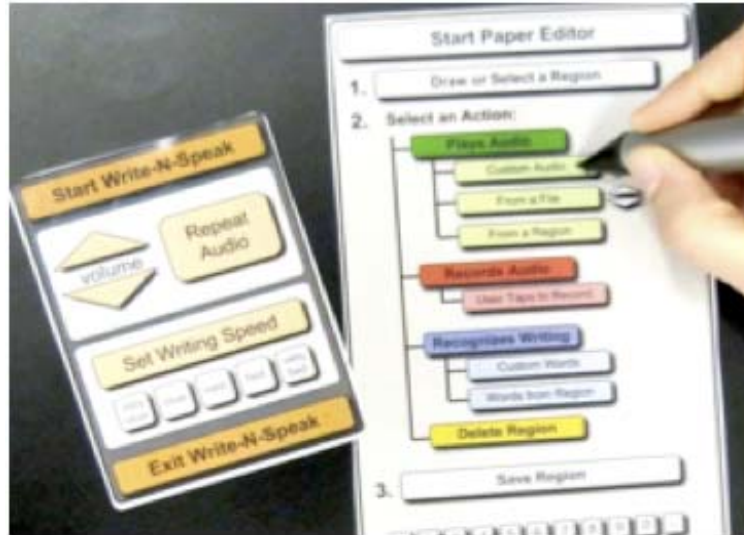


Figure 6. Write-N-Speak paper-based interface and Livescribe pen [14].

3. Methods

3.1 Introduction

The Livescribe pen and Write-N-Speak application are core components in the digital-paper version of the photo album and will therefore be essential tools in my project. The Livescribe pen will enable user to access audio and Write-N-Speak will allow users to attach this audio to specified regions on the photos. Upon completion of the larger audio photo sharing project, additional audio recorded with the pen on this digital-paper version will automatically be updated to the Web-based version by docking the pen to a computer that is connected to the photo album server.

3.2 Specific Aims

The goals of this study are the following: (1) Verify that audio narrations improve the paper-based photo viewing experience, and (2) Determine easy and enjoyable visualization methods of representing audio story-telling on photos, defined as the ability to comment and listen to audio on different regions of photos using a digital pen.

3.3 Procedure

The methods used throughout this study to fulfill these specific aims are rapid iterative prototyping, user-centered design, and self-report data. Rapid iterative prototyping involves the continuous revision of design based on user-centered feedback [16]. While the end product is ultimately the decision of the designer, feedback directly from the user is one of the main steps toward a successful design [8]. For these reasons, this study was conducted based on user responses, including both oral and written comments and behavioral reactions.

3.4 Early Prototype Exploration

My goal for this phase was to brainstorm and prototype various representations of audio narrations. Early prototype exploration involved the representation of particular audio regions of a single image using both tactile and visual manipulations. The different tactile feedback

methods included a clear raised outline, pop-ups, and layering (which was constructed by adding transparent stickers on top of one another.) Visual manipulations included the following: Image shifts, in which the specified audio region was illustrated multiple times side by side; outlines in various colors; different colored backgrounds including black and white, sepia, and blue tinted; and different contrast backgrounds and audio regions, ranging from high to low contrast. In this pilot paper prototype exploration, all ideas were executed regardless of feasibility.



Figure 7. (a) Image shifts. (b) Pop-up.

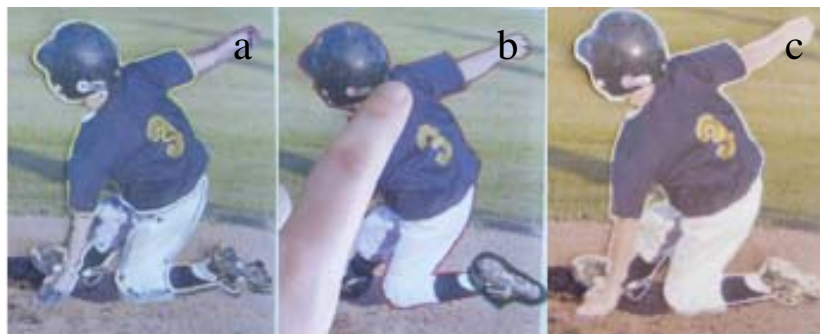


Figure 8. Outlines of audio regions on the player. (a) Rainbow outline. (b) Red outline. (c) White outline.



Figure 9. Visualizing an audio region on the player sliding into second base. (a) Black and white background with colored audio region. (b) Sepia-tinted background with un-changed color audio region. (c) Blue-tinted background with un-changed color audio region.



