Exploring Low-Cost, Internet-Free Information Access for Resource-Constrained Communities

JENNIFER PEARSON, Swansea University
SIMON ROBINSON, Swansea University
MATT JONES, Swansea University

Rural developing regions are often defined in terms of their resource constraints, including limited technology exposure, lack of power and low access to data connections (leading to an inability to access information from digital or physical sources), as well as being amongst the most socio-economically disadvantaged and least literate in their countries' populations. This paper is focused around information access in such regions, aiming to build upon and extend the audio-based services that are already widely used in order to provide access to further types of media. In this article, then, we present an extended exploration of *AudioCanvas* – an interactive telephone-based audio information system which allows cameraphone users to interact directly with their own photos of physical media to receive narration or description. Our novel approach requires no specialist hardware, literacy, or data connectivity, making it far more likely to be a suitable solution for users in such regions.

CCS Concepts: •**Human-centered computing** → **Collaborative interaction; Mobile devices;** *Interaction techniques; Collaborative content creation; Mixed / augmented reality; User studies;* •**Information systems** → *Multimedia and multimodal retrieval; Multimedia content creation;*

Additional Key Words and Phrases: QR codes; photos; audio; developing regions.

ACM Reference Format:
Jennifer Pearson, Simon Robinson, and Matt Jones, 2017. Exploring Low-Cost, Internet-Free Information Access for Resource-Constrained Communities. *ACM Trans. Comput.-Hum. Interact.* V, N, Article A (January 2016), 34 pages.
DOI: http://dx.doi.org/10.1145/2990498

1. INTRODUCTION

In the developed world, we often take for granted the ability to access vast amounts of information from sources all around us. Text-based media (such as flyers, posters and maps) and digital data (from personal devices or public displays) are commonplace, and easily accessed by the majority of the population. In many developing regions of the world, however, low-textual literacy and various resource-constraints (e.g., little technology exposure, poor data connectivity) prevent communities from accessing information from printed or digital sources [United Nations Development Programme 2013].

While access to laptops or other higher-powered devices in many developing regions is low, the majority of the population are likely to own or have access to at least a low-end cameraphone (e.g., featurephone or similar precursor to smartphones, with

[1]See, for example: http://goo.gl/TPbGtc

This work was funded by EPSRC grants EP/J000604/2 and EP/M00421X/1. The authors would like to thank Ram Bhat and Minah Radebe for facilitating and coordinating studies in India and South Africa. We also gratefully acknowledge the help of Anirudha Joshi, Nina Sabnani and colleagues in arranging to meet Kaavad storytellers and coordinating studies in Mumbai.
Authors' addresses: J. Pearson (corresponding author), S. Robinson, and M. Jones, Future Interaction Technology Lab, Computer Science Department, Swansea University, UK; email: j.pearson@swansea.ac.uk.

2016 1073-0516/2016/01-ARTA $15.00
DOI: http://dx.doi.org/10.1145/2990498

a touch screen and access to a limited app store) [Telecom Regulatory Authority of India 2012; Vallina-Rodriguez et al. 2009]. In addition, a large selection of low-end Android smartphones are now available for as little as $20[1], providing many more future users with access to tools and services that were not available on earlier, less sophisticated handsets.

Despite these advancements, however, access to mobile data connections in these regions is still low. In some cases this is due to the unattractive pricing structures and business models, while in others it is the lack of reliable data coverage. A combination of this poor internet connectivity—which prohibits online translation or inquiry—and a lack of textual literacy—which prevents people from reading printed media—means that accessing information in many less-developed regions can prove exceedingly difficult.

One potential solution to this access problem is the introduction of voice-based telephone services known as Interactive Voice Response (IVR) systems (e.g., [Kumar et al. 2007; Patel et al. 2010; Patel et al. 2011]). These services are designed to be an accessible, audio-only version of the Internet, and behave in a similar way to the DTMF-based telephone services often used for customer service enquiries worldwide (e.g., using the phone's number-pad to select navigation options).

These voice services usually contain crowd-sourced information on a variety of topics and are intended to provide spoken information by and for local users, in local languages, accessible via any device without the need for a mobile data connection or textual literacy. To access information, users call via the public telecom network using any type of phone, and navigate to specific locations within the hierarchy by using the phone's keypad.

Voice services are analogous in many ways to websites; however, the hierarchical, audio-based nature of the technology means that the interactions supported by a text- and media-based service over an internet connection are not yet possible with voice services over a phone line. For example, an entirely audio-based infrastructure makes it very difficult to, say, get an overview of a voice space, or navigate through disparate content regions to find the desired information.

Printed media, on the other hand, is entirely visual, providing simple overviews and navigation, but presenting clear barriers with regards to textual literacy. We saw an opportunity to connect the voice and printed media spaces in a novel way, taking advantage of the prevalence of cameraphones, and connecting them to telephone-based audio information services to create interactive audio photos.

Our approach, therefore, is a combination of IVR services and physical objects, providing users with a visual method of accessing a currently audio-only service. The design allows users to take photographs of physical objects—for example, newspapers, flyers or other text-based media—and use the photo itself as a canvas for interacting with a related IVR system (as illustrated in Fig. 1). When a photo is taken, the system automatically dials the remote telephone-based voice service. Once connected, users can then touch specific areas of interest on their own photo— titles, captions, adverts or images, for example—to hear audio narration for their selection. This interaction essentially gives the impression that the system is 'reading' the text under the touch point of the user.

The technique provides a fine level of touch granularity, allowing both broad and more nuanced interaction where appropriate. The novel design uses precisely placed QR codes as reference points, allowing it to transmit the user's touch coordinates via DTMF tones over a standard phone line, and ensuring that the service can be used without an internet connection in the contexts we describe.

Fig. 1. Interacting with the AudioCanvas system. From left to right: the user takes a photo of an item (image 1), which then becomes a canvas to interact with audio content related to the object (image 2). Audio content is provided over the telephone system (image 3); no data connection or application updates are required.

1.1. Motivations

Our motivation for creating AudioCanvas stems from observed and reported difficulties in locating information within IVR services. As such systems are entirely audio-based, hierarchical in nature, and often recorded in a host of different languages and regional dialects, developing usable search and browsing functions has proved particularly challenging for both researchers and service providers alike. It has also been reported that many lower-literate users struggle to understand hierarchical menus [Medhi et al. 2011; Medhi et al. 2013], as well as many literate but novice users [Walton et al. 2002].

Despite this, however, some headway has been made in creating low-level tools for IVR systems. For example, the *T-Web browser* [Agarwal et al. 2008] provides a subset of internet-like functionality by maintaining a history of current user sessions and storing users' bookmarks of 'Voice Sites'. Other research programmes, such as *TapBack* [Robinson et al. 2011] aimed to support more high-end interactions on relatively basic handsets by using back-of-device taps to control navigation through IVR content. Similarly, the *ACQR* approach [Pearson et al. 2013] used a novel audio-based technique to facilitate the sharing of voice service positions between users by mimicking the way URLs can be easily shared on the internet.

Efforts to visualise IVR content such as the work of Joshi et al. [2012], or to combine speech with graphical output (e.g., [Cuendet et al. 2013]) have also been proposed, but to our knowledge there have been no previous attempts to combine paper-based media with IVR content. Our goal in designing AudioCanvas, then, was to enable these widely-used audio systems to be enhanced not only in interactivity, but in their simplicity.

This article is an extended follow-up to our earlier investigations into combining IVR and paper documents. We have previously described the core AudioCanvas concept, including some elements of the studies in Mumbai and Langa [Robinson et al. 2014a; Pearson et al. 2015]. In this article we present a greater amount of detail regarding the user evaluations performed in India and South Africa, along with an additional study performed in the UK, and also discuss how AudioCanvas documents could be created. We also investigate further uses of the technique in related contexts, including an exploration of the system with traditional Indian storytellers, discuss in more depth the impact of the approach, and reflect on the benefits and limitations of cross-site user evaluations.

1.2. Novelty

We believe that the AudioCanvas approach offers several novel aspects over other methods of linking audio to physical documents. The system:

Does not require an internet connection. Our approach of transferring document and positional data via DTMF over a standard phone connection ensures that no internet connectivity is required. This is potentially a major advantage for users in resource-constrained regions where data connections are often unavailable, unreliable or simply unaffordable.

Does not require specialist hardware. Unlike many other approaches that aim to link physical documents to digital content (e.g., [Wellner 1993; Dymetman and Copperman 1998]), our technique requires no specialist hardware (such as mounted projectors or cameras, dedicated reading docks and so on). The AudioCanvas technique could feasibly be used with any phone that has an on-board camera.

Provides client independence. A major advantage of our approach is that the client taking photos is completely independent of the media with which it interacts. This independence—particularly appropriate in the regions we have been designing in—means that no updates or server connections are required to support additional items and, for example, a new edition of a newspaper could be scanned each day using the same client application.

Offers a very fine level of granularity. Other methods of directly linking a physical document to their digital counterparts (e.g., via a single QR code or image recognition) do so in a relatively coarse manner, usually linking a single anchor to an entire resource or page of content. AudioCanvas allows users to select specific pieces of information within a document—such as a sentence within a paragraph—to gain a better understanding of the document itself.

Allows modification of the underlying document. By using two QR codes to calibrate the image, we avoid image processing within the core of the document itself. In fact, as long as the two QR codes remain intact, changes can be made to the underlying document over time, making the design a useful tool for other applications such as storytelling (see Section 6).

1.3. Overview

In the rest of this article we begin by examining related work, before describing in detail the design and implementation of our fully-working prototype. We then move to describe three cross-continent evaluations showing the value of the approach for users with varying levels of textual literacy, smartphone familiarity and access to affordable data-connections. Following this, we describe an additional tool that was created to facilitate canvas creation, and complement this with further use-case scenarios. In the final study of this paper, we explore the application of the system as a low-tech digital storytelling tool, evaluating its potential use with a pair of traditional Kaavad storytellers. We conclude by reflecting on and summarising the findings, with a particular focus on the important issues of trust within systems such as ours, as well as considerations regarding study designs when working in resource-constrained environments.

2. BACKGROUND

A significant amount of previous work has focused on augmenting physical artefacts with digital media. One way in which researchers have approached the notion of merging physical interactions with digital enhancements is by creating desk spaces which are then digitally augmented. An early example of this was the *DigitalDesk* [Wellner 1993], a tangible physical desk with additional digital interaction. The DigitalDesk aimed to

retain the affordances of paper while allowing documents to be manipulated digitally via precisely positioned projectors and cameras. By carefully aligning the physical book on the desk, the system was able to read the paper documents placed upon it, and monitor user interaction with pens or fingers.

Other examples of this type of interaction include Arai et al.'s *InteractiveDesk* [1995], which automatically recognised physical items and linked them directly to electronic files, and Koike and Kobayashi's *EnhancedDesk* [1998], which included real-time manipulation of projected information. More recent desk-based physical-digital amalgamations can capture gestures made by users, either via standard video input (e.g., *WikiTui* [Wu et al. 2008]) or by using depth cameras and fiducial markers (e.g., *QOOK* [Zhao et al. 2013]).

More flexible camera-based approaches include that of Mistry et al. [2009], who created a pico projector system which supported querying everyday physical items for information by framing them with coloured finger tags; and, *d-touch* [Costanza et al. 2010], which used a webcam for the recognition and tracking of fiducial markers to create a simple tangible audio interface, where the positions of markers under a camera were reflected in the sounds that were played. The d-touch approach used a similar calibration system to our technique (i.e., markers placed in the corners of the interactive surface), but used these as calibration areas for a fixed workspace, instead of as alignment points for photographs of objects.

2.1. Digital paper

In addition to these approaches, research has also been conducted into connecting paper pages with digital media. More interactive than straightforward optical character recognition (OCR), these systems aim to create smart links between paper and digital content. One such example of this is the *Intelligent Paper* interface [Dymetman and Copperman 1998] – an early forerunner of the popular commercial Anoto pen.[2] The system supported collaboration and synchronisation between physical and digital documents by allowing communication between a paper page and a digital peripheral. This interaction was accomplished by means of a printed page identifier on the physical sheet, and a pointer input device in the form of a pen that recorded coordinates when pressed to the surface.

Other work has investigated how traditional methods of interaction could be replicated digitally (e.g., [Lucero et al. 2011]), or, how augmentation of physical objects could provide additional information—see, for example, the audiophotography work of Frohlich [2004], or the many commercial mobile apps that are focused around adding digital content to physical items.

