SPEECH ENABLED BROWSER USING MICROSOFT SPEECH API

ABSTRACT: A speech browser is a web browser that presents an interactive voice user interface to the user, useful to those who have difficulty reading or seeing. A voice browser presents information aurally, using text-to-speech conversion to render textual information as audio. It is an Internet browser designed for users who wish to access the Internet in a non-visual or combined auditory and visual way. This includes blind or partially sighted users, people with dyslexia or learning difficulties, and users who are learning new languages. A broad range of users, not only those with disabilities, will have enhanced choice, convenience and control in accessing the information age.

Keywords: Text/Tag Rendering Algorithm (TRA), Text-To-Speech (TTS), Speech-to-Text (STT).

1. INTRODUCTION

Speech technology has now advanced to the stage where it offers great promise for human-computer interaction in a variety of applications. Applications have to be chosen and engineered very carefully, however, with human factors given full consideration, if real gains are to be achieved. In particular, the early, popular belief that speech was somehow a ‘universal’ medium – better in all respects than all other media – is too simplistic. Attention in the research community is turning toward the optimal deployment of speech I/O in multimodal interfaces. Key findings are that speech-only interfaces have some problems but that users have a strong preference for interacting multimodally. At the same time, speech is notable for its absence in current multimedia systems. Speech offers unique advantages over more conventional media.

The ultimate goal is an all-purpose system with which one can interact conversationally as we do with one another. For many easily imaginable applications (e.g. interacting with a system in the dark, while one’s hands are otherwise occupied, or if one is handicapped or at a remote site), direct speech control over what we otherwise control by keyboard or mouse is essential. But even where keyboard or mouse would be accessible, their limited fit to human capabilities means that it may be more natural and comfortable to interact conversationally, by speech.

2. LITERATURE SURVEY

Microsoft speech API: The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. To date, a number of versions of the API have been released, which have shipped either as part of a Speech SDK, or as part of the Windows OS itself. Applications that use SAPI include Microsoft Office, Microsoft Agent and Microsoft Speech Server.

In general the Speech API is a freely-distributable component which can be shipped with any Windows application that wishes to use speech technology. Many versions (although not all) of the speech recognition and synthesis engines are also freely distributable. [6]

Speech synthesis: Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech.[5]

Voice XML: A voice browser is a web browser that presents an interactive voice user interface to the user. Just as a visual web browser works with HTML pages, a voice browser operates on pages that specify voice dialogues. Typically these pages are written in VoiceXML, the W3C’s standard voice dialog markup language, but other proprietary voice
dialogue languages remain in use. A voice browser presents information aurally, using pre-recorded audio file playback or using text-to-speech software to render textual information as audio.

3. SYSTEM DEFINITION

As we enter the 21st century, the interaction between humans and computer is breaking the old barriers and entering a new realm. In the highly technology driven world of today’s computer and internet have become a part of our lifestyles. The use of internet is arising day by day for our general needs. Web content is becoming more popular by different people in different areas of applications. At the present time, access to information on the World Wide Web is difficult for disabled people. They are often able to access the same information as others, but time spent struggling with cumbersome interfaces or non-optimal formatting of information detracts from the amount of time available for the actual task at hand.

The goal of this project is to further develop a World Wide Web browser application which is accessible to disabled people, including those with movement, visual, or learning disabilities. The browser will run under the Microsoft Windows Operating System. Current Web browsers were not designed to accept spoken commands nor were they designed to facilitate human factors experimentation. Hence, significant technical effort must be expended in system design and implementation for our purposes. Many of the problems which arise are novel and challenging: they relate to assumptions made by implementers regarding the way their software products will be used and which do not anticipate our requirements. The ultimate goal is an all-purpose system with which one can interact conversationally as we do with one another. For many easily imaginable applications (e.g. interacting with a system in the dark, while one’s hands are otherwise occupied, or if one is handicapped or at a remote site), direct speech control over what we otherwise control by keyboard or mouse is essential. But even where keyboard or mouse would be accessible, their limited fit to human capabilities means that it may be more natural and comfortable to interact conversationally, by speech.

This speech browser is a web browser that presents an interactive voice user interface to the user, useful to those who have difficulty reading or seeing.

System Architecture:

![Diagram of System Architecture]

Figure: Conversion of TTS and STT

4. IMPLEMENTATION

Technologies:

1. **Voice xml**: A voice browser is a web browser that presents an interactive voice user interface to the user. Just as a visual web browser works with HTML pages, a voice browser operates on pages that specify voice dialogues. Typically these pages are written in VoiceXML, the W3C’s standard voice dialog markup language, but other proprietary voice dialogue languages remain in use. A voice browser presents information aurally, using pre-recorded audio file playback or using text-to-speech software to render textual information as audio.

2. **Speech Grammar Recognition Specification (SRGS)**: 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.

3. Semantic Interpretation for Speech Recognition (SISR): A document format that represents annotations to grammar rules for extracting the semantic results from recognition


5. Speech Synthesis Markup Language (SSML): A markup language for rendering a combination of prerecorded speech, synthetic speech, and music.

6. Microsoft speech API: The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. To date, a number of versions of the API have been released, which have shipped either as part of a Speech SDK, or as part of the Windows OS itself. Applications that use SAPI include Microsoft Office, Microsoft Agent and Microsoft Speech Server. In general all versions of the API have been designed such that a software developer can write an application to perform speech recognition and synthesis by using a standard set of interfaces, accessible from a variety of programming languages. In addition, it is possible for a 3rd-party company to produce their own Speech Recognition and Text-To-Speech engines or adapt existing engines to work with SAPI. In principle, as long as these engines conform to the defined interfaces they can be used instead of the Microsoft-supplied engines.

7. Speech Synthesis: Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware products. A text-to-speech (TTS) system converts normal language text into speech.

5. CONCLUSION

In this paper, the architecture of a special browser for the visually handicapped people (blind) has been described. It is an Internet browser designed for users who wish to access the Internet in a non-visual or combined auditory and visual way. This includes blind or partially sighted users, people with dyslexia or learning difficulties, and users who are learning new languages. A broad range of users, not only those with disabilities, will have enhanced choice, convenience and control in accessing the information age.

REFERENCES