Figure 10. Different photo contrasts.

Preliminary feedback from three undergraduate students at UCSD on this paper prototyping provided initial design considerations. Some specific visualization preferences varied among participants due to personal preferences, but the following assertion was consistent throughout the pilot testing: The original photo should remain untouched with few distractions. One of the participants summarized it best when she stressed that “you want audio *in addition* to the picture.” While the participants did express a need for some sort of representation of audio regions, they wanted minimal modification of the photos. On the topic of these visualizations, one participant remarked, “you don’t want [the visualizations] to be so obvious, so that it still looks like a cohesive picture.” The first participant even suggested having two different pictures, the original and the modified. This way, the user could look at the photo without distracting annotations but still receive information about the location of the audio regions.

The participants were generally in favor of the use of tactile feedback in representing the audio regions, however a number of concerns emerged. While the participants did approve the usage of tactile information, they had a few worries, especially in regards to the sticker layering: (1) Concern for smudging and/or fingerprints from touching the picture and (2) Higher risk of missing a tactile region on a picture than a region represented with visual cues. All of the participants liked the idea of pop-up representations [See Figure 7b] because there were no color alterations or distortions, but one participant noted that it “seems impractical for taking it [the paper] anywhere.” The clear raised outline around audio regions was generally supported, excluding the problem of missing tactile cues. One suggestion mentioned to solve this problem was using a combination of tactile and visual feedback by placing the tactile outline on top of a visual, colored outline. Image shifts [See Figure 7a], on the other hand, were unanimously deemed to be too distracting because they conveyed movement, taking away from the picture itself. One participant remarked that the illustrated movement is good for this specific audio region in this picture, but might not be a good representation in other photos.

The responses of the participants varied in terms of the importance of representing different speaker information. One participant commented that it is “not important to distinguish between speakers because you can hear their voices.” The others, however, stated that outlines around the audio regions to differentiate between narrators (using different colors) would be helpful [See Figure 8b]. Most importantly, they noted that these outlines “do not compromise the

photo.” Regarding outlines, however, all of the participants agreed that black or white outlines (uncolored) were the least distracting and therefore the best representations [See Figure 8c].

One solution to the problem of distracting outlines and annotations is the use of different background colors [See Figure 9]. Paper prototype testing included a normal representation of audio regions placed on top of the following backgrounds: black and white, sepia, and blue-tinted. All of the participants strongly rejected the blue background, commenting that it was too extreme and “ruins the picture as a whole.” Some of the participants liked the sepia version, but there was a concern about whether or not the normal audio representation would be noticeable against the sepia-tinted background. While there was no doubt that the colored audio region stood out in the black and white background, the participants generally thought that it was too different from the original photo.

Another less extreme solution for the annotation problem is the use of different contrasts [See Figure 10]. All of the participants agreed that the different contrasts were less distracting than the different backgrounds and therefore a better visualization. They all concurred that the brighter contrast background was better than the lighter contrast background, saying that the former “still looks like the original picture, just brighter,” whereas the latter is somehow a “lesser image.”

Taking this information into account, it is not surprising that the participants were in unanimous agreement that their preferred visualization was the original picture with a high contrast audio region [See Figure 11]. One participant stated, “it doesn’t compromise the original photo, but the audio region still stands out.” Another summarized the high contrast object as a sort of ‘natural highlighting’ that “stays true to the picture.” This favorite visualization is consistent with the main overall concern of the participants in maintaining the original appearance of the photos.



Figure 11. Most successful visualization. Original picture with high contrast audio region (player #3).

The main finding from this early prototype exploration is that the original photo should not be distorted. In addition, participant feedback suggests that outlines do not compromise the

photo. The results of this early pilot testing have inspired other visualizations, including a separation between the original photo and its annotated audio regions as well as simple outlines around these audio regions.

3.5 Photo-based Narrative

To gain more insight on natural audio narrations involving photos, we conducted observations of a couple that had just returned from a three-week vacation to Costa Rica. We told the couple to select fifteen pictures that were most representative of their trip. We then filmed the couple explaining each of the paper photos to us and audio-recorded their narrations. We extracted the audio from the narrative of one of the pictures, chosen based on audio length and complexity. This allowed us to transform their static photo into an interactive one on the dot-pattern printed photo. Write-N-Speak was then used to interact with the “audio-photo.” We developed a number of different visualizations to represent various audio regions, including (1) outlines around the audio regions directly on the photo [See Figure 12], (2) snapshots below the original photo [See Figure 13], and (3) a timeline below the original photo [See Figure 14]. In order to ensure conclusive feedback on these representations, we conducted usability testing.