Since the DigitalDesk (cf. [Wellner 1993]), which used a projector and camera above a desk, more portable designs (such as *XLibris* [Schilit et al. 1998]) have allowed users to scan paper documents and mark them digitally with freeform annotations on a tablet-like device. Later approaches, such as that of Guimbretiere [2003], lowered the hardware requirements for digital-physical annotation systems, making use of Anoto-marked paper to support notes that could be replicated digitally. Guimbretiere aimed to allow cycles between digital and physical documents, arguing for cohabitation of the two forms of media, rather than replacement of one with the other. We had similar motivations in the design of our technique, aiming to allow cohabitation of (and collaboration around) digital annotations and physical media. Our approach was to use two QR code markers in opposing corners of the document area that identify and align it in a photograph taken of the item. Previous work has used photographs of paper documents in similar ways to extend document interactivity. For example, Seifert et al.

[2]See: anoto.com

[2011] turned photographs of interface designs into interactive prototypes, and Erol et al. [2008] used image recognition of a document and comparison to a ground truth version to detect the regions of a page in a photograph.

Liao et al. [2005] created a gesture-based command interface that facilitated digital document manipulation using paper print-outs as proxies. Their system, known as *PapierCraft*, used a digital pen that allowed users to annotate and create command gestures (including operation indications such as copying and pasting areas) on printed documents, and synchronise these to a customised digital viewer.

Turning to commercial systems, the *LeapPad*[3] was a popular early interactive electronic children's book that featured a touch sensitive tablet and a paper book overlay. This education-focused platform allowed physical books to be read to the user via an interactive 'magic pen' that worked in the same way as a hand-held mouse pointer. The LeapPad design was superseded by the *Tag Reading System*, which uses optical recognition similar to that of the Anoto pen to read uniquely identified dot patterns in specially made books. Using this system, children are able to run their *Tag* pen across a physical book while listening to the corresponding audio book being read digitally.

Other approaches to digital paper, such as *Overlay*[4], strive to combine physical media and audio information by placing paper documents over a standard iPad. This application allows users to use and mark-up a personalised paper document, but also retrieve audio information by tapping on the iPad through the paper laid on top of it.

Unlike AudioCanvas, these examples focus on supporting *direct* interactions with media – manipulating physical objects to control digital interactions. We take a different approach, letting people use their own photos of items to act as canvases to interact with remote audio content.

2.2. Code linking

A simpler approach to connecting physical and digital spaces is by creating a direct link from printed codes. One such example of this is the *Collect* system [O'Hara et al. 2007], which investigated using first-generation cameraphones to photograph 2D barcodes placed on exhibits in a zoo in order to retrieve additional digital content.

A more primitive, yet arguably equally effective method of linking physical documents to their digital equivalents is via short numerical codes that can be entered by the user into a form or other system. Popular in the past for information-on-demand services (which required writing a code on a mail-in form), more recent uses of this technique include marketing materials such as holiday brochures[5], which occasionally make use of this technique to allow their customers quick access to the online booking information for the exact excursion that has caught their interest.

In recent years, a more common approach has been to use QR codes as a way of promoting links and allowing quick sharing of information. One Welsh town has even gone as far as marking-up the majority of its historical buildings and interesting places with QR codes, claiming the title of the world's first 'Wikipedia town' [Warman 2012]. Scanning QR codes, however, typically takes the user to a single specific piece of information and cannot easily offer multiple results without repeated scans or subsequent menus, or provide a particularly fine level of granularity.

There are many mobile augmented reality applications that are able to add overlays or web links to physical artefacts via digital codes or other recognition methods. *Daqri*[6], for example, is an augmented reality application that uses QR codes for tracking; others, such as *Blippar*[7] or *Shortcut*[8], have used image recognition. However, these

[3]Produced by LeapFrog Enterprises (leapfrog.com) from 1999 to 2008 [4]See: overlayapp.com
[5]See, for example: thomson.co.uk [6]See: daqri.com [7]See: blippar.com [8]See: kooaba-shortcut.com

approaches require a data connection and a large upload and download for each interaction. Furthermore, they are entirely focused around augmenting a photo or camera view with visual digital content. Our system uses QR codes to encode a coordinate grid and telephone number on a photographed object. Our aim was to add audio to the experience—without requiring an internet connection—rather than just providing a quick way to enter a website URL. While these previous approaches have required either a *client-side* metadata database or realtime internet access for their functionality, AudioCanvas requires neither, relying instead on a standard telephone connection for audio content delivery.

Previous work in rural developing contexts—the focus for our work—has used QR codes to partially automate tasks such as form-filling. Parikh et al. [2006], for example, allowed users to scan one code at a time to accurately enter form data. Their approach required data to be preloaded into the application, however, unlike our artefact-independent design. Nontheless, Parikh et al.'s design guidelines did highlight the need for linking a mobile device action to a physical process, and found that audio feedback was important for this type of interaction.

Smith and Marsden's Snap 'n' Grab system [2011] allowed riders of South African minibus taxis access to multimedia via photographs. Their prototype used barcodes around a printed icon, which was photographed and then sent to a server phone (also situated in the taxi) to request relevant media. Media was then sent back via Bluetooth. Our approach does not require a local server, as all communication is via a telephone connection to a remote service. We also do not focus on transferring audio content to users, but rather on allowing people to use their own photos to navigate through content from an interactive voice service.

2.3. Annotating with audio

In this work we focus on linking paper documents with remotely-provided audio annotations. Previous work around audio annotations ranges from immersive audio story books (e.g., [Back et al. 2001]) to the audio greetings cards and interactive picture books that have been commercially available for many years. These designs commonly use buttons or basic sensors to start sound playback. Various research approaches have been taken to synchronising audio recordings with physical documents automatically. The *Audio Notebook* [Stifelman et al. 2001], for example, was a custom hardware tablet that allowed users to take paper notes and record audio simultaneously, then skim-review later, referencing the correct recording position from the notes. Erol et al. [2004] took a similar approach to synchronising a slideshow presentation with notes made on a handout – barcodes on the printout automatically linked notes with the correct positions in a later video of the talk.

Our approach uses photos of paper items to link them with digital annotations. AudioCanvas is clearly related to audiophotography (explored in depth by Frohlich [2004]), which is the general area of associating audio with photos. Implementations of audiophotography have ranged from adding short audio transcripts to digital photos on-camera (cf. [Frohlich 2004]), to using overhead image recognition for selecting and browsing audio associated with printed photos (e.g., [Frohlich et al. 2004]). Audiophotographs traditionally associate only one annotation with each photograph, however (e.g., each photo has a single audio annotation); and, more importantly, the audio track is associated with the photograph that has been taken, rather than the object pictured in the photo.

More similar to our approach, then, are systems that link audio annotations with specific places on physical documents, or specific actions with objects. Pfleging et al. [2010], for example, demonstrated a photo-based technique for turning hand-drawn sketches into working interactive prototypes, and Suzuki et al. [2005] used personal photographs of household objects for control and interaction (e.g., take a picture of the

TV to turn it off or change the channel). However, our approach concentrates on audio as an interaction channel, rather than a conversion of photos into digital facsimiles or device controls.

Klemmer et al. [2003] created *Books with Voices*, using barcodes printed in the margins of paper history books to retrieve video interviews with historians talking about the material. West et al. [2007] used Anoto-marked paper in a scrapbook, supporting various user-drawn symbols to associate audio and other content with scrapbook items. Liu et al. [2010] had similar goals, but removed the need for Anoto paper, instead using photos of the page, and recognising preprinted marks that signified media the author had associated with the document. Our approach was originally adapted from that described in [Robinson et al. 2014a], which used digital markers to indicate the edges of the content area. However, our design does not focus solely on retrieving audio content from products and posters; instead, we aim to support both support document annotation and audio narration in a flexible, dynamic manner that allows both the document and annotations to be changed over time.

2.4. Annotation as storytelling

As discussed towards the end of this article, we believe the AudioCanvas design shows strong potential for interactive, evolving physical-digital storytelling. Previous work in this area includes systems such as *KidPad* [Druin et al. 1997], which extended existing sketching software to create a storytelling environment, or *PicoTales* [Robinson et al. 2012], which supported group-based collaborative sketching via gestures and pico projectors. Unlike traditional digital storytelling (which has focused around creating short self-narrated digital films), or these previous approaches, AudioCanvas lets physical sketches or other objects and digital elements coexist as part of the narrative.

Cao et al.'s *TellTable* [2010] supported sketch-based storytelling on an interactive table, aiming to encourage incorporating physical objects into stories via 'capture tools' that inserted photos directly on to the story surface. Other approaches have included those of Jacoby and Buechley [2013], who used conductive ink to allow sketches to be augmented with digital content; de Lima et al. [2014], who used pen and paper sketches to insert characters into a virtual world; or, Wood et al. [2014], whose barcode scanning approach linked digital tales to physical books. Yeh et al.'s approach [2006] aimed to produce a merged digital version of both Anoto-marked paper notebooks and related photos. More similar to our design is the approach of Raffle et al. [2007], who used a drawing tablet and 'stamp,' to associate audio recordings with sketches on the page. However, their design required a custom hardware package, and supported only local playback of the content. AudioCanvas can be used on standard cameraphones, and uses a remote IVR system for audio capture and storage, allowing flexible audio annotation that is accessible by any user who takes a photo of the associated physical counterpart.

3. AUDIOCANVAS

AudioCanvas affords rich experiences with physical objects, linking users' photos directly to audio content from a telephone-based service. Using our prototype, any marked-up object becomes a touch panel to interact with audio resources, as illustrated in Fig. 1 (see Page 3) and Fig. 2 (below). Our approach uses precisely placed QR codes on printed media (flyers, newspapers, invoices, packaging and so on) to detect the position of the user's selections on their photo of an item. To interact with the system, the user positions their cameraphone to take a photo of an object. When both codes are detected, a photo is taken automatically, and the phone dials the voice service in the background. Selecting any region on the photo (with fingers on touchscreens or the joypad on a featurephone) sends the coordinates to the service, which plays back the appropriate audio immediately, removing the need for manual IVR navigation.

Fig. 2. Using AudioCanvas. Left (clockwise from upper left): translating cooking instructions; listening to a spoken version of a newspaper; discovering more information from product packaging; finding out about a movie from its poster. Right: the interaction afforded – touching the photo to hear the audio associated with any section.

We designed AudioCanvas in and primarily for use in regions where internet access is sparse, and where low levels of textual literacy prevent people from reading printed media. It has previously been reported [Medhi et al. 2011] that common text-based interfaces are often completely unusable for lower-literate users, and error prone for literate but novice users, whereas graphical interfaces with audio output are more successful. The AudioCanvas design clearly resonates with Medhi et al.'s [2009] design guidelines for interfaces for non-literate and semi-literate users, and was built with three key goals in mind:

Design Goal 1 (DG1). Allow direct interaction with photographs of physical objects to receive audio-based captioning on demand. This approach allows people to, for example, capture a poster they cannot read to hear to an audio version or a translation, or take a photo of a product to find out about the provenance of its ingredients.

Design Goal 2 (DG2). Provide the service via a standard phone line, ensuring that it can be used where mobile internet is costly or unavailable. This approach also means that the system can be used as a complement to existing voice-based services, such as the Spoken Web [Kumar et al. 2007], Awaaz.De [Patel et al. 2011] or other alternatives.

Design Goal 3 (DG3). Ensure that the AudioCanvas client application is independent of any media it interacts with, and specific physical objects are not 'hard-coded'. This approach allows the system to provide new content without requiring potentially costly (in terms of data usage) application updates, and also acknowledges the fact that many users never update apps once installed.

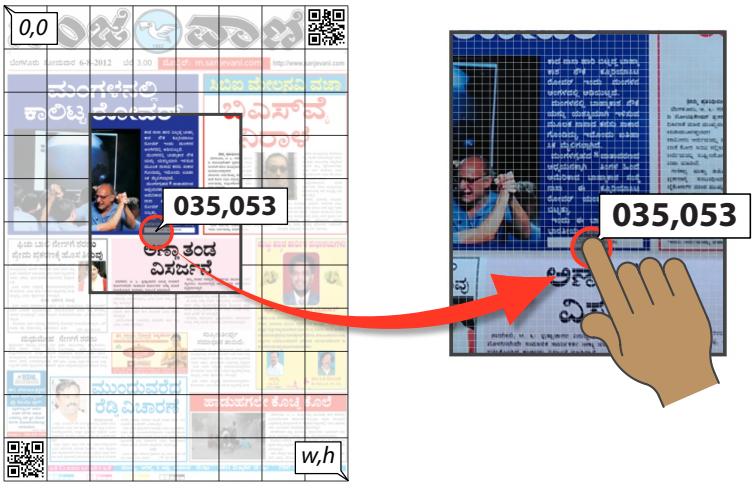


Fig. 3. The AudioCanvas coordinate system. Left: coordinates within the photo that the user takes are calculated automatically, based on the distance between the two QR codes, and their sizes. Right: the resolution of the coordinate grid (overlaid) is high enough to allow zooming and precise selection. Touching a point plays the relevant audio (see Fig. 2). Taps outside the image region pause the audio.

