AN INVESTIGATION INTO VOICE-CONTROLLED WEB BROWSING FOR THE ELDERLY: WEB INTERFACE

Kirti Nathoo

School of Electrical & Information Engineering, University of the Witwatersrand, Private Bag 3, 2050, Johannesburg, South Africa

Abstract: The internet provides users with easily accessible information online. However, the elderly struggle to participate and contribute to the availability of data. This may be due to a lack of interest in technology or physical disabilities which limits them from being able to use computers. Extensive research indicates that speech recognition software removes these limitations by enabling web browsing through voice commands. An investigation into voice-controlled web browsing for the elderly was performed. A web application was developed and numerical and link name referencing techniques were applied to determine the most efficient technique for web applications. Verbal and visual feedback techniques were incorporated to determine which feedback methods are most valuable to users. The web application was tested on elderly users over the age of 55. An analysis of the results indicates that for simple web pages users prefer numerical referencing However further results indicate that link name referencing performs better in comparison for simple web pages. For complex web pages there is no distinct preference between referencing styles but numerical referencing does offer improved performance. Link highlighting and verbal feedback techniques are adequate feedback methods. The existing techniques implemented can be improved through the use of algorithms and combinations of referencing styles.

Key words: Voice-controlled, web, navigation, elderly.

1. INTRODUCTION

The internet offers users a variety of functions that makes data easily accessible. However the elderly struggle to partake in this global data trend. This is attributed to the lack of interaction with computers or physical disabilities that limit the mobility of their hands. Speech recognition technology increases the usability of the web for elderly users through voice controlled web browsing [1]. A voice controlled web application was developed for elderly users. The application incorporates speech recognition into existing browser functionalities. Defined voice commands are used to control browser navigation. The application is used to determine what styles of graphical annotations and feedback techniques would assist elderly users with web navigation. Examples of existing speech recognition and voice controlled browser applications are provided. Application specifications are defined. A high-level design details the application components. The application is subdivided into iterations. The implementation and test results of each iteration is discussed. Societal and environmental impacts are stated and professional and ethical obligations are discussed. Aspects of team work are provided in appendix A. The application is extensively analysed and recommendations for future work is suggested.

2. BACKGROUND

2.1. Contextualisation

The web has drastically evolved from a purely information storage mechanism to an interactive content distribution phenomenon. Data is available in the form of websites, blogs, social networking applications, etc. The web provides a variety of functions for an array of users. These functions include online shopping and banking, current news, encyclopaedias, video calling, etc [2]. However the range of functionalities provided by the web are not explicitly beneficial for elderly users. This may be because elderly users have not been exposed to computers and are reluctant to use these technologies. Additionally, elderly users may be unable to use computers due to physical disabilities which results in difficulty in using a keyboard or mouse. Conditions such as arthritis and rheumatism may result in these difficulties. Speech recognition provides the optimal solution for this problem. Speech recognition enables users to control a computer through the use of voice commands which eliminates the need for keyboard and mouse inputs [1].

Speech recognition technology is a form of assistive technology aimed at improving the quality of life for people with disabilities. This is achieved by using speech to browse content on the internet [3]. Web browsing is one of the most valuable aspects of computers which requires minimal typing. This is a paramount starting point for elderly users with hand disabilities [1].

2.2. Existing systems

Voice controlled web navigation consists of two main components: voice recognition and visual rendering.

2.2.1. Voice recognition

Voice recognition uses voice commands for browser navigation. Examples of existing voice technologies are described in appendix B. Key components of these technologies are provided.

Voice Browsing is a toolbar add-on for Microsoft Internet Explorer (IE). The toolbar enables users to control browser navigation using voice commands. Specified websites are navigated to by saying the name of the website. Conversa is also an add-on for IE which enables website navigation by saying “Link me to” and the name of the website. Conversa incorporates speech recognition into the browser by replacing conventional toolbars with a new “saycons” toolbar. The toolbar consists of icons that can be spoken such as: “Show Favourites,” “Print this page,” etc. Additionally, Conversa enables up, down scrolling and forwards and backwards navigation. Conversa provides

users with verbal feedback. Opera with Voice allows users to interact with the Opera interface through the use of voice commands. All voice navigation commands begin with the prefix “Opera .”

