Designing audio-enhanced paper photos for older adult emotional wellbeing in communication therapy

Anne Marie Piper a,c,*, Nadir Weibel b, James D. Hollan b,c

a Northwestern University, Department of Communication Studies, 2240 Campus Drive, Evanston, IL 60208, USA
b Department of Computer Science, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
c Department of Cognitive Science, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 31 October 2013
Accepted 2 January 2014
Available online 9 January 2014

Keywords:
Emotional wellbeing
Digital pen
Communication
Older adults
Therapy
Audio

ABSTRACT

This article introduces audio-enhanced paper photos enabled by digital pen technology as an interaction paradigm for supporting and understanding emotional wellbeing in late adulthood, particularly in the context of communication therapy activities and related social interactions. We describe the development of a multimodal pen-based system that enables creation of audio-enhanced paper photos and the application of this technology to two domains: older adults with aphasia working to regain expressive and receptive language after a stroke, and older adults wanting to interact with a family member of advanced age with memory loss. Our pen-based authoring software enables caregivers, both therapists and family members, to create audio-enhanced paper photos to encourage meaningful and emotionally appropriate interaction with the older adults they support. We examine the ways in which emotional information is embedded in audio-enhanced photos and how this approach addresses various aspects of emotional wellbeing in late adulthood.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The global trend of an increasing older adult population makes it a priority not only to design interactive experiences that are sensitive to the social and emotional needs of people in late adulthood but also enhance overall wellbeing in late life. The proportion of people over age 60 compared to younger populations is growing at a rapid rate. For the United States, the older population in 2030 is projected to be twice as large as the population was in 2000, growing from 35 million to 72 million, representing nearly 20% of the United States population (NCHS, 2012).

Much research examines the relationship among aging, social interaction, and late life wellbeing. Interpersonal relationships play an important role in emotional wellbeing (Asay and Lambert, 1999; House et al., 1988), yet there is strong evidence that one's social network shrinks as a function of age (Gordon and Gaitz, 1976; Harvey and Singleton, 1989; Carstensen, 1992). This may be due to a limited number of surviving relatives and peers as well as physical or cognitive declines that present barriers to social participation. There are many benefits for maintaining social support and engagement in social activities, including living longer and improved mental and physical health (Stuck et al., 1999; Ferlander, 2007). Frequent social interactions may create an increased sense of wellbeing in older adults (Czaja et al., 1993), and in contrast, a lack of social interaction may contribute to higher rates of disability, slower recovery from illness, and even early death (Lubben and Gironda, 2003; Blazer, 1982).

An important topic of research related to emotional wellbeing is understanding design for the changing social and emotional needs across the lifespan. Here, we focus on design for people in late adulthood. In our work, the end goal is to prevent social isolation and loneliness through communication therapy, which may include structured speech-language therapy, reminiscence or narrative therapy (White and Epston, 1990), and audio-visual prompts for grounding face-to-face interaction. When designing technologies for people in late adulthood, it is important to consider the unique social and emotional qualities of older adults (Carstensen, 1995; Carstensen and Charles, 1998). While social interaction may decrease in late adulthood, many older adults report overall satisfaction with their life as well as sustained or improved satisfaction with interpersonal relationships (Diener and Suh, 1997; Carstensen and Charles, 1998). Older adults are motivated to stay connected with people with whom they already have a strong emotional connection (Carstensen et al., 2003) and contact with family becomes increasingly important (Cicirelli, 1989). Furthermore, Carstensen et al. (2003) argue that aging is associated with an increase in motivation to derive emotional meaning from life.
This paper describes the design and use of audio-enhanced paper photos to support and understand older adult emotional wellbeing. The creation of and interaction with audio-enhanced paper photos is enabled by digital pen technology and custom software (Piper et al., 2012) we developed for the Livescribe pen,\(^1\) a widely available digital pen that enables recording and playback of audio. We introduce this technology into two contexts and present case studies of how caregivers design audio-enhanced paper photos based on the interests and emotional needs of the older adults for whom they are caring. In both case studies we printed paper photos on Anoto digital paper\(^2\) and we used sticky labels pre-printed with Anoto pattern to augment existing paper photos and interact with them through the Livescribe pen. The first case study examines how a speech-language therapist designs custom audio-enhanced paper therapy activities for an older adult with aphasia. In this context, the therapist designs activities that leverage personal artifacts in the older adult’s life. He balances the need to engage the older adult in therapy without incorporating overly emotional content. Negative emotions (e.g., sadness about not seeing a family member) severely impact the older adult’s ability to communicate and willingness to participate in therapy. In the second case study, we describe the creation and use of audio-enhanced paper photos involving a person of advanced age with memory loss and her extended family and care staff. Family and care staff attach personalized audio messages to printed photos of people in the older adult’s life. This approach successfully engages the older adult in reminiscence activities and improves her awareness of people in her social circle. Taken together, these case studies provide new insights into the design of audio-enhanced photos for late life communication therapy, exploring the ways in which this media form supports emotional expression, reflection, and interpersonal communication.

2. Related work

This research brings together literature on aging and the social and emotional goals of older adults, photographs and audio as media that carry and evoke emotion, and technologies for older adult communication, reminiscence, and therapy. We also describe related work on digital pen technology as an interaction technique for older adults.

2.1. Emotion, aging, and social interaction

Literature on emotion and aging has emerged within the past two decades primarily because early models of late-life emotional regulation presumed a decline in this area similar to age-related slowing of physical and cognitive functioning. It is well documented that people in old age interact with others far less than in youth (e.g., Gordon and Gaetz, 1976; Harvey and Singleton, 1989). Early theories in psychology presumed that this decrease in social interaction would have a negative impact on emotional wellbeing in late life. Recent work, however, indicates that older adults are more satisfied with interpersonal relationships and their lives overall (Diener and Suh, 1997; Carstensen and Charles, 1998).

