# ACQR: Acoustic Quick Response Codes for Content Sharing on Low End Phones with No Internet Connectivity

**Jennifer Pearson,[1] Simon Robinson,[1] Matt Jones,[1] Amit Nanavati,[2] Nitendra Rajput[2]**

[1] FIT Lab, Swansea University, SA2 8PP, UK
{ j.pearson, s.n.w.robinson, matt.jones } @swansea.ac.uk

[2] IBM India Research Lab, Vasant Kunj
New Delhi, 110070, India
{ namit, rnitendra } @in.ibm.com

## ABSTRACT
In this paper we introduce Acoustic Quick Response codes to facilitate sharing between Interactive Voice Response (IVR) service users. IVRs are telephone-based, and similar to the world wide web in many aspects, but currently lack support for content sharing. Our approach uses 'audio codes' to let people share their call positions, and allows callers to hold their normal (low-end) handsets together to synchronise. The technique uses remote generation and recognition of audio codes to ensure that sharing is possible on any type of phone without the need for textual literacy or an internet connection. We begin by exploring existing user needs for sharing, then evaluate the technical robustness of our audio-based design. We demonstrate the value of the approach for voice service users over several separate studies—including an eight-month extended field deployment—then conclude with a discussion of future possibilities for such scenarios.

### Author Keywords
Developing regions; mobile sharing; audio codes.

### ACM Classification Keywords
H.5.2 [**User Interfaces**]: Input devices and strategies; Interaction styles; H.5.1 [**Multimedia Info. Systems**]: Audio I/O.

## INTRODUCTION
It is estimated that there are over two-and-a-half billion people worldwide living on less than $2 per day [10, 36]. For many of the people living in these 'developing regions,' the mobile cellular phone is the only digital technology available. In rural areas in particular, these devices are currently relatively low-end handsets [15, 33], which are usually incapable of accessing the internet. Low literacy levels and low access to data connections mean that the majority of people in these regions are unable to access information online.

As an alternative, a class of audio-based voice information services, which are accessible over low-end handsets, have been developed (e.g., [19, 21, 27]). Voice services have been adopted specifically to meet widespread challenges such as cost, textual literacy, data connections and general technology exposure [9, 20]. Their purpose, then, is to provide audio-based information to callers who are unable to access the internet. The structure of these services is hierarchical, similar in design to the Interactive Voice Response (IVR) systems commonly used for customer service telephone enquiries. 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, in that they are accessed via unique addresses (phone numbers rather than URLs) and, in some implementations, contain links to each other (voi-links [1] rather than hyperlinks). Unlike websites, however, many of the interactions supported by a text-based service over an internet connection are not yet possible with voice services over a phone line. One such example, which internet users often take for granted, is the ability to share links to specific pieces of information.

While internet addresses can be shared in a variety of ways—ranging from emailing a link, to searching for keywords, to scanning a QR code—sharing within voice services is far more difficult. The diversity in website sharing methods is possible because of the use of a textual address, and a well-indexed population of websites. In contrast, voice services are audio in nature and recorded in many different languages, which means that searchable indexes are not usually available. This situation makes novel mechanisms to enable sharing within these services particularly important to help the spread of useful content amongst users.

We developed Acoustic Quick Response codes (ACQR; pronounced 'aqua') as a way to allow sharing of voice service content. Our approach, which currently focuses on the Spoken Web (a popular Indian voice service [19]), uses remote generation and recognition of short DTMF-based 'audio codes' using standard mobile phones—not smartphones—allowing content to be shared instantly. The following scenarios illustrate how ACQR can be used to share Spoken Web 'voice sites':

*Vikram finds a community forum on a voice site which he thinks would benefit his neighbour, Akash. Akash is in the next field, so Vikram walks over and they hold their phones together to synchronise. Akash can now listen from the same point and browse independently…*

**Published in Proceedings of MobileHCI 2013, pages 308–317, ACM (2013).**

DOI: 10.1145/2493190.2493195

*Praharsh is a local farming expert who regularly holds teaching sessions with less-experienced farmers. While browsing the Spoken Web he comes across a collection of wheat growing tips from a university researcher. He'd like to share these with his students, so he saves his current call position as a personal bookmark for later. In his next meeting group, Praharsh calls the voice site and plays the bookmark's audio code to the group. Everyone can now save the position, and browse later on…*

*Kissan community radio wants to let its audience re-listen to programmes. So, at the end of each show they play an audio code. Listeners can use the ACQR service to easily visit the station's voice site and listen again…*

The ACQR technique does not require an internet connection, or any specialist hardware, and automatically synchronises receiving devices directly to the correct voice site and the location within it that was shared. Because our approach is audio based, it can also provide fast, low-cost, one-to-one, one-to-several and one-to-many sharing (e.g., to a group, or over broadcast media), allowing people the flexibility to send to multiple recipients at once. Used in conjunction with an existing bookmark feature (detailed in [13]), it allows users to share without the need for potentially error-prone human memory of either navigation steps or the information itself. In addition, our technique allows receivers of content to continue browsing independently from the shared call position.

We begin with a study of current sharing methods that motivated our work, then review related research. The ACQR design, and two potential alternatives we developed, are described in detail. We evaluated the ACQR design in several studies, including a demonstration of the technique's robustness, and a lab study to examine the feasibility of all three systems. We then tested the ACQR service and an alternative with Indian participants who were taking part in an existing Spoken Web trial. Finally, we report on a longitudinal live deployment of the ACQR service on the Spoken Web, and discuss several interesting future expansions of the sharing techniques used.

## EXISTING SHARING METHODS

There is a lack of previous research into how voice service users currently share information that they receive from these services. To gain a better understanding of current sharing habits within the Spoken Web community in particular, we conducted a series of interviews with eight farmers from rural Gujarat, India. Interviews lasted approximately 30 min each, and were conducted by phone. All participants were male, aged from 18 to 43, and had an average of 4.4 years experience using an existing popular Spoken Web voice site. These participants were randomly chosen from the all-male farming communities which commonly use the service. Interviewees' usage rates ranged from once every five days to once per hour, with seven participants using the service at least once a day.

