Conference contribution :

Pearson, J., Robinson, S. & Jones, M. (in press). *PaperChains: Dynamic Sketch+Voice Annotations.*(pp. 383-392). ACM.

http://dx.doi.org/10.1145/2675133.2675138

PaperChains: Dynamic Sketch+Voice Annotations

Jennifer Pearson, Simon Robinson, 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 interface for collaborative creation of evolving audio-visual documents. *PaperChains* allows users to sketch on paper and then augment with digital audio, allowing both the physical and digital objects to evolve simultaneously over time. The technique we have developed focuses on affordability and accessibility in its design, using standard cameraphones and telephone connections, which allow it to be used in regions where literacy, technological experience and data connections cannot necessarily be taken for granted. The main use-case that we focus on in this paper is for collaborative storytelling, an area which has been well studied and previously proven to be of value in resource constrained environments. To investigate the relevance of the approach in these contexts, we undertook two usability evaluations in India and South Africa. Results from these investigations indicate users' ability to both create and interpret stories using the software, as well as demonstrating high overall usability and enjoyment. We end with a discussion of the implications of our design and opportunities for use in other contexts.

Author Keywords
Sketching; audio; photographs; cameraphones; resource-constrained environments.

ACM Classification Keywords
H.5.3 Group and Organization Interfaces: Collaborative computing; H.5.2 User Interfaces: Interaction styles; H.5.1 Multimedia Information Systems: Audio input/output.

INTRODUCTION
In this paper we present PaperChains – a novel method of connecting paper-based material with digital audio using low-tech, easily accessible equipment. We make use of standard cameraphone handsets to photograph physical content, and then allow the photographed item to be augmented with digital audio in multiple locations via interaction with the image. Unlike previous approaches, the design supports evolving physical-digital narratives, allowing sketches, memorabilia, photos and other physical objects to be built up with audio content over time. The technique is unique in that it also allows modification of the underlying physical content, without compromising on recognition – something that similar approaches (e.g., [6, 12, 29] or augmented reality apps such as Blippar or Aurasma) do not support. The core contribution of the paper, then, is the provision of a lightweight technical platform for evolving audio-physical documents, with a wide range of uses.

Our approach uses two precisely placed QR (quick response) codes printed on paper, allowing the PaperChains system to detect the item, its orientation, and its dimensions, with any camera-enabled mobile handset. The system then corrects the photograph for skew and rotation, and lets users add audio snippets anywhere on the image. This method ensures the approach requires no additional specialist hardware (such as a dock or camera-pen) – users interact directly with the photograph using the phone as a proxy. Further, our technique has the additional advantage of being able to utilise an Interactive Voice Response (IVR) system over a standard telephone line as the audio content store, in contrast to the internet connectivity or local private database used by other approaches.

The PaperChains technique has promise in many different scenarios, as we discuss in more depth later. However, we feel it has a particularly strong use case in resource constrained environments. Its appropriateness can be seen in both the interactions it affords and in its low technical requirements, making it an attractive tool for regions where literacy, technological familiarity, power and data connectivity can be low. In particular, using a standard telephone connection to connect with a remote voice server ensures that PaperChains can be used with existing devices and infrastructure. Previous work in this area (e.g., [13, 21]) has shown that IVR systems are a valuable resource, as they are easy to adopt and compatible with any handset. Our technique aims to follow this same strategy by requiring a minimum of technology and experience from the user. PaperChains runs on standard mobile handsets – we studied the design using an Android smartphone, and have also developed a feature-phone version. During a recent ethnographic study in India, we investigated the costs of low-end smartphones in a slum in Mumbai, finding a wide range of so-called "China phones" that support the system, available for as little as $15–35 USD (see Fig. 1 for a sample).

We see one particular application—collaborative storytelling—as an especially attractive and appropriate use for the technique. Previous work has investigated the use of storytelling in resource constrained contexts, focusing in particular on mobile story capture (e.g., [8, 23, 26]). A common theme that emerges from this research has been a desire for collaborative authoring, with a particular focus on the importance of audio. Our approach enables these types of interactions, and makes it possible to build up digital-physical narratives over time. Consider the following scenario, which illustrates the approach:

Twelve year-old Chitra lives in a remote part of rural Gujarat with her parents and younger siblings, but her older brother, Nirav, has recently moved to New Delhi—a 20-hour journey away—to work. Chitra misses him badly, and wants to share some of the recent family experiences with him in a personal way; but neither of them can afford to visit each other often, and she does not have an internet connection. She begins drawing a story, then takes a photograph of the page with her phone, using the on-screen tool to add audio hotspots over each section of her drawing. After recording voiceovers for her pictures she passes the paper on to her sister, Aditi, who draws several sketches. Aditi then takes her own photograph of the page and begins adding audio. Finally, their brother Ram adds his own content. Once the story is complete, Chitra puts it in an envelope and posts it to her brother.

A few days later, Nirav receives the letter. After taking his own photograph of the document, he can hear audio recordings of their voices by touching on each section of the image on-screen, thus preserving both the paper and spoken records of the family story. He can also add his own content and send the letter back to his family.

This scenario reflects the desire of many migrant workers for a sense of connectedness, as reported in [14]. In the context of resource constrained environments such as this example, then, we argue that PaperChains is a novel and compelling way of collaborating to combine the story emerging on paper with the story emerging in the audio.

The rest of this paper is organised as follows. We first describe previous research related to our approach, then detail the technical implementation of our design. To explore the system in resource-constrained environments, we carried out two usability-focused evaluations in India and South Africa. These studies illustrated the value of the approach by demonstrating users' ability to both create and interpret stories (such as the one described in the scenario), as well as showing an appreciation of the system and overall enjoyment of the experience. We end with a discussion of the work, future directions and other applications for the system.