Figure 2 shows a close-up view of the audio browsing interaction afforded, and the prototype in use in several anticipated scenarios. Each of the media items we have developed contains multiple audio 'hotspots,' which are linked to different information based on where the user taps. For example, each area of the food packet (upper left image in Fig. 2) provides a translation of the original text. The other media items support similar functions – translating, speaking text that is too small to read, providing more information than is available in the packaging space, or saying a person's name when their face is touched. This information could, we argue, be beneficial to users who cannot read a piece of text—whether it be due to language or literacy issues—when an internet connection is not available.

3.1. Implementation

There are two components to the AudioCanvas system: a local client (which we focus on in this article), and a remote voice service. The local client is a mobile phone application that is used to take a photograph of an object, allow panning, zooming and selection, and help the user interact with the voice service (*DG1*). The remote voice service is a standard IVR system, where DTMF (i.e., phone keypad) tones over a phone line control the resultant audio interaction (*DG2*).

Our novel client design uses two separate QR codes on printed media to detect the position of the interactive object within a photo taken by the user. The codes are positioned at opposite corners of the object – one at the bottom left and another at the top right. The bottom left code contains the telephone number of the interactive voice service, and an identifier for the item (e.g., the issue and page number of a newspaper, for example). The top right code is used for coordinate calibration and image alignment (we automatically straighten and skew correct the image).

Figure 3 shows the standard layout of marked-up objects, and illustrates how these translate into touch coordinates. The coordinate system that is used for an object is calculated based on the distances between the two QR codes, and their sizes, as shown in Fig. 3. This system allows the exact position of a touch to be converted to a known location on the physical object. That is, a user can touch any position on the object within

Table I. Study participant demographics

Location	Textual literacy	Familiarity with smartphones	Access to affordable data connections
UK	Literate	High	High
South Africa	Semi-literate	Low	Low
India	Illiterate	None	None

their photo, and the client can convert this to a coordinate without requiring knowledge of what the object is (*DG3*). This allows almost any object to be used for interaction, with the system able to calculate the exact coordinates within the media automatically.

When a user touches the picture they have taken, the coordinate of the current touch point is sent as six DTMF tones to the remote service. The coordinate is matched to a database of audio on the server and the relevant audio is played in response. The latency of each touch action—that is, the time taken for the DTMF tone to be generated and the resulting audio to be played—is less than a second. The DTMF sound is intentionally audible to the user, acting as a form of feedback to them to indicate that an action is being performed. The client is not tied to any particular remote service – the phone number of the service to dial and the current object's identifier are indicated by the lower left QR code on the item.

4. EVALUATIONS

To test the AudioCanvas technique, we developed a range of prototype media items (see Fig. 2, above, and Figs. 4, 5 and 7, below), and evaluated them using AudioCanvas on an Android smartphone. We conducted user evaluations in three separate regions, aiming to test over a range of literacies, both textual and technological. Table I shows the broad demographics of the participants recruited for each study.

The first study involved UK participants who were familiar with smartphones, but unable to read the languages of two of the printed media items we provided (Kannada and Korean). In this context, mobile internet connections are common, and the vast majority of users have a very high level of textual literacy in their native language (English). Although literate technology-competent participants did not represent the eventual intended audience for the system, we envisaged certain situations where these users could make use of the technique. For example, literate users may want to use AudioCanvas to translate documents or packaging printed in other languages, particularly when abroad or roaming, or when data packages come at a higher cost than calls.

The second study involved South African township residents. Participants in this evaluation had some level of literacy in isiXhosa (the most widespread spoken language locally), but none in English, which is the primary language used on many local newspapers and packaging (and the language we used for prototype media in the study). In this context, familiarity with phones—particularly touch-screens—is less common, and there is a higher attraction to using a telephone service (rather than a data connection) for audio retrieval. In this study, we were most interested in how literacy levels might affect participants' ability to locate audio hotspots on media items, and whether participants saw value in the system. We anticipated issues relating to the use of a smartphone for interaction, as none of the participants had used smartphones before, but this was not a focus of the study.

Finally, the third study tested the system with rural Indian participants who were largely illiterate and unfamiliar with using mobile devices. In this study we used items written in both English (often used for official documents) and Kannada (a widely-spoken language locally), but participants could read neither. In this context, the fact

that AudioCanvas operates over a standard phone line was crucial, as the availability of data connections is low and their cost at the time put them out of reach. Our aims in this study were similar to those in the South African evaluation, focusing on participants' perception of the system's potential, and not on smartphone usage issues.

The prototype media items used in the studies were modified examples of real documents, packaging and other items. The language(s) used on the printed media varied depending on where the study was being conducted (see below, and Figs. 4, 5 and 7), but audio was always given in participants' native language. Media items were of varying physical sizes (from 8×12 cm to 29×40 cm), aiming to encourage participants to experiment with all aspects of the system, including framing the item for the automatic photograph, panning and zooming around the image, and finding and listening to audio hotspots. In all studies we made it clear that the underlying interface was driven by an IVR system and would therefore cost the same as a normal phone call while listening to audio, and 'tie-up' the phone line for the duration of any use.

It is important to note that none of the prototype objects used in the studies contained any additional indications regarding the locations of audio hotspots. That is, with the exception of the two QR codes we added, the original media remained unchanged, and participants had to use their own intuition to discover where the audio could be found within the media. We did not outline or highlight the interactive areas on the system because we wanted to encourage exploration without the image being obscured by bounding boxes or other markers. This approach also ensured that the communication was only one way, allowing a universal client to interact with a remote server without the need for the server to send coordinates or updates to the client.

Further details about the study procedures and media items used in each evaluation are given in each of the following sections.

4.1. User study 1: Swansea, United Kingdom

The first of our three user evaluations was a lab-based study undertaken in Swansea, United Kingdom. We recruited 22 participants (8M, 14F, aged 20–62) for a 20 min trial of the AudioCanvas system. These participants were literate in English but had no knowledge of Korean or Kannada (the two languages used on prototype media items, in addition to English). They were also technologically literate – all participants owned a mobile phone and had previously used a smartphone.

4.1.1. Media. We created AudioCanvas versions of four media items in three different languages for use during the study (see Fig. 4): a box of tea (English); a ready-meal curry packet (Korean); a daily newspaper (Kannada); and, a movie poster (Kannada). All audio given was in English – the native language spoken by all participants.

We used physical media in both Kannada and Korean to simulate an inability to read text in an unfamiliar language. The box of tea was used as a further probe, where the AudioCanvas audio provided additional information not found on the original artefact. For example, when a user touches the company logo on an object which is in a language they are able to read natively, they might receive an audio description of the history of the company – information that is not available on the item itself. We hoped that this functionality would prove to be a beneficial additional use of the system, and aimed to gather user opinions about the feature during the study.

4.1.2. Procedure. Our main goal for this study was to allow participants to experiment with the system, interacting with various different types of media. As such, the task set used was minimal, asking participants to explore and locate simple information such as the title of a newspaper, or the ingredients in a product. We did not consider task timings to be a relevant metric for determining the success of the interface, as

Fig. 4. Media items used in the UK evaluation. From left: a box of tea (English); a ready-meal curry packet (Korean); a daily newspaper (Kannada); and, a movie poster (Kannada). All audio was given in English. Items are shown at the correct relative physical sizes.

audio is always played immediately after the user touches the photo. Instead, we asked participants to rate several aspects of the system, and gathered qualitative comments.

We also verified with participants that they were not able to read the two non-English languages used. This ensured that they could not gain additional knowledge via the text elements of the media, other than their size and position. Participants had to, in these cases, use their own intuition to discover the audio hotspots within the media.

The procedure of the study was as follows:

(1) After a short briefing and informed consent procedure, the AudioCanvas system was demonstrated to each participant;

(2) Participants then completed three information-finding tasks on each example object, and were also allowed time to explore each item further if they wished. Typical tasks given to the users during the study were:
— *What is the title of the newspaper?*
— *What are the main ingredients of the curry?*
— *Which actress starred in the movie?*
— *Where is the tea sourced from?*

(3) Following this, each participant rated the system on a scale of 1–10 (10 being highest) in terms of usefulness and likelihood of them making use of the system, and on a Likert-like scale from 1 (*extremely easy*) to 7 (*extremely difficult*) on five factors related to ease of use (see Table II);

(4) Finally, a short semi-structured interview was conducted to gather qualitative data from participants about their use of the system. Upon completion of the study, each was given a £5 (approx. $8) gift voucher as a token of our appreciation.

4.1.3. Results. All participants highly enjoyed using the system, and most asked to be able to install it on their own phones. Participants rated the usefulness of the system (out of 10) with an average of 8.2 (s.d. 1.15). The likelihood of participants using the system themselves (also out of 10) was given an average score of 8.1 (s.d. 1.65). Scores and comments given by participants indicated a strong appreciation of the AudioCanvas

Table II. UK study participants' ratings (1–7 Likert-like scale; 7 high).

Ease of use, in terms of	**Average rating (s.d.)**
Focusing the camera and taking the initial photo	6.1 (0.6)
Panning and zooming around the photo	6.1 (0.8)
Selecting specific points of interest within the photo	6.1 (0.8)
Getting the right information back from each section	6.3 (0.5)
Finding the audio hotspots in a photo	5.0 (1.0)
Combined average	**5.9**

design.' For example: *"I thought it was very easy, and a really good idea"* and *"I could definitely see myself using this if it was an app."*

Participants' responses for ease of use were highly positive, with the average for four of the five factors in excess of 6 out of 7 (see Table II). The final criterion, concerning locating audio hotspots within a photo, was ranked slightly lower on average (5.0). This result was somewhat to be expected, however, as there were no visual indications of where audio could be found on any of our prototype artefacts.

Observations during the studies illustrated distinctions between users' ability to locate audio hotspots on different types of media. With the newspaper, for example, participants could generally locate audio content easily by clicking on textual headings or paragraphs, or photographs. For the tea packet, which was written in English, participants tended to read the text before deciding where to touch on the photo; as a result they took longer to find audio hotspots located over photos or logos (having taken time to read the text to search for cues first). In the cases where participants could not read the language on the artefact, they tended to select text before images. This preference depended on context, however, and in some cases, when attempting to locate specific pieces of information, participants ignored non-text content entirely. When asked to locate the actors' names on the movie poster, for example, (achieved by touching the actors' images) only one participant selected an image of a person at their first attempt. Another participant chose to select an actor's image only after touching every text item failed to provide them with the information they required. This finding suggests that the placement of audio on these documents needs to be carefully thought out; or, that items should hint towards audio-annotated areas. It is interesting to note, however, that several participants mentioned the enjoyment of exploring the system to find the right audio location – annotations or hints would need to take this into account, and also bear in mind the simple lightweight integration goals of the system.

When asked about the potential uses of the interface, many participants suggested relatively straightforward use cases, such as when travelling abroad (i.e., reading foreign text) which is closely related to our motivating scenarios for the system. This common sentiment was illustrated by comments such as: *"translation would be invaluable when going abroad on holiday for more information about products or food,"* and *"this would be amazing for a holiday, trying to navigate around foreign transport systems, [or] translating menus."* Several users thought that the system would also be *"really useful for learning languages."* Others identified the benefit of the system as a method of providing audio descriptions of visual content for those with poor eyesight. One participant stated, for example: *"very useful, especially [for] somebody slightly older – getting audio rather than visual is a lot more appealing for older people,"* while another said: *"it'd be useful in supermarkets – when your eyesight starts to deteriorate you can't see ingredients or nutritional information, for example."*

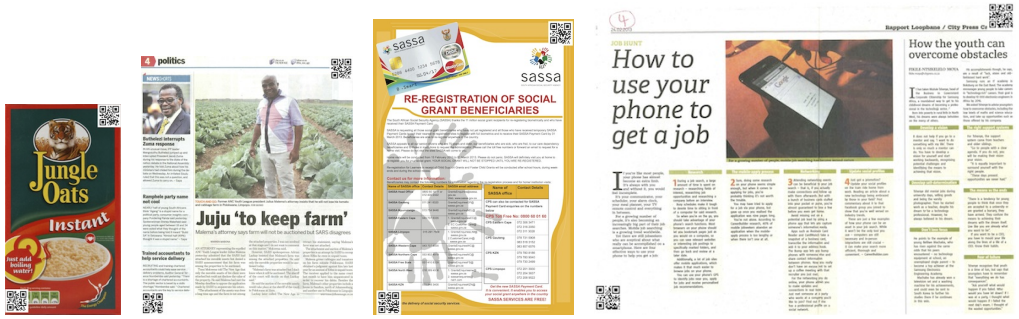


Fig. 5. Example media items used in the South African evaluation. From left: a box of oats; a daily newspaper; a social security information leaflet; and, a weekend magazine article. All items were written in English, with audio given in isiXhosa. Items are shown at the correct relative physical sizes.