When asked about their current sharing habits, it was clear that all eight interviewees had shared content from the Spoken Web with other people, using various methods. While several people used multiple methods, there were four distinct techniques:

**Memorisation** Remember the information and relay it back to others later (2 participants).

**Playback** Play the information directly to others from the sharer's own phone (3 participants).

**Number** Share the voice site number and allow others to find the information within it themselves (4 participants).

**Navigation** Give voice site navigation directions to others so they can navigate to the same position (2 participants).

Participants said that they had encountered problems using the current sharing methods. Two interviewees who had made use of the memorisation method touched on the difficulties of remembering a lot of information, with one stating: *"if it is detailed information, this is hard to do."* In addition, neither recounting to others the information found on a site, nor playing the information in situ, allow the recipient to access the information on their own device. This means that recipients cannot continue to browse the site independently, or review the information later when more convenient.

Providing a voice site number, even when with navigation instructions, can also be problematic, as it can be difficult and time consuming to find a specific point located deep in the hierarchy (see [23]). The navigation approach also involves remembering both the telephone number and the method of locating the exact position within the voice site, and only allows navigation to the start of an audio segment, rather than an exact position within it. For instance, one participant who had used these methods commented: *"the challenge is that it takes too much time to listen and reach the position of interest."*

The enthusiasm for a faster, more reliable method of sharing voice site positions was clear, with comments such as: *"a direct 'go-to' facility would be of great use,"* and that being able to navigate to specific positions within a voice site *"would be a huge benefit, very useful and will make it popular."*

### Design implications

The current techniques employed by Spoken Web users highlight a clear gap in functionality, and an opportunity to improve interaction with voice services. Existing methods of sharing specific pieces of information are fairly crude techniques conceived by users to deal with the lack of a sharing function.

One of the main problems identified with existing methods is that of memorisation; either of the information itself, or instructions for getting back to a specific position. Any new approach should avoid the need for manual recall as much as possible, which could in turn increase the accuracy of locating the correct position and consequently attaining the desired information. A new sharing function should also be fast, to address user complaints about time taken to listen and navigate through voice sites. The new technique should allow direct access to specific positions, but also allow recipients to browse the voice site independently after receiving a position from another user. Finally, a new implementation should be robust, to avoid users having to revert to previous, unreliable methods.

## BACKGROUND

The Spoken Web provides access to audio-based information on any phone, without the need an internet data service or textual literacy. Voice sites (equivalent to websites) can be created for any purpose, the most common being topics such as farming or health information. They are also hierarchical, similar to the IVR systems commonly used for customer service. Navigation through a voice site takes place by listening to prompts and then responding with a choice via either button presses or short spoken inputs. Voice site 'addresses,' then, consist of a number and a navigation structure, similar to the global and local components of web addresses. However, because these parts are entered at different times it is more difficult to remember or share them. Such hierarchies, even when text-free, can affect users' performance, with deeper trees being harder to navigate [23]. To address this problem, previous Spoken Web research (e.g., [13]) has added bookmark capabilities to voice sites, allowing callers to save positions that are meaningful or particularly useful. There is currently no facility for sharing this information with others, however.

Previous research has shown how voice service interaction improvements can have positive impacts and improve user experiences (e.g., [13, 26, 29]). But the areas in which such additions can currently be supported are restricted due to the level of technology available. While there has been a high uptake of basic cellular or feature phones in developing regions, there are currently far fewer smartphones or data-capable handsets that could support many advanced interactions. In India, for example, around 55 % of the population own a mobile phone [31] whereas less than 3 % own a smartphone [2]. This situation is, of course, changing fast, but basic issues such as textual literacy and cost will remain for some time. Very few phone owners subscribe to a 3G data plan (only 2 % in India [30]); consequently, the only channel generally available for voice service communication is still the public telecom network, via standard phone lines on low-end mobile devices. Related work has demonstrated how such low-end devices might support relatively advanced voice site interactions when these are implemented on the service side [29]. Our approach here uses 'audio codes' that are remotely generated and remotely recognised (i.e., on the service side), and that are device-independent, supporting the wide range of phones used to access the service, rather than just the latest models.

## Media sharing

Gunaratne and Brush [17] investigated media sharing during a phone call, aiming to address existing complexities with this process. Their high-end mobile client used Bluetooth to send content from phones to a nearby PC, which synchronised with the other party in the call. Our design does not aim to support media sharing; rather we allow users to share the *position* within an audio call with other potentially interested users. As a result, our approach requires no additional devices to share positions – we use the voice service itself for synchronisation, and the technique can therefore be used by any phone.

Several groups have investigated ways to synchronise music tracks between mobiles, though each of these approaches has concentrated on precise realtime pairing, rather than sharing via the audio itself, as in our technique. Östergren [25] aimed to synchronise audio between cars driving near to each other. Their technique used ad hoc networking to communicate, and played the same music in each car that was within a certain distance radius. Bassoli et al. [6] created a similar system for joint music listening and sharing between mobile devices. Perhaps most relevant to our work is the guidebook system of Aoki et al. [3], which allowed synchronised listening amongst museum visitors by enabling them to 'eavesdrop' on others using the system. However, each of these systems required a background network connection to allow people to share. In our design we focus on allowing users to share audio positions from a remote service, rather than from local devices. No additional infrastructure or hardware is required.

## In situ synchronisation

De Barros et al. [11] used the audio band of a video signal to synchronise recordings from multiple cameras. Their technique used inaudible audio watermarking to mark the audio stream without affecting the quality of the signal. Our approach is intentionally more visible to users – our synchronisation method plays an audible 'jingle' that both synchronises users' devices and provides a cue that the process is taking place. Goodrich et al. [16] demonstrated and highlighted the benefits of using audio in a similar way to authenticate two devices with each other, choosing to use non-hidden audio in order to provide verification by a user who is listening to the exchange. The commercial possibilities of this type of audio-based synchronisation have also been recognised by the very recent launch of Chirp,[1] a smartphone-based media sharing application that uses audio tones to share links to web-based content. Unlike Chirp, however, our design can be used on low-end, non-internet-enabled devices. In addition, our design and evaluations validate the usefulness of this type of sharing, and show its potential for usage in emerging markets.

