Voice Enabled G2C Applications for M-Government Using Open Source Software

Punyabrata Ghatak¹, Neeraja Atri¹, Mohan Singh², Chandan Kumar Goyal², and Saurabh Banga²

¹ Department of Information Technology, Govt. of India, New Delhi – 110003
{pghatak,natri}@mit.gov.in
² Centre for Development of Advanced Computing, Govt. of India, New Delhi - 110016
{smohan,chandang,bsaurabh}@cdac.in

Abstract. M-government is the extension of e-government to mobile platforms. The advancements in mobile communication technology enable a natural transition from the era of e-government to the era of m-government by extending the internet from wired PCs to mobile phones. Since speech is the most natural means of communication, by linking a mobile phone to a VoiceXML gateway we are able to build voice enabled Government-to-Citizen (G2C) applications which are accessible ubiquitously by anyone, anytime. Our implementation of the voice gateway successfully integrates the mobile telephone network with automatic speech recognition, text to speech synthesis for English and Hindi, and web navigation systems based on open standards and using open source software. We describe three voice enabled m-governance G2C applications on the open source Android platform. The platform specific m-governance applications can be downloaded directly on a mobile phone through mobile browsers for their use by citizens.

Keywords: Mobile Computing, Open Source Software, Android, VoiceXML, Automatic Speech Recognition (ASR), Sphinx, Text-to-Speech (TTS), Festival.

1 Introduction

Wireless mobile communication technology has enabled the government to transform from Electronic Government (e-government) to Mobile Government (m-government). Governments can reach a greater number of citizens regardless of the country’s wired infrastructure or the citizens’ economic, educational or social status. This decreases the digital divide among countries and social layers and benefits significantly to citizens and the government. By migrating from traditional paper-based and/or wired internet access based services to the wireless internet, m-government has the potential to provide citizens with the fastest and most convenient way of obtaining government services [1]. The number of mobile phone users in India is far greater than the number of people who use personal computers or the Internet. Wireless mobile communication technology provides citizens with an immediate access to certain government information and services, on anywhere and anytime basis.
To the ordinary citizen, the basic mobile phone is the only easy-to-use medium for information access. The most common m-government G2C applications include information retrieval and update by various users, as well as issuing alerts by governments mainly through SMS. However, most of the mobile phones are not suitable for the transmission of complex and voluminous information and do not have equivalent features and services of wired internet access devices. The user interface of a mobile device (screen size and keyboard) is still far from ideal, limiting the types of services offered. Also, in India, as in other developing countries, with diverse linguistic and cultural groups of citizens, support for different local languages is a crucial issue.

Speech is the most natural means of communication for humans. Also there is no possibility of a virus from a phone call and it is typically much more secure. Voice based services on mobile phones in local languages would allow citizens to get access to government information ubiquitously. However, this requires speech technology to be available in the local languages of the country. Two types of language technology are needed – text to speech (TTS) to deliver information, and automatic speech recognition (ASR) to access it and control its delivery. Of these, TTS is the most essential technology needed because (i) voice services can manage without ASR through the use of touch-screen or DTMF keys, (ii) a single TTS system can cover quite a large region using a neutral dialect.

VoiceXML supports such human-computer dialogs via spoken input and audio output. VoiceXML is an application of the eXtensible Markup Language (XML) defined by World Wide Web Consortium (W3C) that defines dialogs between humans and machines in terms of audio files to be played, text to be spoken, speech to be recognized, and touch-tone input to be collected [2]. A major advantage of VoiceXML is that it provides web content over a simple telephone device or a mobile phone, making it possible to access an application even without a computer and an Internet connection [3]. Comparable to HTML that is interpreted by a Web browser, VoiceXML is interpreted by a voice browser. Audio input is handled by the voice browser’s speech recognizer. Audio output consists both of recordings and speech synthesized by voice browser’s text to speech system. The voice browser runs on a specialized voice gateway server that is connected both to the Internet and to the public switched telephone network (PSTN). The voice gateway connects to the web servers on the Internet using the HTTP protocol. Thus by using VoiceXML applications, we can reach out to more users than is possible by using the Internet.

2 Challenges

Although the ultimate goal of providing access to information using voice is to build a natural language understanding system that understands the query, retrieves information from the Internet and then extracts the relevant answer from the retrieved information, the state of art technology is yet to be developed. However, automatic speech recognition in a domain specific manner with a finite number of words is practically feasible.