4.2. User study 2: Langa, South Africa

The second AudioCanvas study took place in Langa, a township near Cape Town in South Africa. We recruited 36 participants (6M, 30F, aged 25–45) for five hour-long group-based trials of the AudioCanvas system. Groups consisted of 6 or 12 participants (four groups of 6 and one group of 12), and each group was given five AudioCanvas phones for the session which were freely available for any of the participants to use. All participants in the study were able to get hands-on experience with the AudioCanvas system during the sessions.

The participants recruited for the study were fluent speakers of isiXhosa (the most common local language), but could not speak English. Participants were also semiliterate in isiXhosa, but could not read or write English. English-language newspapers and posters are common in the area, however, and are therefore inaccessible to these participants. Data connections in the area can be intermittent and priced out of reach of the participants in this study. The majority of the participants owned a low-end handset, and only one had previously used a smartphone, which had been borrowed from a friend for less than a day.

4.2.1. Media. We used a larger collection of media items during this group-based evaluation. The set of items included: a box of oats; several newspapers; a magazine article; an information leaflet; an advert flyer; and, an advice notice board (see Fig. 5 for four representative examples). These items were predominately written in English; all audio given by the AudioCanvas system was spoken in isiXhosa.

4.2.2. Procedure. We worked with a local researcher—fluent in both isiXhosa and English—to facilitate the evaluation. This researcher was a trusted and respected figure in the township area, and both recruited participants for the study and hosted the study sessions. During the study, she acted as a mediator between the authors and the participants, and helped explain the system's purpose and its operation.

The format of each of the group evaluations was as follows:

(1) After a short briefing and informed consent procedure, we demonstrated to the group the basic functionality of the smartphones they would be using;

(2) Following this, AudioCanvas was demonstrated, and its purpose and use cases explained;

(3) Once every participant was familiar with the system's usage, we handed out several copies of each prototype media item and asked participants to experiment with the system. No specific tasks were given – participants were encouraged to take photos and discover the information on the objects by touching in different locations. We

Fig. 6. Performing the study in Langa, a township in Cape Town, South Africa.

did not explicitly ask participants to discover specific pieces of information as we did in the UK study, as our goal was to elicit natural behaviour. The same method was applied to the study conducted in India (described in the next section);

(4) Following this, each participant rated the system on a scale of 1–10 (10 being highest) in terms of usefulness. We then conducted a semi-structured focus group interview, covering general questions about the AudioCanvas interface, its functionality, and participants' thoughts on the idea. Upon completion of the study, each participant was given R50 (approx. $5) as a token of our appreciation.

4.2.3. Results. We began the interview session by asking each group to tell us about occasions within the last month when they could not read something that they needed to be able to read. The discussions that followed this question typically resulted in stories from participants about job adverts or newspaper articles that they were unable to read; and, sometimes, more serious cases that had required immediate attention (such as financial difficulties). In many of these situations, participants confirmed that they either asked someone else (e.g., a friend or family member) to assist them, or were not able to read the item at all. One participant, for example, had a high enough level of textual literacy in isiXhosa to understand that a letter he had received had been a final bill notice, but was unable to read the bulk of text, and had to ask for help from a family member: *"I got a receipt slip through the post which said 'final final' on it, but I couldn't read it so my daughter had to help me"*. Another participant had needed to fill out a confidential form, but could not read the document in its entirety: *"I had to fill in the form and I didn't understand all of it"*.

These comments illustrated the problems that people often face when trying to interpret text-based media that they are unable to understand. Our aim with AudioCanvas

was to provide this help in a way that would be usable and accessible to these individuals. Firstly, then, we wanted to find out the extent to which these participants felt the system was suitable for their needs. The average score given for usefulness was 8.4 out of 10 (s.d. 2.07), showing a high appreciation of the AudioCanvas method. Participants also gave favourable comments about the system: *"I can see this as something that could really help me and empower people," "for me it's very easy and there's nothing I don't like about it,"* and, *"I like this voice coming out of the phone – I wish all phones could do this"*.

As we anticipated, participants in this study had more trouble using smartphones than those in the UK trials. Specifically, they found it difficult to position the camera in the correct way to take a picture of the entire object that incorporated both QR codes. Although this was a problem initially, the learning curve was small, and by the end of the study all participants were able to frame the image and take the photograph without assistance. Generally, participants chose to take it in turns to hold the phone to take a photo, while also working in groups to listen to the audio provided (see Fig. 6). Despite the lack of smartphone familiarity, there were few problems with tapping, panning or zooming the screen after the initial photograph was taken, and all participants enjoyed using the system.

Participants in this study did not specifically mention any problems in locating audio hotspots. However, this may be due to the fact that we did not ask them to complete any specific comprehension tasks, unlike in the UK trials. In this study, participants used the system purely for experimentation, selecting blocks of text, titles or images of their own choosing. In many cases, participants were genuinely shocked that they were able to hear spoken versions of written text. For example, one participant stated: *"It's really the first time I have seen anything like this – hearing voices coming from the phone like this; it's great, but unheard of!"*; another said: *"I'm asking myself how an article in English can be read in isiXhosa!"*; and, another commented: *"[I am] fascinated by the isiXhosa sound coming out"*.

4.3. User study 3: Devarayanadurga, India

The third AudioCanvas study took place in a rural village near Devarayanadurga (Karnataka), India. The study was run by an independent and experienced local researcher, who we worked with to recruit 25 participants (14M, 11F, aged 22–65) for individual 30 min trials of the AudioCanvas system. Participants in this study were Kannada speaking, but had very low levels of textual literacy.

All except one said that they regularly found themselves in a situation where they wanted to access information from a document they could not read. In addition, general technology exposure in the area is very low. Participants were primarily farmers by occupation and owned (or shared) a low-end phone at most. Data connections in this region, despite being relatively cheap in comparison to UK and South African rates, are also prohibitively expensive for these participants.

4.3.1. Media. We used two media items during the study – a phone bill and a map of Bangalore. These items were suggested by the local researcher who conducted the evaluation, based on his previous experience and familiarity with the user group this work focuses on. In this study, both the phone bill and map were written in Kannada, but, as mentioned above, were inaccessible to participants due to literacy issues. The audio used in the study was spoken in Kannada.

4.3.2. Procedure. Participants took part individually, using the system to explore and experiment with each of the media items. As in the South African trial, the study began with a smartphone demonstration, involving a basic introduction to the phone's interface and general functionality, including touching, swiping and pinching.

The study procedure was as follows:

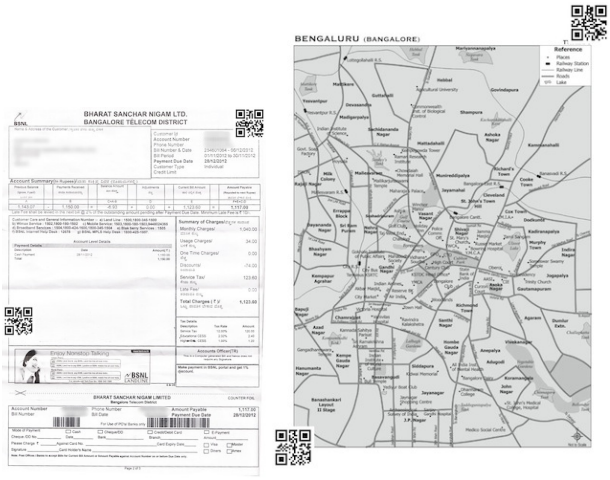


Fig. 7. The media used in the Indian evaluation. Left: a phone bill. Right: a map of Bangalore. Both items were written in English, with audio given in Kannada. Items are shown at the correct relative physical sizes.

(1) After a short briefing and consent procedure, the smartphone was demonstrated to the participant, followed by a demonstration of the AudioCanvas system;

(2) Following this, the participant was asked to use the system to take photos of the media items and interact with their audio content, exploring for as long as they wished. No specific tasks were given – we aimed to allow natural exploration, simulating real usage of the system. This part of the study took around 10 min per participant (though participants were allowed to continue using the system for as long as they wished).

(3) After using AudioCanvas, following the same method as the UK study, each participant then rated the system on a scale of 1–10 (10 being highest) in terms of usefulness and likelihood of them making use of the system, and on a Likert-like scale from 1 (*extremely easy*) to 7 (*extremely difficult*) on five factors related to ease of use (see Table III);

(4) Finally, a short semi-structured interview was conducted to collect qualitative data from participants about their use of the system. Following guidance by the facilitating researcher, no monetary incentive was given to participants for taking part in this study.

4.3.3. Results. Continuing the trend from the first two studies, participants enjoyed using the system and gave highly positive ratings. The usefulness of the system was rated at 8.1 out of 10 on average (s.d. 1.90), and the likelihood of participants using the system personally was rated at 8.2 (s.d. 2.46). Scores and comments regarding the system design and usability show its potential usefulness in resource-constrained regions where textual literacy levels are low. For example, comments included: *"It will be very useful for illiterate people, especially in our village," "It may be useful to me in learning English better," "It is easy to use once you get used to it,"* and *"I can access a lot of information which I otherwise could not have accessed."*

All participants suggested that AudioCanvas would be useful for people unable to read. Other scenarios were also given – for example, prior to the study all participants said that they regularly needed to ask others for help with reading; after using AudioCanvas, one noted its usefulness for helping with sensitive information, such as when filling in a private form.

Table III. Indian study participants' ratings (1–7 Likert-like scale; 7 high).

Ease of use, in terms of	**Average rating (s.d.)**
Focusing the camera and taking the initial photo	4.9 (2.0)
Panning and zooming around the photo	6.2 (1.3)
Selecting specific points of interest within the photo	6.3 (1.2)
Getting the right information back from each section	6.3 (1.2)
Finding the audio hotspots in a photo	5.1 (2.0)
Combined average	**5.8**

The overall ease of use ratings from this evaluation closely mirrored those from the UK study, and were largely positive, with the average for three of five scores in excess of 6 out of 7 (see Table III). As expected, however, the Indian participants who had no experience with smartphones gave lower scores for 'focusing the camera and taking the initial photo' than the more experienced UK users. This mirrors the observations from the South African trial. In this study it was clear that participants found it difficult to frame the image entirely within the camera's viewfinder, accommodating both QR codes in the photo. Many factors could have contributed to this, including perhaps the lack of familiarity with technology. Regardless of cause, this result suggests a need for further refinement of the photo capturing process.

Ratings given for the ease of locating audio hotspots were largely similar to the UK study – and generally lower, on average, than the other attributes. As previously, this result was to be expected, as there were no explicit visual indications of where audio could be found on either of the items. Participants found audio locations more difficult to locate on the map due to the lack of obvious areas for audio feedback. The phone bill, with its structured layout and more familiar context, was more engaging, and participants eagerly explored the image for annotations.

4.4. Discussion

Overall, throughout the three exploratory user studies performed with AudioCanvas, participants greatly appreciated the design and ease of use of the technique. They enjoyed using the system, and were keen to explore documents to find audio augmentations. The simple way of linking photos to audio was also seen as a clear benefit by participants.

One danger in studies like these, using the technology we chose within the contexts we are working, is that the novelty effect can overwhelm participants and skew the results in favour of the new approach. Particularly when we are using devices that are considered high-end for the majority of participants, attention can be focused on the phone hardware as opposed to the technique they are being asked to assess. To reduce this problem, we consistently explained to participants that they were providing feedback on the system itself and not the touch-screen device they were using; and, explained to them that the approach could also be used in lower-end feature phones with buttons in place of touch interaction. Is it worth noting, however, that this is a known problem within the ICTD community [Dell et al. 2012], and is a small but not overriding limitation of our approach.

Turning now to the use of the system, one area in which participants struggled was the requirement to frame both the media itself and two alignment QR codes in the same picture. Participants in the UK study, who were familiar with smartphones and taking digital photos, were easily able to set up the image correctly. Those who had less or no experience with cameraphones found this task trickier. While we expect this issue to be minimal in the long term (as seen in the South African trial, where participants learnt the technique very quickly over the course of the study), there is

perhaps scope for more cues on objects themselves about how to frame the photo, or a simple overlay on the camera screen.