The use of remote recognition of audio is a key feature of our design. Previous non-audio methods of sharing or synchronising have supported pairing via gestures or other techniques using smartphones (e.g., [8, 18, 28]). Pairing is also possible without the need for device-to-device communication – the approach taken by Bump,[2] which compares location and similarity of physical movement, and by Mistry et al.'s gesture-based copy+paste [24]. However, the low availability of smartphones, motion sensors, data connections or GPS-enabled devices within voice services' user bases mean that these remote synchronisation techniques are impractical in this context. Using audio offers a simpler user interaction, and further flexibility. Our method allows one-to-one sharing, which we focus on in this paper, but also supports one-to-several for groups, and one-to-many over broadcast channels.

## Audio recognition

To use our ACQR service, two or more users hold their phones near to each other while an audible code is played. While we use DTMF-based codes, previous work has demonstrated how audio for which a known fingerprint exists can be recognised over a telephone connection. Typke et al. [32] provide a review of the multitude of audio recognition services that have arisen in this area, with the majority focused on matching user input (humming, whistling, etc.) to a known audio track. One of the

[1] See: chirp.io; [2] See: bumpapp.com

most well-known commercial applications of this technique is Shazam[3] [34], a music identification service. In its early development (around 2002), Shazam's service allowed users to dial a short number and hold up their phone to a nearby music source before being sent the recognised song details via SMS message. Shazam has now moved primarily to a smartphone-based service, but its use of telephone-based recognition demonstrates how we envisage our service operating.

Our design uses a DTMF-based approach for two reasons. Firstly, DTMF is specifically designed to be recognisable over telephone connections. Typical noise-prone usage scenarios for voice services (such as rural villages and farms) preclude the use of methods that require low interference levels. DTMF tones are robust, and designed to fit well with the low-quality audio (typically 8 kHz on a low-end device) of a standard mobile phone call. Higher data rates can of course be achieved over an audio connection by, for example, modulating the audio signal (e.g., modems). Alternative methods include those of Bender et al. [7] or Matsuoka et al. [22], who demonstrated methods for watermarking and embedding messages into regular audio material without altering the appearance of the original sound. However, these techniques add additional complexity, processing requirements and a higher bit-rate than is required for our application. They also typically rely on low interference levels for accurate and fast decoding. It is uncertain whether their usage in our scenario—i.e., over two mobile phone lines and the phone-to-phone distance between users—would be sufficiently reliable.

The second reason for using DTMF is that mobiles already have a keypad for generating these tones, providing scope for future expansion of the technique with no extra hardware required. Previous designs in similar contexts have made use of DTMF for the same reasons. Robinson et al. [29], for example, converted back-of-device taps into remote DTMF inputs to provide advanced inputs on existing low-end devices. Agarwal et al. [1] used DTMF to provide an interactive browser for the Spoken Web, while Dubey et al. [14] used keypad inputs to allow farmers to remotely control irrigation flow. Designing our system in this way also allows a caller to, for example, send an audio code via SMS message to allow sharing with people who are not in the same location.

## ACQR: AUDIO SHARING

To address the problems with sharing voice service positions revealed by our initial user study, we designed Acoustic Quick-Response (ACQR) codes to provide an audio-based approach to sharing. Our technique, which is currently implemented on the Spoken Web, allows callers to immediately share their current position within a voice site with other users. We also designed the service to allow callers to save their position (via the service's existing bookmark function) and then share later.

The service is made up of two parts, and is analogous in some ways to a link shortening service. The two components are: an add-on to existing voice sites to allow generation of audio codes; and, a standalone recognition service, which listens for audio codes and then jumps to the shared positions. Callers to

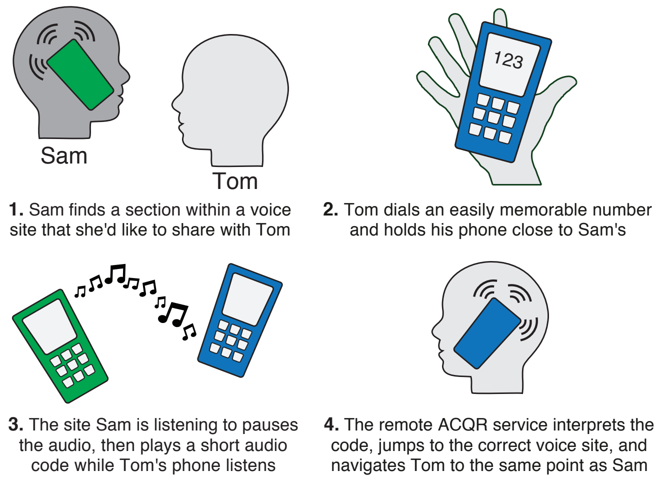


**Figure 1. Audio sharing. Callers listening to a voice site can share their precise position with others by playing them an audio code. After sharing, both users can navigate within the original voice site from the same position. No additional hardware or infrastructure is required, and both the sharer and the recipient may use standard, low-end phones.**

a voice site can press a 'share' key at any point during a call, which causes a unique audio code to be generated. Receiving users call the ACQR service separately, via a short telephone number, then hold their phone near to the sharer's device to synchronise. The audio code played by the original voice site on the sharer's phone is picked up by the microphone of the receiver's phone, and then recognised remotely by the ACQR service. A speakerphone is not required – the audio played through a standard phone speaker is sufficiently loud to be detected remotely. Figure 1 illustrates how this process takes place. In addition, a video accompanies this paper.[4]

The use of audio-based codes makes the ACQR technique accessible to callers with any type of phone. This is a key benefit for those living in developing, rural regions as it ensures that no additional hardware is required to use the service. While smartphone penetration is slowly increasing in these areas, the data connections required to access traditional internet content are not yet available [30]. The fact that the ACQR sharing technique is performed entirely remotely (i.e., at the other end of a phone line) as opposed to locally on the user's phone also means that no internet connection is required for its use, making it accessible over a standard telephone line. Furthermore, the implementation of the sharing function as a component of a remote system (rather than simply playing a linear stream of information to the recipient over a speakerphone) means the recipient can browse independently through a complex tree, possibly sometime after the sharing. We see this as analogous to emailing a link to the person sitting next to you, letting them browse either now or later, rather than forcing them to listen at the point of sharing.

### Implementation

*Sending:* When activated (by pressing #), the call is paused, and the current location within the voice site is stored to the existing Spoken Web bookmark database, then converted to a DTMF-based audio code. Pressing # plays or repeats this audio code whenever the user requires.