2.2.2. Visual rendering

Various applications currently exist which modify the application interface to incorporate speech recognition as discussed in appendix B. Fundamental concepts from these applications are highlighted. Mouseless Browsing (MLB) is a Firefox-extension which enables users to browse the web through the use of a keyboard. In MLB small boxes are placed behind every link. Each box contains a unique identifier which is entered via the keyboard. Numeric and character identifiers can be used.

Pentadactyl is an add-on for Firefox designed to improve the efficiency of web browsing through the use of a keyboard. Pentadactly follows links both backwards and forwards subsequently reducing mouse usage. Pendactyl incorporates the “Hint” mode. “Hint” mode refers to Hit-a-Hint (HaH) mode which provides a unique numeric identifier for each link on a page. The numeric identifier can be entered using a keyboard or through speech recognition software. When the numeric identifier is selected the corresponding element is activated.

2.2.3. Voice controlled web browser

A voice controlled web browser combines speech recognition and visual rendering components. The browser permits users to navigate the web by using voice commands to access links and web page elements. Links are accessed by either saying the text of a link or an associated number. Examples provided are specifically intended for a particular target market. No considerations have been made for elderly users with hand disabilities.

Speech recognition in web browsers eliminates this problem [1]. However, optimal techniques for voice controlled web browsing must first be investigated. Optimal techniques include different types of referencing style and voice commands. Referencing techniques enable users to access links by saying the link text or an associated number. Valuable visual and verbal feedback notification methods for elderly users will be investigated.

3. APPLICATION SPECIFICATIONS

3.1. Functionality requirements

The functionality requirements state the required system functionality. The main requirements are listed and a complete set is provided in appendix C.2.1

- A web application will be developed which combines voice commands and browser navigation.
- The application will be designed to identify efficient referencing techniques for web browsing. Different feedback techniques for users will be investigated.

3.2. Constraints and assumptions

The constraints affect the application design. The main design constrains are listed and a comprehensive list is provided in appendix C.2.2.

- Most elderly people are not computer literate.
- Difficulty in finding elderly users to test that are willingly and have sufficient hearing, vision and hand mobility.
- The system will use voice as a primary means of input.

The following assumptions are presumed:

- Most users are computer illiterate.
- Elderly users have adequate vision to use the computer application.
- Testing environments have minimal noise as ambient noise affects the performance of speech recognition [4].

3.3. Success criteria

- Functionality requirements, constraints and assumptions are met.
- Determine which referencing techniques are preferred by users and perform the best.
- Determine beneficial feedback techniques for elderly users

4. SOFTWARE APPLICATION

4.1. Overview

Two referencing techniques for voice-controlled web browsing were examined to determine which techniques performed best and were preferred by users. These techniques are numerical referencing and link name referencing. In numerical referencing web navigation is controlled by saying numbers. In link name referencing, navigation is controlled by saying selected words.

Simple and complex web applications are designed to aid the investigation. Techniques are applied to both websites to ascertain the most efficient navigation techniques for the respective websites. Visual and verbal feedback methods are incorporated and assessed to determine which methods are preferred by users during browser navigation. Defined voice commands are the primary input to the application and are provided through a microphone input. A button must be pressed to activate speech recognition. Voice commands are interpreted by an online Application Programming Interface (API). The result returned from the API is used to execute navigation commands. The web page represents the GUI of the application through which users interact.

4.2. High-level design

The web application is composed of three primary components: visual rendering, voice recognition and navigation.

4.2.1. Visual rendering

Visual rendering refers to graphical web development. The application provides users with an interactive web page GUI. The web application GUI is designed using HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript languages. HTML is used to incorporate the JavaScript speech recognition functionality into the web pages. HTML and CSS are collectively used to style and develop the application GUI.

HTML is used to format web pages. The HTML code instructs the browser about the layout and structures required by a web page. However, HTML is quite limited. Subsequently CSS has been used to replace certain HTML components such as borders, backgrounds, font settings, etc. CSS provides significantly more customisation capabilities than HTML [5]. CSS enables the style and layout of web pages to be contained in a single document. Web page layouts are automatically edited by editing the CSS document which each web page is linked to. CSS has additionally been used to design a style sheet template for web pages within the web application. Web page designs are illustrated using screenshots of the developed web application and are included in appendix D.2.