Another area for potential improvement is that of cues for audio areas. Currently, the system is entirely independent of the media used, and requires no physical editing of the media except to add two alignment QR codes. Participants found it necessary to explore the image in order to discover where audio hotspots were located. While this is perhaps part of the fun of using the system, in some circumstances it may be beneficial to add visual cues on the physical media itself in order to indicate to users where audio may be found. While this would decrease the attractiveness of the system, reducing its lightweight-ness, another option could be for the mobile client itself to retrieve the coordinates of audio hotspots automatically when first dialling the audio service. In the same way as the client communicates the location of the touched area (e.g., via DTMF tones), the server could send the coordinates of the touchable areas on the current page, which could be highlighted on the client during interaction.

We did not include this functionality in the original design for three reasons. First, we wanted the client to be independent of the server – that is, the user's phone does not need to be aware of hotspot or document information, and therefore does not need to be constantly updated. Secondly, we wanted the interaction to be one-way only, avoiding the need for the server to pass the coordinates of the audio hotspots to the phone at the start of the interaction session. Finally, we chose not to show users where the hotspots were located as we wanted to elicit natural and exploratory behaviour during the evaluations.

A useful area of future work in this area could be to perform a comparison evaluation of the original system against a design with visual hotspots in order to investigate the differences in interaction between the two.

5. CANVAS CREATION

During trials of the AudioCanvas design in the three regions described above, it became clear that participants would appreciate using the system not only for retrieving static audio from printed media, but for *dynamic* content, too. That is, rather than audio being associated with a physical item at the time of creation (by its author), it could be added to items by users or community groups, or entirely crowdsourced. This approach also lends itself to further uses of the tool beyond audio annotation of fixed media. In order to investigate the potential for this usage, we developed a further extension of the tool to allow for the creation of AudioCanvas-augmented documents in situ.

To make an AudioCanvas item, a content creator first prints the two required QR codes to attach to the corners of the item. Alternatively, media can be created with these codes already attached before printing. Both scenarios are supported by our web-based media creation tool[9] (shown in Fig. 8). To create a new AudioCanvas object, the content creator can either open a pre-created background image, or select a paper size to begin from a blank page. The tool then creates a downloadable PDF with the appropriately sized QR codes included in the correct positions. Once the object has been augmented with its two QR codes, audio hotspots can be added, deleted or modified via any smartphone handset.

Figure 9 shows the AudioCanvas creation tool in operation, highlighting the steps taken during the document creation process. The creation mode of this new prototype works in broadly the same way as for browsing audio, with the first step being to take a photo of an object. In this new version of the tool, audio areas are visualised, and can be modified, moved or deleted by the user. Figure 9 (left) shows an item which already has several audio hotspots added (e.g., covering the words 'Instant' and 'Just add boiling water' in the image).

[9]See: enterise.info/codemaker

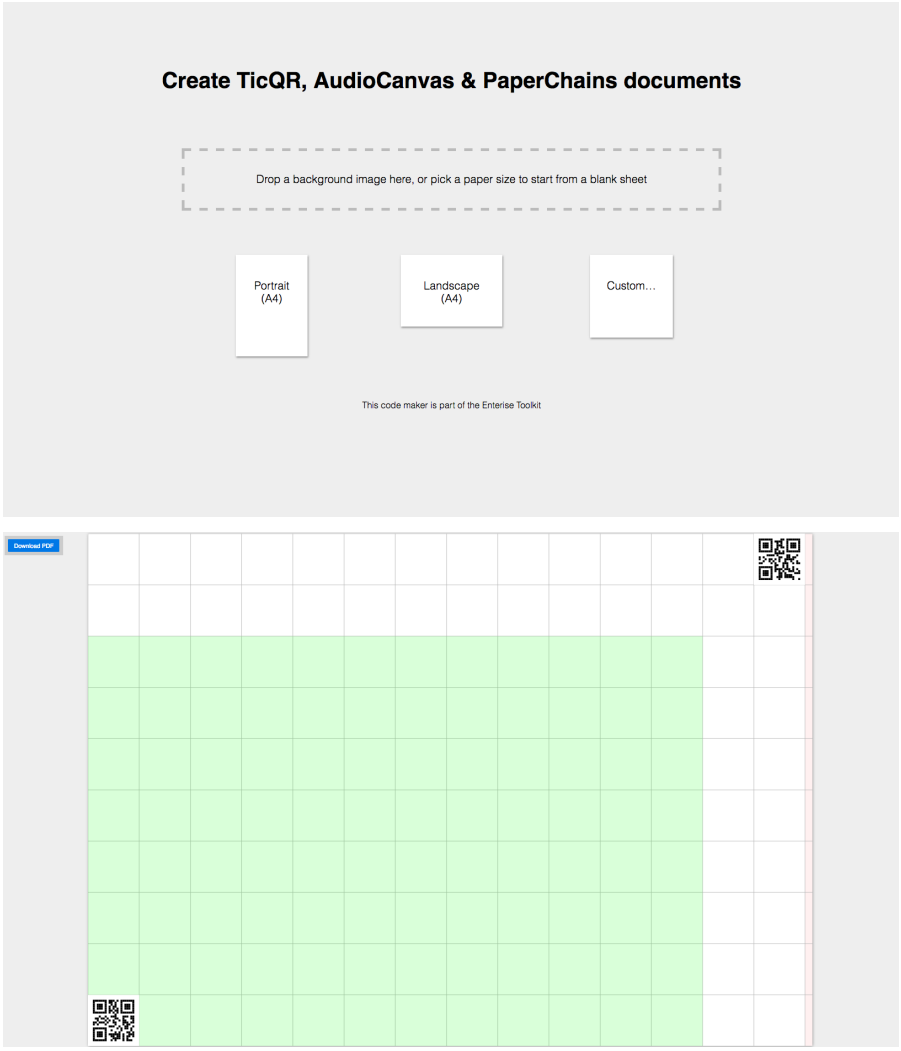


Fig. 8. The web-based authoring tool. Top: select a paper size, or upload an existing document to use as a background. Bottom: position QR codes on the document in order to enable AudioCanvas-based capabilities, then download a pre-made PDF.

To add new audio content, the user 'scribbles' on the image, highlighting the area with which the audio should be associated (Fig. 9 (centre), covering the words 'Jungle Oats'). After scribbling, authors capture their own audio or send pre-recorded audio from the device, as shown in Fig. 9 (right).

We envisage this tool being used primarily in two key scenarios:

(1) As a quick way for companies or organisations to mark-up items that they wish to augment with AudioCanvas audio;

(2) As an audio-based content store centred around physical documents.

[10]See: enterise.info/paperchains

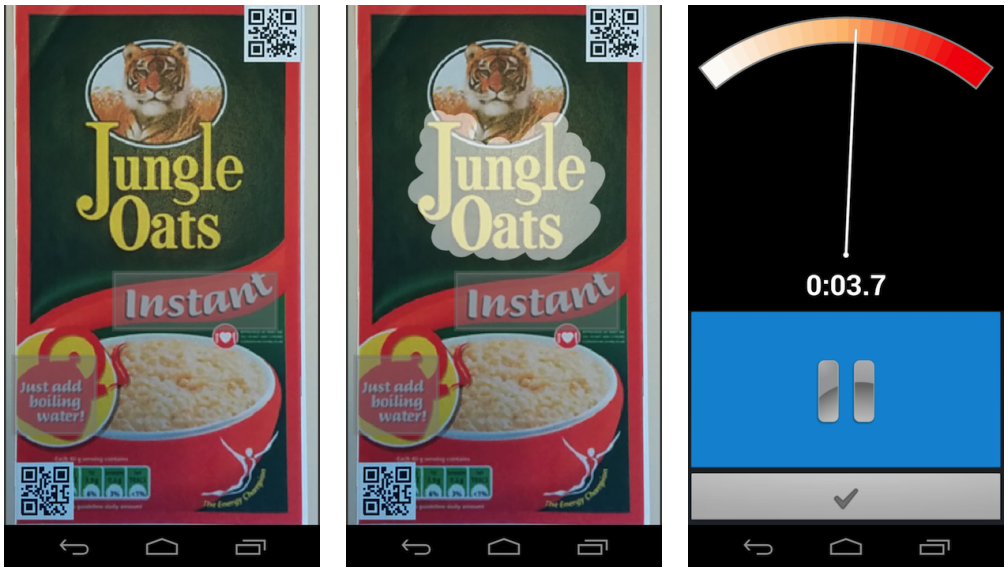


Fig. 9. Creating new AudioCanvas media annotations. From left to right: visualised audio hotspots can be viewed, edited, or added to in situ.

We illustrate the second of these uses in the next Section around storytelling. In addition, we have released an open-source internet-based (e.g., non-IVR) version of the tool for wider community use[10], with the hope that other researchers or organisations can use or modify it for their own benefit. We believe that the AudioCanvas creation tool has many potential avenues for extension beyond the ideas we investigate in this paper, and see a rich area of future work that builds on this initial body of research.

For example, one research question we identified when building the AudioCanvas creation tool was centred around the benefits—and indeed issues—that will come from crowdsourcing audio information. That is, allowing any user, rather than just the original canvas creator, to specify a hotspot and add content. Beyond the common problem of censorship in collaborative environments, there is also the issue of multi-layered audio hotspots. For example, how would the system ensure that the most relevant or useful hotspot is the topmost layer; and, furthermore, what should the system do when two distinct audio hotspots overlap?

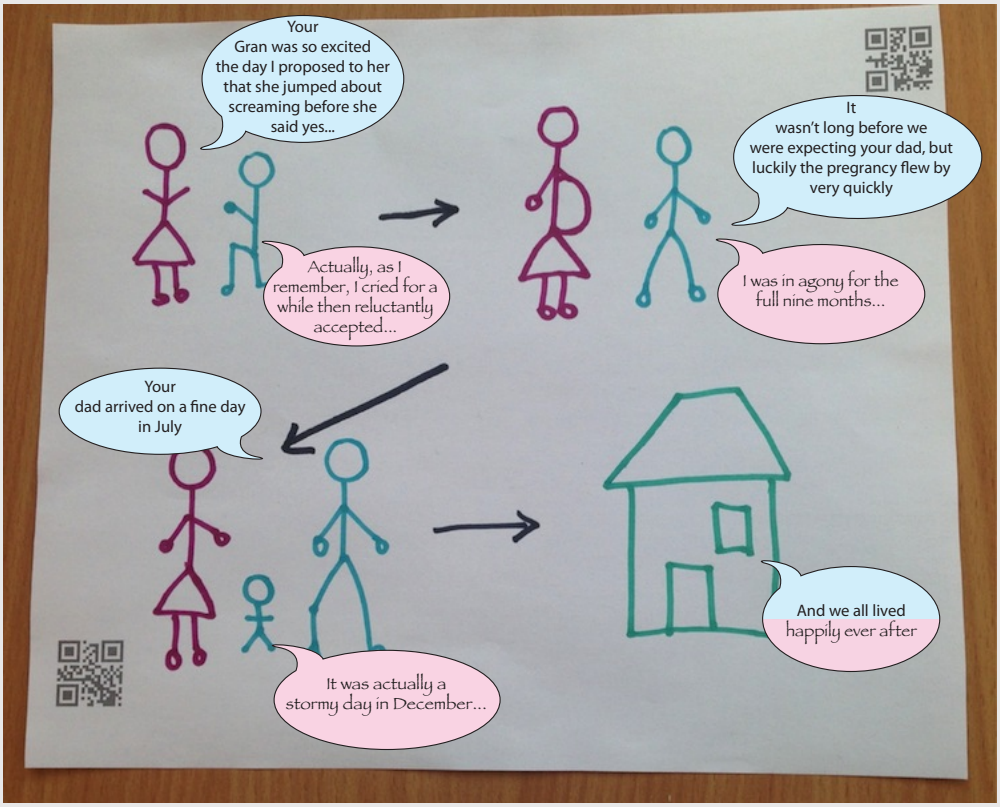
Further research into how this type of system will be used 'in the wild' would also be a beneficial area of future work, looking in more depth into naturalistic use by users in this context over a longer period of time.

6. LOW-TECH DIGITAL STORYTELLING

The most basic usage scenario of the AudioCanvas creation tool is with entirely blank documents, allowing users to sketch or append their own content. Pictures combined with text are a common way of telling stories (e.g., books, magazines, comic strips); but for those unable to read or write, or unwilling to put their thoughts into text, these media are often inaccessible.

In one attempt to address this gap, the digital storytelling movement has allowed users of all backgrounds to combine images with voices to share their stories and experiences. Digital stories are by definition, however, purely *digital*, and do not easily allow users to mix in other media—such as drawings—or to easily associate specific sections of images with particular pieces of audio.

Using the AudioCanvas technique in this scenario allows users to draw or paste images on paper, then augment precise parts with narration. Others can then access the stories by taking a photograph of the page with a cameraphone, and touching different parts to hear the author's narration at each location. Used in this way, the AudioCanvas technique not only connects physical drawings with digital voice-overs, but also allows more precise augmentation with audio, providing a potentially richer environment for storytelling. The following scenario illustrates how we envisage the system being used in this context.