BACKGROUND

The key previous research area related to PaperChains involves augmenting images with sound, animations and all sorts of multimedia. These approaches, now seen in commercial apps, work by taking a picture of an object or a document, processing it to find key elements that allow it to be recognised quickly, and uploading these discriminants to a server. When another picture of the same object is taken, the system can search its database and identify the object and the annotation. However, unlike our design, changes to the original item can mean subsequent attempts to recognise it will fail. These approaches, then, are unable to support PaperChains' core functionality.

Figure 1. A set of low-end touch-screen smart phones that are capable of running PaperChains. These phones were purchased for $15-35 USD each from a shop in Dharavi, a slum in Mumbai.

Augmenting paper with digital content

Many research systems have been developed that attempt to close the gap between the affordances of paper and the benefits of digital content. Early work, such as the DigitalDesk [31], allowed the properties of paper to be preserved while supporting digital annotation and manipulation. More recent work in this area has lowered hardware requirements, adopting technology such as Anoto[1], which uses dot-marked paper and a camera below a digital pen's nib to recognise document areas while writing. Other work has investigated how traditional methods of interaction could be replicated digitally (e.g., [17]), or, how augmentation of physical objects could provide additional information—see, for example, the audiophotography work of Frohlich [7], or the many commercial mobile apps that are focused around adding digital content to physical items.

Since the DigitalDesk, which used a projector and camera above a desk, more portable designs (such XLibris [27]) have allowed users to scan in paper documents and mark them digitally with freeform annotations on a tablet-like device. Later approaches, such as that of Guimbretière [10], lowered the hardware requirements for digital-physical annotation systems, making use of Anoto-marked paper to support notes that could be replicated digitally. Guimbretière 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 PaperChains, 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, Parikh et al. [20] added QR codes next to form fields to improve data entry in rural India; Seifert et al. [28] turned photographs of interface designs into interactive prototypes; and, Erol et al. [6] used image recognition of a document and comparison to a ground truth version to detect the regions of a page in a photograph.

These methods are similar to those taken by the many augmented reality applications that add interactive overlays to physical objects, such as Aurasma[2], Blippar[3] and Daqri.[4] These apps overlay a realtime camera preview with digital content associated with the object in the frame. Other designs for this form of interaction have aimed to add more interactivity to the process – for example, Mistry et al.'s pico projector-based

[1]See: anoto.com; [2]aurasma.com; [3]blippar.com; [4]daqri.com

design [19] allowed users to frame objects with their fingers while wearing coloured tags to request associated interactive digital content. However, none of these types of approaches allow the underlying document to change or evolve while at the same time preserving the annotations that have been added – a key difference between these methods and our design.

Annotating with audio

In this work we focus on linking paper documents with audio annotations. Previous work in this space ranges from immersive audio story books (e.g., [1]) to the audio greetings cards and interactive picture books that have been commercially available for many years. These designs commonly use buttons or simple sensors to start sound playback. Various research approaches have been taken to synchronising audio recordings with physical documents automatically. Audio d-touch [3], for example, used 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 Audio Notebook [30] 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. [5] 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. PaperChains is clearly related to audiophotography (explored in depth by Frohlich [7]), 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. [7]), to using overhead image recognition for selecting and browsing audio associated with printed photos (e.g., [9]). Audiophotographs traditionally associate only one annotation with each photograph, however; and, more importantly, the audio track is associated with the photograph 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. Klemmer et al. [12] 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. [32] used Anoto-marked paper in a scrapbook, supporting various user-drawn symbols to associate audio and other content with scrapbook items. Liu et al. [16] 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. In a more resource constrained context, Smith and Marsden [29] created a client-free mobile platform for media retrieval based on taking a picture of an image representation that was then recognised and returned to the requester via Bluetooth. Our approach is adapted from that taken in [25], which used digital markers to indicate the edges of the content area. However, our design does not focus on retrieving audio content from products and posters; instead, we support document annotation and audio narration in a flexible, dynamic manner that allows both the document and annotations to be changed over time.

	Affordable, accessible	*Modifiable documents*	*Modifiable audio*	*Collaborative authoring*	*Flexible item sizes*	*No data connection*	*Proximate sharing*	*Remote sharing*
PaperChains	✔	✔	✔	✔	✔	✔	✔	✔
Img. recog.	✔	✘	✔	✘	✔	✘	✔	✔
Pen-based	✘	✔	✔	✘	✔	✔	✘	✘
Tablet-based	✘	✔	✔	✘	✘	✔	✔	✘
Table-based	✘	✔	✔	✔	✘	✔	✔	✘

Table 1. A comparison of features between PaperChains and several general classes of previous work. Each broad category is illustrated by, for example: Image recognition (Aurasma[2]); Pen-based [32]; Tablet-based [22]; Table-based [2].