Socioemotional selectivity theory (SST) (Carstensen, 1995) provides a foundation for understanding emotion and social interaction across the lifespan. The theory suggests that a similar set of goals persists throughout the lifetime, but the importance of these goals changes depending on one’s perception of the end of life. Carstensen and colleagues argue “that the realization that time is limited directs social behavior to experiences that are emotionally meaningful” (Carstensen and Charles, 1998). SST also suggests that older adults value self-achievement and the development of self-concept less than younger populations, but they place higher value on regulation of emotion. Older adults are often motivated to stay connected with people in their social network with whom they already have a strong emotional connection (Carstensen et al., 2003). Compared to younger adults, older adults are less concerned with the goal of information seeking, in which novel social partners are often the best sources and are especially valuable in fulfilling this goal (Carstensen, 1995). Contact with family, particularly siblings, takes on new importance in old age and a disruption of those bonds may be related to depression (Cicirelli, 1989). People optimize for socioemotional experience in late life, and this can be characterized as a more complex or multifaceted emotional experience due to the positive and negative emotions associated with nearing the end of one’s life (Carstensen and Charles, 1998).

Aging is also associated with a variety of changes in cognitive abilities, including declines in short-term memory (Gregorie and der Linden, 1997), semantic knowledge (Park, 2002; Schaei, 1996), and verbal ability (Schaei, 1996). Some aspects of cognitive functioning, however, remain relatively stable over the lifespan. Stable aspects include autobiographical memory (Fromholt et al., 2003), implicit memory (accumulation of life experiences and learnings) (LaVoie and Light, 1994), ability to process emotional information (Carstensen et al., 2003), and theory of mind (recognizing the viewpoints of others) (Happe et al., 1998). Our approach to design draws on these strengths of older adults, particularly autobiographical and emotional awareness, through the use of personal photos augmented with custom audio messages.

2.2. Audio-enhanced photographs

Photographs are a powerful tool for social interaction (Hirsch, 1997) and provide anchors for discussion (Sit et al., 2005). Moreover, photographs are both carriers of and triggers for emotion. The medium of audio adds another dimension of emotional information, providing nuanced richness of voice that conveys speaker affect. We couple paper photos with audio recordings through the use of digital pen technology. Our implementation of audio-enhanced photos builds on the idea of audiophotography (Frohlich and Tallyn, 1999). Enhancing a photo with audio has the potential to bring the image to life and to aid one’s memory of events (Frohlich and Tallyn, 1999; Frohlich, 2004). Similarly, Dib et al. (2010) explore the notion of a “sonic souvenir” for family remembering. Compared to photos, sounds were found to be more varied, familial, and creative as well as being evocative and generating reflective narrative.

Central to our design for older adults is the introduction of augmented paper photos rather than digital images. The term “Kodak culture” (Chalfen, 1987) refers to generations of people who grew up with photos printed on paper, largely today’s older demographic. Paper photos have immediate affordances understandable by everyone (Petrelli and Whittaker, 2008), whereas digital photo sharing alternatives require a level of technical understanding, which often limits use by older generations (Sarvas et al., 2008). Frohlich and Fennell (2007) recommend audio photographs rather than video clips as a new media form, emphasizing the importance of paper photos in contrast to a screen-based viewing experience.

2.3. Tools for reminiscence and therapy

Much related work describes the development of technologies for older adult communication, reminiscence, and related therapies. This includes studies of older adults’ attitudes about keeping

\(^1\) http://www.livescribe.com  
\(^2\) http://www.anoto.com
in touch (e.g., Lindley et al., 2009) and the ways in which physical and digital artifacts might be passed down among generations (Odom et al., 2012). Several projects introduce multimedia tools for older adult communication and reminiscence, particularly related to seniors with memory loss. For example, remote photo- and video-sharing supports interactions between individuals with dementia and their therapists (Kuwahara et al., 2006), and a multimedia tool stimulates long-term memory and encourages conversation involving adults with dementia (Gowans et al., 2004). Smith et al. (2009) explore the idea of multimedia biographies to support communication and reminiscence between individuals with memory loss and their families. Webster (2011) and colleagues introduce the Portrait system to serve as a communication bridge between people with dementia and their caregivers. As we focus on in this paper, the majority of these technologies support interaction between older adults and family or caregivers.

2.4. Digital pen interaction for older adults

Much of computing is moving off of the desktop and into the world in the form of lightweight, portable devices, giving rise to new interaction techniques that are well-suited for older adults. Our approach investigates pen-based computing as a way for older adults to access, interact with, and modify audio-enhanced paper photos. Prior work examines the affordances of paper compared to purely digital media (Sellen and Harper, 2001; Luff et al., 1992; Mackay, 1999; Nomura et al., 2006), revealing the importance of paper in collaboration, information sharing, and communication. Digital pens are often used to link paper and digital media and create hybrid interactions that span the physical and digital divide (Mackay et al., 2002; Stifelman et al., 2001; Guimbretière, 2003). Applications of digital pen technology are numerous and include document editing (Conroy et al., 2004), linking handwritten and audio notes (Chiu and Wilcox, 1998), and interactive scientific Notebooks (Yeh et al., 2006). Closely related to our approach is the Memento project, which introduces a hybrid paper-digital scrapbooking architecture (West et al., 2007). Memento uses an Anoto pen to allow users to annotate a paper photo album and link these annotations with audio recordings and video files on a nearby computer. While the system does record and replay audio, the digital pen must be coupled with a computer, requiring users to setup and manage Bluetooth connections and ensure Internet connectivity.

Recently, Piper et al. (2010, 2011, 2013) have investigated digital pens and associated paper-based interfaces as a platform for inclusive design, particularly with respect to older adults. A digital pen serves as an effective pointing device for older adults (Piper et al., 2010, 2013), a user group who often feels a lack of comfort and control over computers (Czaja and Sharit, 1998). Older adults experience an overall slowing of movement and difficulty with fine motor activity and coordination (Holmes et al., 2009) as well as disability due to osteoarthritis (Hootman and Helmick, 2006). Pen-based computing has been shown to be an accessible interaction technique for older adults with arthritis and even partial paralysis due to a stroke (Piper et al., 2010, 2011). Disability in late life is commonly the result of multiple minor impairments, including physical, visual, and hearing loss, and digital pens provide redundancy in interaction through visual and auditory channels.

