

# A Browser for a Public-Domain SpeechWeb

R. A. Frost

School of Computer Science  
University of Windsor  
Windsor, Ontario, Canada  
1-519-2534232 ext 2990

rfrost@cogeco.ca

X. Ma

School of Computer Science  
University of Windsor  
Windsor, Ontario, Canada  
1-519-2534232 ext 2990

Y. Shi

School of Computer Science  
University of Windsor  
Windsor, Ontario, Canada  
1-519-2534232 ext 2990

## ABSTRACT

A SpeechWeb is a collection of hyperlinked applications, which are accessed remotely by speech browsers running on end-user devices. Links are activated through spoken commands. Despite the fact that protocols and technologies for creating and deploying speech applications have been readily available for several years, we have not seen the development of a Public-Domain SpeechWeb. In this paper, we show how freely available software and commonly used communication protocols can be used to change this situation.

## Categories and Subject Descriptors

H.5.4 [Hypertext/Hypermedia]: Architectures, H.5.2 [User Interfaces]: Voice I/O.

## General Terms

Human Factors.

## Keywords

Public-Domain SpeechWeb, X+V.

## 1. INTRODUCTION

The web has hundreds of millions of users. However, for many people, and in many circumstances, the value of the web is limited owing to its dependence on vision.

Frost [1] has argued that the conventional web should be augmented with a Public-Domain SpeechWeb consisting of hyperlinked applications, that are designed specifically for natural-language interaction, created and deployed by non-expert as well as expert users, and which are accessible through readily-available voice browsers. (Hemphill and Thrift [3] made a similar argument over ten years ago but little has happened since then).

As explanation for the lack of progress on the development of a Public-Domain SpeechWeb, Frost et al [2] argue that commonly-used architectures for distributing speech applications are not conducive to such an endeavor. They propose an alternative: the **LRRP** architecture (see next page), in which **L**ocal grammar-based **R**ecognition of speech is carried out by a voice browser on the user device, text translation is sent through the Internet for **R**emote **P**rocessing by an application residing on a conventional web server.

Although Frost et al have demonstrated the viability of using the LRRP architecture; they did not solve the problem of how to facilitate the involvement of non-expert users in the development and use of a Public-Domain SpeechWeb. In particular: 1) The prototype speech browsers, which they created, were based on proprietary software and could not be distributed. 2) The browsers required a substantial runtime environment on the local device. 3) Communication with remote applications involved non-standard protocols.

We have solved these problems by making use of the X+V voice markup language, in a novel way, as discussed in section 2. We have created a small prototype speechweb which we discuss in section 3, where we also explain how to access it. In section 4 we explain how others can create their own speechweb applications, deploy them on conventional servers, and access them using our X+V browser. We include the URL of a video demonstration and of an extended version of this paper.

The major objective of this paper is to illustrate how easy it is to build a Public-Domain SpeechWeb which can be accessed by freely-available browsers, and which can be extended by adding applications created by people with little or no knowledge of speech-recognition technology.

## 2. AN X+V BROWSER

X+V is a markup language, which combines XHTML with a subset of VXML, and which is intended to be used for creating multi-modal hyperlinked applications [5].

Our use of X+V is a little different. Rather than implementing applications in X+V, we use a single X+V page as an interface (browser) into a set of remote hyperlinked speech applications, which can be written in any programming language.

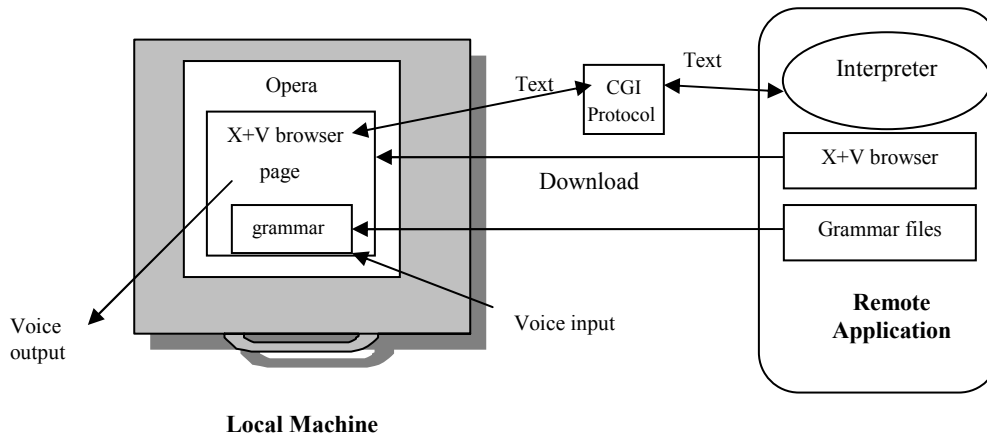
When our X+V browser is directed to a remote application, it begins by downloading a recognition grammar from that application (which is used to tailor the local speech-recognition engine). Spoken user input is converted to text and sent to the remote application, which processes it and returns a text response, which is output as synthesized voice. If the user asks to be connected to another application, the input is recognized as such by the current application, which returns the appropriate URL to the X+V browser.