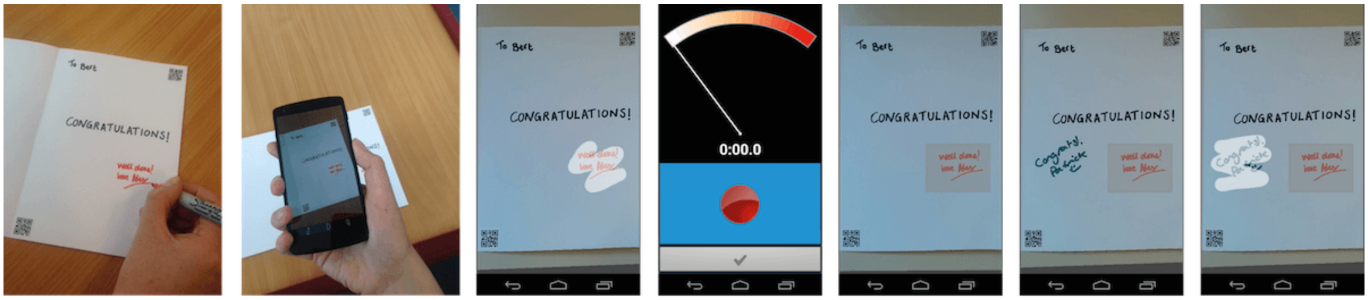
Annotation as storytelling

We believe the PaperChains design shows strong potential for interactive, evolving physical-digital storytelling. Previous work in this area includes systems such as KidPad [4], which extended existing sketching software to create a storytelling environment, or PicoTales [24], 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, PaperChains lets physical sketches or other objects and digital elements coexist as part of the narrative.

Cao et al.'s TellTable [2] 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 [11], who used conductive ink to allow sketches to be augmented with digital content; de Lima et al. [15], who used pen and paper sketches to insert characters into a virtual world; or, Wood et al. [33], whose barcode scanning approach linked digital tales to physical books. Yeh et al.'s approach [34] 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. [22], who used a drawing tablet and 'stamp,' to associate audio recordings with sketches on the page. However, their design required both a custom hardware package, and supported only local playback of the content.

A comparison of PaperChains to previous work

PaperChains is a lightweight platform for evolving, collaborative, audio-physical documents, and thus differs from previous work in several ways. Table 1 compares the PaperChains approach with general examples of previous work. The approach is *affordable* and *accessible*, which we define as being compatible with existing or low-cost devices and suitable for users with low technology experience. PaperChains supports *collaborative authoring* – audio is created by users of the system (rather than content producers); and, interaction with the audio uses a photograph of the page that is content-agnostic. Because of this, both the *documents* and the *audio* are modifiable, unlike most other systems. The few other approaches that do allow modification require specialised hardware (e.g., Anoto pen or overhead camera), which prevents easy sharing of the content, and makes group-based creation



(a) (b) (c) (d) (e) (f) (g)

Figure 2. The PaperChains system, illustrating the process of scribbling and recording audio hotspots. From left to right: (a) annotating on paper; (b) taking a photograph of the document; (c–e) adding audio; (f–g) other users adding additional audio.

and collaboration more difficult. PaperChains supports *flexible item sizes* and multiple annotations on the same item, whereas many other methods (e.g., most augmented reality apps) are limited to a single position for an annotation on any one physical object. It also needs *no data-connection*, because the system is able to utilise IVR systems for content storage. Finally, the technique supports both *proximate* and *remote sharing* of audio-augmented documents, which, as Table 1 shows, is offered by only one of the alternative methods.

TECHNICAL FEATURES AND IMPLEMENTATION

PaperChains provides a simple way to merge paper and audio annotations, by using a photograph of the annotated document as a proxy for adding and browsing audio hotspots. The approach affords rich interactive experiences with physical items, allowing any marked-up object to be augmented with audio. PaperChains documents can be created or edited by any user, allowing collaborative and shareable audio annotations.

Creating PaperChains documents

The process of creating a PaperChains document begins with a user adding a physical annotation—anything from sketching to writing to pasting items—on a QR code-augmented object (Fig. 2a). Then, using a standard smartphone, they take a photograph of the entire object, making sure to include both QR codes in the frame (Fig. 2b). The picture is captured automatically once both QR codes are detected. This photograph can now be used as a collage to add audio to the physical annotations. Audio is added by 'scribbling' over the part of the object that the user wishes to narrate (Fig. 2c and 2d), creating a bounding-box audio 'hotspot'. Once audio recording has finished, the new audio area is visible on the photo (Fig. 2e).

Further hotspots can be added, or other users can add their own physical and digital annotations at any time, by repeating this procedure. The process is identical for every user – when a photo is taken, the system recognises the same document as before, and allows access to previous users' annotations (Fig. 2f). Others can add their own audio messages to the canvas via the same scribbling technique (Fig. 2g). This process can be repeated as often as required.

Browsing PaperChains documents

To listen to the audio associated with PaperChains documents, the same process is used – first a photo is taken of the document. Touching anywhere on the photograph plays the audio for that location. When listening to the audio there are no visual cues as to where the hotspots are located, to avoid interfering with the natural browsing and exploration of the document. Users wishing to interpret someone else's Paper-Chain document do so in an exploratory manner, discovering audio areas by touching different regions of the photograph.

Technical implementation

PaperChains consists of two main components: a client application and a remote telephone service. The client application is used to take the initial photograph and provide tools such as pan, zoom, selection and editing of the audio content. The remote telephone service is an IVR system that stores the audio recordings and all information associated with the underlying physical object. One particularly attractive aspect of this setup is that the client application is completely independent of the information with which it is interacting – all data, including the audio recordings, are stored on the telephone server and retrieved via DTMF requests over the phone line (see below). This ensures that no internet connection is required for listening, editing or retrieving updated audio recordings.

The main technical aspect of the system, which allows the flexibility of photographing and augmenting items without any additional hardware, is the placement of the two QR codes on the physical item. These QR codes, which are placed in the corners of PaperChains-augmented objects, identify the item and allow boundary detection and image correction, in the same way as [25]. When a user holds the camera over an item, the system takes a photo automatically as soon as it detects the two QR codes. The image is corrected for skew and rotation, and the object's dimensions are inferred according to a simple grid coordinate system. The grid system is based on the size and positions of the QR codes in the image, as in the example shown in Fig. 3. This grid system is then used to determine the exact position of a user's touch on the photo. This method makes recognition quick and accurate, and—more importantly—allows the underlying document to be modified while still being recognisable.