3. Authoring audio-enhanced paper documents

Audio-enhanced paper photos are physically printed photos that become interactive. Audio messages are linked to particular regions of a printed photo and can be played back by using a Livescribe digital pen.1 Livescribe pens come equipped with a microphone and an integrated speaker that allow users to record and playback audio. The digital pen can also be programmed—through a special SDK—with dedicated software that can further process the captured written information and recorded audio.

We leveraged the Livescribe SDK to create TAP and PLAY (Toolkit for Authoring Pen and Paper Language Activities) (Piper et al., 2012), a software application that is deployed on the pen and enables the creation of links between regions on paper and audio messages. It enables non-technical users to create multimodal paper-based applications by programming paper documents to respond to particular user interactions using the Livescribe digital pen. Users can print photos on special dot-pattern paper and interact with them using the digital pen. Transparent or opaque stickers printed with dot pattern can be attached to any other material, including existing traditional printed photos.

The Livescribe pen is based on Anoto technology,2 which employs an infrared camera in the pen to decode the printed pattern in real-time and detect the pen tip position. TAP and PLAY exploits this precise localization on paper and the on-board recording and playback facilities for creating multimodal applications to support language activities.

3.1. Creating paper-based multimodal applications

The TAP & PLAY authoring software allows users to create multimodal applications for interactive paper documents and use the developed applications to support various activities based on writing, speaking, and recording. In order to create such applications, authors use a printed interactive control panel (Fig. 1). Tapping with the digital pen on the paper control panel initiates a series of voice prompts that guide users through how to create interactive regions.

Fig. 1. TAP and PLAY control panel. Top: the Livescribe Pulse Digital Pen. Right: the main control panel, allowing authors to use the full functionality of the toolkit. Left: additional paper controls to manage applications settings, get further help, and quickly define new interactive areas. In the main control panel authors tap on the Start Here button to start creating their own multimodal paper-based applications. The toolkit guides them then throughout the process of designing a new application depending on the kind of actions and interactions that authors want to define. The Save My Voice to a Page widget enables quick creation of interactive areas that play back recorded audio.
Multimodal paper-based applications can leverage combinations of the following actions:

1. **Record** – Create regions on paper that, when tapped, record audio and store it on the internal memory of the pen.
2. **Playback** – Plays back a specific audio message from the provided audio library, or previously recorded on the digital pen. This can also be interactively defined as part of the developed application by using a Record action to first record a new audio message in a specific region, and then a Playback action to play it back.
3. **Hand-writing recognition** – Create a region where the toolkit on the pen interactively parses written text and recognizes written letters and words in real-time. Recognizable words are based on a user-defined dictionary, and words can be added to the dictionary through the printed control panel and a printed keyboard.

Multiple actions can be added to the same area on paper by specifying different user interactions. These include the following: **tap and hold**, **tap and release**, and **double tap**. Effectively, a single region can have three different interactions. Consider, for example, a printed word on a page, where **tap and hold** spells the word, **tap and release** reads the word, and **double tap** reads the word within a sentence. By default the programmed action executes when the pen touches the paper, and this is the nature of interaction with the audio-enhanced photos described in the case studies below. Our prior work indicates that this form of basic interaction is necessary for many older adults (Piper et al., 2010); however, TAP & PLAY also provides advanced users with additional ways of specifying and layering interactive content. For instance, authors can create overlapping regions that are linked to different functionality. For example, a user may add audio describing a group of people in a photo, and tapping on the background plays this audio. Then they may add individual audio descriptions of each person in the photo, perhaps by circling individual faces. This provides a layering effect, and the most recently added layer is the one that becomes active when in contact with the digital pen. Fig. 6 (top) provides an example of this interaction, where a description of the photo’s context is played when the user taps in the background and a message from the people in the photo is played when the user taps on their torsos.

3.2. **Audio-enhanced paper photos**

One particular use of our technology is to augment printed photos with audio messages. As shown in Fig. 2, by using the pen-based authoring software we developed, users can enhance photos with audio messages and create a custom interactive experience to support communication and social interaction. Audio-enhanced paper photos carry and evoke emotions through the use of personal photos combined with the nuanced richness of voice. Audio messages contain content but also have other properties related to affect such as voice intonation, volume, pitch, and changes in these qualities throughout the recording. The next section presents two case studies of audio-enhanced photos that examine the design and use of this media form for older adult emotional wellbeing.

4. **Case studies**

This article is the first to synthesize findings from our larger research project on end-user authoring of audio-enhanced paper documents, and through this we find that audio-enhanced paper photos are a powerful interaction paradigm for engaging older adults in communication therapy and related social interaction. The contribution of this paper is a new examination of audio-enhanced photos in two respects. First, we examine how emotional information may be embedded in this media form as a way of triggering memories of one’s past, supporting communication, and encouraging new social interactions. Second, we analyze the audio that is archived as part of authoring interactive paper photos as a new data source about designing for older adult emotional wellbeing.

In each case study below, the older adult’s caregiver received a digital pen equipped with the content authoring software and a variety of paper materials to use with the pen, including photos of people or keepsake items printed on dot pattern paper. The first case study describes a 12-week deployment of the technology, and the second case study describes a five-month deployment. For each case study, researchers collected field notes and video data of technology use, conducted interviews, received email and written diary-like reports from participants, and logged all pen-based interactions internally on the devices.

