Sharing Voice Sites on the Spoken Web

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

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

[2] IBM India Research Laboratory
Institutional Area, Vasant Kunj
New Delhi, 110070, India
namit@in.ibm.com

ABSTRACT

In this work-in-progress we examine methods for sharing content from the Spoken Web – an Indian hierarchical, audio-based information service, that is analogous in many ways to the internet. We explore current information sharing methods via a preliminary study with local farmers who already use the service, and present two prototype sharing designs for future use, using both SMS messaging and dynamically-generated QR codes. Both methods allow users to share a Voice Site—and the exact position within the site—with others, either locally or remotely. We present the results of a short feasibility study of our two interface designs, and conclude by discussing potential directions for future work in this area.

INTRODUCTION AND MOTIVATION

Internet use in developing regions is generally considerably lower than that in more developed countries. There are many challenges faced by people in such regions that hinder their ability to access the web. For example, many potential users have low written language skills, while even users who are more literate may still find much of the internet inaccessible, as it is predominantly in English (rather than their local language). The low incomes of many people in these regions also mean they are less likely to be able afford the equipment or mobile data packages needed to access the web.

Despite the low penetration of internet-enabled devices in developing countries, however, there has been a high uptake of basic cellular phones [1] – devices that are the main focus of the India-based Spoken Web [2]. The Spoken Web was developed to create a voice-driven alternative to the internet that addresses the challenges of data access, textual literacy and cost. Its main focus is to provide users of low-end phones with access to cheap, audio-based information, and without the need for textual literacy or an internet connection.

The backbone of the Spoken Web service is a series of interconnected *voice sites*. A typical voice site is hierarchical in structure and allows the user to navigate to specific audio segments using their phone's keypad, in a similar way to traditional IVR systems. Prompts and previews help the user to traverse the menu structure and, after a series of these menus, find what they are searching for. Sharing this information with other people can be difficult, however. When using the internet, users can simply copy the URL of a specific website and send it electronically to a recipient. In contrast, users of the Spoken Web currently have no reliable method of transferring direct access to specific voice site information to other people.

CURRENT SHARING METHODS

To investigate existing sharing methods, and issues surrounding the sharing of voice sites, we conducted telephone interviews with eight farmers in Gujarat, India, all of whom are heavy users of an existing voice site (see [3]). The results of these interview sessions confirmed a desire to share content from the voice site by all eight participants, and also revealed several current methods of doing so.

Three existing sharing methods were used by the people we interviewed. Some used the loudspeaker on their phone to share with a small group of local farmers; others simply tried to remember all the relevant information and relay it to others later. Neither of these existing methods allow the recipient to access the information on their own device, however, meaning that they cannot continue to browse the site independently, or review the information later when more convenient. If the information that is shared is complex or lengthy this will become increasingly problematic. Two of the people we interviewed reported that they had occasionally memorised the navigation steps required, and then recited these button presses (and the original voice site phone number) to the recipient – a process that is error-prone and time consuming.

PROTOTYPES

From the results of our small study of existing sharing methods, we can see that sharing facilities for the Spoken Web are lacking at present. To address this issue we developed two prototypes, aiming to test ways for providing users of the Spoken Web with the facility to quickly and reliably share content with others, in much the same way as URLs can be shared on the internet.

Our prototypes were developed on Android-based smartphones, using mock-ups of Spoken Web voice sites. Developing using simulated voice sites, rather than for a real system, allowed us to test and perfect technological approaches with a subset of users before releasing the sharing methods to actual voice site callers. Both of our prototype systems have the same basic keypad display, from which the user can dial into a voice site. At any point in the call the user can press a key (currently #) to "share" a specific voice site position. The two sharing methods used were:

SMS: When a user decides to share a position within a voice site, they are prompted to enter the phone number of the person they would like to share with. The system then sends an SMS message to this number with the exact voice site telephone number and position within it (see left in Fig. 1).

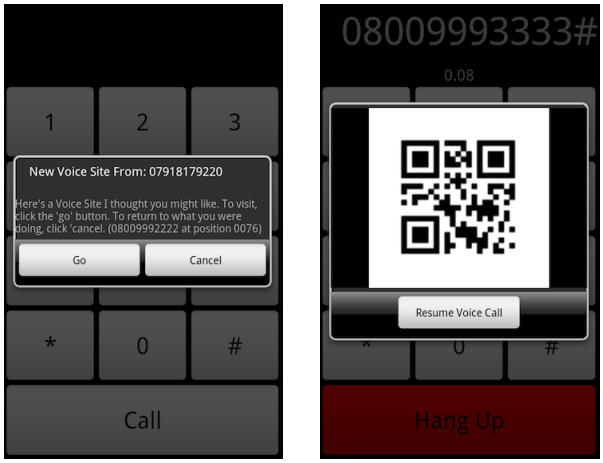

Figure 1. Our audio sharing prototypes. Left: SMS interface, receiving an sharing message from another user. Right: QR interface, sharing a precise voice site position with another user.

