AudioCanvas: Internet-Free Interactive Audio Photos

Simon Robinson, Jennifer Pearson, Matt Jones
Future Interaction Technology Lab, Swansea University, SA2 8PP, UK
{ s.n.w.robinson, j.pearson, matt.jones } @swansea.ac.uk

ABSTRACT
In this paper we present a novel interaction technique that helps to make textual information more accessible to those with low or no textual literacy skills. AudioCanvas allows cameraphone users to interact directly with their own photos of printed media to receive audio feedback or narration. The use of a remote telephone-based service also allows our design to be used over a standard phone line, removing the need for data connections, which can be problematic in developing regions. We show the value of the technique via user evaluations in both a rural Indian village and a South African township.

Author Keywords
QR codes; camera phones; audio; developing regions.

ACM Classification Keywords
H.5.2 User Interfaces: Input devices and strategies; Interaction styles; H.5.1 Multimedia Information Systems: Audio I/O.

INTRODUCTION
In many developing, rural regions, low textual literacy rates prevent large proportions of the population from accessing information from flyers, posters, maps or other text-based media. In addition to this, while mobile phones are relatively widespread, internet or data connections, and hence online information access, are not widely the norm. There are several reasons for this, varying from region to region. In some places, the pricing structures and business models are not attractive to users. Users may also appreciate "internet-free" methods due to the lack of a reliable data signal, or their greater comfort and familiarity with making conventional telephone calls. As a result of both low textual literacy, which prevents people from reading printed media, and poor internet connectivity, which prohibits online translation or information enquiry, accessing information in many underdeveloped regions can prove exceedingly difficult.

To address these issues, several voice-based telephone services have been proposed (such as [5]). These audio-only tools are intended to provide spoken information in local languages, accessible via low-end devices. However, the use of purely audio makes it difficult to, for example, get an overview of a voice space, navigate through disparate content regions, or quickly locate a certain item. Printed media, on the other hand, is entirely visual, providing simple overviews and navigation, but has clear issues with regards to textual literacy.

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

We saw an opportunity to connect these two spaces in a novel way, taking advantage of the fact that while access to internet-connected devices is minimal, in many areas people own at least a low-end cameraphone or featurephone, and smartphones are increasingly price accessible. This popularity of camera-capable phones, coupled with local-language voice services, offers potential for creating interactive audio photos.

AudioCanvas is a fusion of Interactive Voice Response (IVR) services and physical objects, as illustrated in Fig. 1. Our design allows users to take a photograph of printed media and use the picture as a canvas for interaction with related audio. The system automatically dials a remote telephone-based voice service, and allows users to touch areas of interest on their own photo—titles, captions, adverts or images, for example—to hear audio narration for their selection. Our 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.

BACKGROUND
Previous work has augmented physical artefacts with digital media. DigitalDesk [12], for example, used precisely positioned projectors and cameras to retain paper's affordances but allow documents to be manipulated digitally. More recently, Mistry et al. [7]'s pico projector system allowed objects to be queried for information by framing them with coloured finger tags. Other designs, such as Audio d-touch [2] used custom markers to create user-friendly tangible interfaces. Unlike AudioCanvas, these examples focus on *direct* interactions with objects; e.g., manipulating physical items to control digital interactions. Our approach differs, letting people use their

own photos of items to act as canvases to interact with audio. Other augmentation designs, such as Listen Reader [1], OCR methods, or commercial options,[1] have required either a *client-side* metadata database or realtime internet access for their functionality. A major advantage of our design, then, is that the client is completely independent of the media used. This independence—particularly relevant in the regions where we created AudioCanvas—means that updates or internet access are not required to support additional items in the future.

Previous work has used digital codes for partially automating paper-based tasks. Parikh et al. [8] focused on QR codes in rural developing contexts, allowing form fields to be scanned one-by-one to enter data. Klemmer et al. [4] used barcodes in page margins to recall audiovisual interview transcripts. However, these approaches required preloaded content, unlike our artefact-independent method. Snap 'n' Grab [10] provided South African taxi users with access to media via photos, having similar internet-free goals to our design. The prototype uses barcodes around a printed icon, which are photographed and sent via Bluetooth to a server phone (also in the taxi), which then returns relevant media. Our approach does not require a local server, as all communication is via a normal telephone connection to a remote voice service.

