# MySpeechWeb: Software to Facilitate the Construction and Deployment of Speech Applications on the Web

Richard A. Frost, Ali Karaki, David A. Dufour, Josh Greig, Rahmatullah Hafiz, Yue Shi, Shawn Daichendt, Shahriar Chandon, Justin Barolak, and Randy J. Fortier
School of Computer Science, University of Windsor
Windsor ON Canada
01-519-253-4232
rfrost@cogeco.ca

## ABSTRACT
Few voice-in/voice-out applications are available on the web. One problem appears to be the lack of appropriate open-source tools. More speech applications would increase the functionality of the web for people with visual, cognitive, and motor disabilities. Our research group has developed open-source tools for the creation and deployment of speech applications by non-expert as well as expert users, and an open-source software platform to deploy those applications on the web. In addition, a suite of exemplar speech applications, together with documentation, has been built to facilitate the creation and deployment of similar applications by others.

## Categories and Subject Descriptors
H.5.4 [Hypertext/Hypermedia]: Architectures, H.5.2 [User Interfaces]: Voice I/O.

## General Terms: Human Factors

## Keywords
Speech application, voice-application, X+V, SpeechWeb.

## 1. INTRODUCTION

### 1.1 The Problem
Few speech applications exist on the web. One problem is the lack of open-source tools to facilitate the construction of speech applications. Another problem is that the three predominant architectures for providing voice access to applications on the web all have shortcomings as a basis for public involvement in developing and deploying speech applications:

*1) Voice access to web pages provided by various commercial screen readers* enables important non-visual access to conventional web content. However, a) this approach does not support question/answer or conversational applications as the screen reader does not know the language of allowable questions/queries/utterances, b) screen readers are expensive, and c) their software components are not readily accessible to application developers, especially non-experts.

*2) The telephone call-centre architecture* is not appropriate for public involvement in creating and deploying their own speech applications.


*ASSETS'08,* October 13–15, 2008, Halifax, Nova Scotia, Canada
ACM 978-1-59593-976-0/08/10.

3) *Creation of speech applications as hyperlinked VXML pages* has the potential to contribute significantly to the growth of speech applications on the web. However, there are two disadvantages to this approach: (a) the applications must be written in the voice markup language VXML rather than the developer's favorite programming or scripting language, and (b) sophisticated applications, such as those which require substantial natural-language processing and access to large knowledge sources, would require significant client-side computing power if written in VXML. However, we show that VXML, X+V, and other speech markup languages can be used (although in an unconventional way) to facilitate the deployment of speech applications on the web.

### 1.2 Our Contributions
We have developed an architecture, software, and an open-source platform which overcomes all of the limitations above:

1) Our architecture allows speech applications to be created in any programming language, deployed on the web in a few minutes, and be accessible to end-users anywhere through voice using the freely-available Opera browser (see section 2).

2) Our software allows non-expert developers to create question/answer type speech applications in minutes through the completion of web-accessible forms and the deployment of those applications through the click of a button (see section 4).

3) Our software also enables natural-language speech database query processors to be constructed as executable specifications by developers with college or junior-level university programming abilities (see section 4).

4) Our website [3] contains links to exemplar speech applications (see Section 3) and full documentation on how to develop and deploy such applications:

The architecture and speech browser have been described in previous publications, [1] and [2]. The new work reported in this paper, to be demonstrated for the first time at the ASSETS'08 conference, includes the complete open-source implementation of the architecture, the suite of exemplar speech applications, and a web site with extensive end-user and developer documentation.

The major contribution of this work is that by demonstrating the ease which speech applications can be created and deployed, we may help overcome a "perceived-technological barrier" which may be discouraging others from contributing to the development of a Public-Domain SpeechWeb.

## 2. THE MySpeechWeb PLATFORM

Our open-source platform is based on the Local speech Recognition and Remote Processing (LRRP) architecture [2], in which a speech application consists of three components stored in a conventional web directory on a server connected to the Internet: 1) an executable text-in/text-out program written in any programming language supported by the server, 2) a file containing a grammar defining the application's input language, and 3) a file containing a small script written in a speech-markup language such as VXML or X+V.