[3]See: shazam.com; [4]See: goo.gl/iycnF

| Environment | Ambient noise (dB (s.d.)) | Rate (%) |
|---|---|---|
| Inside, quiet room | 26 (2.80) | 99.5 |
| Outside, quiet park | 46 (4.72) | 99.5 |
| Inside, information radio show | 62 (1.85) | 98.5 |
| Inside, busy marketplace | 69 (1.66) | 98.0 |
| In car while driving (50 mph) | 77 (1.47) | 98.0 |
| Outside, alongside busy highway | 79 (3.44) | 97.0 |
| Inside, loud music | 82 (2.36) | 98.0 |
| Inside, very loud music | 101 (1.91) | 84.5 |
| **Average** | | **96.6** |

**Table 1. Recognition rates in each of the test environments. The ACQR design was tested 200 times in each case. A sound level meter was used to monitor the average ambient sound level for each trial session.**

*Receiving:* A custom voice site is set up to listen for audio code DTMF digits and redirect callers. To access, the recipient dials in via a short telephone number, and the service begins listening immediately (i.e., no IVR menu navigation is required). When a code is detected, the listening site parses and decodes the tones, then redirects the caller to the correct voice site and exact position the sender requested. If recognition fails, the user is prompted to replay the code.

This DTMF-based approach helps alleviate many of the potential problems with ambient noise, interference and speed of processing of alternative audio recognition techniques. However, while DTMF is specifically designed for robustness in such situations, background noise or speech can still hamper recognition. In our approach we use a four-digit location hash, with no repeating characters, and with two Reed-Solomon error correction characters (which provide error recovery for one digit). With 40 ms tone length and 20 ms gap, the tones are short, so we play the tone twice to allow for further error correction. For recognition we use the subband NDFT recognition method detailed by Bagchi et al. [4] for efficient parsing under varying conditions. As a result, our design is robust enough to allow sharing under many of the typical noise conditions encountered by Spoken Web callers.

### Technical robustness evaluation

To quantify the accuracy and robustness of the ACQR service, we tested its recognition in various environments with differing levels of ambient noise. The environments selected reflected various common scenarios in which the service is likely to be used, and also several louder potential sharing environments. The ACQR service was used to perform 200 sharing tasks in each environment. Table 1 shows the recognition success rates achieved in each case (mean: 96.6 %), and also the mean ambient noise level, recorded before and after every ten sharing tasks (measured using a sound level meter). When excluding the very loud music environment—an extreme case, included here to show the point at which accuracy diminishes—the mean recognition rate was 98.4 %. This is a promising result for the ACQR technique's robustness.

In addition to 1-1 simulation, we also tested 1-many usage of the technique, following the same method but sharing to four phones simultaneously. Sharing 200 times (i.e., $N = 800$) at a mean ambient noise level of 65 dB, the accuracy was 96.25 %.

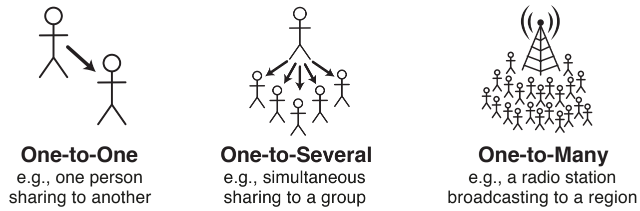


**Figure 2. Sharing relationships possible with the ACQR technique.**

### Benefits of the ACQR design

The ACQR design aims to provide a robust, easy to use method of sharing specific points within the Spoken Web hierarchy; as a result is it accessible to users with low-textual literacy who do not have access to smartphones or an internet connection. The use of audio—a location-limited channel [5]—offers further benefits, however. Key amongst these is that audio codes can be "heard" by more than one device simultaneously.

Figure 2 illustrates three possible relationships that can be provided. The first, and probably most common, of these relationships is *local* one-to-one sharing, where one user sends a voice site position to another, and both users are in the same place (as detailed in Fig. 1). The second, also used *locally*, is a one-to-several relationship, which involves one user sharing information with a larger group. Here, a single caller could play an audio code to the group via speakerphone while other members are connected to the ACQR service on their individual devices. This allows the code to be played a single time but shared to multiple people simultaneously.

Finally, the service supports a one-to-many relationship, which can be used to share information from one person to a large numbers of recipients *remotely* via, for example, radio broadcasts. In this case, a radio station could insert an audio code at the end of a broadcast, and prompt listeners to interpret it using the ACQR service. We distinguish between one-to-several and one-to-many relationships by means of the location of the recipients. That is, we see the one-to-several relationship to mean local, in situ sharing, and the one-to-many situation to be remote sharing over broadcast channels.

## ALTERNATIVE APPROACHES

We felt it important to investigate other approaches to voice service sharing, in addition to the ACQR technique, in order to validate our design's effectiveness and suitability for sharing. Consequently, we implemented two alternative methods – SMS (text) messaging and standard (visual) QR codes. Both were carefully designed based on the prerequisite that the base communication channel must be a standard telephone line, rather than an internet-enabled data connection, due to the requirements and needs outlined above. SMS is a clear choice for an alternative sharing method, being widely used and available on all mobiles, but we also wanted to account for the predictable future popularity of smartphones in regions where voice services are common. Our QR code approach, while smartphone dependent, is still usable without a data connection, however. Both approaches provide the same basic functionality as the ACQR service (i.e., sharing a voice site position), but differ in interaction during sharing and receiving.

| | Low-end devices | No text required | No data connection | Local sharing | Remote sharing | Realtime sharing | One-to-one | One-to-several | One-to-many |
|---|---|---|---|---|---|---|---|---|---|
| **ACQR** | ✔ | ✔ | ✔ | ✔ | ✘ | ✔ | ✔ | ✔ | ✔ |
| **SMS** | ✔ | ✘ | ✔ | ✔ | ✔ | ✘ | ✔ | ✘ | ✘ |
| **QR** | ✘ | ✔ | ✔ | ✔ | ✘ | ✘ | ✔ | ✘ | ✘ |

**Table 2. Comparison of features between the three sharing designs.**

| Classification | User base |
|---|---|
| Lab-based feasibility study | Literate, English speaking, non-Spoken Web users from the UK |
| In situ study via local translator | Mixed literacy, Kannada speaking, Spoken Web users from Karnataka, India |
| Field deployment (8 months to date) | Mixed literacy, Gujarati speaking, Spoken Web users from Gujurat, India |

**Table 3. Summary of user evaluations conducted.**

### SMS messaging

*Sending:* Upon initiating the sharing process (via #), the caller is prompted to enter the recipient's phone number. An SMS message containing the voice site number and a locator code (for example, '123,1865') is then generated by the service and sent to the recipient's phone via the cellular network.