4.1. **Aphasia therapy and rehabilitation**

As the first case study, we present research on audio-enhanced paper photos in older adult aphasia therapy. This work is informed by ten months of field research, including six months of observations and interviews dedicated to understanding therapy and a 12-week field deployment of digital pen technology with one therapist and older adult with aphasia in a clinical setting. Piper et al. (2010) describe this formative research on aphasia therapy, and Piper et al. (2011) detail system usability and usage data from the field deployment. In this analysis, we examine the emotional needs of an older adult with aphasia and discuss how audio-enhanced paper photos support and provide new data about emotional wellbeing in the context of structured therapy.
4.1.1. Coping with aphasia

This case study focuses on clinical treatment and related interactions involving a single speech-language therapist (JB) and one older adult with aphasia (AF, age 88). JB is a licensed and practicing therapist (officially called a speech-language pathologist) who works at multiple senior care facilities. AF experienced a stroke in the left side of her brain, resulting in Broca’s aphasia, one year before we began field observations. Aphasia is an acquired language disorder that makes it difficult to remember words, speak, read, and/or write. Approximately one million Americans have aphasia, which often results from a stroke, brain injury, or neurological condition (NINDS, 2012). Sources estimate that 25–40% of stroke survivors acquire aphasia (Nat. Aphasia Assoc., 2012). There are different forms of aphasia, and here we focus on Broca’s aphasia from a stroke. Similar to other people with Broca’s aphasia, AF has difficulty recalling certain words (i.e., word finding), constructing coherent sentences, and articulating her speech. As an example of AF’s speech, JB held up a spoon and said, “What is this?” AF replied, “It’s a spear. This is a peter... It’s a jar... It’s a spoon.” Although AF has expressive language difficulties, she is still able to say many words and generally able to communicate her needs. AF is an avid reader, and her stroke had minimal impact on her ability to process written information. She lives alone in the independent living portion of her retirement community.

Given the sudden and dramatic loss of language associated with aphasia, it is not surprising that AF’s condition has a negative impact on her emotional wellbeing. The relationship between aphasia and depression is well studied (see Code et al., 1999, for a review). The inability to communicate effectively as one was once able is a vast source of frustration. People may grieve over what they have lost, and AF is no exception to this. Individuals with Broca’s aphasia may be keenly aware of their deficits and easily become frustrated by the change in their abilities.

Speech-language therapy helps individuals with aphasia improve language comprehension, practice speaking, and develop strategies to cope with communication challenges. Additionally, therapists may provide emotional support for individuals with aphasia. Is it not uncommon for a person with aphasia to cry during therapy, and our observations of AF evidenced this. Speech-language therapists use a variety of techniques and devices to motivate treatment and provide a comfortable working environment for the person with aphasia. Piper et al. (2010) note the importance of paper documents to both the therapist’s existing work practices but also to older adults’ comfort and control over an activity. Of particular importance is the ability to customize paper therapy materials to leverage the older adult’s interests and encourage meaningful participation. We note that touch-screen systems could also provide such customization, but our prior work indicates that many available touch-based interfaces are too expensive, complex, and overwhelming for older adults with aphasia (Piper et al., 2010). Therapists may use traditional audio recorders to capture and replay a client’s speech, but as we discuss below, this stands in contrast to interaction with audio recordings that are linked dynamically and accessed directly through TAP & PLAY.

Maintaining social support for people with aphasia is especially important, particularly when people with aphasia tend to communicate with fewer friends and have smaller social networks (Davidson et al., 2008). Individual therapy goals vary depending on the person. For AF a primary goal is to improve functional communication that allows her to converse fluently with friends and family.

4.1.2. Audio-enhanced photo design and usage

Speech-language therapy is self-motivated, and motivation has long been considered as an important variable in treatment outcome (Shill, 1979). Engaging the client’s interests is a valuable strategy for encouraging participation in therapy, and the customization afforded by our pen-based authoring software enables creation of audio-enhanced paper materials designed around an individual’s interests. Data from our field research provides many examples where JB tailored paper-based therapy activities around AF’s interests. A predominant therapy activity involved object naming, which addresses word recall and pronunciation challenges. For this type of activity, JB used printed photos of AF’s living room, keepsake objects within AF’s home, her cat, and jewelry she made, all of which are sources of enjoyment for AF (see Fig. 3). He augmented the paper photos with custom audio recordings to support the object naming task. For example, JB recorded the names of various parts of AF’s cat and made these areas interactive with the digital pen (e.g., tapping the pen on the cat’s ear played the word “ear” aloud). JB also added audio recordings to objects within a photo of AF’s living room (see Fig. 4). When preparing these materials, it was common for JB to record an object name, play it, and revise his audio recording. He did this on multiple occasions, especially when first learning to...
use the digital pen and authoring software. His goal was to create loud, clear audio recordings that AF could understand. For instance, he recorded the word “candle” three times, each time the recording was progressively louder and more articulate. The audio recordings he created and linked to other paper photographs had similar acoustic qualities.

JB instructed AF to use the digital pen to explore these images, listen to the words, and then practice saying the vocabulary words on her own. While the activity of tapping the pen on an image and hearing JB’s voice played aloud was an appropriate way to introduce audio-enhanced photos, AF said this activity was “too easy.” JB has the challenge of designing activities that are an appropriate level of difficulty for AF. We observed her resist for this activity that appeared too simple but then become frustrated with a task that was too difficult or where the instructions were too complex. As the content designer, JB must carefully balance the level of difficulty in order to engage but not discourage AF.

A far more difficult object naming activity required AF to verbally generate the names of objects and, with JB’s assistance, save her voice recordings on the digital pen. For this activity, JB used photos of items found within AF’s apartment, such as candlesticks (Fig. 3, left). For each photo, JB started the pen’s recorder and asked AF to say the name of the object that was pictured. This task proved to be difficult for AF given her expressive language challenges, and she needed to create multiple recordings before successfully labeling the object. Below is a transcription of AF’s multiple successive attempts at saying the word “candlestick,” each of which was saved on the digital pen.

AF1: … thats a, thats a f f fla, i, its a stick its a st, a cand, cando, cland, its a slip of a, an author of a including, including plastic, a plastic hole
AF2: … this is a silver flas, flas, flasernage flas of with a glass erger
AF3: clineg does manic of closic, as clean, for avine ((sighs)) im sorry it freezes me
AF4: this is a clandis, no, classical, it is with a cale on the corner. is that better?
AF5: this is a candlestick
AF6: its a candlestick

The audio recordings that are saved on the digital pen provide a record of AF’s attempts and eventual success during the object naming task. AF’s tone changes throughout this series of audio recordings. Her speech sounds tentative in the first four recordings, and she pauses and restarts her speech frequently. When she is successful in the fifth and the sixth recording, the tone of her speech sounds more like that of JB’s audio recordings on the pen. Her speech is both louder and more assertive. Perhaps the change in AF’s tone indicates her confidence in her speech, or possibly her tone is mirroring that of JB’s when he labels similar objects within a photo. These recordings also capture other indicators of AF’s emotional state. At one point AF pauses, sighs, and then apologizes (attempt 3). Later she asks JB if her response is better (attempt 4).