3.6 Usability Testing

Usability testing included a free exploration period with the digital pen and paper, an interview session to determine how each participant would access audio on the three different visualizations, and a personal paper prototype opportunity. The visualizations were given names based on their components: the highlight version [See Figure 12], the snapshot version [See Figure 13], and the timeline version [See Figure 14].



Figure 12. Highlight version.

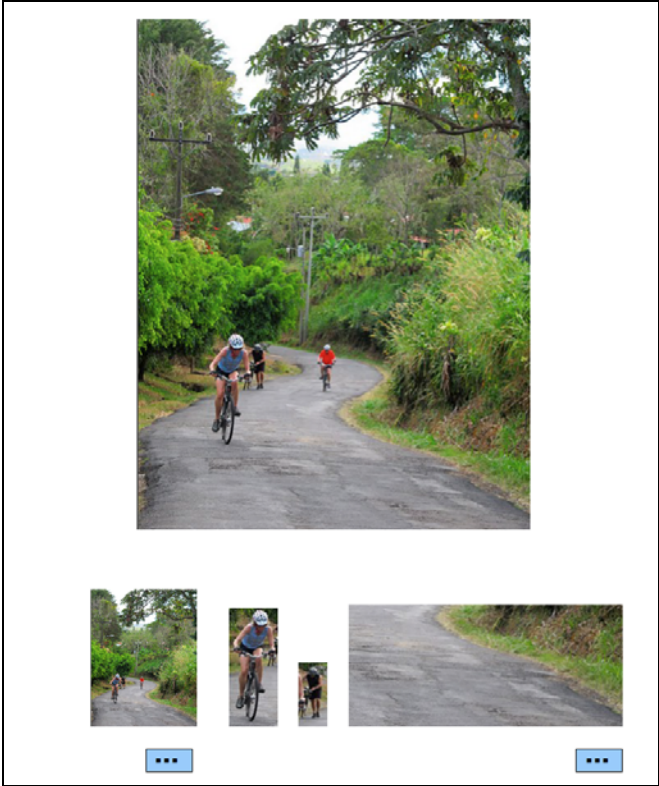


Figure 13. Snapshot version.

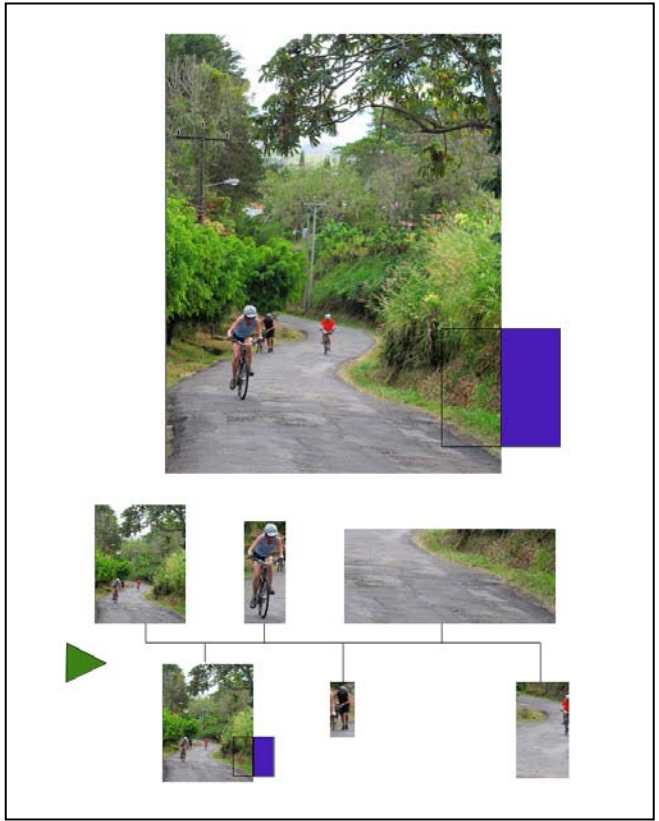


Figure 14. Timeline version.