*Receiving:* Once the SMS message arrives to the recipient, they can dial the given phone number and code to navigate directly to the correct position within the voice site. Most mobile phones offer a shortcut for dialling numbers received in SMS messages, automating this dialling process.

### QR codes

*Sending:* When the caller presses #, the service generates a standard QR image encoding the voice site number and a locator code. This QR code is then sent (via Multimedia Messaging Service – MMS message) to the sender's phone.

*Receiving:* The recipient uses a barcode reader application on their own phone to scan the QR code on the sender's phone screen. This is then decoded and displayed, and the recipient is offered the option to dial the voice site and code automatically, navigating them to the correct position.

### Comparison of designs

Table 2 shows a comparison of the basic features of each design. Both the ACQR and SMS approaches support low-end devices, whereas the QR code design does not. Clearly, then, the QR design is less relevant for regions where smartphones are currently scarce. However, we argue that it is worth considering for future use as smartphone adoption increases. Only the SMS design supports remote (i.e., non co-located) sharing of content positions without the use of a broadcast channel. However, some level of textual literacy is required. While it is possible that users with low textual literacy levels could, after some experience, recognise the message format and still use the service, this is a downside that is not present in the ACQR and QR designs. Remote sharing can be achieved with the ACQR design by transmitting audio codes over radio, television or tannoy, as illustrated in our earlier scenario.

All three designs allow sharing one-to-one and, to some extent, to more than one person. However, the ACQR design provides this capability in realtime, simultaneously to multiple users, and in a much simpler manner than the alternatives. Remote sharing with the SMS and QR designs would need to be achieved by, for example, a radio station maintaining a list of phone numbers and sending users messages, or by dictating a one-time code for users to enter. Playing a single audio code can provide many people with access to the same information at the same time, which would be time consuming to achieve with QR codes, and costly (to the provider) with both SMS and QR approaches. Clearly, this method of sharing via publicly-played audio has the potential to violate privacy. That is, it is possible that others could intercept content by simply listening using the ACQR service. In the context of the Spoken Web, however, we do not perceive this to be a significant privacy infringement, as all content is currently public.

As can be seen from Table 2, the ACQR design rates well in its basic functionality when compared to the alternative approaches. To further validate this, however, additional evaluation is required to investigate how the designs compare when voice sites are shared by users. We explored the effectiveness of the ACQR service via three separate studies, each exploring a different aspect of its usage. First, we measured the feasibility of the interface compared to the two alternative sharing methods in a lab-based study, with simulated voice sites and experienced phone users. We then explored Spoken Web users' thoughts, interactions and possible sharing in a field study of the ACQR and SMS designs. Finally, we deployed the live ACQR service on the Spoken Web for eight months (to date) to assess the usefulness and actual usability of our approach. Table 3 shows a summary of the studies performed.

## FEASIBILITY STUDY

We recruited 24 (13 M, 11 F; ages 20–31) UK, English speaking, non-Spoken Web users to take part in a within-groups feasibility evaluation of our three designs. The study used smartphones and simulated voice sites for interaction. Studying textually literate users who were familiar with smartphones allowed us to let participants make use of the three systems themselves, and gather first-hand, subjective feedback on their usability. We argue that it would be unfair, and a potential waste of participants' time, had we involved Spoken Web users in an initial trial of designs that might not be feasible. Testing with Spoken Web users would also have made it significantly more difficult to gather accurate and useful information on the technological capabilities of our alternative interfaces due to language, experience and literacy issues. Clearly, however, if any of the approaches are too demanding for users who are familiar with mobile technology, they will be equally or even more problematic for users who have less experience.

### Tasks and procedure

Participants attended in pairs, and took turns to be the sender or receiver of voice site positions. Pairs completed 18 tasks (six for each system; three sending, three receiving per system; counterbalanced), sharing positions from one participant to the other. To avoid any network delays in sending and receiving we simulated a cellular connection using Bluetooth so that

| | Time taken (s) | Ease of sending | Ease of receiving |
|---|---|---|---|
| **ACQR** | 7.5 (3.5) | 5.4 (1.0) | 5.6 (1.1) |
| **SMS** | 20.6 (10.2) | 5.4 (1.2) | 6.1 (1.2) |
| **QR** | 11.4 (5.7) | 5.9 (1.0) | 6.0 (0.8) |

**Table 4. Timings and user ratings for each of the three sharing designs. Times are given in seconds; ease of sending and receiving are on a 1–7 (7 high) Likert-like scale. Standard deviations are shown in parentheses.**

messages arrived immediately. For both the QR and SMS designs, participants used the phone's automatic dialling shortcut (as described previously). The sharing participant listened to the audio until they heard a distinctive beep (randomly positioned), which was their cue to share. Once the recipient had synchronised positions with the sender, they swapped to become the sender and repeat the process. After completing the tasks, participants answered a short questionnaire to rate and rank each system. Following this there was a short semi-structured interview and discussion session with each pair.

We collected both subjective scores and task timings as metrics for determining the designs' feasibility. The time taken to share positions was measured automatically by each system. This metric is defined as the difference between when the sender presses the # button to share and when the receiver has used the ACQR, SMS or QR system to arrive at the same audio position. In addition to the timed results, we also gathered subjective opinions about which design was easiest to use. We asked participants to rate each system on a Likert-like scale (1–7, 7 being the highest) for: ease of *sending* an audio position; and, ease of *receiving* an audio position. We also asked each participant which system they would use out of choice. Each study took, on average, around 30 minutes to complete, and participants were given a gift voucher in return for their time.

