

VoiceList: User-driven Telephone-based Audio Content

ABSTRACT

We present VoiceList, a telephone-based user-generated audio classifieds service. The aim of the system is to provide an infrastructure for a user-driven community service where there is minimal connectivity to the Internet. Our approach takes advantage of prior knowledge and the ubiquitous presence of telephones in developing nations, overcoming literacy and connectivity barriers.

Categories and Subject Descriptors

H.5 [Information Interfaces and presentation]: User Interfaces – User-centered Design, Voice I/O, Interaction Styles,

General Terms

Human Factors, Design

Keywords

Mobile Devices, Audio-based User Interfaces, User-generated Content, ICT4D

1. INTRODUCTION

The recent move towards user-generated content in social media has resulted in the proliferation of online content production services such as wikis, blogs, and photo sharing websites. One of the most successful technologies in this realm is online classifieds services, such as Craigslist [4]. This service allows users to post advertisements, such as rentals, jobs, consumer goods, tickets, automobiles, along with the means to contact the poster of the advertisement. Such user-driven websites allow anyone with access to the technology to create and edit content, for free.

However, in Developing countries, such as those in Africa and India, fraught with issues of bad connectivity, spectrum management, and policy hurdles, these online services are yet to achieve noticeable penetration. Moreover, the typical user in these regions has low literacy skills, and little or no computing experience. There are also differences between the social and the cultural norms of the designer and the user populations. This, in turn, results in a mismatch between the cognitive and the conceptual models that the users have and the traditional interaction metaphors and Windows, Icons, Menus, Pointers (WIMP) interfaces. The United Nations estimates the total number of illiterate adults to be 799 million worldwide, 270 million of which are located in India alone [8].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobileHCI 2008, September 2–5, 2008, Amsterdam, the Netherlands
Copyright is held by the authors/owners.

By contrast, mobile phones are pervading all social and economic strata in countries previously un-connected. It is estimated that over four billion will have access to mobile phones by 2010, increasing coverage of mobile phones to 90 percent from the current 80 percent [6].

We aim to leverage the widespread popularity and usage of mobile phones to increase the adoption of user-driven content sites. To achieve this objective, we built VoiceList, a system that combines telephony with voice command input and user-generated content. Voice activated telephonic systems have recently begun to be introduced in general usage. Some examples include Speakeasy[1], a web and telephone-based language service, Goog411 and Let's Go Public [7][2], that are automated telephone directories. However, none of these allow the user to create and access their own content solely using the telephone. VoiceList, in contrast, allows its users to *create* and *access* classifieds using the telephone, a device familiar to many illiterate and semi-literate users, that comprise a large percentage of the populations in Developing countries. By creating content through voice, VoiceList breaks free from WIMP conventions, allowing these users to harness the full benefit of the system. The following sections describe the VoiceList system in terms of its implementation and functionality and outline future steps.

2. APPROACH

VoiceList is a telephone-based, user-generated audio classified service. It provides a means to post and access voice content to a central database. VoiceList is informed by the traditional forms of posting services and advertisements, such as classifieds and word-of-mouth. It customizes the classifieds based on the user's location. By remaining independent of the literacy skills and the extra operational operating costs, our hope is that VoiceList would merge into an existing set of practices. Since VoiceList can be accessed by simply dialing a phone number, it does not impose additional technical requirements.

3. IMPLEMENTATION

3.1 Interaction Design

The interaction model of VoiceList is based on a dialogue between the user and the system. The user first dials a number, then VoiceList identifies the location of the user. Next, the user can either create or listen to the classifieds. The user can either speak out the search query such as “Honda Accords made in 2005” or listen to the general categories such as Housing, Jobs, and Services. When the user selects a category, the appropriate filters are applied to the options. For example, when the user selects Cars, the next level set of options are “by Price”, “by Model”, “by Year Manufactured”, “by Technical Features”, “by Aesthetics”, and “by Company”. If there are no further options in a level, the system does not enter the level. At any point, the user can choose to backtrack to a previous state. Following is a dialogue between the user and VoiceList:

S: Which city are you in?
 U: Chennai.
 S: Please select a category: community, housing, for sale, or jobs?
 U: Jobs.
 S: What type of jobs you would like to hear about? office, services and hospitality, medical and health, non profit, professional, skilled trade, or domestic labour?
 U: Skilled trade.
 S: Please state your trade.
 U: Carpenter.
 S: There are 8 jobs for Carpenters. Listing number one: Carpenter with Electrician knowledge required. Say "One" for further details. [The system then continues with the remaining listings. The user can say or type the number of the category that he/she is interested in at any point.]
 U: One.
 S: [The system plays the entire advertisement, where the poster can explain in further detail about the type of job. At the end of the advertisement, the system lists out the contact options for contacting the poster.] Say 'SMS' to send an SMS to this person or say 'Call' to automatically dial this poster's number.
 U: SMS.
 [The system sends an SMS to the user, with the phone number and the keywords related to that listing, that were chosen by the poster. The user can choose to continue to listen to the remaining listings.]
 Similarly, while creating a classified, the user pronounces keywords describing the classified, which are used to categorize the advertisement. The user then speaks out the classified text itself, which is played out without any recognition.

