
Remote Web Browsing via the Phone with TeleWeb

Yevgen Borodin

Computer Science Department
Stony Brook University
Stony Brook, NY 11794-4400, NY
borodin@cs.sunysb.edu

TeleWeb demo video is available at:
<http://www.sbhearsay.net/>

Copyright is held by the author/owner(s).
CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract

TeleWeb is an assistive voice-enabled application empowering users to remotely access the Web through the most ubiquitous device – the phone. The uniqueness of the technology is that it enables users to gain access to information from almost anywhere via a plain, old-fashioned telephone. TeleWeb users will be able to call their own personal numbers, authenticate themselves, and then use speech and phone key-pad to remotely browse the Web on their own PCs. TeleWeb may especially appeal to people with vision loss, as well as older adults who may find the phone interface to be more familiar and easier to use. In this paper, I describe the TeleWeb approach and the interface.

Keywords

HearSay, Blind Users, Web Browser, Screen Reader, Macro Recorder, Macro Player, Non-Visual, Audio Interface, TeleWeb.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia – *architectures, navigation*

General Terms

Design, Human Factors

Introduction and Technology Overview

The World Wide Web has come to play an important part in our lives – it has developed into an indispensable source of information, communication, education, and work. As people are becoming busier and more mobile, they can benefit from accessing public and personal information on the move, *e.g.*, while driving a car, walking, *etc.*

Although smart-phones (*e.g.*, BlackBerry, iPhone, Palm, *etc.*) can be used as computers to access the Internet, they typically require visual interaction, which is inconvenient or even unsafe in many real-life scenarios such as driving. Worse still, visual interaction critically hinders access for persons with vision impairments and blindness, a population that is growing ever larger.

Being the dominant assistive technology used by persons with print disabilities, *screen-readers* [11, 16, 19] have no built-in intelligence, which forces users to discover content by serially listening to or skipping through web pages. At a cost of up to \$1,000, screen-readers are inefficient and provide no automation, while requiring lengthy training to learn hundreds of keyboard shortcuts. Despite these problems, people with vision impairments are leading active lives: they are studying at universities, commuting to work, traveling, *etc.* Just like anyone else, they need quick and affordable access to information. This clearly underscores the need for more accessible and efficient assistive tools with alternative computer interfaces that could benefit any mobile users.

Blind computer users who want to be mobile often use specialized applications that provide remote access to users' home computers [16]. Plain remote desktop

tools [15] can also be used by blind users if the remote computer they connect to has a screen reader. WebAnywhere [2] makes any web browser accessible as soon as the WebAnywhere website is loaded in the browser. However, these technologies require that the user have access to an Internet-enabled computer.

The newest smart-phones are beginning to support some screen-reading functionality [13], involving in some cases touch and gesture interfaces [19]. This technology, which can be expensive, often finds its early adopters among the congenitally-blind users who have grown up with various assistive technologies. However, older adults, who are more prone to vision loss, often find smart-phones too complex; so they continue using cell-phones with basic functions, or even landlines. In addition, some areas may not have cellular Internet, or even cellular service, available.

Voice-driven phone-based services targeted at the general population are a new emerging technology that can also help blind people access information. The 1-800-2ChaCha [7] service enables users to search the Web over the phone. Google has recently released its Voice Search [10] service, which is essentially a voice interface to Google maps. IBM is developing the Telecom Web [12], which is envisioned as a network of interconnected Voice Sites hosting voice-driven applications created by users themselves. Tellme Networks, Inc. [18] provides online voice services such as directory assistance and search via mobile phones. There are also telephony services for accessing email via phone [8], Wikipedia [17], a specific website [9] and even the general Web [14]. However, none of these applications provides remote access to a customizable web browser as TeleWeb does.

TeleWeb Approach

To meet the needs of mobile users, both with and without vision impairments, we developed TeleWeb – a free assistive voice-enabled application empowering users to remotely access the Web through the most ubiquitous device – the phone. The uniqueness of the technology is that it enables users to access information from almost anywhere, using a plain telephone, cell-phone, or smart-phone. TeleWeb users will be able to call their own personal phone numbers, authenticate themselves, and then remotely search the Web, check their email, and perform other web browsing tasks on their home computers by speaking commands and/or using the phone key-pad.

To operate the TeleWeb, users will have to install a free Skype client application and subscribe to a paid (\$60/year) Skype-In service for a personal phone number (www.skype.com). Users also have to install the HearSay browser [6] with the TeleWeb module, which interacts with the Skype client application by sending the text-to-speech output to Skype and forwarding the incoming audio stream to the DTMF¹ and speech recognizers. HearSay-TeleWeb will then provide a “bridge” to the Firefox browser via a Firefox add-on and can be used directly like a desktop screen reader, or remotely via the phone. **Figure 1** illustrates the TeleWeb concept.

In contrast to other approaches offering public telephony access to web-based information [10, 12], TeleWeb does not require maintaining its own dedicated

telephony service and can instead use third party telephony services such as Skype-In (www.skype.com). In this setup, information processing, text-to-speech generation, and speech recognition – all happen on the end-user computers. Compared to other solutions for remote access to computers [2, 16], users do not require Internet connectivity to access the web browser at home.

