Working of Voice Based Browser

Bharat Shrivastava[1], Dr.T.N.Sharma[2]
[1]*Research Scholar, CSE, Vivekananda Global University, Jaipur, bharat.champ.17@gmail.com*
[2]*Professor, Global Institute of Technology, Jaipur, tnsharma@rediffmail.com*

***Abstract*—**Voice Browsing can be defined as a browser which is using speech to navigate an application. These applications might be written using several parts of the Speech Interface Framework. W3C considers voice access to be one piece of more general "multimodal" access, where users can use combinations of means to interact: voice input, speech feedback, electronic ink, touch input, and physical gestures (such as those used in some video games).

Presently voice browser is available for very small set of domain like mobile applications, simple instruction etc. There are huge scopes of work for voice browser like support desks, order tracking , automated telephone ordering services, airline arrival and departure information, cinema and theater booking services etc. This papers presents concept of voice browsers with some of the technologies which can be used for development of voice browser.

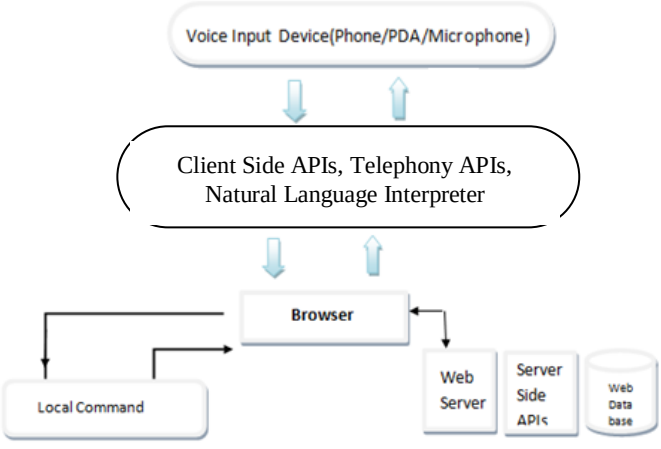
Keywords*—Voice Browser, Voice XML, Voice framework.*

I. INTRODUCTION

Need of browser can be described as a browser which is using speech base interface. Visual browser are very common these days. Visual browser is not always practical, especially for the visually impaired persons. In the voice browsers, the command input and the delivery of web contents are entirely in voice. [1]

II. TECHNOLOGIES

For development of Voice browser framework, lots of technologies were proposed. The whole flow of application can be designed as following figure



In the flow diagram, main component is technologies for voice browser which is combination of various different technologies. Some of these technologies are described below:

A. VOICE INPUT DEVICE

Voice input device are general devices which can accept input in form of speech. For example these might be Phone/PDA/Microphone. These input will later be processed and communicated to the next layer of framework. These devices also can work as output devices for the given input.

B. VOICEXML

A language for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations. Some of its versions are:

- VoiceXML 1.0: designed for creating audio dialogs.
- VoiceXML 2.0: uses form interpretation algorithm (FIA).
- VoiceXML 2.1: 8 additional elements.
- Voice XML 3.0: relationship between semantics and syntax.[7]

C. SPEECH RECOGNITION GRAMMAR

Specification (SRGS) 1.0 : A document language that can be used by developers to specify the words and patterns of words to be listened for by a speech recognizer or other grammar processor.[7]

D. SPEECH SYNTHESIS MARKUP LANGUAGE

A markup language for rendering a combination of prerecorded speech, synthetic speech, and music. Some of its versions are:

- Speech Synthesis Markup Language (SSML) 1.0
- Speech Synthesis Markup Language (SSML) 1.1[7]

E. AUTOMATED SPEECH RECOGNITION (ASR)

The speech recognition component is responsible for managing the associated application grammars and recognition state, and processing the spoken utterances, attempting to recognize the spoken utterances to a set of known valid inputs, which drive the flow and logic of the application.[9]

F. TEXT TO SPEECH (TTS)

The text-to-speech component is responsible for turning textual output into synthesized audio that can be played back to the user as if it was spoken by a human. Text-to-speech is

useful when dynamic content does not lend itself to pre-recording.[9]

VoiceXML Interpreter can be explained as the module used to retrieve VoiceXML pages from Internet for parsing and execution. The lower layer APIs such as speech recognition API, Prompt API, or Telephony API were called to complete the execution.