Once a photograph has been taken and the dimensions calculated, the client application decodes the QR codes' information to retrieve an IVR telephone number and identifier unique to the item. It then calls this number to connect to the remote

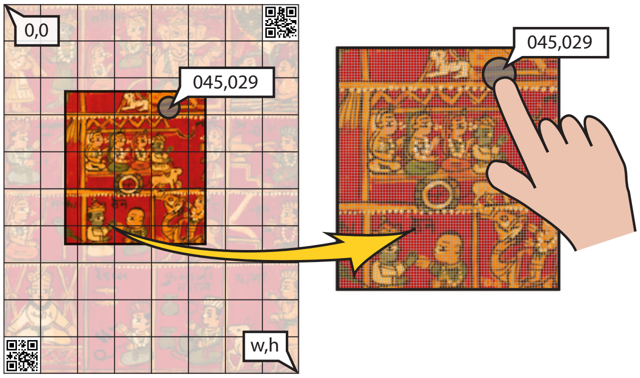


Figure 3. The grid system used by the PaperChains technique. The QR codes detected on the item are used to calculate a coordinate grid from the upper left [0, 0] to the lower right [width, height] of an object, as shown in the overlay (the grid is not visible during usage). Touch point positions are calculated based on this grid. When the image is touched, a six-digit DTMF-encoded request is sent to the server over the phone line, and the requested audio is played in response. As shown in the inset image, the resolution is high enough to allow very precise touch positions.

service. This live telephone call is now "listening" to any touches made on the photograph. Touching a position on the photograph triggers the application to determine the exact coordinate of the touch, convert this into a DTMF tone, and play it down the live telephone call, much like a touch tone keypad. The listening server then interprets this information and performs the correct action. For example, if the application is in creation mode, it will ask the user to narrate an audio snippet for that location, and associate this with the location selected. Similarly, in browsing mode it will check to see if there are any audio recordings in that location, then play them in real time down the phone line.

Viability in resource constrained environments

Table 1, discussed earlier, compared the benefits of the PaperChains approach with general examples of previous approaches in this area. Clearly, there are further considerations we must be aware of when designing for resource constrained environments, such as the cost and availability of the hardware and consumables required to make use of the design.

Although affordable smartphones are becoming widely available (see Fig. 1), we are well aware that many people in these areas own lower-end phones that do not include a touchscreen. With this in mind, we have implemented a Java ME version of PaperChains, compatible with lower-end, camera-enabled feature phones. This version works in the same way as the original application, but uses the phone's joystick or directional pad to navigate the image – a slightly clumsy but viable approach to facilitate use on lower-end devices.

Turning now to consumables, PaperChains requires items to be augmented with two QR codes to function. These QR codes can be created with our simple browser-based tool,[5] and printed onto standard paper using a low-end desktop printer. Several previous systems in this area have demonstrated how partnership with local organisations, community computer centres or NGOs can support systems requiring consumables (e.g., [8, 18]). Another possible use case, which does not require new QR sheets each time, is the notion of a reusable PaperChains surface. Consider, for example, a community noticeboard made of a wipe-clean material—such as a typical school whiteboard—augmented with PaperChains QR codes in each corner. Users could mark-up the board, augment it with audio, and then erase the marks when necessary to make way for a new mark-up session.

USABILITY EVALUATIONS

In order to test how well users from resource constrained contexts can make use of PaperChains, we undertook two group-based user evaluations of the system. The purpose of these studies was to explore the usability of the system, and find out how PaperChains could support storytelling from a functional point of view.

The first study took place in Mumbai, India; the second in Langa, a township in Cape Town, South Africa. Both studies followed the same procedure. Eighteen people took part in each location, with six taking part in each session (i.e., three groups per site). Each of the people in the groups of six were friends or relatives, to ensure they had shared experiences with one another. Video and audio equipment was used throughout all stages of the studies to record usage of the system, with consent obtained from all participants prior to commencement. Local facilitators ran both the Indian and South African studies and helped with communication with participants.

Procedure

After a short introduction and obtaining informed consent, the six people in each group were given an overview of the study procedure and a demonstration of the PaperChains system. Once each participant had been given the opportunity to try using the system, they split into two groups of three to undertake storytelling tasks separately. Each group sat separately to the other so they could not hear the other group.

Each group was provided with a blank A3 PaperChains sheet (i.e., a blank sheet of paper with QR codes in two corners), a pack of 12 fibre-tipped coloured pens and a PaperChains phone. The groups were then asked to sketch and discuss a shared experience, documenting it using the PaperChains system. This step involved taking turns drawing parts of the story and augmenting them with audio using the phone provided. A member of our research team observed each group and was available to answer any questions they might have throughout the study.

Once each group was happy with their creations, the groups swapped their story sheets with each other. Each group then attempted to interpret the other's story, first taking a new photograph, then using it to listen to the audio hotspots. Following this, the groups came back together and we conducted a semi-structured focus group interview, asking participants to rate the system on several subjective Likert-like scales (1–10, 10 high). These questions probed difficulty (*"To what extent was the system easy or difficult to use?"*), enjoyment, (*"To what extent was the system enjoyable to use?"*), and also ability to, first, tell their own stories (*"To what extent did the system support*

[5] Available for use and download at http://enterise.info

Figure 4. A group of users collaboratively authoring a PaperChains document in the Indian evaluation.