Although JB wanted a single one-word answer from AF, the ability to record, replay, and revise the audio-enhanced photo allowed AF to make multiple attempts at pronouncing the word in a low-risk way. A few minutes later, JB asked AF to say the word “candlestick” again and recorded her response with the digital pen. Instead of saying the word as she did before, AF said, “…when I think of it… It’s glass, it’s fabric, it’s a design, it’s very pretty…” The medium of audio supports free-form verbal expression such as this, and the pen-based authoring software enables in-the-moment capture of these rich descriptions when viewing the photograph. In this case, AF may be unable to say the word “candlestick” consistently, but the pen-based authoring software allows her to incorporate other verbal descriptions into the experience that are easier for her to generate.

In this context we examine the use of audio-enhanced photos that contain meaningful objects and a pet, but we did not observe the use of photos or dialog about people in AF’s life as a planned therapy activity. Field observations and interview data indicate that JB avoided talking about AF’s family because it brought up negative emotions. AF became sad after receiving a phone call or visit from a relative. When she mentioned these events during therapy, she became noticeably upset and needed several minutes to compose herself before she was able to proceed with therapy activities. According to JB, her emotional response was a combination of frustration over not being able to communicate well and sadness over not seeing family as she once did. He also explained that this emotional stress caused further breakdowns of AF’s communication abilities. The customizable and personal nature of audio-enhanced photos allowed JB to design therapy experiences that were motivating and engaging for AF without being overly emotional.

4.1.3. Impact on emotional wellbeing

An important aspect of AF’s emotional wellbeing is giving her the confidence to attempt speech, make mistakes, and learn to correct them. It is through this practice that she will rebuild her ability to converse more fluently with family and friends in her community. The above examples and the cooperative, iterative authoring process involving JB and AF are characteristic of how audio-enhanced photos were used in this context. JB and AF generated multiple audio recordings for each image, revising the audio until they were satisfied with the final product. The process of speaking, reflecting on one’s speech, and then attempting improved speech is an important aspect of therapy. Learning to evaluate and correct one’s speech is a central learning goal for AF, and ideally this practice will improve her ability to interact socially with people in her daily life.

Another aspect of AF’s emotional wellbeing involves her perception of herself as an active participant in the therapy process. JB designed and used audio-enhanced photos as a tool to legitimately engage AF in the therapy process, which is a valuable strategy for motivating her to attend and continue therapy. AF commented on the approach, “And this is for people, for me, to, to, to understand what I should sound like.” The ability to record the client’s voice and attach it to related images is key. JB noted the importance of including AF’s voice as part of this experience:

It just gives…personalization to the task. It’s her voice. She’s listening to it and she’s doing the task… If you’re part of the process, you know, you will understand. You’ll find it more rewarding because you helped make it… You’re empowering her.

In the context of speech-language therapy, the older adult plays a critical role in adding verbal descriptions to the printed images, and this is a point of contrast with the second case study where the older adult is minimally involved in the design process. The audio recordings contributed by AF in this context generate a new and important source of data about late life emotional wellbeing. Both positive (e.g., laughter) and potentially negative affect information (e.g., hesitation and uncertainty in speech) is evident in the audio recordings.

4.2. Family memories and reminiscence

The second case study details the design and use of audio-enhanced photos by a large extended family to stimulate and engage Ethel (age 105). Our analysis is informed by six months of
field research, including a five-month deployment of the authoring system with Ethel, her extended family, and her care team. Full details of this case study, including system usability and usage data, are provided in Piper et al. (2013). While other approaches such as photo and video sharing (Kuwahara et al., 2006) and multimedia biographies (Smith et al., 2009) have been used to support reminiscence for people’s names and memory loss, we examine the media of audio messages coupled with printed photos as a tool for reminiscence. We focus on enabling content generation by family and care staff and empowering the older adult to take an active role in interacting with this media. We briefly describe the context and participants in this study, and then we present excerpts that highlight the ways in which emotion is embedded in and triggered by audio-enhanced paper photos.

4.2.1. Photos as communication and memory aids

Ethel is the oldest member of her family. She is experiencing normal cognitive and physical declines for her advanced age. She has trouble remembering people’s names and her relationship to them. Ethel often forgets who extended family members are, whether appearing in a photo or in person. She is also unlikely to know which day of the week it is and generally her ability to understand the passage of time has declined.

Ethel lives in a retirement community along with two of her children. In addition to onsite nursing staff, Ethel’s four children (in their 70s and 80s) help care for her and visit her daily. Alice (Ethel’s oldest daughter, age 78) has been Ethel’s primary caregiver over the past 10 years. While many older adults rely on other nonkin residents in a retirement community for social support (Stacey-Konnert and Pynoos, 1992), Ethel has become socially withdrawn and interacts primarily with family:

[Ethel] was 101 when she went into the nursing home and she didn’t want to interact with any of the people there, she considered them ‘old.’ She also didn’t interact with the people taking care of her. If she had a complaint she waited until I [Alice] came or my brother came and then told us. She didn’t want to eat in the dining room for any of her meals.

Alice is the family member who interacts with Ethel most frequently, and as such she is a key informant for this research. She comments on the challenges of engaging her mother in conversation:

I come in [to Ethel’s room], say a few things. She doesn’t really interact in the conversation, but if you start talking about who someone is in a picture, and “do you remember that person?” It gives a focal point for communicating… It gets real easy just to go in with a book and sit with her.

Ethel has a strong interest in her family and memories captured in photos. She has a bulletin board in her room filled with paper photos of her family. Alice says, “[Ethel] wants to get me engaged in doing something with the pictures. It’s her way of communicating.” Paper-based photos are familiar media for this user group, as Ethel and her children are all part of the “Kodak culture” (Chalfen, 1987), people who grew up with and value photos printed on paper. We leverage the ways in which Ethel’s family currently uses paper photos to foster communication with her and study this family more closely as they design audio-enhanced paper photos as a reminiscence activity for Ethel.