QR: In this prototype, when a user shares, the system generates a QR code and displays it on-screen (see right in Fig. 1). The recipient then uses their barcode scanner to read the code, decoding the voice site number and position automatically.

We undertook a short user evaluation to test the two prototypes and explore potential sharing alternatives, gathering user feedback and performance metrics for each of the interfaces.

USER EVALUATION

A short, lab-based pilot study was conducted in the UK with six pairs of users (12 participants total). Each session lasted around 40 minutes. At the start of each session the pair were given a short overview of the interfaces, and then asked to perform six sharing tasks on each prototype, sharing voice site positions back and forth between each other. To avoid bias from ordering and learning effects, we counterbalanced the order in which the interfaces were used.

Following the sharing tasks, participants completed individual written questionnaires, and completed the with a semi-structured interview. We collected both subjective scores and task timings as metrics for determining the most appropriate of the two interfaces for sharing voice site positions. The time taken to complete a sharing task is defined as the difference between when the sender presses the button to share and when the receiver has arrived at the correct position within a voice site. Timings were logged automatically by the prototypes.

In addition to the timed results, we also wanted to determine which interface was easiest to use. We asked participants to rate each system on a Likert scale (1–7, 7 being the highest) for: ease of sending an audio position to their partner; and, ease of receiving an audio position from their partner.

Results

Sharing voice site positions between users took 21 s and 12 s for the SMS and QR systems, respectively. Statistical testing using a paired t-test shows that these results are highly significant ($p < 0.0005, t = 4.654, DF = 35$). The QR system yielded higher average subjective scores for the Likert scale questions (sending: 6.25; receiving: 6.35) than the SMS prototype (s: 5.67; r: 6.33). However these results were not statistically significant (Wilcoxon signed rank test). It is clear from the results, however, that both interfaces were ranked highly overall, suggesting that both are relatively easy to use.

When asked which interface they would use if given a choice, eight of the 12 participants chose the QR prototype, while only four selected the SMS version. When those who chose the QR prototype were questioned further about their choices, the reasons given included that "it was fun" to use, or just that they had not used QR codes before, making them more inclined to use this version. Interestingly, two of the participants who chose SMS gave familiarity as the reason for their preference, stating that they had no experience using QR codes and so found them challenging to use.

AUDIO SHARING CHALLENGES

It is clear from the results of our pilot study that the audio sharing methods we have tested are potentially viable for real usage on actual voice sites. There are several obvious benefits and drawbacks to both of these methods, however.

SMS-based sharing can be used on basic cellular handsets, and allows users to share remotely. However, this method requires the receiving user to have some level of textual literacy (or recognition of the message's context). It also requires memorisation of the voice site number and position code, and is not free to use (the sender, or the operator of the voice site service, must pay for the message). QR-based sharing is free to send and receive, and does not require textual literacy, but the code must be scanned by a cameraphone and can therefore only be shared between users who are in the same physical location. Sending QR codes via MMS is an alternative method, but would incur further cost, and also require some way to decode a QR code from an image already on the receiving device (rather than via its camera).

Our two current sharing methods are text- and visual- based. However, the Spoken Web is entirely audio-based. One future sharing technique might be, then, to initiate a call to the recipient from the service, rather than send an SMS message. This brings further costs to the service operator, however. Further work is needed to investigate this and other ways of sharing audio content that are attractive to service users.

CONCLUSIONS

In this work-in-progress we have explored the need for sharing precise locations within audio-based information services, and described two early enabling prototypes. A pilot study using simulated Spoken Web sites generated positive reactions from UK-based participants to the techniques. Our next research step is to integrate these and other tools into the Spoken Web itself, and undertake a longitudinal study of real-world use.

REFERENCES

1. *ICT Development Index*. ITU, 2002.
2. Kumar, A., Rajput, N., Chakraborty, D., Agarwal, S., and Nanavati, A. WWTW: The world wide telecom web. In *Proc. NSDR workshop '07*, ACM (2007).
3. 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.