Initially targeted at people with vision impairments, TeleWeb’s simple telephony interface controlled by a combination of speech commands and telephone keypad shortcuts may also appeal to people with a variety of abilities and needs, including older adults, individuals with cognitive and motor impairments, and anyone who wants to access the Web on the go.

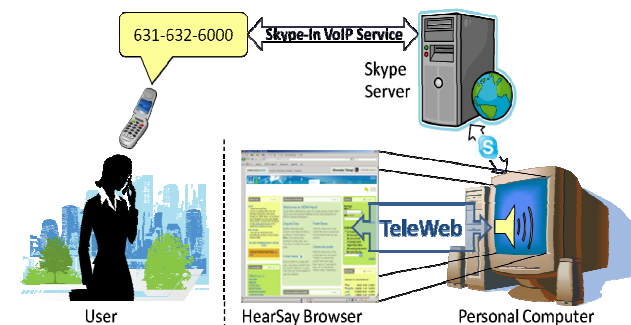


Figure 1. Using TeleWeb

¹ *DTMF*, or Dual Tone Multi-Frequency, is used to create the touch tones corresponding to telephone keys.

Interface

The TeleWeb interface was designed through consultations with a blind expert screen-reader user. To make the TeleWeb interface more flexible and to enable efficient control over web browsers, we prototyped an interface that can use phone keys, speech commands, and voice-enabled macros.

We found that the touch-tone interface had the lowest latency (50-100ms) compared to voice commands. However, since most of today's phones have only 12 keys that generate touch-tone signals, we carefully mapped the available keys to the most frequently used actions (**Figure 2**).

| | | |
|---|--|---|
| Previous heading *1: Main menu Edit: Previous character | Previous segment *2: Bookmarks Edit: Home | Next heading *3: Macro Edit: Next character |
| Previous sentence *4: Previous page Edit: Previous word | Pause/Resume *5: View changes Edit: Copy/Paste | Next sentence *6: Next page Edit: Next word |
| Previous link *7: Recognizer on/off Edit: Previous line | 8: Next segment *8: Change Voices Edit: End | Next link *9: Editing mode Edit: Next line |
| Function key for combos ***: Contextual Help | Cancel, close *0: Documentation Edit: Delete | Select, Click, Enter *#: Enter URL Edit: Start/Stop selection |

Figure 2. TeleWeb Key Mappings

Based on the observation of browsing strategies of screen-reader users [1, 4], we identified that the shortcuts most frequently used by these users include navigation among headings, links, lines, and paragraphs. These functionalities closely match navigation among headings, links, sentences, and segments in the HearSay browser. In addition, screen-reader users frequently use the "Ctrl" key to pause screen readers, the "Esc" key to close dialogs, and the "Enter" key to follow links, press buttons, start editing, etc. We chose to map these functionalities to the telephone keys.

To minimize memorization, semantically opposite actions involving navigation to the previous/next item were assigned to keys located on the opposite sides of the phone keypad. Specifically, navigating to the previous/next heading was mapped to 1/3, the previous/next sentence to 4/6, the previous/next link to 7/9, and the previous/next segment to 2/8. To preserve the conventional semantics of the pound (#) key, we reserved it for selecting, clicking, or pressing "Enter." And from the remaining phone keys, we chose 5 to pause/resume reading and 0 to close/cancel/escape.

To increase the number of actions that can be performed with the keypad, we introduced a telephony menu, where actions can be selected from a list, while reserving Star (*) for key combinations. Pressing *1 opens the main menu, which allows users to listen to and quickly select available actions. Pressing *2 and *3 opens bookmarks and macros [3] selection lists, respectively. Pressing *4 and *6 takes users to previous and next web pages, respectively. *5 allows users to review the differences between web pages [5].

Pressing *7 turns on/off the speech recognizer which may be disruptive in noisy environments. Pressing *8 changes text-to-speech voices, as voices can be less or more intelligible for different environments and different users. Pressing ** provides contextual help for the current state of the TeleWeb interface. Pressing *0 opens TeleWeb documentation. And, finally, pressing *# allows users to edit the browser address bar.

Although it may be easier to set up all the bookmarks for frequently used websites and macros for logging into e-mail on the desktop computer, the touch-tone interface, if necessary, can also be used for entering text. Lacking a full keyboard, many phones map several letters to the same key, which requires that users press the same key several times to get to a letter, *e.g.*, pressing 4 twice types 'G'. TeleWeb also supports this mode of text entry. Pressing *9 toggles between entering text and navigating in the edit mode so that users could press 4 to type 'H' or shift the cursor to the previous word. In the edit-navigation mode, pressing 1/3 shifts the cursor to the previous/next character, 4/6 to the previous/next word, 7/9 to the previous/next line, and 2/8 to the beginning/end of the link. Pressing 0 deletes the character to the left of the cursor. Finally, pressing # starts/stops selection, whereas pressing 5 copies or pastes text, depending on whether something is selected or not.