Ethel’s family began authoring audio-enhanced photos by placing the interactive photo album out as a guest book at Ethel’s 105th birthday. Over the five month field deployment, 36 people from Ethel’s family and care staff added 102 photos to the album, attached 104 audio recordings to various printed images, and accessed these audio recordings 1757 times (see Piper et al., 2013, for full details).

4.2.2. Audio-enhanced photo design and usage

In this context, family members and care staff were the designers of the audio-enhanced photos and Ethel’s subsequent interactive experience. Typically, participants looked through the album for photos in which they were pictured and added audio narrations to these images. A common convention was for a person to designate their face as an interactive region that plays an audio message directed toward Ethel. Fig. 5 illustrates an example of this.

Audio-enhanced photographs convey affect information in multiple ways. Foremost, the photos themselves are evocative for the family and care staff as content creators. Family members contributed their own personal photographs to the album, and many of these images evoked an emotional or experiential response from family members. With the digital pen able to record audio, family members offered verbal descriptions of the context of the photo, the surrounding environment, their mood

![Fig. 5. Left: Ethel’s family members create an audio-enhanced photo. Right: Paper photo and linked audio comment.](image)

---

This is a picture of you. I don’t remember what birthday it is, but you’re old already here, and now you’re even older, 105.

We’re in Savannah. We just took a ride on this carriage, looking at all the old buildings and having a really great day.

We’re out at the beach and I know it was one of your birthday parties. We had a big fish fry I’m sure. We’re probably all filled with fried fish, and everybody’s getting a little sunburn, and the wind’s blowing. Remember all of those wonderful birthdays you had out here...

I can still smell the pine cones. I can remember the water. I can remember us standing and taking this picture...
that day, and other details related to the image. Fig. 6 provides examples of photos and linked audio descriptions.

These descriptions may help Ethel understand the photo and remember its context. Family members also include other information about the photo such as the purpose of the photo or markers of time to indicate when the photo was taken. In this way, the audio-enhanced photos may trigger emotions such as sadness when a speaker calls attention to a person in the photo who has passed away (e.g., “...this is [Minnie] and myself quite a few years ago before [Nicole] passed away...”). Here, the tone of voice in the recording can convey sadness and grief over a deceased loved one.

Another way in which family members embedded emotional information in audio messages is through direct statements about affection and well wishes. These messages were directed towards Ethel. Common examples include: “I love you, Grandmother,” and “I hope you have a happy birthday.” Four messages from great-grandchildren are songs directed toward Ethel. In this way, the audio recordings allow children to capture and save a form of expression (i.e., singing) that is comfortable for them. We observed one child (age 2) who was unwilling to record a spoken message, but with her mother’s encouragement, she sang a song for Ethel and added it to the photo album.

Audio-enhanced photos also capture more indirect or contextualized affect information. For example, many of the audio recordings include laughter by the speaker or speakers. Several recordings even include instances of banter and teasing between family members. As part of a single audio recording, two of Alice’s daughters argue jokingly about who is the favorite child, and the message ends in laughter among Alice and her daughters. Fig. 7 illustrates another instance of teasing involving Ethel’s youngest daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s daughter.

Fig. 7. An instance of teasing among siblings which was recorded and linked to a photo of Ethel and her four children.

Audio recordings even include instances of banter and teasing between family members. As part of a single audio recording, two of Alice’s daughters argue jokingly about who is the favorite child, and the message ends in laughter among Alice and her daughters. Fig. 7 illustrates another instance of teasing involving Ethel’s youngest daughter. It appears that such teasing may not have been meant for Ethel directly, as it occurred between Ethel’s children and grandchildren when authoring photos as a group. There are subtleties to voice that allow for different interpretations by different audiences, just as photos are subject to individual interpretation. A single audio-enhanced photo may evoke different emotions depending on the viewer’s perspective and relationship to the content.

Adding personalized audio recordings to printed photos enables a richer expression of emotion by Ethel’s family members. Audio recordings carry emotional information triggered by an individual’s memory of a photo as well as spontaneous, in-the-moment interaction that contains laughter among siblings. This interaction during the photo viewing experience, also called phototalk (Crabtree et al., 2004; Frohlich et al., 2002), is now archived and associated with the image. Both Ethel and her family members may review the phototalk that is now part of the album, and in turn this may encourage further narratives that build on the previous dialog. We note that the introduction of the digital pen changes the nature of the phototalk, as much dialog was carefully crafted to be included as a recording for Ethel. Audio recordings were often refined until the author was satisfied, changing the phototalk experience from a conversation that might have occurred more naturally within the context of viewing the photo and interacting with other co-present individuals (Crabtree et al., 2004; Frohlich et al., 2002) towards a more performative, intentionally designed experience.

The vast majority of audio-enhanced paper photos were created by Ethel’s family, but a few photos were added by Ethel’s care staff with the help of Alice. These photos picture the care staff member and include an audio recording that typically states the caregiver’s name and the ways in which they care for Ethel. Several recordings created by care staff express positive affect information toward Ethel. One nurse’s audio recording includes, “…I enjoy working with [Ethel] and she’s a great person.” Another nurse’s recording states, “I help take care of [Ethel]... We spend a lot of time together, and I enjoy it very much... She is an extremely excellent person... and I love her.” Interestingly, the nurses’ language is not directed towards Ethel, but instead it appears that her message was meant for Ethel’s family or other photo album viewers.

4.2.3. Impact on emotional wellbeing

Throughout the five-month field deployment, we document several ways in which the experience of interacting with audio-enhanced paper photos improves Ethel’s emotional wellbeing. Overall, receiving and using the interactive photo album were a positive experience for Ethel. Pen-based interaction was an appropriate and accessible method of interacting with audio-enhanced photos (see Piper et al., 2013 for more details). Beyond system usability, however, the media of audio-enhanced paper photos provides a rich interactive experience for Ethel to enjoy independently or in cooperation with others. Importantly, this experience is guided by a caregiver but self-paced by the older adult. That is, Alice helped Ethel get started with the digital pen, but then Ethel’s actions are required for progression through the book. Alice reported to the research team:

“I took it to [Ethel] today and started it up for her and she just took off. She tapped on the pictures and listened with delight, smiling a big smile. I sat with her for thirty minutes while she did it all herself and then she continued for another hour... She was the center of attention and loved it. Earlier in the day she was dozing in her chair and not focusing on anything. This brought her to life.