Figure 5. Users in the Indian evaluation (left) and in the South African evaluation (right) experimenting with methods for framing the image for the photograph early on in the study session.

telling a story?"), and second, interpret others' stories (*"To what extent could you understand other people's stories?"*). The groups were then thanked and each person was given an incentive of 200 Rupees (India) / 120 Rand (South Africa).

We chose to use slightly higher-end, touch-screen smart phones in both studies rather than the lower-end ones we found on sale in the Mumbai slum. Our decision to use these more expensive handsets rather than the $15 versions, or even lower-end camera-enabled feature phones, was two-fold. Firstly, we wanted to ensure the system was as robust as possible for the usability evaluations – something we could not currently guarantee on the sometimes counterfeit devices sold in these areas. Secondly, there is strong evidence to suggest that the trajectory of smart phone use in these areas over the next 5–10 years is heading towards the type of handset we were testing on. It was our desire, therefore, to explore how well participants would cope with not only the PaperChains system, but the increasingly sophisticated hardware they are likely to have access to in the near to mid-term future.

Study 1: Mumbai, India

Our first study involved 12 male and 6 female participants aged 20–45 from Mumbai, India. Education levels of the group varied from none at all (i.e., no literacy), to education to a 7th grade level (11–12 years; i.e., a basic level of literacy in the local Marathi language). Access to affordable data connections was very low for this set of users, as was higher-end technology exposure. Basic feature phones or lower-end 'dumb' handsets were the dominant device within the cohort (17 of 18 people). One participant owned a touch-screen phone, and only 6 of the remaining 17 had tried using one before.

Results

All participants were able to collaboratively create stories on paper, and augment them with audio recordings (see Figs. 4 and 6). Participants were also able to use the system to interpret the other group's stories effectively, and provided useful feedback on general use of the system.

Ratings and observations: The results of the Likert-like ratings across both studies are summarised in Table 2. Due to the low technological familiarity of the groups, we anticipated

Question	India (s.d.)	SA (s.d.)	Avg
Ease of use	7.0 (2.3)	8.9 (1.0)	8.0
Enjoyment	9.7 (0.6)	9.6 (1.1)	9.7
Telling stories	7.3 (2.0)	9.4 (0.9)	8.4
Understanding stories	8.8 (1.8)	8.0 (1.9)	8.4

Table 2. Summary of PaperChains ratings across the two studies (range 1–10, 10 high; standard deviations in parentheses).

encountering problems with using smartphones. At the start of the study, each participant was given the opportunity to use the system until they felt comfortable using it. Although a minority had problems in framing the photograph initially (see Fig. 5), participants quickly learned how to use the system effectively, rating it 7.0 on average for ease of use.

It was also clear from the observations made during the study sessions, and the post-study interviews, that users enjoyed the experience. Participants rated the system 9.7 on average for how fun the system was to use, with comments such as: *"it was amazing because if I had to tell one of my friends I would never be able to attach audio to the picture before"*, *"I found it really fun"* and *"we like that we can draw our own pictures."* Conversely, there were some complaints about the volume of the audio playback; one participant stated she felt uncomfortable sketching as she did not see herself as a capable artist; and, another had reservations about the possibility of the system being used for unfavourable purposes: *"it is 99 percent good but 1 percent problem because it shouldn't be misused."*

When asked how well they were able to convey their stories to others, participants gave an average score of 7.3. In terms of how participants created stories collaboratively, we observed several different behaviours within the six groups. The majority of the time, group members took it in turns to draw on the paper then add their own voice to that section. This action would then repeat with the next person sketching and adding their own voiceover. At times, we did observe participants making additions to other participants' drawings (drawing over or around existing illustrations) as well as attempting to add additional audio on top of existing hotspots. It was also clear from the studies that participants liked to listen back to their own voices after recording. Other behaviours

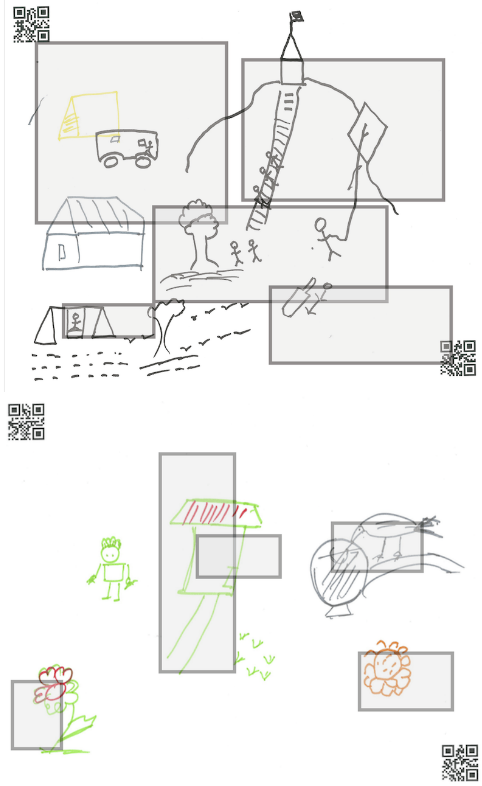


Figure 6. Two example stories from the Indian evaluation. The upper story is a group narration about childhood, while the lower scene consists of individual anecdotes about each of the items pictured. Grey shaded areas indicate where participants added audio to their sketches.

included all participants talking on the same recording, as well as a single person taking charge of all sketching for the group. However, all participants appeared to be collaboratively working together in order to help create a single, unified story.