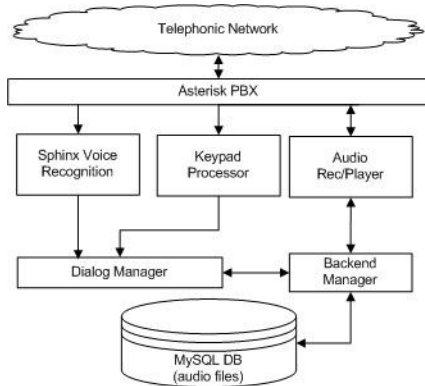


Figure 1. Architecture of the VoiceList system

3.3 System Architecture

With this approach in mind and with (low literacy) users in Developing nations as our target user population, we propose a system implementation that combines reliability with low cost. To focus our efforts on delivering better functionality, the system uses two well-established Open Source technologies: CMU Sphinx [5], that enables speech recognition and synthesis, and Asterisk[3], a PBX telephony engine and toolkit. Figure 1 shows the architecture of VoiceList. When the user calls VoiceList, he or she interacts with the system by voice commands or by key-based input. The voice commands are interpreted by *CMU Sphinx*, which analyzes the voice signal and translates it into text, passing it as an argument to the *Dialogue Manager*. The keystrokes are interpreted by the *Key Processor* and passed as the appropriate arguments to the *Dialogue Manager*. The *Dialogue Manager* controls the interaction between the user and the system, keeping track of the requests of the user, either by spoken sentences or

keystrokes. The *Dialogue Manager* would use NLP techniques used in [2] for understanding the requests of the user in the form of sentences. The *Backend Manager* retrieves the appropriate audio sequence, if the user wishes to listen to an advertisement, or records the user's message if he/she chooses to place an advertisement. The *Audio Recorder/Player* component plays a given audio sequence or records the user's voice input, outputting it to Asterisk. The audio sequences are stored as binary objects in a MySQL database.

4. CURRENT PROTOTYPE

Currently, we have implemented a prototype that allows for the user to issue voice commands and key input. As a next step, we hope to integrate the system with the telephony platform provided by Asterisk. We also hope to integrate NLP techniques to parse long sentences with incomplete and grammatically incorrect phrases. We also plan to integrate functionality for parsing multiple languages and test the system on various accents. Once we have a fully functional system, we will proceed to conduct user studies in a real setting and evaluate its effectiveness. We hope to understand how the categorization, navigation and content make sense to the user.

5. CONCLUSION

We have presented a VoiceList, an user-generated audio-based mobile interface telephone system for user-generated classified advertisements. By combining the existing telephone networks with voice input, we have VoiceList provided a novel user experience to generate classified information advertisements. VoiceList facilitates the inclusion of an otherwise excluded user-population – low literacy users from Developing nations.

6. ACKNOWLEDGMENTS

We would like to thank Arup Chakraborty for his insights into the design of VoiceList project.

7. REFERENCES

- [1] Hirsch, T. and Liu, J. 2004. Speakeasy: overcoming barriers and promoting community development in an immigrant neighborhood, *Designing Interactive Systems*, Cambridge, MA
- [2] Raux, A., Langner, B., Black, A., Eskenazi, M. (2005) Let's Go Public! Taking a Spoken Dialog System to the Real World, *Interspeech 2005 (Eurospeech)*, Lisbon, Portugal.
- [3] Schwarz, B. 2004. Asterisk open-source PBX system. *Linux J.* 2004
- [4] Craigslist, <http://www.craigslist.org>
- [5] CMU Sphinx whitepaper, <http://cmusphinx.sourceforge.net/sphinx4/doc/Sphinx4Whitepaper.pdf>
- [6] Dev Sood, Aditya., <http://www.cks.in/mdr/Mobile%20Development%20Report.pdf>, 2006
- [7] Goog411, <http://www.google.com/goog411/>
- [8] UNESCO 2004, <http://www.uis.unesco.org/>