Other family members also noted Ethel’s interest in the album. After going through the album with Ethel, her granddaughter Sharon said, “She becomes animated and excited. She really enjoyed listening to the narrations.” Email reports from Alice indicated that Ethel preferred to review the photo album in the evenings instead of watching television as she used to do. Ethel’s enjoyment of the photo album documents another dimension of her emotional wellbeing.

The photo album supports Ethel in reflecting on past memories and remembering people in her life. Interestingly, six of the seven most frequently accessed recordings were related to photos of Ethel’s past birthday parties. According to Alice, Ethel’s birthday parties have served as a family reunion for many years, and the memories made during her birthday event are positive. Alice said of the photo album, “...It brings back the past and it seems very happy memories.” While many memories captured in the photo album were positive, feelings of sadness may be evoked when reviewing a photo of a deceased family member. Ethel’s photo album contains photos of her late husband and other family
members who have passed away, and video data of Ethel interacting with these photos conveys sadness over the loss of family members. Additionally, an instance of phototalk between Alice and Ethel when viewing a photo of Ethel's husband is captured on the digital pen. This recording has a noticeable shift in Ethel's tone from laughter at the beginning to a more somber tone when she mentions that her husband died.

Another important dimension of emotional wellbeing is Ethel's ability and interest in interacting with other people in her community. At six weeks into the field deployment, we surveyed Ethel's close family members (n = 5) and her nursing staff (n = 4) to assess any changes they observed given the introduction of the photo album. All nine respondents said they saw improvement in the quality of conversations they had with Ethel as well as her interest in socializing with others. Eight people reported improvement in each of the following dimensions: overall happiness, responsiveness, alertness, and willingness to interact with others. Beyond survey data, we received numerous accounts from Ethel's family and care staff about changes in her social behavior. Alice describes one example:

Last evening one of the residents came to [Ethel's] door in her wheelchair and wanted to come in to visit. We had just started to use the photo album... [Ethel] was so pleased to show her the pictures and tap on them. She told her who some of them were... She actually talked to her, which is unusual because she doesn’t usually interact with the other residents.

Ethel’s awareness of people in her social circle has an impact on her ability to interact with others and her subsequent emotional wellbeing. Alice said of her mother, “She doesn’t point to the bulletin board as she used to do, asking me if I knew all those people...” Ethel's nurse commented that she “will show me family pictures on the wall and tell me who they are. [Ethel] never did that before.” Ethel's daughter Mary explained, “The good thing about the album was that it helped her to recognize those family members she does not see often.” At the end of the study, Alice reported via email:

By repeating the process of looking at the pictures and hearing the recorded comments, she gradually learned who the people were... Now when she sees some of these people, she has something to attach the current people to the pictured people and remembers names much better.

Ethel used the audio-enhanced photos created by her family and care staff during the last year of her life, as she passed away shortly after this study. The photo album was a positive resource for social interaction, reminiscence, and relearning the names and faces of people in her social network. While we focus mainly on Ethel's needs, the introduction of the photo album also had an impact on Alice. As Ethel's caregiver, Alice sought activities to socially engage and stimulate her mother. The photo album successfully achieved this, but at the same time it also helped provide meaning to the time Alice spent with her mother and impacted their one-on-one relationship. The activity of reminiscing was shared between Ethel and Alice, and our data suggests that this interaction caused by the photo album, both through its creation and use, had positive consequences for Alice's wellbeing and closeness with her mother. At the end of Ethel's life, however, there are questions about who owns the photo album content and who should have access to it in the future. We begin to address issues of content ownership and preservation in the following section.

5. Discussion

This paper presents two contrasting case studies that describe the design and use of audio-enhanced paper photos for older adult emotional wellbeing. In both case studies caregivers are the primary content designers; however, the two case studies differ in terms of the goals for introducing audio-enhanced photos and the older adult's involvement in content creation. We examine the ways in which emotional information is embedded in and triggered by this media form, and discuss considerations for how these artifacts should persist over time.

5.1. Caregivers as designers

Our pen-based authoring software enables caregivers to rapidly create, revise, and customize audio-enhanced paper photos for the older adults they support. Caregivers intimately know the needs of the older adults for whom they are caring and leverage that expertise in their design of audio-enhanced photos. While the caregiver is the primary designer of content, they are also the main support person who enables interaction with the photos. In neither case study did the older adult use the digital pen and associated photos without the support of a caregiver. This has the benefit of the caregiver observing the older adult's interaction needs and then adjusting the design and use of audio-enhanced photos to meet those needs. For example, Ethel had difficulty reaching her arm out to tap regions at the top of a photo, so Alice began to put interactive regions lower on the page and closer to Ethel. Alice also labeled interactive regions with the words “tap here” to guide Ethel. Beyond designing to address usability issues, the caregivers understand the type of content and interactive experience that might be most emotionally and socially appropriate for the older adult. JB selected photos of objects within AF's home and her cat, which are sources of enjoyment for AF. Alice helped Ethel's family and care staff assemble a photo album full of biographical information, and Ethel's use of the photo album served as a memory aid and a catalyst for new social interactions. The TAP & PLAY pen-based authoring system enables therapists, family, and care staff to design custom interactive paper-based content that is tailored to the emotional and social needs of the older adults for whom they support.