### Results

Table 4 shows the timings and ratings for each system. The ACQR technique was the fastest method of sharing. Statistical analysis using a paired t-test confirms these results are highly significant ($p < 0.0001, t = 56.42, DF = 2$). It is also important to bear in mind that cellular connections were simulated – here we studied the best case scenario; actual usage could only worsen the time required to send and receive with the SMS and QR systems. Table 4 also shows participants' ratings for ease of sending and receiving. These results are not statistically significant (Friedman test). However, high average ratings were given for all systems (all participants gave a minimum of 5 out of 7 for both sending and receiving), suggesting that each was usable for the sharing task, and that all the designs are technologically viable choices for sharing on voice services.

Five participants chose the ACQR design as their preferred sharing method, with eight choosing SMS. Almost half of the participants (11 out of 24 people) selected visual QR codes as their preferred design. However, the reasons given for this selection were predominantly based around the novelty of the interface. That is, the majority of the users who made this choice (8 of 11) stated that there was little-to-no difference between the QR and ACQR interfaces in important aspects such as usability. These participants selected the ACQR design as their second choice, but preferred using the visual system because it was, as one person put it: *"more fun."* This fun is, of course, an important consideration for future designs that cater to smartphone owners, but for people who only have access to low-end phones the ACQR design is a respectable alternative.

## INDIAN EVALUATION

As illustrated earlier, in Table 2, there are several key differences in the features of the three potential sharing methods. Our UK-based feasibility study provided strong evidence that all three designs are feasible methods for synchronising audio positions, as well as showing that the novel ACQR interface is the fastest of the three options. However, it is clearly also important to evaluate the designs with people who are familiar with voice sites and current sharing methods.

To better understand how sharing might be received by actual Spoken Web users, we performed a study in situ with 30 male Kannada-speaking farmers (aged 18–65) from Karnataka, India, who are currently participating in a Spoken Web service trial. All participants were familiar with other IVR-based services and owned (or shared) a mobile phone. No participants owned or had used a smartphone, however, so for this study we chose to study only the two designs that could currently be accessed by Spoken Web users (i.e., ACQR and SMS).

The study used demonstrations of the systems, rather than asking participants to use the smartphone-based simulations. As the smartphone in this study was not integral to the design of either of the systems (i.e., it was used as a design-stage prototyping method), and none of the participants were familiar with smartphones, we chose not to add the further challenge of learning how to use a touchscreen phone as well as understanding the sharing systems themselves. This methodology follows that of previous voice service trials (e.g., [35]), for early evaluations with users who are unfamiliar with the types of devices or interactions used. The evaluation itself was conducted by an independent Kannada-speaking local researcher who was not part of the design team. The researcher did not associate with either design; and, because we developed both systems (both appearing unfamiliar and novel to participants) it was not the case that there was an obvious design that the researcher associated with [12].

### Procedure

At the start of the study the local researcher demonstrated both the ACQR and SMS methods several times in a group setting, using an example voice site read in Kannada (see Fig. 3). Following this demonstration, participants were interviewed individually, and asked to rate each technique with a percentage (a form of rating familiar to participants), choose a preferred design, and make additional comments on the usability and value of each approach in their daily lives.

### Findings

Interview responses from farmers confirmed a strong desire for a facility that allows voice site sharing, as indicated in the exploratory study reported at the start of this paper. When questioned about the value of such a system on the Spoken Web, participants were united in their view that it would

**Figure 3. Conducting the Indian evaluation in Karnataka, India.**

provide a worthwhile and practical solution. Comments made by participants about a potential sharing service included: *"very helpful, 100 % useful – organic farming is a new trend and we want to share experiences and knowledge on it [the Spoken Web]," "this is [an] extremely useful medium where we can share with another very fast,"* and *"I am confident that if you launch such a thing it will be a success across India."*

Participants felt the ACQR service would be easier to use than the SMS approach, with average ratings of 81 % and 62 % respectively. Statistical analysis using a paired t-test confirmed the significance of these scores ($p = 0.015, t = 2.573, DF = 31$). The ACQR service was also generally preferred by participants, with 16 of 30 selecting it as their first choice, 10 choosing SMS and four selecting both with equal preference.

Many of the participants who thought the ACQR service would be easiest to use stated low literacy levels as their main concern with the SMS approach, suggesting that this method would be difficult to use for those with low written language skills. Others appreciated the speed of audio-based sharing: *"[the ACQR service] saves a lot of time, it is efficient and good for sharing between farmers."* Two of the participants who preferred the ACQR service, but were also adept at text messaging, saw the potential of the SMS version, suggesting that it would be very useful for people who are not geographically close to them: *"illiterate people have mobile phones and struggle initially but can learn – [it would be] very useful for me."*

## LONG-TERM FIELD DEPLOYMENT

Our evaluation of the ACQR and SMS designs with Spoken Web users suggested that the audio-based technique would be preferable for ease of use, speed and literacy reasons. After these successful trials, then, we chose the ACQR design for deployment. The service was deployed on the same farming information voice site as in the study reported at the start of this paper. This extended field investigation of the design allowed us to gather more naturalistic usage data over an extended period of time. At the time of writing, the service has been continuously available to voice site callers for eight months.

### Procedure

To inform existing users about the addition of the ACQR sharing service, an announcement was added to the main welcome area of the voice site. This spoken prompt informed callers that a sharing capability had been added, and gave them general usage instructions for sharing and receiving. After choosing to share or receive, a separate prompt instructed callers about how to hold their phones together to share. To help determine usage rates and user behaviours while sharing, the voice site logged all interactions with the service.

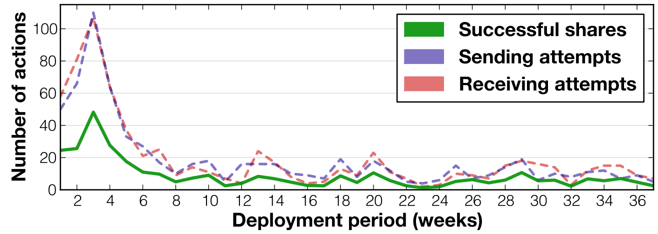


**Figure 4. ACQR usage to date, showing the number of sending and receiving actions attempted, and the number of successful shares.**

We also called six people directly to explain how to share. These users were locally-expert farmers, respected by the rest of the community – our hope was that they would later use the service and help others understand its usage. The expert farmers were told about the new functionality and instructed step-by-step how to share with others using audio codes. With the exception of these six people, however, the only way to learn about the new sharing service was via the introductory announcement, or by word of mouth from other users.