AudioCanvas storytelling scenario: *Humphrey has a story from his childhood that he wants to leave for his young grandson. Although his literacy levels are low, Humphrey is good at sketching and decides to draw a series of pictures illustrating the story on AudioCanvas-augmented paper. After making the sketch, Humphrey takes a photograph of the page with his phone, and uses the on-screen tool to scribble over each part of his drawing. He then records voice-overs for each section to describe what is going on in each portion of the image. After he has told his version of events, he then passes the paper to his wife, Maude for her to add audio notes to it as well. By taking his own photograph of this image, their grandson, Timmy, can see the physical drawing of his grandparents' history, and hear audio recordings of their voices by tapping on specific sections of the image on the screen, thus preserving the family story.*

It is this type of scenario that we focus on in this section. We argue that the AudioCanvas approach could allow storytellers to combine the benefits of physical media for remote retelling, with the advantages of audio narration for personalised recitals.

To investigate the use of the AudioCanvas creation tool for low-tech digital storytelling, we performed evaluations in several resource-constrained environments to determine the benefits of our approach for this use in these contexts.

The first of these was a group-based evaluation in two locations. We have previously reported these studies in [Pearson et al. 2015]; here we summarise key aspects of the procedure and results. Full details can be found in the original paper.

The second study, which we have not previously reported, was undertaken with two Kaavadiyas—traditional travelling storytellers—in order to explore the benefit of the system for their profession.

6.1. Group-based evaluations

We performed two experiments with the AudioCanvas creation tool with groups of resourced constrained users. The first of these was in Mumbai, India, and the second in Langa, South Africa. The purpose of these studies was to explore both the situations in which groups currently share stories with one another; and, how our system could support or enhance this behaviour.

Both studies followed the same procedure, and involved 18 mixed-literacy participants in each location performing storytelling tasks in six groups of three people. Each group was given a blank A3 AudioCanvas sheet, an AudioCanvas enabled phone, and a set of drawing equipment. They were then asked to take turns at drawing parts of a story and augmenting these with audio narration using the phone provided.

When each group was happy with their story, they were asked to swap sheets with another group of three, and to interpret the other group's story by first taking a new photograph, then using AudioCanvas to listen to the audio hotspots. We then performed semi-structured focus-group interviews with the groups (i.e., six participants at once), and also asked participants to provide ratings for usability and usefulness.

6.1.1. Results. The results from both studies indicated that the AudioCanvas technique is viable and valuable for storytelling in resource-constrained contexts. All 36 participants were able to effectively augment AudioCanvas documents with voice annotations and retrieve annotations left by others.

Despite many of our participants having low-technological familiarity, practice framing the photographs resulted in effective use of the system, and yielded average ease of use ratings of 7 and 9 out of 10 for the Indian and South African studies respectively.

Both sets of participants tended to use a structured turn-taking approach to using the system – that is, choosing to draw and record audio one-by-one, with individuals occasionally adding to or modifying others' images.

High subjective scores were given in both evaluations for how well participants felt they were able to convey stories to others (7.3 out of 10 in Mumbai; 9.4 out of 10 in Langa), as well as how well they were able to interpret other people's stories using the system (8.8 out of 10 in Mumbai; 8.0 out of 10 in Langa).

Overall, users in both sites were very positive about the technology, with many users describing scenarios where they feel it would be particularly beneficial to them. For example, a common theme amongst participants was to use AudioCanvas to keep in touch with loved ones who were not physically present, with others suggesting using the system as a way of communicating directions on a map, or even to mark-up homework assignments.

Fig. 10. A Kaavad shrine. The panels at each side fold into the main box for portability. Each image on the panels is pointed to in turn as the Kaavadiya recites the story.

6.2. Exploratory study with traditional Kaavad storytellers

Suggestions from users about how the system could be used in their daily lives help to support the suitability of AudioCanvas in these contexts. Of particular interest to us was the use of the platform to enhance or extend existing storytelling practices. As such, we wanted to explore the usage of the system in more natural storytelling settings. In order to do this, we recruited two veteran Kaavad storytellers to use the AudioCanvas system.

The Kaavad storytellers of Rajasthan, India follow a 400-year-old tradition using an elaborately painted portable wooden temple or shrine (see Fig. 10) as a prop to recite stories to patrons [Sabnani 2014]. Kaavadiyas are expert storytellers, narrating their patrons' caste genealogy, singing praises of their ancestors or reciting epic stories from folk tradition that are depicted on the Kaavad temple, while pointing to individual painted panels with a peacock feather.

Patrons—those in the audience who have made a special donation to the storyteller—are often painted onto the temple and brought back year after year as elements of different stories. Typically each storyteller will have 30 to 50 patrons, widely spread across Rajasthan, who will each place a donation into the shrine during their annual visit.

The experience is audio-visual but requires no written language skills on the part of the storyteller or the patron. Both the physical aspect and the spoken story are customised over time based on the patron in question. Stories are personalised using different combinations of panels, and the images painted on each panel are modified by the Kaavadiya to include new people or activities before each visit. For example, if a family member of a patron has previously avoided making a donation, the storyteller will make this known by subtly turning some panels on the Kaavad upside down and stating that this image is their family member. This encourages the embarrassed patron to donate on his family's behalf, and avoids the unpleasantness of the storyteller stating the problem directly.

The Kaavad storytellers were of particular interest to us for two reasons. Firstly, they use an artefact that changes physically over time, with stories that alter at each evolution. Secondly, the highly-experienced storytellers use the box—touching it gently with a feather—as a prop for their physically collocated, situated storytelling (see Fig. 11). We felt that studying Kaavadiyas' methods and introducing our approach

Fig. 11. The ritual of reciting the genealogy being performed by a Kaavadiya. The two Kaavadiyas take it in turns to tell different stories, passing the shrine between them when necessary.

would offer valuable pointers to the benefits and drawbacks of our technique in real storytelling situations.

Two experienced Kaavadiyas (both male) were invited to recite stories for our research team; and, afterwards, to use the AudioCanvas system in their storytelling. We aimed to understand how the Kaavadiyas might use AudioCanvas to recite their own stories, as well as to gather their feedback on the suitability of the technique for storytelling. With this in mind, we produced several customised Kaavad panels with AudioCanvas QR codes embedded to allow the storytellers to record their narration using the system.

The format of this exploratory study with the Kaavadiyas was unstructured and exploratory. We began by explaining why we built the system, before moving on demonstrate how it could be used to sketch and annotate documents. The storytellers were then given time to record their own audio stories using the customised Kaavad panels. The session concluded with a group discussion of AudioCanvas and its benefits.

6.2.1. Results. It was immediately apparent that the Kaavadiyas were excited by the concept. *"This is a new way of doing our business!"* and *"Where can I buy this?"* were the first comments given. Both men stated that they would like to be able to use the system, and one suggested the specific use case of remote patrons: *"people like you [our research team] and other patrons could use it – those who live abroad"*. As might be expected, however, there were some reservations around the possibility of redundancy of the storytellers themselves: *"if I gave the Kaavad to those who I visit every year they would just say: 'fine, you don't need to come any more'!"* One suggested that a possible solution to this problem would be to restrict the number of times a patron

could listen to the story: *"we'd need a way of making sure the patron only views it once [. . .] unless the patron pays again!"*

We were particularly interested in whether the Kaavadiyas perceived the introduction of technology as something that would affect the traditions of the storytelling method. Their response was that it did not feel right using their fingers to recite the stories – the tool has to be a peacock feather. Another important aspect regarded how the flow of stories is laid out. For example, should the digital version be constructed in comic-strip style with a clear progression from one image to another, or should it be more flexible, as in the traditional Kaavad? When using the system, the Kaavadiyas chose to first summarise the story using the entire shrine, then move on to select individual images and tell specific parts one-by-one. There was a clear flow in the stories told during these recitals. However, they chose not to leave any sort of clue to future listeners as to which parts to select and in which order.

6.2.2. Summary. The Kaavad storytellers provided us with an insight into how traditional storytelling could evolve in the digital age, including the simultaneous use of props and narration in the story. Kaavadiyas accomplish their craft independently of any written language, and recite their body of stories, from memory, to multiple listeners. Patrons pay for the storyteller to recite to them, and expect a certain level of exclusivity in return. Indeed, each story and genealogy is unique when told, despite the Kaavad shrine having a finite set of visual panels. The storyteller is able to tailor a patron's story specifically to the audience by using a different combination of visuals and a variant of the narrative. This combination of custom visuals and audio is crucial to this storytelling process. Removal of either of these elements would significantly reduce the Kaavad storytelling experience.

We believe that our design was well received by the Kaavad tellers because of its ability to allow individual parts of drawings (or in this case panels) to be augmented with recorded audio. In certain scenarios—for example, the remote patron, as suggested by one of the storytellers themselves—the design has the potential to be used for Kaavad storytelling. The added benefits of requiring no literacy and no internet connection also mean that the general technique is particularly suitable for this context.

7. REFLECTIONS

The AudioCanvas technique that we have discussed in this article provides users with a straightforward audio accompaniment to physical media. Taking a photo of an AudioCanvas-augmented object turns it into an internet-free interactive sound surface. Six separate experiments have demonstrated the benefits of the design in different contexts and for a range of applications. Here we reflect upon general themes that arose from these studies.

7.1. Trusting the system

In our work so far, we have focused on using the AudioCanvas technique for providing audio versions of text that users would otherwise be unable to read. However, the benefits of the approach go beyond simple translation of content, which could, given the appropriate technology and infrastructure, be accomplished by approaches such as Google Goggles[11] (assuming the required language was supported).

Our design aims to provide support for interpretation rather than translation – a feature that fits well with the use of proximates[12] in many low-literate communities.

[11]See: google.com/mobile/goggles

[12]In this context, proximates are intermediaries who have skills in literacy or technology, that act as go-betweens for people in the community [Chipchase 2012; Sambasivan et al. 2010; Walton et al. 2012]

Indeed, it was clear from discussions with participants in both India and South Africa that the role of proximates was not only to relay information, but also to describe and interpret content in a way that could be more easily understood and related to.

The issue of trust is also a major factor in this context [Patel et al. 2012]. When relying on others to interpret information from important sources—such as government forms, medical treatments and other confidential documents—it is vital that there is a high level of trust in the person relaying the content. Clearly, when the information being relayed is crucial for, say, avoiding fines or receiving the correct medical treatment, it is critical that the information given is correct.

Mistakes in translation or malicious intent on the part of the relayer are common fears when relying on others for this task. The use of proximates—typically older or more educated members of the community who are well known and trusted—is common to provide others with the trusted translations they require. Hearing the voice of a familiar and respected community member reading the information, then, can increase the trust given by users of the system. Using AudioCanvas in this situation also means that the respected translator need only interpret a document once, rather than repeating the translation for each potential user.

7.2. Design reflections

We have also suggested the possibility for using the AudioCanvas technique as a form of interactive editable audio artefact, allowing any user to record their own audio at points they touch on an object. This crowd-sourcing of recorded translations could create, in effect, 'audio wikis' on physical objects. If such a system were to become widespread, however, we would also need to consider more organisational aspects of the design. For example, we have not yet considered how to deal with overlapping audio regions, content that is in poor taste, or multiple conflicting versions of the same content.

One way of dealing with these types of scenarios would be to overlay visual cues on the photograph to indicate where audio clips are located. One of the major attraction points of the current design—and one which was praised by participants—is its ease of use, requiring only a single photo and touches anywhere on the photo to access related audio. The introduction of a visual indication as to where hotspots are located could well reduce the overall lightweight-ness design of the system itself. Whether this would reduce or enhance the user experience of the system is not clear, and is an area of work to be investigated in future.

The current version of the technique uses QR codes attached to objects to act as document identifiers and alignment points. Future versions of the general technique could use less-prominent image identifiers to remove the need for any object modification at all (e.g., a technique such as *Aestheticodes* [Meese et al. 2013]). We chose to use QR codes as alignment identifiers for two reasons. Firstly, they are an already established and robust marker scheme, and one which is easy to recognise using a smartphone camera. Secondly, the construction of QR codes—with two codes giving six fixed-position alignment points—makes them particularly useful for correcting image skew and rotation.

As has been previously pointed out (e.g., in [Robinson et al. 2014b]), one of the most beneficial qualities of the QR code design is the fact that they are visible. QR codes are often chosen, then, because they are highly recognisable, and clearly afford a specific action. In the same way that knobs afford the action of turning or buttons afford the action of pushing, QR codes afford the action of scanning. A future version of our design might, then, use its own code designs that are specifically recognisable as AudioCanvas markers, giving the user a cue to take a photo that includes both codes in order to listen to its content.