The average score given for the ease of understanding other people's stories was 8.8. Groups were engaged while attempting to interpret stories, with multiple members touching to select drawings to listen to, and huddling together to be closer to the microphone on the phone. It was evident during the sessions that locating the actual audio hotspots within the image was sometimes challenging – several groups missed some of the hotspots from the other group's story. This did not seem to distract from the overall storytelling experience, however. In fact, many participants commented that they enjoyed the exploratory nature of the process, discovering the hotspots for themselves. Despite this, further development of the design of the interface could include some form of visual cue as to where hotspots exist to aid with this process.

Comparison to other methods: When comparing the PaperChains approach to current methods of storytelling that they used, participants appreciated the PaperChains approach, with several stating that they preferred the system to simply reciting a story over a telephone (something they regularly did). This was illustrated well by one participant, who said: *"it is better than using the telephone – it's better to show the picture than just explaining what I saw."* Participants were also excited about how they could make use of the technology in their day-to-day lives. For example, activities such as documenting accidents, sharing with children, and using PaperChains for directions to inform others on how to get to a particular place were all mentioned as possible uses for the system.

Figure 7. A group of users collaboratively creating a story in the South African evaluation. (The third participant is off-camera to the right.)

Study 2: Langa, South Africa

Our second study involved 9 male and 9 female participants aged 18–37 from Langa, a township near Cape Town in South Africa. Users from this cohort were moderately educated with mixed levels of literacy in both English and the local isiXhosa language. Access to affordable data connections was slightly higher in Langa than in Mumbai, but still beyond the reach of many. For example, data connections are relatively cheap on Blackberry phones (R60 for 30 days) but far more expensive on other devices. Technological exposure was also higher in Langa than Mumbai – although the majority of participants owned only feature phones, most had experience with using touch screens and all have access to a local internet cafe.

Results

As with the Indian evaluation, all groups managed to create and view each others' stories successfully (see Figs. 7 and 8). It was also clear, based on the ratings, that the groups enjoyed using the system to tell stories – the groups gave an average score of 9.6 regarding enjoyment. Participants' comments regarding the experience strengthened the findings here: *"the whole in general was iyoba! [fantastic]"*, *"it was an interesting fun activity"*, and *"it's really dramatic and exciting."* That said, however, some negative points were also raised, with one participant stating: *"it's a little childish – imagine a business man doing this."* As in the Indian evaluation, there were some participants who felt uncomfortable drawing and, this time, two who suggested less-educated people might find the system harder to use. There were similar initial issues with framing the photograph as in the previous study, but these were overcome quickly, leading to an average ease of use score of 8.9.

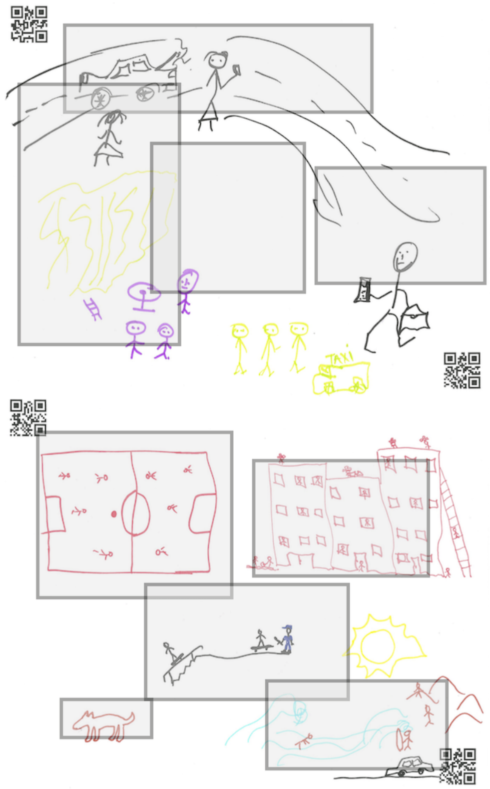


Figure 8. Two example stories from the South African evaluation. As in the first study, groups varied in the methods and layouts chosen. The upper story here is a collective effort about road issues and a car breakdown; the lower image is a collection of individual stories about the sketched items. Shaded areas indicate audio hotspots.

The average ratings for ease of telling and understanding stories were 9.4 and 8.0, respectively, and the methods of creating and interpreting stories were broadly similar to those in the first evaluation. Participants tended toward taking turns to draw and speak, with individuals occasionally adding to or modifying others' images, and often multiple people speaking on the same recording. Similar issues with finding audio hot spots when interpreting others' stories were observed in this study, indicating that a visual cue solution would be beneficial.

After participants had used PaperChains, we asked what they thought the role of the audio was in the storytelling experience. Comments such as: *"it's more descriptive with audio; you get a sense of who is talking – it works nicely together"*, *"[the audio] is very helpful because you can always refer back to the recoding; it's great"* and *"[the audio] makes it easier for those who can't read and write"* strongly indicate that the audio annotations complement the paper sketches well. Some even felt that the audio was an essential part of the experience, and without it the drawing would have been very difficult to interpret: *"with the help of the audio it was easy to understand, but without it it would have been very hard to understand"*.

Comparison to other methods: After some discussion about how the PaperChains approach differed from participants' usual methods of telling stories, it became apparent that most participants felt that they were able to go into more depth with PaperChains than usual: *"this is more detailed than face-to-face – more practical,"* *"[PaperChains] lets you elaborate more and be more creative,"* and *"for me the drawing approach is a way of exercising my mind more, and I'm able to express myself more this way."* One individual went further, commenting that the system enabled him to share his thoughts and feelings when he usually could not: *"[PaperChains] takes my mind off things because as an introvert I think too much, so it was good fun. I can share my emotions and my feelings to people now,"* suggesting that PaperChains could be a powerful tool to support sharing emotional content.