Usability testing consisted of five participants, all UCSD undergraduate students or recent graduates. The specific procedure began with a three-minute period of free exploration with the Livescribe pen and the three visualizations. This time allowed the participants to become accustomed to using a digital pen as well as experiment with audio recording placement on each of the visualizations. After these three minutes of free exploration, the individuals were asked to access the audio in different ways to see if the designs were clear. The following five questions were asked during this portion of the testing:

1. How would you listen to the whole audio, from start to finish, on the timeline version?
2. How would you listen to the audio about the woman biking in the highlight version?
3. If you wanted to hear more about the road section on the snapshot version, how would you do that?
4. How would you listen to the background audio of the photo on the snapshot version?
5. Did the audio make the photo more enjoyable for you?

These questions were condition-ordered and randomly assigned for each participant. After these questions, the participants were asked to paper prototype their own visualization of audio regions. Each participant was provided the following materials: a glue stick, scissors, a pencil, a sharpie, a four-colored pen (black, red, blue, green), a highlighter, and different size paper copies of a personal photo, which was made available prior to the experiment. The personal picture was printed twice in normal photograph size (5x7) and in different smaller sizes iteratively by one-inch.



Figure 15. Sizing of the personal photo for paper prototype.

4. Results

4.1 Free Exploration

Even though audio from the pen is accessible by tapping the pen on different regions of the photos, two of the users initially tried to drag and make drawing motions with the pen. After a few attempts, one of these users realized that tapping was more effective in accessing audio,

however the other user consistently dragged the pen throughout the free exploration period. Tapping with the digital pen was therefore intuitive for 3/5 users.

This free exploration process also provided insight into placement of audio regions because 4/5 of the users attempted to tap on the main picture in the snapshot and timeline versions [See Figures 13 and 14]. However, no audio had been attached to these main pictures in an effort to abide by preliminary feedback from early paper prototype exploration and leave the original picture as untouched as possible.



Figure 16. Participant free exploration.

4.2 Accessing Audio

The following figure (Figure 17) is provided as a quick reference for the questions about accessing audio.



Figure 17. (Left) Highlight version. (Center) Snapshot version. (Right) Timeline Version.

Question 1: How would you listen to the whole audio, from start to finish, on the timeline version [See Figure 14]?

Response: 4/5 clicked on the green play button. 1/5 clicked on the main picture.

Question 2: How would you listen to the audio about the woman biking in the highlight version [See Figure 12]?

Response: 5/5 managed to access the audio of the woman biking. Specifically, 4/5 clicked inside the tracing of the woman biking and 1/5 clicked directly on the outline of the box around the woman biking.

Question 3: If you wanted to hear more about the road section on the snapshot version [See Figure 13], how would you do that?

Response: 1/5 clicked on the “...” beneath the small road picture as the design intended. 4/5 clicked on the snapshot of the road at first, but upon further explanation and emphasizing the word "more" in the question for these 4 participants, 2/4 then clicked on the “...” button beneath the small road picture. Out of the remaining two, 1/4 clicked on the small picture of one of the bikers because “she remembered there being information about the road on that one from the free exploration” and the other 1/4 clicked on the “...” underneath the main, small picture.

Because of the initial confusion and extreme variation in responses, this question and its responses do not have sufficient weight as evidence to influence the final audio visualization design.

Question 4: How would you listen to the background audio of the photo on the snapshot version [See Figure 13]?

Response:

3/5 clicked on the main picture. 1/5 clicked on the smaller, main picture. 1/5 clicked on the “...” under the smaller, main picture.

Question 5: Did the audio make the photo more enjoyable for you?

Response: 4/5 confidently answered, “Yes.” 1/5 responded “Sure.” Although 5/5 of the responses were positive, this question may have been misleading. It is possible that the participants only responded positively to please the experimenter. In order to remove the possibility for emotional pressure to influence participant responses, a follow-up survey was conducted. In addition to behavioral responses to the visualizations already collected during usability testing, this survey functioned as a way to gain personal feedback from the users.

4.3 Survey

This follow-up survey [See Appendix] was created using SurveyMonkey⁵. This survey served as a way to gain insight about user preferences. The survey consisted of six questions, rated based on a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree.) The ordering of choices in each question was randomized for each respondent. In addition, copies of the three different visualizations were provided as an attachment to accompany the survey for participant reference.

⁵ <http://www.surveymonkey.com>

The following graphs [Figures 18-23] show the number of positive (agree/strongly agree) and negative (disagree/strongly disagree) values for each question. Neutral answers are not represented.