They also explained JAVAScript Interpreter, Speech Recognition API, Prompt API , and Telephony API. This API set defines the interface between the VoiceXML interpreter and the Telephony interface. When telephony related function such as waiting for call, disconnecting a call, recording a voice file are required, the interpreter calls the Telephony API to complete this job.[3]

Some of the related technologies can be discussed as follows

i. VoiceXML**:** a language for creating audio dialogs that feature synthesized speech, digitized audio, recognition of spoken and DTMF key input, recording of spoken input, telephony, and mixed initiative conversations. Some of its versions are:

- VoiceXML 1.0: designed for creating audio dialogs.
- VoiceXML 2.0: uses form interpretation algorithm (FIA).
- VoiceXML 2.1: 8 additional elements.
- Voice XML 3.0: relationship between semantics and syntax.

ii. Speech Recognition Grammar Specification (SRGS) 1.0: a document language that can be used by developers to specify the words and patterns of words to be listened for by a speech recognizer or other grammar processor.

iii. Speech Synthesis Markup Language (SSML): a markup language for rendering a combination of prerecorded speech, synthetic speech, and music. Some of its versions are:

- Speech Synthesis Markup Language (SSML) 1.0
- Speech Synthesis Markup Language (SSML) 1.1.[7]

III. CONCLUSION

This paper presented general overview of voice browser. The paper presents various technologies which are useful for development of voice browser. Even after a huge need of voice browser in social and business aspects still the voice browsers are not very common in use. Still the work is going on for betterment of these technologies.

IV. REFERENCES

[1] Josiah Poon et al (2001), Browsing The Web from a Speech-Based Interface, Proc. Of Human-Computer Interaction (INTERACT01), p.p. 302-309

[2] IBM (2001), Developing Voice Application An IBM White Paper, p.p. 01-19

[3] Chin-Hsing Hsu et al (2006), On The Construction Of A VoiceXML Voice Browser

[4] Zan Sun et al (2006), Dialog Generation for Voice Browsing

[5] Muhai Hu et al (2010), Voice Browsing Approach to E-Business Access: A Blind's Perspective, Computer and Information Science, Volume4, p.p.4

[6] K.Sireesha et al (2011), Voice Recognition browser for reduced vision and vision loss Learners, International Journal Of Scientific & Engineering Research, Volume 2, Issue 12

[7] Khushbu et al (2012), THE NEW ERA OF BROWSING - VOICE BROWSING, International Journal Of Engineering And Computer Science, Volume1 Issue 2, p.p. 59-62

[8] Nizar Banu P K et al (2013), Voice Based Search Engine And Web Page Reader, International Journal Of Computer Engineering Research (IJCER)

[9] http://www.dialogic.com/webhelp/CSP1010/VXML1.1CI/WebHelp/intro%20-%20Features%20of%20the%20Vocalocity%20Voice%20Browser.htm

[10] Sachin Kapoor et al (2014), Hands Free Browser-'An Interactive Speech Browser for Visually Handicapped', International Journal Of Advanced Computer Research, Volume4 Issue 14

[11] Sungjae Han et al (2014), A Voice-Controlled Web Browser to Navigate Hierarchical Hidden Menus of Web Pages in a Smart-TV Environment

Bharat Shrivastava is MCA from Rajathan Technical University in 2014 and now pursuing his research work in Vivekananda Global University, Jaipur. His area of research is "Application of heuristic algorithm in the development of voice based browser".

(Prof).Dr.T.N.Sharma is having experience of about 20 years in the various sectors of IT. He is author of 18 international papers and 3 international books. Presently he is working as Professor "Computer Science and Engineering" in Global Institute of Technology (An Engineering College).