5.2. Involving older adults in content creation

The older adult's role in authoring audio-enhanced paper photos differed greatly between the two case studies. AF played a central role in the design of audio-enhanced photos. She pointed out meaningful objects within her home, and these objects were photographed and included as part of the therapy materials. Given the focus on improving AF’s expressive language, including her speech as part of the audio-enhanced photo was paramount, and the vast majority of audio recordings include AF's voice. With JB's support, AF reviewed and revised the audio recordings linked to photos. Involving AF as a legitimate participant in the creation of audio-enhanced photos was critical to her understanding of the task and has consequences for her emotional wellbeing. However, the process of authoring the photos was challenging for AF, and JB reported that just the act of following multi-step instructions was a valuable part of her therapy. Ethel’s audio-enhanced photos were evolved in a much different way. Ethel's family selected photos of significance and created audio recordings for Ethel. Only a few recordings (seven of 104) include Ethel's voice, and these were instances of phototalk between family members and Ethel. For Ethel, audio-enhanced photos were designed primarily to archive and revisit her past rather than document her current abilities, as with AF.
5.3. Embedding emotional information

The two case studies differ in how emotional information was embedded in and triggered by the audio-enhanced photos. With AF, the goal was to promote emotional stability and to avoid negative emotions related to her condition or family situation. Her photos pictured personally meaningful objects but not people in her life. Since AF’s voice was part of many audio-enhanced photos, her voice recordings provide additional data about her emotional state. For example, AF’s recordings include pauses, sighs, and questions to JB about her progress that may indicate her affective state. Ethel’s situation was precisely the opposite, where the goal of introducing audio-enhanced photos was to engage her in reminiscence activities. Ethel’s family and care staff added elabo-rate audio narratives that describe family relationships, contextual reminders, and sensory information about the photo. Voice provided increased expression for Ethel’s great-grandchildren who were more comfortable singing a song than speaking a message. Ethel’s family also included many messages that seemingly had multiple purposes and intended audiences. A single message may evoke humor for one viewer and sadness for another person. The flexibility of audio-enhanced photos enables this richness of expression as well as a straightforward mechanism for interacting with these new artifacts.

5.4. Pen and paper as a recording and playback tool

Digital pens and paper have several advantages compared to other recording devices, such as a standard handheld audio recorder that JB previously used in therapy. With traditional audio recorders, the caregiver is typically in control of the experience, starting and stopping the recordings; whereas with the system we present, the older adult is empowered to interact with previously recorded audio somewhat more independently. Digital pens also afford a certain precision and make it possible to specify and link a particular region of a photo that is relevant to the dialog. Moreover, all recordings are stored on the digital pen and can be accessed in any desired order and in context of visual aids that are integrated into the interaction. This stands in contrast to the effort required to access specific points of an audio file captured with a traditional recorder and gives users a unique freedom to explore the material and to define ad hoc trails within and across the available audio-enhanced photos.

5.5. Ethical considerations

End-user creation of an emotionally rich content base such as the audio-enhanced photos we describe naturally generates questions about who owns the content, who should have access, and how it should exist after the research study concludes. In the first case study, JB maintained ownership of the digital pen, recorded audio files, and paper-based content for several months after the study. Audio-enhanced photos in speech-language therapy, as we observed, provided a new way to document AF’s progress. In this context, it was appropriate for the therapist to keep and continue accessing this material as needed. Toward the end of the study, JB noticed that AF had several smaller strokes. He mentioned that indicators (i.e., changes in AF’s speech) may be captured on the digital pen and reviewing these might give additional insight into her condition.

In contrast to the first case study, Ethel and her family had a much stronger attachment to the content they created. The interactive photo album captured detailed biographical information about Ethel and her family and was used frequently by Ethel during the last year of her life. Ethel’s family maintained ownership of the photo album after the study, although they discontinued use of it after Ethel passed away. The family viewed the photo album as an important artifact that documented family history. One of Ethel’s daughters suggested using audio-enhanced photos for documenting and describing family heirlooms, whereas they previously did this through video recordings.

5.6. Limitations and future work

One limitation of the current approach is that content is maintained on individual digital pens. Content sharing between pens requires development of additional software. It is necessary to consider ways in which content created with digital pens can persist and be preserved beyond the technology’s lifecycle. In addition, we envision web-based additions to the current software so that both content creation and usage can be extended. While this opens new design dimensions, involving a wider ecology of devices and cloud-based storage increases the complexity of ethical considerations.

A natural question arising from this research is how pen- and paper-based audio-photographs compare to similar media presented in a purely digital form on a tablet computer. A tablet-based system may be more sustainable over time, providing a platform for the creation, use, and long-term storage of audio-photographs. However, as we found in our earlier work, tablet computers may be too complex and overwhelming for older adults with physical and cognitive impairments (Piper et al., 2010). Audio-enhanced paper photos have certain properties that make them ideal for this population: they are large format, always “on”, tactile, and in a form that is familiar and generally appreciated by older adults. Also, navigating a paper photo album was intuitive and comfortable for the older adults we studied. The medium of paper and pen allowed asynchronous content creation and exploration of the messages (Sellen et al., 2006), yielding a certain freedom in interaction directed by one’s interests.

Another area of future work involves applying computational techniques to analyze archived audio recordings in an effort to better understand emotional wellbeing. The digital pens deployed as part of the two described field studies archived all audio recordings created by participants. While our current system does not include automated analysis of audio files, the capture of audio data enables further analysis and potentially new insights into the older adult’s emotional state. Computational techniques to detect affective state are especially interesting. For example, work by Scherer et al. (2013) on analyzing and categorizing emotion in audio recordings provides a promising approach for identifying vocal indicators that reliably differentiate various emotions. In future work we plan to analyze the archived audio files from each case study to aid understanding the rich emotions captured by audio-enhanced photos.

6. Conclusion

This article presents pen-based authoring software that enables caregivers to create custom audio-enhanced paper photos to support the emotional and social needs of the older adults for whom they are caring. We analyze the design and use of audio-enhanced paper photos in two contexts, and through this analysis provide insight into the ways in which this media form carries and evokes emotion. Printed photos on paper are familiar to older adults and allow caregivers to flexibly incorporate a variety of meaningful images into communication therapy activities. The nuanced richness of voice adds another expressive dimension to static printed images, and the open-ended nature of audio recordings allows people, from children to older adults, to express themselves in multiple ways. Through two contrasting case studies, we describe how audio-enhanced paper photos aid in documenting and improving emotional wellbeing for older adults.
Acknowledgments

We thank our study participants, and we thank Livrescribe for donating digital pens.

References