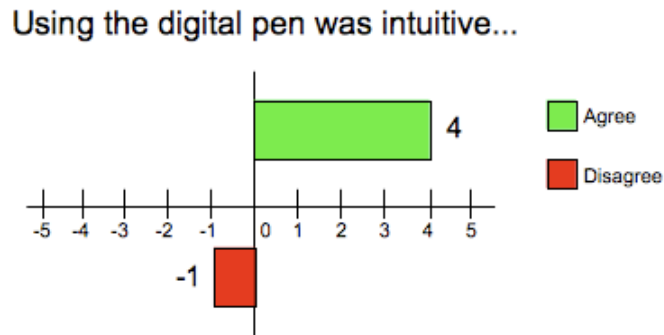


Figure 18. Intuitiveness of the digital pen.

According to the survey, users generally thought that the Livescribe pen was intuitive. This self-report feedback is consistent with the behavioral responses observed in the free exploration, which suggests the user-friendly interface of the digital pen and paper.

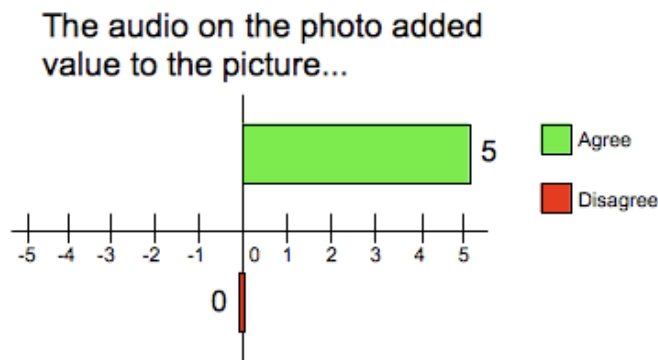


Figure 19. Added value of picture with audio.

Answers to this particular survey question adds confidence to the argument that audio does add value to images because all five of the participants agreed or strongly agreed with this statement. While this evidence is encouraging, the term 'added value' is extremely vague. For this reason, more specific questions about the audio on the picture were asked, including whether or not it was enjoyable or annoying.

The audio on the photo was enjoyable...

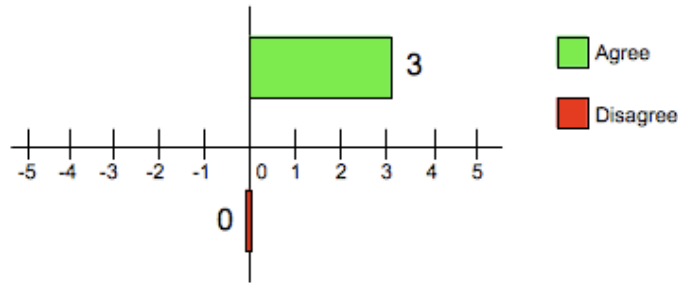


Figure 20. Enjoyable nature of the audio on the photo.

In terms of the enjoyable nature of audio narrations on the picture, there was no negative self-report feedback; answers were either neutral or positive. Because the majority of the users agreed that the audio on the photo was enjoyable, it can be concluded that listening to audio narrations was a positive experience for the users.

The audio on the photo was annoying...

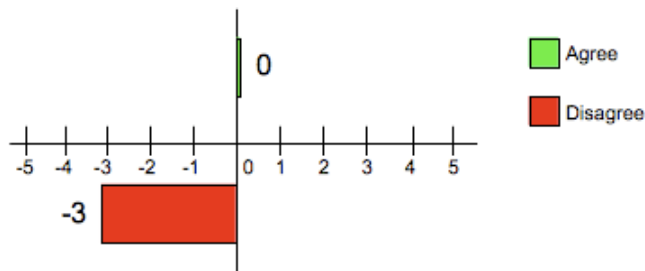


Figure 21. Annoyance ratings of audio on the photo.

Adding confidence to the argument that listening to these photo narrations was a positive experience is the fact that none of the users thought the additional audio was annoying. Results for the annoyance ratings of audio were either neutral or negative, suggesting a positive attitude regarding an interactive audio photo sharing toolkit.

The yellow box outlines on the highlight version were helpful...

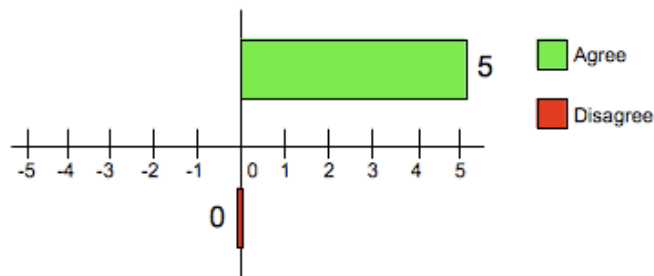


Figure 23. Rating of helpfulness of the yellow box outlines on the highlight version.