Table IV. The average rating (out of 7) given in both India- and UK-based studies; and, the results of a two-tailed Mann-Whitney *U*-Test to measure significance.

Ease of use, in terms of	India (avg.)	UK (avg.)	p-value	z-score
Focusing the camera and taking the initial photo	4.9	6.1	0.114	1.579
Panning and zooming around the photo	6.2	6.1	0.250	-1.154
Selecting specific points of interest within the photo	6.3	6.1	0.147	-1.446
Getting the right information back from each section	6.3	6.3	0.226	-1.215
Finding the audio hotspots in a photo	5.1	5.0	0.384	-0.875

7.3. Study considerations

Evaluating AudioCanvas in three separate geographically and culturally distant sites (Sections 4.1 to 4.3) revealed several general insights with regards to conducting user studies. While it is typically considered to be best practice to perform evaluations with target users (e.g., [Kjeldskov and Stage 2004; Esbjörnsson et al. 2006]), from our experiences in this work, we argue that in certain cases, simulating the interaction scenario with different users can yield similar results. While clearly not suitable for all cases, in this example, testing with users in the UK and simulating usage conditions proved to be a valuable and useful alternative.

In our case, we mimicked a lack of textual literacy in the first user study in the UK. Participants in this region were highly literate in English, but could not speak or read the languages on the prototype media items used. For this aspect of the trials, then, users' ability to read the information provided was identical across all three sites – none were able to read the media given, and all relied heavily on the audio feedback to find the desired information.

Clearly, the UK participants had completely different technological backgrounds to the eventual target user groups, which could, of course, play a large role in how they engaged with the prototype. For example, the lack of familiarity with smartphones was clearly evident in the initial stages of both the South African and the Indian evaluations, with many users being unfamiliar with gestures such as swiping, or requiring instruction in how to frame photographs.

What we observed in these cases, however, was that the learning curve of using the devices themselves (including standard input gestures such as touch/swipe/pinch) was surprisingly minimal. After an initial demonstration and a short hands-on learning session, participants were able to use the basic phone functions, and understood how best to frame a photograph using the AudioCanvas system. We argue, then, that for systems relying on completely new interactions—techniques that would be unfamiliar to participants already familiar with the base hardware—it can be possible to evaluate some aspects without needing to account for every possible existing factor.

To test our assumption, we compared the ratings given to the system in the UK study with those from the India study, which was conducted in the same manner. Our hypothesis was that the results from a study using foreign language media in the UK would yield similar results to the same study in India using media participants were unable to read. We evaluated the ratings gathered from both sets of 22 participants using a 2-tailed Mann-Whitney *U*-Test, the results of which are shown in Table IV.

We could find no evidence of significant differences between the two samples for any of the five separate measures recorded.

In some cases, then—perhaps at an early stage of design—it may be possible for certain factors attributed to participants to be simulated with a different user group. We would argue that in situations such as the example we have described here, it is beneficial to mimic the issues encountered by a certain set of users and achieve very similar results, avoiding the difficult, time consuming and potentially expensive pressure of studies with users 'in the field' at every stage of the design process. Instead,

studies of this type could be reserved until later in the design process, where they can be more impactful.

Previous research has studied this issue of field studies in comparison to lab experiments. In some cases there has been little difference in results (e.g., [Kjeldskov et al. 2004; Sun and May 2013]). Others understandably disagree (e.g., [Nielsen et al. 2006]). Furthermore, while we have been successful in achieving similar results between diverse groups when focusing on audio and the readability of a foreign language, clearly there are many elements of field studies it is not possible to mimic – for example, differences in power dynamics, cultural etiquette, comfort with technology and so on [Medhi et al. 2010].

Our argument, then, is not that it is always necessary to perform lab and field studies between different locations and populations, nor that it is always unnecessary. Rather, the issue is that a blanket view of whether studies of this type are required is not the best way to approach a design; instead, it depends on the system in question, and indeed the question being asked – and a more nuanced approach (along with further study) is needed.

8. CONCLUSIONS

In this article we have explored AudioCanvas – a novel photo interaction system that provides audio content on-demand for printed physical media. Our design allows users to interact directly with personal photographs of physical items by simply touching regions within them. The AudioCanvas design requires no specialist hardware, and uses a remote voice service accessed via a standard phone line to ensure it can be accessed without a potentially costly data connection, making it more accessible to those in more resource-constrained regions of the world.

We conducted a broad set of user evaluations with the core AudioCanvas design, recruiting participants in three considerably different locations around the world. These separate studies allowed us to gather subjective responses from users in diverse contexts, ranging from highly textually and technologically literate with access to fast data connections (UK); semiliterate, with little technology exposure and intermittent data connections (South Africa); and, finally, illiterate, with low levels of technology skill and no access to data connections (India).

In addition to detailing the basic technique, we have also described an accompanying authoring tool, and provided further usage scenarios for the AudioCanvas approach. A study with traditional storytellers demonstrated the potential usage of the approach for flexible, interactive storytelling. The studies we conducted show a strong appreciation and desire for the AudioCanvas design, and provide evidence that it would be particularly beneficial in areas where literacy and data connection access are low.

References

Sheetal Agarwal, Arun Kumar, Amit Nanavati, and Nitendra Rajput. 2008. The World Wide Telecom Web Browser. In *Proc. WWW '08*. ACM, New York, NY, USA, 1121–1122. DOI:http://dx.doi.org/10.1145/1367497.1367686

Toshifumi Arai, Kimiyoshi Machii, Soshiro Kuzunuki, and Hiroshi Shojima. 1995. InteractiveDESK: A Computer-augmented Desk Which Responds to Operations on Real Objects. In *Proc. CHI '95*. ACM, New York, NY, USA, 141–142. DOI:http://dx.doi.org/10.1145/223355.223470

Maribeth Back, Jonathan Cohen, Rich Gold, Steve Harrison, and Scott Minneman. 2001. Listen Reader: An Electronically Augmented Paper-based Book. In *Proc. CHI '01*. ACM, New York, NY, USA, 23–29. DOI:http://dx.doi.org/10.1145/365024.365031

Xiang Cao, Siân E. Lindley, John Helmes, and Abigail Sellen. 2010. Telling the Whole Story: Anticipation, Inspiration and Reputation in a Field Deployment of TellTable. In *Proc. CSCW '10*. ACM, New York, NY, USA, 251–260. DOI:http://dx.doi.org/10.1145/1718918.1718967

Jan Chipchase. 2012. Imperialist Tendencies. See: http://janchipchase.com/content/essays/imperialist-tendencies/. (January 2012).

Enrico Costanza, Matteo Giaccone, Olivier Kueng, Simon Shelley, and Jeffrey Huang. 2010. Ubicomp to the Masses: A Large-scale Study of Two Tangible Interfaces for Download. In *Proc. UbiComp '10*. ACM, New York, NY, USA, 173–182. DOI:http://dx.doi.org/10.1145/1864349.1864388

Sebastien Cuendet, Indrani Medhi, Kalika Bali, and Edward Cutrell. 2013. VideoKheti: Making Video Content Accessible to Low-literate and Novice Users. In *Proc. CHI '13*. ACM, New York, NY, USA, 2833–2842. DOI:http://dx.doi.org/10.1145/2470654.2481392

Edirlei Soares de Lima, Bruno Feijó, Simone D.J. Barbosa, Antonio L. Furtado, Angelo E.M. Ciarlini, and Cesar T. Pozzer. 2014. Draw Your Own Story: Paper and Pencil Interactive Storytelling. *Entertainment Computing* 5, 1 (2014), 33–41. DOI:http://dx.doi.org/10.1016/j.entcom.2013.06.004

Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. "Yours is Better!": Participant Response Bias in HCI. In *Proc. CHI '12*. ACM, New York, NY, USA, 1321–1330. DOI:http://dx.doi.org/10.1145/2207676.2208589

Allison Druin, Jason Stewart, David Proft, Ben Bederson, and Jim Hollan. 1997. KidPad: A Design Collaboration Between Children, Technologists, and Educators. In *Proc. CHI '97*. ACM, New York, NY, USA, 463–470. DOI:http://dx.doi.org/10.1145/258549.258866

Marc Dymetman and Max Copperman. 1998. Intelligent Paper. In *Proc. EP/RIDT '98*. Springer-Verlag, Berlin, Heidelberg, 392–406. DOI:http://dx.doi.org/10.1007/BFb0053286

Berna Erol, Emilio Antúnez, and Jonathan J. Hull. 2008. HOTPAPER: Multimedia Interaction with Paper Using Mobile Phones. In *Proc. MM '08*. ACM, New York, NY, USA, 399–408. DOI:http://dx.doi.org/10.1145/1459359.1459413

Berna Erol, Jonathan J. Hull, Jamey Graham, and Dar-Shyang Lee. 2004. Prescient Paper: Multimedia Document Creation with Document Image Matching. In *Proc. ICPR '04*, Vol. 2. IEEE, Washington, DC, USA, 675–678. DOI:http://dx.doi.org/10.1109/ICPR.2004.1334349

Mattias Esbjörnsson, Barry Brown, Oskar Juhlin, Daniel Normark, Mattias Östergren, and Eric Laurier. 2006. Watching the Cars Go Round and Round: Designing for Active Spectating. In *Proc. CHI '06*. ACM, New York, NY, USA, 1221–1224. DOI:http://dx.doi.org/10.1145/1124772.1124955

David Frohlich. 2004. *Audiophotography: Bringing Photos to Life with Sounds*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

David Frohlich, Tony Clancy, John Robinson, and Enrico Costanza. 2004. The Audiophoto Desk. In *Proc. 2AD demos*. The Appliance Design Network, Bristol, United Kingdom, 139–139.

François Guimbretière. 2003. Paper Augmented Digital Documents. In *Proc. UIST '03*. ACM, New York, NY, USA, 51–60. DOI:http://dx.doi.org/10.1145/964696.964702

Sam Jacoby and Leah Buechley. 2013. Drawing the Electric: Storytelling with Conductive Ink. In *Proc. IDC '13*. ACM, New York, NY, USA, 265–268. DOI:http://dx.doi.org/10.1145/2485760.2485790

Anirudha Joshi, Arnab Chakravarty, Abhishek Shrivastava, Nagraj Emmadi, Amrutha Krishnan, Surbhi Bindlish, Saurabh Srivastava, and Nitendra Rajput. 2012. Usability Evaluation of Visual IVR Systems. In *Proc. APCHI '12*. ACM, New York, NY, USA.

Jesper Kjeldskov, Mikael B. Skov, Benedikte S. Als, and Rune T. Høegh. 2004. Is it Worth the Hassle? Exploring the Added Value of Evaluating the Usability of Context-Aware Mobile Systems in the Field. In *LNCS*, Vol. 3160. Springer-Verlag, Berlin, Heidelberg, 61–73. DOI:http://dx.doi.org/10.1007/978-3-540-28637-0_6

Jesper Kjeldskov and Jan Stage. 2004. New Techniques for Usability Evaluation of Mobile Systems. *International Journal of Human-Computer Studies* 60 (2004), 599–620. DOI:http://dx.doi.org/10.1016/j.ijhcs.2003.11.001

Scott Klemmer, Jamey Graham, Gregory Wolff, and James Landay. 2003. Books with Voices: Paper Transcripts as a Tangible Interface to Oral Histories. In *Proc. CHI '03*. ACM, New York, NY, USA, 89–96. DOI:http://dx.doi.org/10.1145/642611.642628

Hideki Koike and Motoki Kobayashi. 1998. EnhancedDesk: Integrating Paper Documents and Digital Documents. In *Proc. APCHI '98*. IEEE, Washington, DC, USA, 57–62. DOI:http://dx.doi.org/10.1109/APCHI.1998.704149

Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty, Sheetal Agarwal, and Amit Nanavati. 2007. WWTW: The World Wide Telecom Web. In *Proc. NSDR workshop '07*. ACM, New York, NY, USA. DOI:http://dx.doi.org/10.1145/1326571.1326582

Chunyuan Liao, François Guimbretière, and Ken Hinckley. 2005. PapierCraft: A Command System for Interactive Paper. In *Proc. UIST '05*. ACM, New York, NY, USA, 241–244. DOI:http://dx.doi.org/10.1145/1314683.1314686

Qiong Liu, Chunyuan Liao, Lynn Wilcox, Anthony Dunnigan, and Bee Liew. 2010. Embedded Media Markers: Marks on Paper That Signify Associated Media. In *Proc. IUI '10*. ACM, New York, NY, USA, 149–158. DOI:http://dx.doi.org/10.1145/1719970.1719992

