Voice Driven Web System

Christine Niyizamwiyitira
Blekinge Institute of Technology
S-371 79 Karlskrona SWEDEN
Christine.Niyizamwiyitira@bth.se

May 17, 2013

1 Introduction

Voice driven web system can play a significant role in providing efficient customer service. It can increase customer satisfaction, lower costs and offer new services that usually were inaccessible through voice channel. The return on investment (ROI) on this system is quite amazing, as it is an exemplary innovation in the area of voice assisted browsing and data retrieval from web through analog and digital telephone, the web source data contains information of interest and has straight relevance to the user.

Voice driven web is a Voice based application that provides automated interaction for callers to retrieve information from WorldWide Web (WWW) through telephone keypad or voice. It controls and responds to callers by utilizing speech technologies. In fact, an impressive growth of the World Wide Web and new special customers' needs information from web requires new ways to access web information. Moreover, people with low computer literacy or who have no Internet accessibility can use their mobile phones to access information through voice driven web. The system proposed in this assignment, is especially for fetching the most updated information about weather forecast for different cities of Blekinge, among them we can mention Karlskrona, Ronneby, and Karlshamn. This information is traditionally presented in graphical user interface (GUI) as web feeds known as Really Simple Syndication (RSS). Our system presents the later in Voice user interface (VUI)[1] through telephone channel.

2 System Description

Figure 1 shows the system that provides weather information; in the traditional way, the user browses using his computer and gets visual information. Whereas in voice driven web, the user calls in, the system replies with a greeting message and options to get the specific region weather information from web RSS feeds. The user has two input options to select the location; either input city number as digits on the phone keypad or saying the city name.

The voice server plays the role of intermediate between the Internet and telephone; it hosts the voice browser which acts as interface between the telephone and the voice application. That voice application is a Voice extensible markup language (VXML) documents that contains the voice grammar, frame based dialogue presentation, and the element that fetches web feeds from RSS. Later on, the voice browser interprets the voice document, verbalize the contents for the

[1]Voice user interface (VUI) term is used interchangeably with Vxml document as well as voice application, technically VUI is referred to a class of human computer interface (HCI).

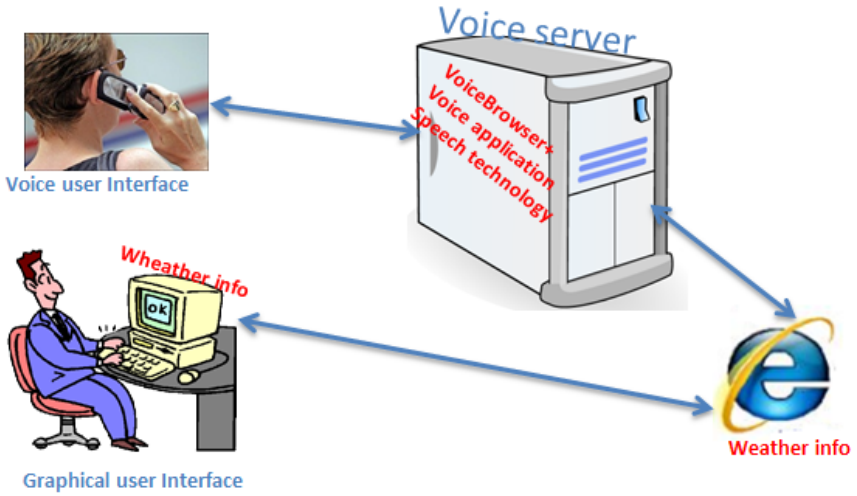

Figure 1: A Voice Driven Web System

caller, and then accept voice input from the user, thereafter, it transmits the request to the voice document which process the web feeds fetching according to the request. The server also contains speech technologies which are automatic speech recognition (ASR), dual multitone frequency (DTMF) that plays a role of recognizing the input when it comes as voice and tone keypad respectively, and text to speech (TTS) that reads the text to output the speech to the user. A set of voice browser and speech technology can be on the same platform otherwise they can be implemented separately.

3 System Requirements

1. The system shall be compatible with multiple telephone line, i.e. support both on analog and digital.
2. The system shall support Text to Speech (TTS) and Automatic Speech recognition (ASR).
3. The system shall be able to playback pre-recordings when the web is not available.
4. The system shall be able to collect tone keypad.
5. The system should be able to distinguish spoken output for voice callers.
6. The system should provide voice input and output information.
7. The system shall provide a voice user interface to manage various interactions.
8. The system shall provide help to the user when required.
9. The system‘s output should be meaningful to the user, and delivered in an explicit way.
10. The system shall personalize the requested content for different users.
11. The system shall speak text to users i.e., read prompts.

12. The system shall access and play media files via a highly reliable, distributed media server such as ASR and TTS.
13. The system shall seamlessly be able to integrate with any third-party components.

3.1 Quality Requirements

1. The system shall allow callers to get access to information without human intervention.
2. The system shall be able to deal with incorrect input (voce and tone).
3. The system shall be easy extendible with new types of information and sources of content, i.e., existing web services.
4. The functionality of the system should be quick to learn and easy to use.
5. The system shall support 64Kbps voice bit rate, 20 msec to 30 msec sample speech time, G711 codec.
6. The system shall be able to service 5000 simultaneous users (average) without affecting response times.

3.2 Assumptions

1. The server is assumed to have computer telephone interface (CTI) otherwise it must be hosted on the VoIP service provider in order to handle telephone calls.
2. Internet connectivity between the server and the websites is assumed to be seamless.
3. The voice application for retrieving information from web must have a telephone number which users call to get information.
4. TTS /ASR technologies work perfectly.
5. Internet connection is assumed to be seamless.
6. End devices are assumed to have at least 30 MB storage dedicated for TTS processing.