The highlight version was the most successful visualization in terms of both ease of use and aesthetics. According to the self-report feedback, all five of the participants agreed that the yellow outlines around the audio regions were helpful. This unanimous agreement is consistent with behavioral observations during the usability testing because all of the participants accessed the audio correctly during the question section. In addition, every single participant successfully accessed audio on this highlight version on the first attempt during the free exploration period.

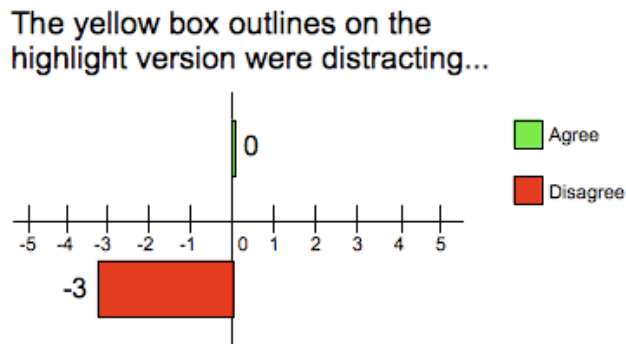


Figure 22. Distraction ratings of yellow box outlines on the highlight version.

The highlight version was also the most aesthetically pleasing visualization. While one user did not have an opinion, marking 'Neither Agree nor Disagree' on the survey, one of the users forgot to answer this question. This means that 3/4 of the users did not think that the outlines were distracting.

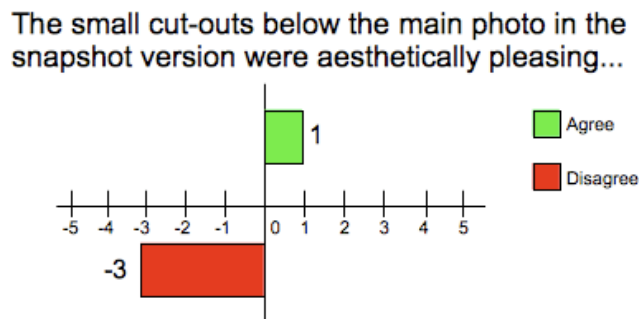


Figure 23. Aesthetics of small cut-outs below the main photo in the snapshot version.

The snapshot and timeline versions, on the other hand, were not as successful as the highlight version in terms of either ease-of-use or aesthetics. Answers about accessing audio during the question session had a lot of variation, suggesting that the design is confusing and difficult to use. Furthermore, participants did not like the aesthetics of the design as can be seen from the self-report responses. One person did not have an opinion on the matter, but the majority of the users (3/5) thought that the small cut-outs below the main photo were not aesthetically pleasing.

While answers to these survey questions do not provide definitive evidence about any particular visualization because the participant pool was small (only five participants), they do provide provisional figures and add to the compilation of collected data. Unanimously positive

reviews from this survey on the highlight version add weight to the positive behavioral responses observed during usability testing.

4.4 Paper Prototype

Participant paper prototypes functioned as a way to gain insight on personal user preferences. In addition to providing feedback on the three visualizations, this section of the usability testing gave them the opportunity to design their own way of representing audio regions. This personal representation could draw from parts of the three visualizations from usability testing or be completely novel.



Figure 24. Participant paper prototyping.

Adding confidence to the success of the highlight version already determined from survey responses and behavioral observations, 2/5 of the participants copied the highlight version exactly in their own paper prototypes. Furthermore, a total of 4/5 of the participants used some aspect of the highlight version in their paper prototypes. More specifically, 1/5 used a copy of the main picture cut into individual audio regions (similar to a geometric puzzle) below the main picture with highlights for necessary additional audio [See Figure 25b]. 1/5 used different size versions of the photo (un-cut) with audio intended to go in a clockwise motion [See Figure 25d]. 1/5 created two separate paper prototypes – a highlight version with different color box outlines and a timeline version using different sizes of the original picture as snapshots with additional highlights [See Figure 25e]. The fact that 4/5 of the personal paper prototypes included some feature of the highlight version points to the success of this visualization.



Figure 25. Final participant paper prototypes.

4.5 Analysis/Expected Results

One goal of this larger project, as shown by other work [10], is to make images more engaging. Our results confirm that the digital pen is a promising way of experiencing photos that makes the process more enjoyable.