4.2.2. Voice recognition

The voice recognition component enables users to interact with the application GUI through the use of voice commands. An online open source speech recognition API was used [6]. The API allows speech recognition functionality to be easily added to any web page through the use of JavaScript and flash. No server resources are required as the API is freely hosted. The basic mode JavaScript API was used which provides simple methods and callbacks. For additional details on the voice recognition component see [7].

4.2.3. Navigation

The navigation component refers to website navigation and integrates both the visual rendering and voice recognition components. Voice commands input by the user are interpreted by the speech API and the results returned are linked to corresponding web navigation functions. For example, a user requests to return home and says "home." This command is sent to the API which returns a string result indicating the corresponding navigation function linked to the home command. The navigation function is then executed on the web page.

4.2.4. Architecture

The above components are categorised according to a three-tier architecture model. The model is composed of three layers: presentation, business and data as shown in figure 14 in appendix D.3. The visual rendering component forms part of the presentation layer. The application GUI is the primary means through which users interact with the application. The voice recognition component also forms part of the presentation layer as voice commands are input through the GUI and verbally returned to the user.

The business layer links the presentation and data layers. The business layer allows the presentation layer to access information from the data layer. The navigation component forms part of the business layer as it interprets the results returned from the API and appends the corresponding navigation function. The voice recognition component forms part of the data layer. Voice commands are interpreted by the API and the results are returned and accessed by the business layer. The three layers collectively work together to ensure the web application functions correctly.

4.2.5. Software life cycle

The Rapid Application Development (RAD) method is followed. RAD is an agile method wherein requirements are prioritised into iterations to ensure essential requirements are met first [8]. Three iterations are defined and the main requirements of each iteration are indicated in appendix D.4. Each iteration is a small project with requirements, design, implementation and testing processes. Time boxes for each iteration are defined and shown in appendix G.2. A total time frame of 7 weeks was defined for the complete project.

5. APPLICATION STRUCTURE

5.1. Iteration 1

A simple website composed of questions and answers has been developed and used to investigate the performance of numerical and link name referencing techniques on simple websites. Additional feedback techniques are incorporated and assessed.

The website is designed as a set of 16 questions. Each question is based on an animal fact. To answer a question, the correct animal category must first be selected. Upon selecting the category, four facts are listed and only one fact is the correct answer. The website is divided into five sections: numerical referencing, link name referencing, two visual feedback sections and one verbal feedback section. Each of the referencing sections consist of the same five questions. The feedback sections are implemented to determine which feedback techniques are preferred by users. Each feedback section consists of two questions which are repeated for each subsection. The visual feedback section consists of pop up and link highlighting sections.

Questions in the numerical referencing section are answered by saying the corresponding number of the animal category and fact number. To answer questions in the link name referencing section, the green highlighted text must be said. To initialise speech recognition, the space bar button on the keyboard must first be pressed and released. After answering a question the button must be pressed again to notify the API that the user has completed speaking.

5.2. Iteration 2

A large number of errors were recorded for the first referencing technique in iteration one. To remain unbiased to a particular referencing style, the website in iteration one was restructured and tested again to determine the performance of referencing techniques on simple websites.

The website has been restructured as a set of 14 animal fact related questions. The questions are equally divided into two sections: numerical and link name referencing. The feedback sections have been removed as sufficient data has been collected from testing in iteration one. Each section is preceded by a tutorial to help guide the user. Each tutorial consists of two questions which illustrates how to answer questions in each section. Similar to iteration one, questions in the numerical referencing sections are answered by saying the link number. In the link referencing section, questions are answered by saying the text highlighted in green.

5.3. Iteration 3

A facsimile of a local news website has been designed for iteration three. Numerical and link name referencing techniques have been applied to the website to determine the performance of these techniques on complex and realistic websites. Two versions of the website facsimile have been created. The numerical referencing technique is applied to the first version wherein links are accessed by saying the associated number. Link name referencing is applied to the second version of the website wherein links are accessed by saying the text highlighted in green. To activate speech recognition for either of the websites, the ctrl button on the keyboard must be pressed before speaking. The start and stop listening buttons have been removed to increase the usability of the websites.

