An Audio Wiki for Publishing User-Generated Content in the Developing World

Pratik Kotkar
Indian Inst. of Technology, Guwahati
Guwahati, India
pratik@iitg.ernet.in

William Thies
Massachusetts Inst. of Technology
Cambridge, MA, U.S.A.
thies@mit.edu

Saman Amarasinghe
Massachusetts Inst. of Technology
Cambridge, MA, U.S.A.
saman@mit.edu

ABSTRACT

The power of the Web revolution came with making everyone a publisher, rather than just a consumer of information. However, as the prevailing interfaces to the Web require a high degree of literacy and full-fledged computer access, much of the world's population lacks the skills or resources needed to create online content. In this work, we aim to lower the barrier to sharing local information via an Audio Wiki: a repository of spoken content that can be accessed and modified via a low-cost telephone. Because content is in purely audio form (from entry to playback), the system is accessible to the illiterate and naturally supports any local language. By enabling poor and illiterate users to create local stores of information, we envision broad applications in agriculture, health, government, and entrepreneurship.

1. INTRODUCTION

In the developed world, most user interfaces involve the computer as a display and text and images as the main forms of information. While the availability of computers and high literacy rates in developed countries allows for such interfaces, large displays and text-based user interfaces (as demanded by Web browsers) are a mismatch for the developing world. In the developing world, a large part of the population is illiterate and economic constraints do not allow for widespread use of computers. Furthermore, over 60% of today's Web pages are in English, while less than 10% of the world's population speaks English as their native tongue. Even when interfaces are based in local languages, the illiterate population of developing countries stays out of their reach. The illiterate population accounts for 40% of India and over a billion people worldwide. This highlights the need for voice-based content and user interfaces that would reach out to the entire population of the developing world.

The Audio Wiki fills this need by hosting all content in audio form: users can create, edit, and listen to content without having to read any text. In addition, the Audio Wiki is accessible via a low-cost phone, which is the most widespread

computational device in developing countries. In 2005, the mobile phone penetration was about 10% in India while the computer penetration was only 2% [1]. Similar figures for Ghana reveal that about 15% had access to mobile phones but only 0.75% had access to a computer. This underscores the need to migrate the prevailing user interface from a computer to a phone.

There are many interesting challenges in designing a usable interface for an Audio Wiki. Our system departs from the hierarchical format of Wikipedia entries and instead relies on a flat set of topics, each of which is open to comments by users. In the following sections, we describe our approach to organizing, editing, linking, and searching content in such a framework. We have completed an initial prototype of the Wiki and are currently expanding its functionality and ease-of-use. We envision that it can serve as a general platform for researching new approaches to voice-based user interfaces.

2. SYSTEM DESCRIPTION

The basic setup of our system is based on a central server which will be interfaced with a speech recognition engine, an information storage database and an audio recording and playback system. We use an Asterisk-based server [2] for this purpose because it easily interfaces with databases and the prevalent telephony systems. The system allows users to listen to content on any topic added by other users as well as edit the content or leave their own views about the topic.

2.1 User Interface

The mobile phone serves as the primary interface to the Audio Wiki. While all user-generated content on the Wiki is held in audio format, the navigation system is a hybrid of keypad navigation and speech recognition (in progress). The Sphinx 2.0 speech recognition engine [3] will be used since it interfaces directly to the Asterisk server.
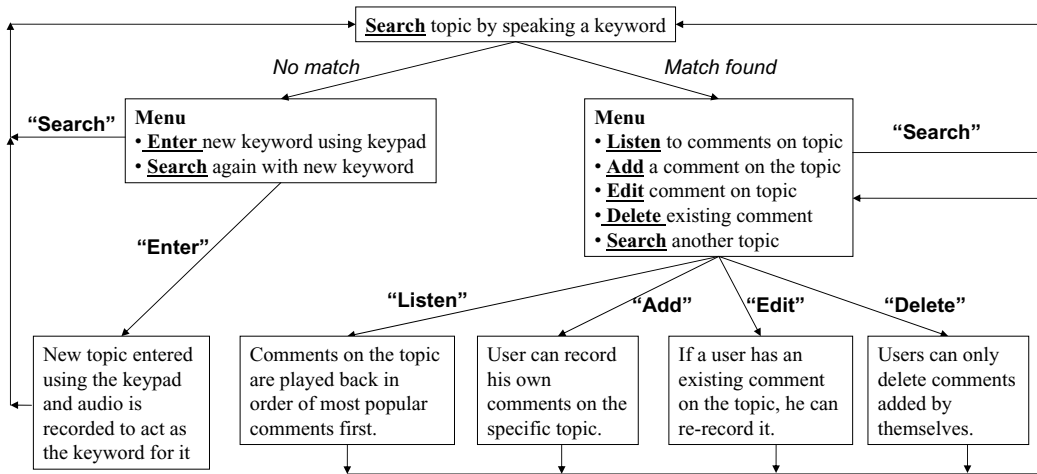
2.1.1 Organizing content