Andrés Lucero, Jussi Holopainen, and Tero Jokela. 2011. Pass-them-around: Collaborative Use of Mobile Phones for Photo Sharing. In *Proc. CHI '11*. ACM, New York, NY, USA, 1787–1796. DOI:http://dx.doi.org/10.1145/1978942.1979201

Indrani Medhi, Ed Cutrell, and Kentaro Toyama. 2010. It's Not Just Illiteracy. In *Proc. IHCI'10*. British Computer Society, Swindon, UK, 1–10. http://dl.acm.org/citation.cfm?id=2227347.2227348

Indrani Medhi, S. N. Nagasena Gautama, and Kentaro Toyama. 2009. A Comparison of Mobile Money-Transfer UIs for Non-Literate and Semi-Literate Users. In *Proc. CHI '09*. ACM, New York, NY, USA, 1741–1750. DOI:http://dx.doi.org/10.1145/1518701.1518970

Indrani Medhi, Meera Lakshmanan, Kentaro Toyama, and Edward Cutrell. 2013. Some Evidence for the Impact of Limited Education on Hierarchical User Interface Navigation. In *Proc. CHI '13*. ACM, New York, NY, USA, 2813–2822. DOI:http://dx.doi.org/10.1145/2470654.2481390

Indrani Medhi, Somani Patnaik, Emma Brunskill, S.N. Nagasena Gautama, William Thies, and Kentaro Toyama. 2011. Designing Mobile Interfaces for Novice and Low-literacy Users. *ACM Trans. Comput.-Hum. Interact.* 18, 1, Article 2 (May 2011), 28 pages. DOI:http://dx.doi.org/10.1145/1959022.1959024

Rupert Meese, Shakir Ali, Emily-Clare Thorne, Steve Benford, Anthony Quinn, Richard Mortier, Boriana Koleva, Tony Pridmore, and Sharon Baurley. 2013. From Codes to Patterns: Designing Interactive Decoration for Tableware. In *Proc. CHI '13*. ACM, New York, NY, USA, 931–940. DOI:http://dx.doi.org/10.1145/2470654.2466119

Pranav Mistry, Pattie Maes, and Liyan Chang. 2009. WUW – Wear Ur World: A Wearable Gestural Interface. In *Proc. CHI EA '09*. ACM, New York, NY, USA, 4111–4116. DOI:http://dx.doi.org/10.1145/1520340.1520626

Christian Monrad Nielsen, Michael Overgaard, Michael Bach Pedersen, Jan Stage, and Sigge Stenild. 2006. It's Worth the Hassle!: The Added Value of Evaluating the Usability of Mobile Systems in the Field. In *Proc. NordiCHI '06*. ACM, New York, NY, USA, 272–280. DOI:http://dx.doi.org/10.1145/1182475.1182504

Kenton O'Hara, Tim Kindberg, Maxine Glancy, Luciana Baptista, Byju Sukumaran, Gil Kahana, and Julie Rowbotham. 2007. Collecting and Sharing Location-based Content on Mobile Phones in a Zoo Visitor Experience. *Computer Supported Cooperative Work* 16, 1 (2007), 11–44. DOI:http://dx.doi.org/10.1007/s10606-007-9039-2

Tapan Parikh, Paul Javid, Sasikumar K., Kaushik Ghosh, and Kentaro Toyama. 2006. Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India. In *Proc. CHI '06*. ACM, New York, NY, USA, 551–560. DOI:http://dx.doi.org/10.1145/1124772.1124857

Neil Patel, Deepti Chittamuru, Anupam Jain, Paresh Dave, and Tapan Parikh. 2010. Avaaj Otalo: A Field Study of an Interactive Voice Forum for Small Farmers in Rural India. In *Proc. CHI '10*. ACM, New York, NY, USA, 733–742. DOI:http://dx.doi.org/10.1145/1753326.1753434

Neil Patel, Scott Klemmer, and Tapan Parikh. 2011. An Asymmetric Communications Platform for Knowledge Sharing with Low-end Mobile Phones. In *Proc. UIST '11 Adjunct*. ACM, New York, NY, USA, 87–88. DOI:http://dx.doi.org/10.1145/2046396.2046436

Neil Patel, Kapil Shah, Krishna Savani, Scott R. Klemmer, Paresh Dave, and Tapan S. Parikh. 2012. Power to the Peers: Authority of Source Effects for a Voice-based Agricultural Information Service in Rural India. In *Proc. ICTD '12*. ACM, New York, NY, USA, 169–178. DOI:http://dx.doi.org/10.1145/2160673.2160696

Jennifer Pearson, Simon Robinson, and Matt Jones. 2015. PaperChains: Dynamic Sketch+Voice Annotations. In *Proc. CSCW '15*. ACM, New York, NY, USA, 383–392. DOI:http://dx.doi.org/10.1145/2675133.2675138

Jennifer Pearson, Simon Robinson, Matt Jones, Amit Nanavati, and Nitendra Rajput. 2013. ACQR: Acoustic Quick Response Codes for Content Sharing on Low End Phones with No Internet Connectivity. In *Proc. MobileHCI '13*. ACM, New York, NY, USA, 308–317. DOI:http://dx.doi.org/10.1145/2493190.2493195

Bastian Pfleging, Elba Bahamondez, Albrecht Schmidt, Martin Hermes, and Johannes Nolte. 2010. MobiDev: A Mobile Development Kit for Combined Paper-Based and In-Situ Programming on the Mobile Phone. In *Proc. CHI '10 Extended Abstracts*. ACM, New York, NY, USA, 3733–3738. DOI:http://dx.doi.org/10.1145/1753846.1754047

Hayes Raffle, Cati Vaucelle, Ruibing Wang, and Hiroshi Ishii. 2007. Jabberstamp: Embedding Sound and Voice in Traditional Drawings. In *Proc. IDC '07*. ACM, New York, NY, USA, 137–144. DOI:http://dx.doi.org/10.1145/1297277.1297306

Simon Robinson, Matt Jones, Elina Vartiainen, and Gary Marsden. 2012. PicoTales: Collaborative Authoring of Animated Stories Using Handheld Projectors. In *Proc. CSCW '12*. ACM, New York, NY, USA, 671–680. DOI:http://dx.doi.org/10.1145/2145204.2145306

Simon Robinson, Jennifer Pearson, and Matt Jones. 2014a. AudioCanvas: Internet-Free Interactive Audio Photos. In *Proc. CHI '14*. ACM, New York, NY, USA, 3735–3738. DOI:http://dx.doi.org/10.1145/2556288.2556993

Simon Robinson, Jennifer Pearson, and Matt Jones. 2014b. A Billion Signposts: Repurposing Barcodes for Indoor Navigation. In *Proc. CHI '14*. ACM, New York, NY, USA, 639–642. DOI:http://dx.doi.org/10.1145/2556288.2556994

Simon Robinson, Nitendra Rajput, Matt Jones, Anupam Jain, Shrey Sahay, and Amit Nanavati. 2011. TapBack: Towards Richer Mobile Interfaces in Impoverished Contexts. In *Proc. CHI '11*. ACM, New York, NY, USA, 2733–2736. DOI:http://dx.doi.org/10.1145/1978942.1979345

Nina Sabnani. 2014. *Kaavad Tradition of Rajasthan: A Portable Pilgrimage*. Niyogi Books, New Delhi, India.

Nithya Sambasivan, Ed Cutrell, Kentaro Toyama, and Bonnie Nardi. 2010. Intermediated Technology Use in Developing Communities. In *Proc. CHI '10*. ACM, New York, NY, USA, 2583–2592. DOI:http://dx.doi.org/10.1145/1753326.1753718

Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price. 1998. Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations. In *Proc. CHI '98*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 249–256. DOI:http://dx.doi.org/10.1145/274644.274680

Julian Seifert, Bastian Pfleging, Martin Hermes, Enrico Rukzio, and Albrecht Schmidt. 2011. Mobidev: a tool for creating apps on mobile phones. In *Proc. MobileHCI '11*. ACM, New York, NY, USA, 109–112. DOI:http://dx.doi.org/10.1145/2037373.2037392

Graeme Smith and Gary Marsden. 2011. Providing Media Download Services in African Taxis. In *Proc. SAICSIT '11*. ACM, New York, NY, USA, 215–223. DOI:http://dx.doi.org/10.1145/2072221.2072246

Lisa Stifelman, Barry Arons, and Chris Schmandt. 2001. The Audio Notebook: Paper and Pen Interaction with Structured Speech. In *Proc. CHI '01*. ACM, New York, NY, USA, 182–189. DOI:http://dx.doi.org/10.1145/365024.365096

Xu Sun and Andrew May. 2013. A Comparison of Field-based and Lab-based Experiments to Evaluate User Experience of Personalised Mobile Devices. *Advances in Human-Computer Interaction* 2013, Article 619767 (2013), 9 pages. DOI:http://dx.doi.org/10.1155/2013/619767

Genta Suzuki, Shun Aoki, Takeshi Iwamoto, Daisuke Maruyama, Takuya Koda, Naohiko Kohtake, Kazunori Takashio, and Hideyuki Tokuda. 2005. u-Photo: Interacting with Pervasive Services Using Digital Still Images. In *Proc. Pervasive '05*. Springer-Verlag, Berlin, Heidelberg, 190–207. DOI:http://dx.doi.org/10.1007/11428572_12

Telecom Regulatory Authority of India. 2012. The Indian Telecom Services Performance Indicators: October – December 2011. See: http://goo.gl/g8xSl. (April 2012).

United Nations Development Programme. 2013. *Human Development Report*. United Nations Development Programme, New York, NY, USA.

Narseo Vallina-Rodriguez, Pan Hui, and Jon Crowcroft. 2009. Has Anyone Seen My Goose? Social Network Services in Developing Regions. In *Proc. CSE '09*. IEEE, Washington, DC, USA, 1048–1053. DOI:http://dx.doi.org/10.1109/CSE.2009.276

Marion Walton, Gary Marsden, Silke Haßreiter, and Sena Allen. 2012. Degrees of Sharing: Proximate Media Sharing and Messaging by Young People in Khayelitsha. In *Proc. MobileHCI '12*. ACM, New York, NY, USA, 403–412. DOI:http://dx.doi.org/10.1145/2371574.2371636

Marion Walton, Vera Vukovic', and Gary Marsden. 2002. 'Visual Literacy' As Challenge to the Internationalisation of Interfaces: A Study of South African Student Web Users. In *Proc. CHI '02 Extended Abstracts*. ACM, New York, NY, USA, 530–531. DOI:http://dx.doi.org/10.1145/506443.506465

Matt Warman. 2012. Monmouth to be world's first 'Wikipedia town'. See: http://www.telegraph.co.uk/technology/wikipedia/9274591/Monmouth-to-be-worlds-first-Wikipedia-town.html. (May 2012).

Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. *Commun. ACM* 36, 7 (1993), 87–96. DOI:http://dx.doi.org/10.1145/159544.159630

David West, Aaron Quigley, and Judy Kay. 2007. MEMENTO: A Digital-physical Scrapbook for Memory Sharing. *Personal and Ubiquitous Computing* 11, 4 (2007), 313–328. DOI:http://dx.doi.org/10.1007/s00779-006-0090-7

Gavin Wood, John Vines, Madeline Balaam, Nick Taylor, Thomas Smith, Clara Crivellaro, Juliana Mensah, Helen Limon, John Challis, Linda Anderson, Adam Clarke, and Peter C. Wright. 2014. The Dept. Of Hidden Stories: Playful Digital Storytelling for Children in a Public Library. In *Proc. CHI '14*. ACM, New York, NY, USA, 1885–1894. DOI:http://dx.doi.org/10.1145/2556288.2557034

Chih-Sung Wu, Susan Robinson, and Ali Mazalek. 2008. Turning a Page on the Digital Annotation of Physical Books. In *Proc. TEI '08*. ACM, New York, NY, USA, 109–116. DOI:http://dx.doi.org/10.1145/1347390.1347414

Ron Yeh, Chunyuan Liao, Scott Klemmer, François Guimbretière, Brian Lee, Boyko Kakaradov, Jeannie Stamberger, and Andreas Paepcke. 2006. ButterflyNet: A Mobile Capture and Access System for Field Biology Research. In *Proc. CHI 06*. ACM, New York, NY, USA, 571–580. DOI:http://dx.doi.org/10.1145/1124772.1124859

Yuhang Zhao, Yongqiang Qin, Yang Liu, Siqi Liu, and Yuanchun Shi. 2013. QOOK: A New Physical-virtual Coupling Experience for Active Reading. In *Proc. UIST '13 Adjunct*. ACM, New York, NY, USA, 5–6. DOI:http://dx.doi.org/10.1145/2508468.2514928