Interactive photos

Our design turns photographs into interactive touch canvases for communicating with telephone-based services. Frohlich [3]'s related technique allowed interaction with paper photos to request audio. The design supported only one track per image, however, and also required a separate projector. Suzuki et al. [11] demonstrated the use of photos of physical objects as interaction tools, but required a marker on *each* object (cf. [8]). More similar to our design is Seifert et al. [9]'s technique which turns hand-drawn sketches into interactive prototypes. However, our approach focuses on audio interaction, rather than a conversion of photos into digital facsimiles.

Many mobile augmented reality applications are able to add overlays or web links to physical artefacts. Daqri,[2] for example, uses QR codes for tracking; others, such as Blippar[3] or Shortcut,[4] have used image recognition. However, these approaches require a data connection and a large download for every scan. Furthermore, they focus entirely on augmenting a photo or camera preview with visual content. Our system uses QR codes to encode a coordinate grid and phone number on a photographed object, instead. The aim is to add audio to the experience—without requiring an internet connection—rather than just providing a different way to enter a website address.

AUDIOCANVAS

AudioCanvas affords rich experiences with physical objects, allowing any marked-up item to become a touch panel to interact with remote audio resources (see Fig. 1). Our approach uses two precisely placed QR codes on printed media to support detection of the position of the user's selections on their photo of an item. To interact with the system, the user positions their phone to take a photo. 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 feature phone) passes the coordinates to the voice service, which plays back the appropriate audio immediately, removing the need for IVR navigation. A video of the technique accompanies this paper.

[1]leappad.com; [2]daqri.com; [3]blippar.com; [4]kooaba-shortcut.com

Figure 2. Using AudioCanvas. Left: (1) translating instructions; (2) finding out about a movie from its poster; (3) listening to a newspaper. Right: the interaction afforded – touching the photo to hear the audio.

Figure 2 shows the audio browsing interaction afforded, and our prototype in use in several scenarios. The system was designed specifically for use where internet access is sparse or unaffordable, and where low textual literacy levels prevent people from reading printed media. The interaction afforded clearly resonates with Medhi et al.'s design guidelines for interfaces for non-literate and semi-literate users (see [6]), allowing audio-based captioning on demand without any connection beyond a standard phone line.

AudioCanvas is not meant to be a competitor to automatic translation systems. This is partially because such services can require extensive online interaction. More importantly, though, the aim here is to provide a technique through which the content provider or community members can produce appropriate interpretations or comments to supplement the media item in both textual and visual elements. For example, touching on form fields might explain in useful terms the reason this data is required and what it might be used for, in the same way a proximate might guide in person.

Implementation

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

Our novel client design uses two separate QR codes on printed media to detect the position of the interactive area within a photo taken by the user. The codes are positioned at opposite corners of the item – 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

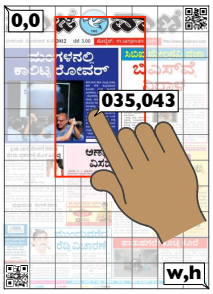

Figure 3. The AudioCanvas coordinate system. Touch point coordinates on the user's photo of an object are calculated based on the distance between and relative sizes of the two QR codes detected on the item. The resolution of the coordinate grid is high enough to allow zooming in to the image and very precise selections, if necessary. Touching a point causes a six-digit DTMF-encoded request for the relevant audio to be sent to the voice site. The audio is played over the phone line in response. The grid is not visible during usage (see Fig. 2).

(e.g., the issue and page number of a newspaper). The top right code is used for coordinate calibration and image alignment (we automatically straighten and skew correct the image).

Figure 3 shows the layout of marked-up objects, and illustrates how touch coordinates are interpreted. Based on the sizes and positions of the two QR codes, the exact position of a touch on the photo can be converted to a location on the physical object without the client application requiring any knowledge about the object itself. This design allows almost any object to be used for interaction, with the system able to calculate 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 DTMF tones to the remote telephone service indicated by the lower left QR code. The voice service then plays the relevant audio in response.