Based on the early paper prototype exploration feedback, the highlight version was not expected to be the most successful visualization. The visualizations made below the original photo were predicted to be the most successful method for representing audio narrations because snapshots do not interfere with the original picture. However, participants repeatedly attempted to listen to audio on the original versions of the photo even when snapshots were placed below them. Additionally, the highlight version had the quickest learning curve during the free exploration and everyone accessed the audio. Furthermore, the highlight version was incorporated into the majority of the personal paper prototype creations. Although these results were not predicted, they still do agree with early user feedback from initial prototype exploration. According to these participants, outlines did not compromise the picture as a whole. For this reason, the success of the highlight version is not too surprising.

5. Conclusion

This preliminary study serves as only a piece of the final, audio photo-sharing toolkit project as it focuses on only a slice of the target population. This study has provided good

evidence about preferential audio visualization methods among UCSD students and affiliates. The target population for this interactive audio photo-sharing toolkit includes all age groups and backgrounds, beyond the selected participants. For this reason, more studies need to be done on different groups.

All evidence from this initial group, however, points to the success of audio supplements with photos. Audio narration in combination with pictures is enjoyable and socially engaging for the user. The probability of success of this interactive audio photo album is therefore high. In terms of visualizations, all feedback about the highlight version was positive. In addition to the easy learning curve from the behavioral responses and positive survey responses about the highlight version, the personal participant paper prototypes confirm this preference. For this reason, the digital version of the project is based on similar aspects.

6. Future Work

While the initial aim of this larger project is to enhance story-telling abilities across general users, further steps include customizability of the toolkit to make it more person specific. Upon fabrication of this final audio photo album, it will be easy to change and add settings to make it beneficial for people with specific needs. A later feature in the photo album creation, for example, could enable users to print the audio photos on digital paper in a variety of ways depending on their impairments. For example, people with aphasia often have difficulty viewing objects on their right hand side [9], so photos could be printed on the left side. Another feature could allow for audio speed controls, automatically slowing speech and prolonging pauses to make comprehension easier for people with aphasia. The versatility and mobility of this digital-paper photo album in combination with the visual and audio aspects makes it the perfect assistive device for maintaining social relationships for people with and without impairments.

In addition to testing additional population groups that vary in terms of age and background, the next steps in this larger research project include syncing up a digital version of the photo with its paper-digital version so that people can easily upload and access photos and audio online through their Web browsers. Upon completion of this syncing, information will then be sharable online or printed on digital-paper. A computer user can make audio regions on uploaded photos by using a four-sided geometric figure (square or rectangle), like in the digital-paper highlight visualization. The user can then select the region to record audio on it. The primary goal of this toolkit is to make story-telling between people easier, especially across long distances. The digital side of this project will facilitate this process. The success of this toolkit will result in more enjoyable socialization methods.

6. References

- [1] Al Mahmud, A., Limpens Y., & Martens, J.B. (2010). Expressing Through Digital Photographs: An assistive Tool for Persons with Aphasia. In *proceedings of the 5th Cambridge Workshop on Universal Access (UA) and Assistive Technology (AT), CWUAAT'10*. University of Cambridge, UK. Springer-London (Springer Book Title: Designing Inclusive Interactions).
- [2] Chapey, Roberta et. al. <http://www.asha.org/public/speech/disorders/LPAA.htm>
- [3] Choi, S. H., & Walker, B. N. (2010), Digitizer Auditory Graph: Making graphs accessible to the visually impaired. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)*, Atlanta, GA, (10-15 April).

