

AudioCanvas: Interactive Audio Photos

Simon Robinson, Jennifer Pearson, Matt Jones (Swansea University, UK)

© Simon Robinson, Jennifer Pearson, Matt Jones 2014. This is the authors' version of the work. It is posted here for your personal use. Not for redistribution. The definitive version was published in *Advances in Printing and Media Technology* XLI(I):133–136 (Proceedings of IARIGAI '14).

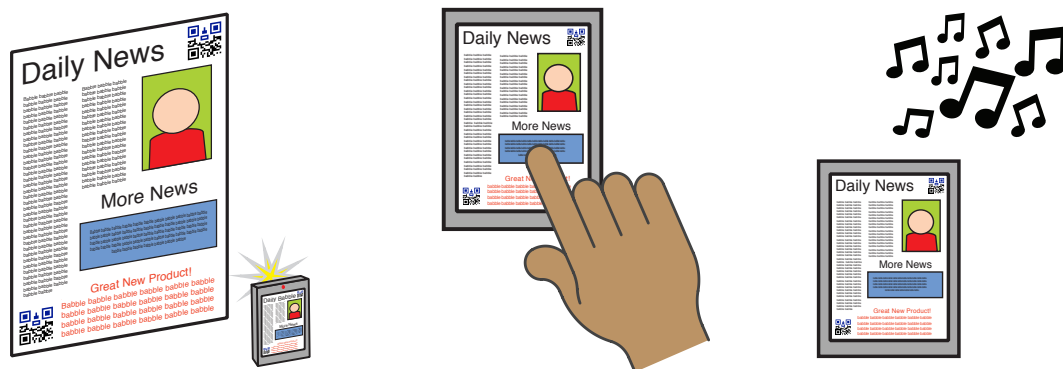
Abstract

We present a novel interaction technique that helps to make printed information more accessible to those with low or no textual literacy skills. AudioCanvas allows cameraphone users to interact directly with their own photos of media items to receive audio feedback or narration on demand.

Keywords: QR codes; camera phones; audio.

1. Introduction

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 the need for an internet connection.



1. Take a photo of the item using the phone's camera
2. Use the photo as a canvas to select areas of interest
3. Listen to information about the selected region

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.

2. Background

Previous work has augmented physical artefacts with digital media. DigitalDesk (Wellner, 1993), for example, used precisely positioned projectors and cameras to retain paper’s

affordances but allow documents to be manipulated digitally. Other designs, such as Audio d-touch (Costanza et al., 2010) 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 (Back et al., 2001), 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 means that updates or internet access are not required to support future media items.

2.1 Interactive photos

Our design turns photographs into interactive touch canvases for communicating with telephone-based services. Frohlich's related technique (Frohlich, 2004) 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. (Suzuki et al., 2005) demonstrated the use of photos of physical objects as interaction tools, but required a marker on each object (cf. Parikh et al., 2006). More similar to our design is Seifert et al.'s technique (Seifert et al., 2011), 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 Blippar³ 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, rather than just providing a different way to enter a website address.

¹ e.g., leappad.com; ² daqri.com; ³ blippar.com; ⁴ kooba-shortcut.com

3. 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 first positions their phone to take a photo of an item. When both codes are detected, a photo is taken automatically, and the phone dials a 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.



Figure 2. Example AudioCanvas media. Clockwise, from top left: curry packet; newspaper; box of tea; movie poster.

Figure 2 shows several example media items that we have tested for use with AudioCanvas. 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.

AudioCanvas is not meant to be a competitor to automatic translation systems, however. 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 information is required and what it might be used for, in the same way a proximate might guide in person.

3.1 Implementation

AudioCanvas has two components: a standard remote voice service (not covered in this paper) and a local client. 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 (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.