Due to the fact that current versions of the freely-available multi-modal web browser Opera [4] can be easily configured to incorporate the IBM speech-recognition technology and execute X+V pages, our browser can be used by anyone who has installed Opera on their computer.

### 3. A SMALL PROTOTYPE

We have constructed a small example speechweb consisting of three hyperlinked applications: Monty and Judy can answer a few questions such as “where do you live?”, “how old are you?” etc. and link requests, such as “can I talk to solarman”. Solarman can answer around 160,000 questions such as:

who discovered phobos?  
 did hall discover every moon?  
 which planet is orbited by nereid?  
 which planet is orbited by two moons?  
 how many moons were discovered by kuiper?



Despite the differences in these applications, they are all hosted on the same web server, and are all accessible through our SpeechWeb browser using common communication protocols.

### 4. HOW TO ACCESS OUR SPEECHWEB

Begin by downloading Opera from [www.opera.com](http://www.opera.com) and then configure it for voice capability: click on “Tools” from the toolbar, and choose “Preferences”. Then open the folder “Advanced”. Choose “Voice” from the list on the left. Then check the “Enable voice-controlled browsing” box near the top of the page. Choose “male” or “female” for speech-recognition, then the “scroll lock” option for the key to press when talking. Choose the “hold key while talking” option. Now you can access our speechweb applications by directing Opera to the following URL:

```
http://luna.cs.uwindsor.ca/~speechweb/
    p_d_speechweb/monty/monty.xml
    or ...judy.xml
    or ...solarman.xml
```

Our browser will open automatically in Opera. Hold down the Scroll Lock key while speaking into the microphone. Try these:

“hello there”      “what can I say”      “where do you live”  
 “can I talk to judy” OR “monty” OR “Solarman”

For Solarman, try questions such as those in section 3. You can interrupt the response by pressing the scroll key.

A video demonstration of a session with our example speech web is accessible at:

<http://davinci.newcs.uwindsor.ca/~speechweb/movie.mov>

### 5. HOW TO CREATE APPLICATIONS

You need to create three files and put them into a web directory which supports the .cgi protocol:

- 1) You need a copy of our X+V browser, which can be obtained from the URL given in section 4.
- 2) A grammar file should then be created defining the input language of your application.
- 3) The interpreter for your application can be created as any text in/text out program written in any language. A detailed description of our browser, and complete information on how to create your own voice applications, is available at:

[http://luna.cs.uwindsor.ca/~speechweb/p\\_d\\_speechweb/speechweb\\_instructions.pdf](http://luna.cs.uwindsor.ca/~speechweb/p_d_speechweb/speechweb_instructions.pdf)

### 6. ACKNOWLEDGEMENTS

The authors acknowledge the support of NSERC, the Natural Science and Engineering Research Council of Canada.

### REFERENCES

- [1] Frost, R. A. (2005) A call for a public-domain SpeechWeb. *Commun. ACM* 48, 11, 45–49.
- [2] Frost, R., Abdullah, N., Bhatia, K., Chitte, S., Hanna, F., Roy, M., Shi, Y., and Su, L. (2004) LRRP SpeechWebs. In *Proceedings of the Conference on Communication Networks and Services Research*, Fredericton, N.B., Canada, May 19–21. IEEE Computer Society, 91–98.
- [3] Hemphill, C.T. and Thrift, P. R. (1995) Surfing the Web by Voice. *Proceedings of the third ACM International Multimedia Conference (San Francisco 1995)* 215 – 222.
- [4] Opera: <http://www.opera.com/>
- [5] VoiceXML Forum (2004) <http://www.voicexml.org/specs/multimodal/x+v/12/>