- [4] Daemen, E., Dadlani, P., Du J., Li, Y., Erik-Paker, P., Martnens, J-B., and de Ruyter, B. Designing a Free Style, Indirect, and Interactive Storytelling Application for People with Aphasia. Eindhoven, The Netherlands: Philips Research.
- [5] Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. *Proceedings of the International Conference on Auditory Display (ICAD 2008)*, Paris, France (24-27 June).
- [6] Duchan, Judith Felson and Sally Byng, eds. Challenging Aphasia Therapies: Broadening the Discourse and Extending the Boundaries. Hove [England]; New York: Psychology Press, 2004.
- [7] Harper, R. & Sellen, J. A. The Myth of the Paperless Office. MIT Press, November 2001.
- [8] Holtzblatt, K., Wendell, Jessamyn B., Wood, S. Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design (Interactive Technologies). Elsevier Inc., 2005.
- [9] LaPointe, Leonard L. Aphasia and Related Neurogenic Language Disorders. New York: Theime, 2005.
- [10] Luan, Q., Drucker, S., Kopf, J., Xu, Y. & Cohen, M. *Annotating Gigapixel Images. (Proceedings of UIST 2008)*, Monterey, CA (19-22 October).
- [11] Nees, M. A., & Walker, B. N. (2008). Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data. *Proceedings of the International Conference on Auditory Display (ICAD 2008)*, Paris, France (24-27 June).
- [12] Nees, M. A., & Walker, B. N., (2009). Auditory Interfaces and Sonification. In Constantine Stephanidis (Ed.), *The Universal Access Handbook* (pp. 507-521). New York: CRC Press Taylor & Francis.
- [13] Papathanasiou, Ilias and Ria De Bleser, eds. The Sciences of Aphasia: From Therapy to Theory. Amsterdam; Boston: Pergamon, 2003.
- [14] Piper, A.M., Weibel, N., and Hollan, J. Write-N-Speak: A System for Authoring Multimodal Paper-Digital Materials for Speech-Language Therapy. ACM Transactions on Accessible Computing (TACCESS), under review.
- [15] Piper, AM., Weibel, N., Fouse, A., Hollan, J. "SHB: Small: A Paper-Digital Infrastructure for Health Care Applications." NSF Grant Proposal, under review.
- [16] Preece, J. Rogers, Y. Sharp, H. Interaction Design: Beyond Human Computer Interaction. John Wiley & Sons, Ltd., 2011.
- [17] Spreen, Otfried and Anthony H. Risser. Assessment of Aphasia. Oxford; New York: Oxford University Press, 2003.
- [18] Stifelman, L., Arons, B., Schmandt, C. The Audio Notebook: Paper and Pen Interaction with Structured Speech. *Proceedings of the SIGCHI conference on Human factors in computing systems (ACM March 2001)*.
- [19] Tesak, Juergen and Chris Code. Milestones in the History of Aphasia. Hove [England]; New York: Psychology Press, 2008.
- [20] Whitworth, Anne et.al. A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia. Hove [England]; New York: Psychology Press, 2005.

Appendix

Appendix 1. Survey Results

1. Using the digital pen was...							Create Chart	Download
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count	
Difficult	0.0% (0)	50.0% (2)	25.0% (1)	25.0% (1)	0.0% (0)	2.75	4	
Intuitive	0.0% (0)	20.0% (1)	0.0% (0)	60.0% (3)	20.0% (1)	3.80	5	
answered question							5	
skipped question							0	

2. The yellow box outlines on the highlight version were...							Create Chart	Download
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count	
Distracting	25.0% (1)	50.0% (2)	25.0% (1)	0.0% (0)	0.0% (0)	2.00	4	
Helpful	0.0% (0)	0.0% (0)	0.0% (0)	60.0% (3)	40.0% (2)	4.40	5	
answered question							5	
skipped question							0	

3. The audio on the photo...							Create Chart	Download
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count	
Added value to the picture	0.0% (0)	0.0% (0)	0.0% (0)	40.0% (2)	60.0% (3)	4.60	5	
Was enjoyable	0.0% (0)	0.0% (0)	40.0% (2)	40.0% (2)	20.0% (1)	3.80	5	
Was annoying	0.0% (0)	60.0% (3)	40.0% (2)	0.0% (0)	0.0% (0)	2.40	5	
answered question							5	
skipped question							0	

4. The chronological ordering of audio narrations in the timeline version was...

[Create Chart](#) [Download](#)

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count
Helpful	0.0% (0)	0.0% (0)	60.0% (3)	40.0% (2)	0.0% (0)	3.40	5
Important	0.0% (0)	20.0% (1)	20.0% (1)	40.0% (2)	20.0% (1)	3.60	5
answered question							5
skipped question							0

5. The small cut-outs below the main photo in the snapshot version were...

[Create Chart](#) [Download](#)

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count
Aesthetically pleasing	0.0% (0)	60.0% (3)	20.0% (1)	20.0% (1)	0.0% (0)	2.60	5
Confusing	40.0% (2)	0.0% (0)	40.0% (2)	20.0% (1)	0.0% (0)	2.40	5
answered question							5
skipped question							0

6. I would like...

[Create Chart](#) [Download](#)

	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	Rating Average	Response Count
Listening to audio on my family and friend's photos	0.0% (0)	20.0% (1)	20.0% (1)	20.0% (1)	40.0% (2)	3.80	5
Sharing photos with audio	0.0% (0)	20.0% (1)	60.0% (3)	0.0% (0)	20.0% (1)	3.20	5
answered question							5
skipped question							0

Appendix 1. Survey questions, responses, and average ratings (from www.surveymonkey.com).