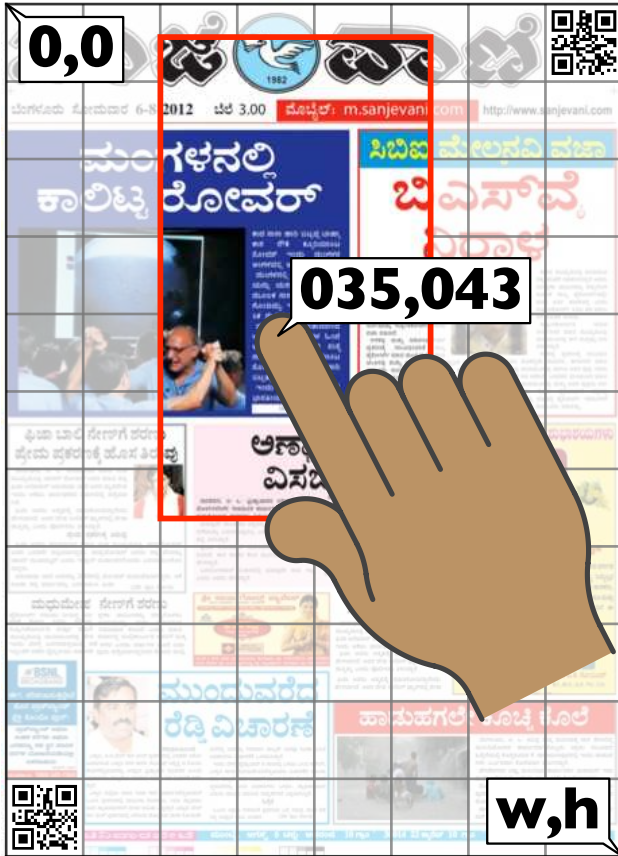


Figure 3. The AudioCanvas coordinate system. Touch point coordinates on the user's photo of an object are calculated based on the distance between and relative sizes of the two QR codes detected on the item. The resolution of the coordinate grid is high enough to allow zooming in to the image and very 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.

4. Discussion and Conclusions

In this paper we have presented AudioCanvas – a novel photo interaction method that pairs audio with printed media. The design allows interaction 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. We believe, as illustrated by our experiments detailed elsewhere (e.g., Robinson et al., 2014), 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.

Acknowledgements

This work was funded by EPSRC grant EP/J000604/2.

References

M. Back, J. Cohen, R. Gold, S. Harrison and S. Minneman (2001). 'Listen reader: an electronically augmented paper- based book'. In: Proc. CHI '01, ACM, pp. 23–29.

E. Costanza, M. Giaccone, O. Kueng, S. Shelley and J. Huang (2010). 'UbiComp to the masses: a large-scale study of two tangible interfaces'. In: Proc. UbiComp '10, pp. 173–182.

Frohlich, D. (2004). Audiophotography. Springer.

T. Parikh, P. Javid, S. K, K. Ghosh and K. Toyama (2006). 'Mobile phones and paper documents: evaluating a new approach for capturing microfinance data in rural India'. In: Proc. CHI '06, ACM, pp. 551–560.

S. Robinson, J. Pearson and M. Jones (2014). 'AudioCanvas: Internet-free Interactive Audio Photos'. In: Proc. CHI '14. CHI '14, ACM, pp. 3735–3738.

J. Seifert, B. Pfleging, M. Hermes, E. Rukzio and A. Schmidt (2011). 'Mobidev: a tool for creating apps on mobile phones'. In: Proc. MobileHCI '11, ACM, pp. 109–112.

G. Suzuki, S. Aoki, T. Iwamoto, D. Maruyama, T. Koda, N. Kohtake, K. Takashio and H. Tokuda (2005). 'u-Photo: Interacting with Pervasive Services Using Digital Still Images'. In: LNCS. Vol. 3468. Springer, pp. 190–207.

P. Wellner (1993). 'Interacting with paper on the DigitalDesk'. In: Communications of the ACM 36.7, pp. 87–96.