Our initial prototype was created using an Android phone, allowing direct touch interactions with photos. It caters directly to the increasing number of users in developing regions who have or will own smartphones. The use of QR codes, rather than more complex image recognition, means that lower-end feature phones are capable of providing AudioCanvas interaction. A feature phone compatible J2ME version of the design is in development. While this type of handset does not offer the same affordances as a smartphone (i.e., directly touching photos), the image is easily navigable via the phone's joypad. We have also developed a mobile tool to allow easy creation of new items by, for example, NGOs or companies who wish to mark-up their printed material or products with AudioCanvas content. This tool works in a similar way to the client, and allows product designers or authors to highlight parts of the photo to create an audio area. Audio annotations can then be created on the fly, or added from a local file and sent to the voice server via the phone line.

EVALUATION

We conducted two user evaluations in separate locations to trial the AudioCanvas technique, aiming to test over a range of literacies (both textual and technological). Our main aim in trials was to observe usage of the system and gather qualitative feedback from users. Experienced local researchers (who were not associated with the system) were employed in both cases to manage studies and translate information from participants.

The first study took place in a village near Devrayandurga, Karnataka, India, with Kannada-speaking participants who were textually illiterate and largely unfamiliar with mobile devices. Twenty-two participants (11M; 11F, aged 22–65) were recruited for individual 30 min trials in this location.

The second study took place in Langa, a township in Cape Town, South Africa, with isiXhosa-speaking participants who had low-to-moderate literacy in isiXhosa, but none in English. Thirty-six participants (6M; 30F, aged 18–45) were recruited for six 60 min group trials (4–8 participants per session). Participants requested focus group evaluations in Langa as this setup was more comfortable – many came to the session with friends or other family members, and were reluctant to participate individually. We provided four AudioCanvas-enabled phones during these sessions, feeling that group-based participation was the best compromise to allow us the opportunity to evaluate the system with this potential user base.

Media

The media items used in the Indian study were a phone bill and a map of Bangalore. Text was in both English and Kannada (participants could read neither) and audio was given in Kannada. Media used in the South African study included a newspaper, product packaging and several local flyers. These items were written in English, with audio provided in isiXhosa.

The items used in both evaluations were of various physical sizes, from A3 to A6, 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 areas. It is important to note that none of the objects contained indications as to where audio was available. That is, with the exception of the two QR codes, the original media was unchanged, and participants had to discover audio areas independently.

Procedure

Studies began with a demonstration of the system (individually in India; groups in South Africa). Participants were not familiar with smartphones, so part of this demonstration involved a basic introduction to touchscreens. Following this, participants were asked to use AudioCanvas to take photos of the media items and interact with the audio content, exploring for as long as they wished. No specific tasks were given; we aimed to allow exploration rather than forced usage. At the end of the study, participants rated usefulness (1–10; 10 high – a familiar feedback scale in both cases) and ease of use (five attributes – see Table 1; India only, due the group-based setup in South Africa). The study concluded with a semi-structured interview.

Results

All participants enjoyed using the system, and many expressed a desire to be able to have the app pre-installed on the next generation of their phones. Feedback given by participants indicated a strong appreciation of the design, and shows its potential usefulness in regions where textual and technological literacy levels are low. The average usefulness scores given were 8.1 (s.d. 1.90) in India and 8.4 (s.d. 2.07) in South Africa.

These results are strengthened further by user comments during the tasks. For example, participants in the Indian evaluation commented: *"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*

Ease of use, in terms of:	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 information back from each section	6.3 (1.2)
Finding the audio areas in a photo	5.1 (2.0)

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

could not have accessed." Participants in the South African group studies felt similarly: *"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," "it works like music, now I can listen to articles"* and *"it's perfect and easy."*

Over both studies, all except one participant said they regularly found themselves in a situation where they needed information they could not read. When in this situation, participants confirmed that they either asked someone else (e.g., a friend or other proximate) to read text for them, or were not able to read it at all. In post-study interviews, all participants suggested that AudioCanvas would be useful for people unable to read. Other scenarios were also given – for example, after using AudioCanvas, one noted its usefulness for helping with sensitive information, such as when filling in a private form without the need to resort to others; and several participants pointed out opportunities for helping with language learning.

