Voice Technologies in the Public Sector for E-Government Services

V

Mark J. Headd
University of Delaware, USA

INTRODUCTION

Spurred by the development of new Extensible Markup Language (XML)-based telephony markup languages and the adoption of requirements for universal access to government Internet content (CITA, 2005), more and more governments are making use of a new family of open-standard development languages to make their visual Web applications accessible by telephones and cell phones. This article will explore the development of the factors driving this trend, and explore possible future trends as these new technologies mature.

BACKGROUND

Early in the development of the Internet as a tool for governments to provide information and services to citizens, researchers recognized the fundamental dilemma of e-government service delivery; access to the Internet and the ability to utilize Web browsers and Web-based services is unevenly spread (Novak, 1998). In the United States (U.S.), researchers identified several key factors that help explain access to Internet technology generally and to Internet-connected computers specifically. Income and other factors unevenly spread throughout the population can influence access to the Internet (NTIA, 1999). Additionally, researchers have identified age as a key factor determining ability to access Internet resources (Fox, 2004). These factors can present significant issues for governments, which operate under a universal service delivery mandate.

At approximately the same time during the 1990s that the Internet began to be used by more and more governments, researchers at AT&T/Bell Laboratories began work on a project to develop a phone markup language to support telephone access to Internet content. Researchers at AT&T, Lucent (following the company's separation from AT&T) and Motorola continued to pursue the development of a telephony markup language and incorporated the framework of XML, which became a formal standard in 1998 (W3C, 2002). The work conducted by these organizations on phone markup languages was eventually submitted to the World Wide Web Consortium (W3C) for consideration as a formal Web standard, and the Voice Extensible Markup Language (VoiceXML) Forum was created to advance the use of voice markup languages.

These efforts resulted in the development of VoiceXML. In simplest terms, VoiceXML is a Web technology that turns any telephone, even a rotary phone, into an Internet device. VoiceXML is a non-proprietary, Web-based markup language for creating vocal dialogs between humans and computers. VoiceXML is similar to another common markup language—Hypertext Markup Language (HTML), the basic language of most visual Web pages.

EVOLUTION OF THE VOICE WEB

Just as a Web browser renders HTML documents visually for display on a computer monitor, a VoiceXML interpreter renders VoiceXML documents for vocal representation (W3C, 2004). In this respect, one can think of a VoiceXML interpreter as a telephone-based voice browser. As with HTML documents, VoiceXML documents have

Figure 1. An HTML Web page

```
<html>
<head>
<title>NIST.gov</title>
</head>
<body>
<h1>Welcome!</h1>
<h2>
This is the National Institute of
Standards and Technology.
</h2>
<p> NIST is a non-regulatory federal
agency within the U.S. Commerce
Department's Technology
Administration.
</p>
</body>
</html>
```

Figure 2. A VoiceXML dialog

```
<vxml>
 <form id="welcome">
 <block>

<prompt>
Welcome!
<break size="medium"/>
This is the National Institute of
Standards and Technology.
NIST is a non-regulatory federal agency
within the U.S. Commerce Department's
Technology Administration.
</prompt>

 </block>
 </form>
 </vxml>
```

uniform resource identifiers (URIs) and can be located on any traditional Web server. However, instead of pointing a client-side Web browser at a specific URI, citizens can access a VoiceXML application by calling a telephone number from any ordinary telephone—cellular or traditional, touch-tone or rotary.

VoiceXML belongs to a family of open Web standards referred to collectively as the speech interface framework (SIF). The underlying communications protocol used by VoiceXML and the SIF is the hypertext transfer protocol (HTTP), the same basic communications format used by visual Internet applications (Rehor, 2001).

BENEFITS OF THE VOICE WEB FOR GOVERNMENT SERVICES DELIVERY

Governments exist to provide services and information to constituents regardless of race, gender, socio-economic status or other factors. However, many governments continue to struggle with Internet access issues and ways to provide e-government services to different groups of citizens. For example, research on the ability of different groups in the U.S. to access the Internet shows clear differences between whites and other groups of citizens, with whites being more likely to have access than other groups (Novak, 1998).

There are also differences in rate of use of the Internet among older citizens as opposed to younger ones. Only about 22% of seniors in the U.S. report being able to access the Internet (Fox, 2004), and research suggests that this group faces unique barriers to traditional Internet access because of the more likely incidence of factors like poor vision and other disabilities. Research from other countries suggests a similar trend outside the U.S. (Paul, 2005). VoiceXML may be a useful technology to help address the challenges that face certain populations in accessing Internet-based government content.

VoiceXML and technologies in the SIF can help make government Internet content and applications available through any traditional or cellular telephone. In some cases, VoiceXML content and applications (when coupled with widely available speech recognition technologies) can even be made available through rotary telephones. This has the effect of dramatically widening the audience for government Web content or services.

In addition, VoiceXML and the SIF can help governments address some of the most fundamental issues impacting the ability of citizens to access online government resources. Speech interfaces can be more natural than visual Web forms, particularly for those with visual disabilities or literacy issues. Additionally, the skills required to operate a telephone are more evenly spread than that of a computer or visual Web browser. In addition, while many governments still struggle to make their Internet content accessible to individuals with physical limitations, VoiceXML can enhance the accessibility of Web content and services to those with visual or other impairments.

VoiceXML can also provide a number of operational benefits for governments that make use of it as opposed to proprietary interactive voice response (IVR) technology. Because VoiceXML is a Web-based markup language, and since the vast majority of public-sector entities have infrastructures to support Web sites (West, 2004), governments can leverage existing knowledge in other markup languages like HTML to develop voice applications more quickly. Also, because the underlying

Table 1. SIF components

Component	Purpose
VoiceXML	XML language for constructing dialogs with callers.
Speech Recognition Grammar Specification (SRGS)	XML language for specifying grammars that define allowed input to a speech application.
Speech Synthesis Markup Language (SSML)	XML language for specifying the rendering of synthesized (computer-generated) speech to a caller.
Call Control Extensible Markup Language (CCXML)	XML language for specifying call control functions, like call conferencing and session management.