While existing systems such as Wikipedia are often hierarchical, in an audio context hierarchies are hard to navigate when interleaved with content. Thus we envision a flat organization, where the analog of a webpage is a single *topic* with a set of user-generated *comments* (see Figure 1). Users can add a new topic if it does not exist in the database. During audio playback, comments are played one at a time and users can freely skip to the next comment. The most popular comments (i.e., those that are skipped least) propagate to the front of the list so that users hear them first.

Figure 1: User interface for the Audio Wiki. Content is organized by topic. Users may create topics, add spoken comments to topics, or edit / delete their previous comments.

2.1.2 Editing content

A major advantage of sharing information on a Wiki is the ability to edit content. The editing on Wikipedia is at the granularity of individual words and phrases, which can become very difficult in a speech-based system. With the Audio Wiki, editing and adding content to a topic is done at the granularity of comments, analogously to paragraphs on a webpage. Users can record one comment per topic, and can update their entry by re-recording the entire comment. This ensures smooth and coherent audio playback.

2.1.3 Searching content

Full-text interfaces break up the search phrase into words which are searched. For a speech-based system this is not feasible because segmenting words and extracting important ones from spoken entries is difficult. Hence, with the Audio Wiki we will use keyword-based search where the speech engine will only have to recognize a set of keywords. We also enable a hybrid key-entry facility which resolves conflicts in case there is a failure to recognize or a conflict between two keywords. Another interesting possibility is to use a set of queries (such as 20 questions [4]) to hierarchically resolve the topic that the user is interested in.

2.1.4 Linking content

Linking one topic to another is an important part of an information system. Perhaps the simplest approach is to allow users to provide a set of links at the end of each comment, rather than interleaving the links with comment itself. However, we also hope to evaluate an interleaved approach where a tone is played in the background to signify a link to the word just spoken; the user could follow the link by pressing a key.

2.1.5 Adaptive interfaces

To minimize the amount of time that experienced users spend in navigating menus, the Audio Wiki will automatically adapt the user interface to suite the experience of the user. Frequent users require a smaller number of prompts from the system while new users require a detailed explanation of the system to understand its functionality.

3. RELATED WORK

The MobileED group has developed an audio wiki and evaluated its use in schools in South Africa [5]. Their system supports audio playback of existing Wikipedia content as well as user-generated audio annotations. Our system (conceived independently) aims to apply a similar concept to a broad set of users in agriculture and health in addition to education. We also hope to develop innovative interfaces that can apply across audio wiki projects.

The World Wide Telecom Web (WWTW) [6] aims to allow users to create "voice sites" that are linked together analogous to Internet websites using telephones. Though the interface developed in this case would be similar to the Audio Wiki, the major difference lies in that proposed sites on the WWTW are organized by user (analogous to a home page), while sites on the Audio Wiki are organized by topic with many users contributing to a single site.

4. CURRENT STATUS

A prototype of the Wiki has been developed in Boston, USA using an English interface for testing purposes. In parallel, we are developing the interface in Hindi to be deployed in India. The prototype relies on keypad input to navigate the system menus. Our current focus is on integrating a speech recognition engine to simplify navigation for the user.

5. REFERENCES

[1] ITU report, http://itu.int/newsroom/wtdc/2006/stats/.

[2] J. Meggelen, J. Smith, and L. Madsen, *Asterisk - the Future of Telephony*. O'Reilly Media, 2005.

[3] K. Lee, H. Hon, M. Hwang, S. Mahajan, and R. Reddy, The Sphinx Speech Recognition System, *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-38(1), 35-44.

[4] 20 Questions, http://www.20q.net/.

[5] T. Leinonen, E. Sari, and F. Aucamp, Audio wiki for mobile communities: Information system for the rest of us, *ACM Conference on Human Computer Interaction with Mobile Devices and Services*, 2006.

[6] A. Kumar, N. Rajput, D. Chakraborty, S. K. Agarwal, and A. A. Nanavati, WWTW : The World Wide Telecom Web, *Workshop on Networked Systems for Developing Regions*, 2007.