Navigation commands such as up, down, home and backwards have been included. Verbal feedback has been incorporated where commands spoken by the user are verbally repeated back to the user. User confirmation has been integrated into the application for complex navigation methods such as selecting a link or going backwards. After an action is selected by the user, yes or no verbal confirmation must be provided. In the event of recognition errors, users are requested to repeat the command. Link highlighting has been provided as visual feedback. Upon selecting a link, the colour of the selected link changes to red to notify the user that the element is selected. Visual feedback and user confirmation ensure that correct elements are selected which improves the performance and accuracy of the voice navigation.

6. TEST RESULTS

Each iteration has been extensively tested on elderly users over the age of 55. A series of use cases have been designed and used to test users. User and application errors have been recorded. User errors are mistakes made by the user during testing. Application errors are caused by the speech API and include commands that are not interpreted correctly, not recognised or not accepted by the API. Individual surveys have been designed for each iteration to obtain additional information about the user, views of the application and recommendations to guide future implementations. The surveys have been included in appendix E. Comprehensive analysis of the test results is provided in appendix F.

6.1. Iteration 1

Seven elderly users have been tested. In total 48 application errors were recorded for the referencing sections. 57.14% of users prefer using numerical referencing for a simple website. 28.57% prefer link name referencing and the remaining 14.29% of users are fond of both techniques as shown in figure 1 in appendix F.2.1. However, the perceived performance of the referencing techniques was contradictory to the above data as 57.14% of users that felt that link name referencing performed better in comparison to the remaining 28.57% of users whom felt that numerical referencing performed better as observed in figure 2 in appendix F.2.1. 71.42% of users felt that link highlighting was the most valuable form of visual feedback as illustrated in figure 3 in appendix F.2.1.

6.2. Iteration 2

Iteration two was tested on five elderly users and the number of application errors minutely reduced from 48 to 44. This indicated that the appended tutorial sections were not significantly helpful to users and the first referencing technique in iteration one was adequately tested. 60% of users tested prefer using numerical referencing in iteration two. 20% prefer link name referencing and the remaining 20% are fond of both techniques as shown in figure 4 in appendix F.2.2.. However, 60% of users that felt that link name referencing performed better in comparison to the 40% of users whom felt that both referencing techniques performed well. None of the users felt that numerical referencing performed well as indicated in figure 5 in appendix F.2.2. 20% of users questioned prefer numerical referencing to be applied to a website. Whereas 40% of users would prefer link name referencing and the remaining 40% were unsure as they have never used a news website. This is due to the fact that many elderly users are computer illiterate. From figure 6 in appendix F.2.2 it was evident that users prefer saying a particular word for link name referencing in comparison to part of or the complete link sentence.

6.3. Iteration 3

A total of eight people were tested on iteration three. 50% of users prefer using numerical referencing and the other 50% prefer link name referencing for complex websites as shown in figure 7 in appendix F.2.3. 62.5% of users were happy with the current numerical referencing style. The same percentage of users would like different sections of web pages to be annotated using different colours. When questioned about user confirmation, 62.5% of users felt that confirmation was not necessary. The remaining 37.5% felt that confirmation was necessary to ensure the application performed as expected. 50% of users felt that is was unreasonable to expect elderly users to press a button to activate speech recognition. However the remaining 50% thought it was an acceptable requirement

which indicates that a single button press may be acceptable.

6.4. Results analysis

Users were questioned to determine which of the referencing techniques they preferred using. For simple websites most users preferred using numerical referencing. For complex websites no distinct preference between referencing styles was noted as seen in figure 9 in appendix F.2.4. This maybe because numbers are easy to pronounce and sequential in comparison to random words that are selected for link name referencing. Even though users preferred using numerical referencing for simple websites, figure 8 in appendix F.2.4 illustrates that link name referencing actually performs better based on the number of application errors recorded. This may be due to the similarities between numbers that are spoken and possibly the inaccuracy of the speech API. Additional results have been concluded and are provided in appendix F.2.5.