End users access speech applications using any browser which has the capability of executing the speech markup script. When the script is downloaded to the end user device, the browser executes it. The script begins by speaking out an application-dependent message using the browser's text-to-speech capabilities. The script then contacts the remote server and downloads the grammar file which is used to tailor the speech-recognizer on the end-user device. Finally, the script waits for the user utterance, and if it recognizes it as being in the language defined by the grammar, sends the text translation to the remote application for processing. If the user did not ask to be transferred to a hyperlinked speech application, the result returned is output as voice, and the script cycles back waiting for the next utterance. Otherwise, if the user asked to be connected to another application, the script redirects the browser to the appropriate URL, a new speech markup script is downloaded (corresponding to the new application), and the new script downloads a new grammar file defining the input language of the new application. A more complete description of the LRRP architecture, including relative advantages and disadvantages, is given in [2].

The current implementation of the LRRP architecture uses X+V scripts [1] and the freely-available Opera browser and IBM speech plug-in for end-user access.

## 3. THE EXEMPLAR APPLICATIONS

Our exemplar voice-in/voice-out applications [3] include:

a) Question/answer applications for geography, etc.

b) A sophisticated NL interface to a solar system database.

c) A natural-language calculator, and a units converter.

d) A "read a book" application.

e) A simple English/French word translator.

f) A "time and weather status" for requested cities in Canada,

g) An arithmetic skills tester, and several multiple choice test applications, which illustrate how our architecture supports spoken-dialogue through user-dependent sessions.

h) We also have two multi-modal applications: 1) one provides voice and visual access to conventional web pages. This application demonstrates the way in which our architecture allows speech applications to be seamlessly integrated with the conventional web, with varying combinations of visual and non-visual features, and 2) a game illustrating use of images as well as voice.

## 4. SOFTWARE FOR DEVELOPERS

In addition to providing exemplar speech applications which can be used as templates by others, we have also developed software to assist in the creation of speech applications:

a) An on-line form, accessible from our web site [3], allows users with no programming skills to create simple question/answer speech applications. After entering the questions and associated answers, the user clicks a "submit" button and their application is created and deployed on the MySpeechWeb server. The URL for the application is displayed and the user can immediately begin talking to their application.

b) Multiple-choice question applications can also be created through a form-based interface available on our web site.

c) Two of our exemplar natural-language applications have been created as executable specifications of attribute grammars, using a library of operators that we have created in a functional programming language. The library is for use by developers with college or first/second-year university programming skills. As example of the benefit of constructing applications in this way, the arithmetic calculator was built in three days by a student just completing first-year of a university CS program.

Our website contains manuals for end-users and for developers. The documentation includes instructions for developers who want to host their speech applications on their own web servers.

## 5. THE DEMONSTRATION

The demonstration at ASSETS'08 will include access to the exemplar voice-in/voice-out applications. Conference participants will also be able to use our software to create and deploy their own speech applications, which will be immediately accessible through voice by anyone who downloads the Opera browser, IBM plug-in, and has access to the Internet.

## 6. CONCLUDING COMMENTS

The conventional web grew at remarkable speed for several reasons, one of which was the availability of tools to create and deploy web content. Another, was the adoption of those tools by developers with a variety of skills. We believe that our tools augment technologies such voice interfaces to web content, and speech markup languages such as VXML. In particular, our tools enable non-experts to create and deploy speech applications. The next step is to promote these tools and encourage others to use them. We hope that the demonstration at ASSETS'08 may provide good exposure for this. We are also planning to give demonstrations at local schools, and at universities in North America. We are also working with Disability Services at the University of Windsor in order to identify and build useful applications for their clients.

## REFERENCES


[1] Frost, R. A., Ma, X. and Shi, Y. (2007) A browser for a public-domain SpeechWeb. *WWW 2007*: 1307-1308.

[2] Frost, R. A. (2005) A call for a public-domain SpeechWeb. CACM 48, 11, 45–49.

[3] www.myspeechweb.org , cs.uwindsor.ca/~speechweb