Participants in the Indian trials also gave ratings for ease of use (see Table 1). The average for three of the five scores was high, in excess of 6 out of 7. However, due to a lack of familiarity with cameraphones, some participants did find it difficult to frame and take the initial photograph, causing lower scores for this attribute, and comments such as: *"getting the barcodes in view is challenging"* and *"[it was] difficult for me because I couldn't fit in the barcodes."* Many participants tried framing the image in a skewed or slanted manner, only just managing to accommodate the QR codes. The majority of these initial difficulties with the camera were overcome during the study with additional time spent using the application.

The ease of locating audio areas in a photo was also ranked lower on average. This result was somewhat to be expected, as there were no explicit visual indications of where audio could be found on any of the items. Participants found audio areas more difficult to locate on the map due to the lack of obvious places for feedback. The phone bill, with its structured layout, and more familiar context, was more engaging, and participants eagerly explored photos for audio. Given that there were no extra visual indications of audio areas, the rating given is very encouraging. Similar behaviours were also observed in South Africa, with participants keenly exploring local flyers, but less able to find audio on more image-heavy items.

DISCUSSION AND CONCLUSIONS

In this paper we have presented AudioCanvas – a novel photo interaction method that pairs audio with printed media. The design allows users with low textual literacy levels to interact directly with self-taken photos of physical items by touching regions on-screen. A remote voice service provides audio content, accessed via a standard phone line to ensure it can be used without a potentially costly network data connection.

Our evaluations with both Indian and South African participants who have low technology familiarity, and mixed levels of generally low literacy, show promising results for the ease of use and popularity of our design. The majority of participants enthusiastically embraced the AudioCanvas design, and saw value in the concept despite some initial difficulties in using the technology. Comments made with regards to the usefulness and novelty of the interface were also extremely encouraging. This evidence suggests that AudioCanvas can provide users who are unable to read text—either due to literacy or language issues—an alternative means of access to printed media via mobile phones. We believe, as illustrated by our experiments, that the technique could be particularly beneficial to those in impoverished areas where low-literacy is common, and where data connections are not yet available or attractive.

ACKNOWLEDGMENTS

This work was funded by EPSRC grant EP/J000604/2. We thank Ram Bhat and Minah Radebe for managing studies.

REFERENCES

1. Back, M., Cohen, J., Gold, R., Harrison, S. and Minneman, S. Listen reader: an electronically augmented paper-based book. In *Proc. CHI '01*, ACM (2001), 23–29.
2. Costanza, E., Giaccone, M., Kueng, O., Shelley, S. and Huang, J. Ubicomp to the masses: a large-scale study of two tangible interfaces. In *Proc. UbiComp '10*, ACM (2010), 173–182.
3. Frohlich, D. Audiophotography. Springer (2004).
4. Klemmer, S., Graham, J., Wolff, G. and Landay, J. Books with voices: paper transcripts as a physical interface to oral histories. In *Proc. CHI '03*, ACM (2003), 89–96.
5. Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S. and Nanavati, A. WWTW: the world wide telecom web. In *NSDR workshop, SIGCOMM '07*, ACM (2007).
6. Medhi, I., Gautama, N. and Toyama, K. A comparison of mobile money-transfer UIs for non-literate and semi-literate users. In *Proc. CHI '09*, ACM (2009), 1741–1750.
7. Mistry, P., Maes, P. and Chang, L. WUW – wear ur world: a wearable gestural interface. In *Proc. CHI '09: Extended Abstracts*, ACM (2009), 4111–4116.
8. Parikh, T., Javid, P., K, S., Ghosh, K. and Toyama, K. Mobile phones and paper documents: evaluating a new approach for capturing microfinance data in rural India. In *Proc. CHI '06*, ACM (2006), 551–560.
9. Seifert, J., Pfleging, B., Hermes, M., Rukzio, E. and Schmidt, A. Mobidev: a tool for creating apps on mobile phones. In *Proc. MobileHCI '11*, ACM (2011), 109–112.
10. Smith, G. and Marsden, G. Providing media download services in African taxis. In *Proc. SAICSIT '11*, ACM (2011), 215–223.
11. Suzuki, G., Aoki, S., Iwamoto, T., Maruyama, D., Koda, T., Kohtake, N., Takashio, K. and Tokuda, H. u-Photo: interacting with pervasive services using digital still images. In *LNCS*, vol. 3468. Springer (2005), 190–207.
12. Wellner, P. Interacting with paper on the digitaldesk. *Communications of the ACM* 36.7 (1993), 87–96.