Participants also offered suggestions about how they would make use of the system. Common examples included using the system for people who are not as literate, or for older people to help them communicate. One participant commented that he would use the technique to draw a map and describe directions for someone, while another suggested it could be useful to mark up school assignments. Another common scenario was to use the system for people who are not physically in the same location. For example: *"I would use it a lot to communicate with my child who does not live with me,"* *"Valentine's day for a long distance relationship – an original way to create for lover,"* *"I will be able to send to my mum if I'm away and I can share my experience with her"*, and *"[PaperChains] makes a person who is not around feel like they are around"*.

CONCLUSIONS AND FUTURE WORK

We began this paper with a strong motivation from existing research, highlighting the lack of techniques that allow both the audio and the visual component of a story to evolve simultaneously. This incentive to create sketch+voice narratives that emerge and grow over time led to the development of PaperChains – an audio annotation system specifically designed with high affordability and accessibility in mind. We now discuss the system's benefits, the limitations of this work, and areas for future research.

PaperChains' novel technique makes use of an interactive voice response (IVR) backend, using standard, widely-recognised DTMF codes to communicate with client handsets over a normal phone line, removing the need for a potentially costly data connection. In addition, the approach works with even the lowest-end of the smartphone market (or camera-enabled feature phones), making it particularly appropriate for users in resource constrained environments.

Our usability evaluations of PaperChains have demonstrated its ability to support the creation and interpretation of dynamic audio-annotated documents. All groups in both Mumbai and Langa were able to use the system to sketch and annotate their own stories, and to understand other groups' stories. During the studies, participants mentioned other scenarios that they imagined the system being useful for, including specific aspects of the expressivity that the system enables that made it particularly useful. There were also strong statements about its usability, and the majority of participants stated that they enjoyed using the system.

Clearly, there are opportunities for further work to refine the system's operation. When sketching and annotating, users

Figure 9. Noticeboards—such as this local advert board—could easily be augmented with PaperChains content after adding two small QR codes.

tended to sketch in more areas than they eventually annotated, or annotated only a small section of the sketch rather than the entire object. It was apparent in both evaluations that the exploratory nature of the audio discovery process that this annotation behaviour led to was enjoyable. However, further development of the design could investigate ways to draw users toward audio areas in an unobtrusive manner.

While the results from our studies have been highly encouraging, the experiments we have conducted to date have been in a lab-like environment. As a result, usage of the system so far has not been contextualised in the type of storytelling scenario we imagine the system being used. Our next task is to deploy the PaperChains system in a more naturalistic environment in order to gather feedback on real-world experiences from users. One possible way to achieve this could be to distribute PaperChains-augmented paper via a local NGO or community group within the regions we have been working in. Monitoring uptake and the types of documents created would provide useful information about long-term usage of the system.

A further alternative for a more longitudinal and in-the-wild investigation in this area would be to explore the use of annotated community notice boards (see Fig. 9). If we consider the nature of group notice boards, it is clear that they are built-up and modified by multiple users over time, making them an ideal use-case for the PaperChains technique. An additional benefit of such an application, as mentioned earlier, is that a notice board is static and can therefore be reused, reducing the consumable materials required to use the system.

Stepping back from resource constrained environments, we also believe that our approach has benefits for more developed contexts. For instance, in areas where data-costs are higher than usual (such as when data roaming abroad), an IVR-based rather than internet-based method could significantly lower costs for this type of interaction. Alternatively, modifying the system to use an internet-based backend would better suit areas where data connectivity is cheap. In this context, we imagine using the technology for the archive of personal media. For example, this could include building up scrap- or baby-books over time, including various physical keepsakes such as photographs or memorabilia augmented with personal anecdotes from family and friends. A specific instance of this idea, then, could be to use the PaperChains technology as a method of creating interactive greetings cards, which could be particularly useful when there are multiple well-wishers who would like to leave a message, as explored in the following scenario:

Liam is leaving the office for several months to take up a summer internship. To wish him well, his friends buy a PaperChains greetings card, and over the next few days the card is passed around the lab for everyone to sign. At their turn, each person writes in the card, then uses their phone to take a photo, scribbling on-screen over their handwritten message to leave an audio greeting. The space fills up over time, and soon the card is crowded with penned notes and paired audio messages.

Liam takes the card with him while away. At any time during the trip, he can take a photo of it and touch the handwritten notes to hear the accompanying voice messages. When Liam gets back he writes his own note on the card and adds audio comments about the experience, then pins the card to the lab noticeboard. Anyone in the office can take a photo of the card and hear about his summer visit – the card is now an interactive memento of the experience.

As illustrated in the scenario, we see benefits of the PaperChains approach in wider situations than that demonstrated in this work. The aim of this paper has been to explore the notion of evolving collaborative digital-physical artefacts in resource-constrained contexts. We hope that this work stimulates discussion on ways to provide advanced services with low technological requirements, and motivates further work of this type in these contexts

ACKNOWLEDGMENTS

This work was funded by EPSRC grant EP/J000604/2. We thank Minah Radebe (South Africa), and our colleagues at IIT Bombay (India), for their help running usability evaluations.

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. Cao, X., Lindley, S. E., Helmes, J. and Sellen, A. Telling the whole story: anticipation, inspiration and reputation in a field deployment of TellTable. In *Proc. CSCW '10*, ACM (2010), 251–260.
3. Costanza, E., Giaccone, M., Kueng, O., Shelley, S. and Huang, J. Ubicomp to the masses: a large-scale study of two tangible interfaces for download. In *Proc. UbiComp '10*, ACM (2010), 173–182.
4. Druin, A., Stewart, J., Proft, D., Bederson, B. and Hollan, J. KidPad: a design collaboration between children, technologists, and educators. In *Proc. CHI '97*, ACM (1997), 463–470.