After the first two weeks of deployment, we conducted individual, semi-structured phone interviews with 14 farmers. Participants in these interviews were randomly selected from both users and non-users of the technique (following the approach of [29]). Interviews lasted on average 20 min each, and provided qualitative feedback about the sharing service.

### Results

The ACQR service has been used by 358 individuals during the deployment to date (35 % of 1033 unique callers). Of these, 246 have attempted sending, and 267 have tried receiving, with many trying both functions. The send and receive actions have been performed 1202 and 1004 times, respectively. However, some of these interactions, particularly those early in the deployment, were the result of curiosity or experimentation on the part of users and therefore did not result in actual sharing. To date, there have been a total of 294 successful sharing actions using the service, as illustrated in Fig. 4.

#### *Farmers' feedback*

Five of the farmers interviewed at the end of the first two weeks of deployment had successfully shared with the ACQR service. Their comments suggested that they were excited about its capabilities: *"it works really well," "this is very useful,"* and *"I want the farmers to use it more."* At the end of this two-week period there had been a total of 47 successful sharing actions. Interviews with some participants indicated that there were initial difficulties in understanding how to share, however. Some farmers had assumed that the ACQR method allowed them to share remotely; in this case sharing had failed. Others had struggled with the timing of holding their phones together. We improved the instruction prompts as a result of this feedback, and saw an increase in usage (see Fig. 4, week 3) that is perhaps attributable to the better guidance available to callers.

### Discussion

Over 350 Spoken Web callers have used or experimented with the ACQR service during its field deployment to date, sharing 294 hierarchy locations with other callers. After a short initial period of experimentation and curiosity, usage of the sharing functionality has remained relatively stable, with up to 20 sending attempts per week (see Fig. 4). This level of usage is encouraging for a new interaction method that is so different from the current ways of sharing content from voice sites.

Early user comments suggested some difficulty understanding how to share voice site locations; as a result we refined the spoken prompts to better explain how to use the ACQR method. However, despite this adjustment, the number of unsuccessful sharing attempts has remained higher than we would have liked throughout the deployment. This could perhaps be due to experimentation by new users, confusion about how to share, or the result of unsuccessful attempts to share remotely, as discovered in our earlier interviews. We plan to leave the ACQR sharing service in place on this voice site for the foreseeable future, and will continue to monitor usage and improve interaction and functionality where possible.

## SUMMARY OF FINDINGS

We began this paper by demonstrating the potential benefits of a sharing service to voice site callers looking to send or receive specific information from the complex and currently unsearchable hierarchy of the Spoken Web. Interviews with Spoken Web users from rural Gujarat confirmed that manual, human memory-based schemes for sharing voice site content are already used, but also highlighted the inadequacy of these methods for sharing exact audio positions, and strongly indicated that a more reliable and usable solution was required. With these requirements in mind, we designed and implemented a novel audio-based sharing technique that allows users to quickly and efficiently share voice site positions without an internet connection or extra equipment.

We measured the robustness of the ACQR service, demonstrating high accuracy in a wide range of background noise situations. We also explored its feasibility against two alternative designs, showing that the audio-based approach is faster than both SMS messages and visual QR codes, even when network effects are removed. While there were no significant results to suggest any one method was any easier or harder to use than any other, user ratings were positive for all designs.

The ACQR design was then tested against the SMS-based approach with rural Indian farmers, yielding favourable reactions to a sharing interface for the Spoken Web, and highly positive comments about the use of audio-based sharing. Finally, we deployed the ACQR service on a popular Spoken Web voice site. Early users' reactions were favourable, and voice site callers have used audio-based sharing for the past eight months. The ACQR service is still in place on the voice site, and we expect to report in the longer term about its continued usage.

## CONCLUSIONS AND FUTURE WORK

In this paper we have presented a novel audio-based method for sharing specific positions within Spoken Web voice sites. The evidence gathered from multiple evaluations of the ACQR service suggests that the technique provides a robust, fast and easy to use method of sharing within voice services.

We have explored the range of benefits of three possible sharing designs, and shown how the ACQR technique, which does not require any specialist hardware or an internet connection, can be used by normal callers to a popular Spoken Web voice site. The audio-based design is robust, and faster than the two alternative sharing methods we compared it against. It was also well received by Spoken Web users from rural Gujarat, India, with an average ease of use score significantly higher than an SMS-based design. The technique offers further benefits beyond an SMS method that address the requirements of the farmers interviewed at the start of this work. For example, particularly on lower-end phones, the ACQR method allows users to simply replay the audio code at will, rather than memorising an access number and having to hang up and re-dial if an error (such as forgetting the code) occurs.

Our studies of the ACQR technique have provided strong evidence to support its usage for sharing audio positions, but we also envisage several areas where the audio-based approach could provide useful functionality outside the scenarios explored in this paper. For example, audio codes could be used to replace standard QR codes for visually impaired or blind users. Another possibility could be to incorporate audio codes into sound-based greetings cards or toys to direct users to personalised messages across a phone line. Synchronising positions within other media such as music or video files is also a potentially useful extension. For example, a person watching a video on a mobile device could use an audio code to transfer to a PC to continue watching from the same position.

The aim of this paper is to explore the notion of sharing amongst users with limited resources. We have investigated three designs comprehensively, particularly the use of audio-based sharing for voice services. However, we also hope that this paper stimulates discussion on more sophisticated information interaction services in low resource settings for very large communities of underserved users, and provokes more researchers to design for these contexts.

## ACKNOWLEDGEMENTS

This work was funded by EPSRC grant number EP/J000604/2, and also supported by IBM Research India. We thank Gururaj Mahajan for his vital help during the study in Karnataka.

## REFERENCES

1. Agarwal, S., Jain, A., Kumar, A., and Rajput, N. The world wide telecom web browser. In *Proc. ACM DEV '10*, ACM (2010), 4:1–4:9.
2. Ahonen, T. *Phone Book: Statistics and Facts on the Mobile Phone Industry*. 2010.
3. Aoki, P., Grinter, R., Hurst, A., Szymanski, M., Thornton, J., and Woodruff, A. Sotto voce: Exploring the interplay of conversation and mobile audio spaces. In *Proc. CHI '02*, ACM (2002), 431–438.