TeleWeb also supports voice commands. A voice interface, although slower than the touch-tone interface, is more efficient for complex tasks and those tasks that are performed relatively rarely such as opening bookmarks; however, every action has one or more synonymous voice commands associated with it. With improving speech recognition accuracy, it may

soon become possible to use voice for text entry. Voice macros described in [3] enable recording and then replaying complex browsing tasks thus improving the efficiency of the interface.

Conclusion and Future Work

In this paper, I described the architecture and interface of the TeleWeb prototype – a system for remote web browsing via the phone. The current prototype will mostly benefit blind people; however, older adults and individuals with cognitive and motor impairments may also benefit from the telephony access to the Web. And, with improving speech recognition and the need for mobile access to information, TeleWeb may eventually appeal to a wider population.

TeleWeb is currently being beta tested by several blind and sighted users. Following beta testing, we will conduct a user study to evaluate the usability of the proposed interface and to identify possible differences in the learnability and navigational preferences between blind and sighted users.

Some further directions of work include: improving security of the TeleWeb interface, supporting other telephony services in addition to Skype, supporting full keyboards on smart-phones, among others.

Acknowledgements

I thank my advisor I.V. Ramakrishnan for helpful feedback and Glenn Dausch for helping to design and test the TeleWeb interface. We would like to thank NSF (Awards: IIS-0534419, IIS-0808678, CNS-0751083) and Sigma Xi (Award: G20071013028487872) for their support of this project, and also Cepstral.com for donating their realistic synthetic voices.

References

- [1] Bigham, J.P., A.C. Cavender, J.T. Brudvik, J.O. Wobbrock, and R.E. Lander, *WebinSitu: a comparative analysis of blind and sighted browsing behavior*, in *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*. 2007, ACM: Tempe, Arizona, USA.
- [2] Bigham, J.P., C.M. Prince, and R.E. Ladner, *WebAnywhere: a screen reader on-the-go*, in *Proceedings of the 2008 International Cross-Disciplinary Conference on Web Accessibility (W4A)*. 2008, ACM: Beijing, China.
- [3] Borodin, Y., *Automation of repetitive web browsing tasks with voice-enabled macros*, in *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*. 2008, ACM: Halifax, Nova Scotia, Canada.
- [4] Borodin, Y., *Bridging the web accessibility divide*, in *Computer Science*. 2009, Stony Brook University: Stony Brook.
- [5] Borodin, Y., J.P. Bigham, R. Raman, and I.V. Ramakrishnan, *What's new?: making web page updates accessible*, in *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*. 2008, ACM: Halifax, Nova Scotia, Canada.
- [6] Borodin, Y., J.P. Bigham, A. Stent, and I.V. Ramakrishnan, *Towards one world web with HearSay3*, in *Proceedings of the 2008 International Cross-Disciplinary Conference on Web Accessibility (W4A)*. 2008, ACM: Beijing, China.
- [7] ChaCha. *ChaCha mobile search*. 2009 [cited 2009]; Available from: <http://www.chacha.com>.
- [8] email2phone. *Email phone service from Across Communications*. 2009 [cited 2009]; Available from: <http://email2phone.net/>.
- [9] Ferreras, C.G., D.E. Mancebo, and V.C. Payo, *From HTML to VoiceXML: A First Approach*, in *Proceedings of the 5th International Conference on Text, Speech and Dialogue*. 2002, Springer-Verlag.
- [10] GOOG-411. *Google*. 2009 [cited 2009]; Available from: <http://www.google.com/goog411/>.
- [11] JAWS. *Screen reader from Freedom Scientific*. 2009 [cited 2009]; Available from: <http://www.freedomscientific.com/products/fs/jaws-product-page.asp>.
- [12] Kumar, A., N. Rajput, D. Chakraborty, S.K. Agarwal, and A.A. Nanavati, *WWTW: the World Wide Telecom Web*, in *Proceedings of the Workshop on Networked Systems for Developing Regions*. 2007, ACM: Kyoto, Japan.
- [13] MobileSpeak. *Mobile screen reader from Codefactory*. 2009 [cited 2009]; Available from: <http://www.codefactory.es/en/>.
- [14] NetEcho. *Internet Speech*. 2009 [cited 2009]; Available from: <http://www.internetspeech.com>.
- [15] PCNow. *WebEx*. 2009 [cited 2009]; Available from: <http://pcnow.webex.com>.
- [16] SaToGo. *Screen reader from Serotek*. 2009 [cited 2009]; Available from: www.satogo.com.
- [17] Sherwani, J., D. Yu, T. Paek, M. Czerwinski, Y.C. Ju, and A. Acero, *VoicePedia: towards speech-based access to unstructured information*, in *INTERSPEECH-2007*. 2007. p. 146-149.
- [18] TellMe. *Voice services*. 2009 [cited 2009]; Available from: <http://www.tellme.com>.
- [19] VoiceOver. *Screen reader from Apple*. 2009 [cited 2009]; Available from: <http://www.apple.com/accessibility/voiceover>.