5. Erol, B., Hull, J., Graham, J. and Lee, D.-S. Prescient paper: multimedia document creation with document image matching. In *Proc. ICPR '04*, vol. 2. 2004, 675–678.
6. Erol, B., Antúnez, E. and Hull, J. J. HOTPAPER: multimedia interaction with paper using mobile phones. In *Proc. MM '08*, ACM (2008), 399–408.
7. Frohlich, D. Audiophotography. Springer (2004).
8. Frohlich, D. M., Rachovides, D., Riga, K., Bhat, R., Frank, M., Edirisinghe, E., Wickramanayaka, D., Jones, M. and Harwood, W. StoryBank: mobile digital storytelling in a development context. In *Proc. CHI '09*, ACM (2009), 1761–1770.
9. Frohlich, D., Clancy, T., Robinson, J. and Costanza, E. The audiophoto desk. In *Proc. 2AD Demos*, 2004.
10. Guimbretière, F. Paper augmented digital documents. In *Proc. UIST '03*, ACM (2003), 51–60.
11. Jacoby, S. and Buechley, L. Drawing the electric: storytelling with conductive ink. In *Proc. IDC '13*, ACM (2013), 265–268.
12. 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.
13. Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S. and Nanavati, A. WWTW: the world wide telecom web. In *Proc. NSDR workshop at SIGCOMM '07*, ACM (2007).
14. Lang, X., Oreglia, E. and Thomas, S. Social practices and mobile phone use of young migrant workers. In *Proc. MobileHCI '10*, ACM (2010), 59–62.
15. Lima, E. S. de, Feijó, B., Barbosa, S. D., Furtado, A. L., Ciarlini, A. E. and Pozzer, C. T. Draw your own story: paper and pencil interactive storytelling. *Entertainment Computing* 5.1 (2014), 33–41.
16. Liu, Q., Liao, C., Wilcox, L., Dunnigan, A. and Liew, B. Embedded media markers: marks on paper that signify associated media. In *Proc. IUI '10*, ACM (2010), 149–158.
17. Lucero, A., Holopainen, J. and Jokela, T. Pass-them-around: collaborative use of mobile phones for photo sharing. In *Proc. CHI '11*, ACM (2011), 1787–1796.
18. Maunder, A., Marsden, G. and Harper, R. Making the link—providing mobile media for novice communities in the developing world. *International Journal of Human-Computer Studies* 69.10 (2011), 647–657.
19. Mistry, P., Maes, P. and Chang, L. WUW – wear ur world: a wearable gestural interface. In *Proc. CHI EA '09*, ACM (2009), 4111–4116.
20. 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.
21. Patel, N., Chittamuru, D., Jain, A., Dave, P. and Parikh, T. S. Avaaj Otalo: a field study of an interactive voice forum for small farmers in rural India. In *Proc. CHI '10*, ACM (2010), 733–742.
22. Raffle, H., Vaucelle, C., Wang, R. and Ishii, H. Jabberstamp: embedding sound and voice in traditional drawings. In *Proc. IDC '07*, ACM (2007), 137–144.
23. Reitmaier, T., Bidwell, N. J. and Marsden, G. Field testing mobile digital storytelling software in rural Kenya. In *Proc. MobileHCI '10*, ACM (2010), 283–286.
24. Robinson, S., Jones, M., Vartiainen, E. and Marsden, G. PicoTales: collaborative authoring of animated stories using handheld projectors. In *Proc. CSCW '12*, ACM (2012), 671–680.
25. Robinson, S., Pearson, J. S. and Jones, M. AudioCanvas: internet-free interactive audio photos. In *Proc. CHI '14*, ACM (2014), 3735–3738.
26. Sawhney, N. Voices beyond walls: the role of digital storytelling for empowering marginalized youth in refugee camps. In *Proc. IDC '09*, ACM (2009), 302–305.
27. Schilit, B. N., Golovchinsky, G. and Price, M. N. Beyond paper: supporting active reading with free form digital ink annotations. In *Proc. CHI '98*, ACM Press/Addison-Wesley Publishing Co. (1998), 249–256.
28. 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.
29. Smith, G. and Marsden, G. Providing media download services in African taxis. In *Proc. SAICSIT '11*, ACM (2011), 215–223.
30. Stifelman, L., Arons, B. and Schmandt, C. The audio notebook: paper and pen interaction with structured speech. In *Proc. CHI '01*, ACM (2001), 182–189.
31. Wellner, P. Interacting with paper on the DigitalDesk. *Communications of the ACM* 36.7 (1993), 87–96.
32. West, D., Quigley, A. and Kay, J. MEMENTO: a digital-physical scrapbook for memory sharing. *Personal and Ubiquitous Computing* 11.4 (2007), 313–328.
33. Wood, G., Vines, J., Balaam, M., Taylor, N., Smith, T., Crivellaro, C., Mensah, J., Limon, H., Challis, J., Anderson, L., Clarke, A. and Wright, P. C. The dept. of hidden stories: playful digital storytelling for children in a public library. In *Proc. CHI '14*, ACM (2014), 1885–1894.
34. Yeh, R., Liao, C., Klemmer, S., Guimbretière, F., Lee, B., Kakaradov, B., Stamberger, J. and Paepcke, A. ButterflyNet: a mobile capture and access system for field biology research. In *Proc. CHI '06*, ACM (2006), 571–580.