4. Bagchi, S., and Mitra, S. Efficient robust DTMF decoding using the subband NDFT. *Signal Processing 56*, 3 (1997), 255–267.
5. Balfanz, D., Balfanz, D., Smetters, D. K., Stewart, P., Stewart, P., and Wong, H. C. Talking to strangers: Authentication in ad-hoc wireless networks. In *Proc. NDSS '02*, ISOC (2002).
6. Bassoli, A., Moore, J., and Agamanolis, S. TunA: Local music sharing with handheld Wi-Fi devices. In *Proc. Wireless World '04*, University of Surrey (2004).
7. Bender, W., Gruhl, D., Morimoto, N., and Lu, A. Techniques for data hiding. *IBM Systems Journal 35* (1996), 313–336.
8. Castelluccia, C., and Mutaf, P. Shake them up!: A movement-based pairing protocol for CPU-constrained devices. In *Proc. MobiSys '05*, ACM (2005), 51–64.
9. Cervantes, R., and Sambasivan, N. Voicelist: User-driven telephone-based audio content. In *Proc. MobileHCI '08*, ACM (2008), 499–500.
10. Chen, S., and Ravallion, M. The developing world is poorer than we thought, but no less successful in the fight against poverty. World Bank, 2008.
11. De Barros, R., Russomanno, T., Brenzikofer, R., and Figueroa, P. A method to synchronise video cameras using the audio band. *Journal of Biomechanics 39*, 4 (2006), 776–780.
12. Dell, N., Vaidyanathan, V., Medhi, I., Cutrell, E., and Thies, W. "Yours is better!": participant response bias in HCI. In *Proc. CHI '12*, ACM (2012), 1321–1330.
13. Dhanesha, K., Rajput, N., and Srivastava, K. User driven audio content navigation for spoken web. In *Proc. MM '10*, ACM (2010), 1071–1074.
14. Dubey, V., Dubey, N., and Chouhan, S. Wireless sensor network based remote irrigation control system and automation using DTMF code. In *Proc. CSNT '11*, IEEE (2011), 34–37.
15. Gitau, S., Marsden, G., and Donner, J. After access: challenges facing mobile-only internet users in the developing world. In *Proc. CHI '10*, ACM (2010), 2603–2606.
16. Goodrich, M., Sirivianos, M., Solis, J., Tsudik, G., and Uzun, E. Loud and clear: Human-verifiable authentication based on audio. In *Proc. ICDCS '06*, IEEE (2006).
17. Gunaratne, J., and Brush, A. J. Newport: enabling sharing during mobile calls. In *Proc. CHI '10*, ACM (2010), 343–352.
18. Holmquist, L. E., Mattern, F., Schiele, B., Alahuhta, P., Beigl, M., and Gellersen, H.-W. Smart-its friends: A technique for users to easily establish connections between smart artefacts. In *Proc. UbiComp '01*, Springer-Verlag (2001), 116–122.
19. Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S., and Nanavati, A. WWTW: The world wide telecom web. In *NSDR workshop, SIGCOMM '07*, ACM (2007).
20. Lerer, A., Ward, M., and Amarasinghe, S. Evaluation of IVR data collection UIs for untrained rural users. In *Proc. ACM DEV '10*, ACM (2010).
21. Manjunath, G., and Revathi, M. Rural healthcare on lowend phones. Tech. rep., HP Laboratories, 2011.
22. Matsuoka, H., Nakashima, Y., Yoshimura, T., and Kawahara, T. Acoustic OFDM: Embedding high bit-rate data in audio. In *Proc. MMM '08*, Springer-Verlag (2008), 498–507.
23. Medhi, I., Lakshmanan, M., Toyama, K., and Cutrell, E. Some evidence for the impact of limited education on hierarchical user interface navigation. In *Proc. CHI '13*, ACM (2013).
24. Mistry, P., Nanayakkara, S., and Maes, P. Sparsh: Passing data using the body as a medium. In *Proc. CSCW '11*, ACM (2011), 689–692.
25. Östergren, M. Sound pryer: Adding value to traffic encounters with streaming audio. In *Proc. ICEC '04*, Springer-Verlag (2004), 541–552.
26. Patel, N., Chittamuru, D., Jain, A., Dave, P., and Parikh, T. Avaaj otalo: A field study of an interactive voice forum for small farmers in rural India. In *Proc. CHI '10*, ACM (2010), 733–742.
27. Patel, N., Klemmer, S., and Parikh, T. An asymmetric communications platform for knowledge sharing with low-end mobile phones. In *Proc. UIST '11 Adjunct*, ACM (2011), 87–88.
28. Pearson, J., Robinson, S., Jones, M., and Nanavati, A. Sharing voice sites on the spoken web. In *MWB workshop, MobileHCI '12*, ACM (2012).
29. Robinson, S., Rajput, N., Jones, M., Jain, A., Sahay, S., and Nanavati, A. Tapback: Towards richer mobile interfaces in impoverished contexts. In *Proc. CHI '11*, ACM (2011), 2733–2736.
30. Telecom Regulatory Authority of India. Highlights on telecom subscription data as on 31st March 2012. See: http://goo.gl/cPBlC, May 2012.
31. Telecom Regulatory Authority of India. The Indian telecom services performance indicators: October – December 2011. See: http://goo.gl/g8xSl, April 2012.
32. Typke, R., Wiering, F., and Veltkamp, R. A survey of music information retrieval systems. In *Proc. ISMIR '05*, University of London (2005), 153–160.
33. Vallina-Rodriguez, N., Hui, P., and Crowcroft, J. Has anyone seen my goose? Social network services in developing regions. In *Proc. CSE '09*, IEEE (2009), 1048–1053.
34. Wang, A. An industrial strength audio search algorithm. In *Proc. ISMIR 2003*, The Johns Hopkins University (2003), 7–13.
35. White, J., Duggirala, M., Kummamuru, K., and Srivastava, S. Designing a voice-based employment exchange for rural India. In *Proc. ICTD '12*, ACM (2012), 367–373.
36. World Bank. World development indicators. See: http://goo.gl/bCaiP, January 2012.