7. DEVELOPMENT ENVIRONMENT

7.1. Operating system

The application has been designed for execution on Windows systems. Windows is a common operating system which most users are familiar with as most Personal Computer (PC) and laptop manufacturers install Windows onto machines prior to retail.

7.2. Programming languages

JavaScript, HTML and CSS are used to develop the web application. JavaScript is used to incorporate speech recognition into the application GUI. HTML and CSS are collectively used to develop the application GUI as discussed in section 4.2.1. Notepad++ was used as an Integrated Development Environment (IDE) for both JavaScript and HTML programming.

7.3. Version Control

Git is an open source distributed version control system. A Git repository has been set up which allows both developers to share and access source code and other project files. The repository provides extensive revision tracking capabilities and is not dependent on a network connection [9]. Git provides an efficient GUI which simplifies the execution of commands and instructions.

7.4. Design and implementation challenges

Initially the Eclipse IDE was used to develop the application. However countless problems with the IDE were encountered and subsequently Notepad++ was used to replace Eclipse. The application GUI was originally designed for larger screen sizes. When the GUI was tested on smaller screens, some of the web page components overlapped. As a result the automatic mode in HTML was used to ensure components automatically adjusted to corresponding screen sizes.

Initially the web application was deployed from the file directory of each developer. However upon integration, the file directories conflicted and the application was unable to deploy. Subsequently Windows IIS was then used to locally host the application. On open day – 24 October 2011 – the online speech recognition API was non-functional and users were unable to test the application. Fortunately a video of the functional application was recorded prior to the day.

8. TIME MANAGEMENT

Time administration and key dates concerning project implementation are provided in detail in appendix G. The application has been decomposed into iterations and each iteration has been assigned a time frame. The iterations and time frames are shown in table 2 in appendix G.2. In total, 220 hours was spent on project implementation. Most implementation was performed on weekdays and minimal work was produced over weekends. Daily tasks and time spent per task has been illustrated in figure 3 in appendix G.3.3. Minutes of meetings with the project supervisor have been included in appendix H. The application was divided into components and fairly distributed between developers. The work division between developers has been indicated in appendix A.2.

9. SOCIETAL AND ENVIRONMENTAL IMPACTS

Voice-controlled web browsing improves the usability of computers for the elderly [1]. This increased usability may lead to a substantial increase in the number of elderly computer literate users which may place constraints on current network infrastructure. In addition, data costs in South Africa (SA) are high and certain users may prefer an inexpensive non-internet based application. With a growing group of computer users a substantial amount of electricity will be consumed. This contributes to the current energy crisis in SA.

10. PROFESSIONAL AND ETHICAL IMPLICATIONS

During testing developers ensured that testing methods were not strenuous for elderly users. Test cases were kept short and users where guided throughout the procedure. If at any point users were tired, testing was stopped. Elderly users were not discriminated against due to their age or disabilities. All users were respectfully treated.

The web application would fail if no network connectivity is available or if the online speech API is non-functional. In these cases – although speech recognition is not available – users that do not have hand disabilities can still browse the application using a mouse or keyboard. The software tools used were free and open source. The content used for the website facsimile in iteration three was extracted from the News 24 website and was strictly used for testing purposes and not for unauthorised and malicious content distribution. [10]. The website was not intended for commercial use and strictly developed for academic purposes. The Association of Computer Machinery (ACM) Code of Ethics and Professional Conduct has been adhered to during the investigation process [11]. The Institute of Electrical and Electronics

Engineers (IEEE) Code of Ethics has been abided by during project implementation [12].

11. CRITICAL ANALYSIS

Three iterations of the application have been successfully implemented and tested. The results obtained in the first iteration were questioned due to a large number of errors recorded for the first referencing style. Subsequently the website was restructured and retested. No discrepancies in iteration one's results were observed. Even though a small sample group of users were tested and most of the users were computer illiterate, sufficient data was collected. The data indicates that for simple web pages link name referencing performs better and for complex web pages the numerical referencing technique performs best. This is contradictory to user preference noted for the specific web page styles. These results indicate that for complex web pages, numeric identifiers are easy to use and result in less application errors. This can be attributed to the limited vocabulary of numeric identifiers. The performance of numerical referencing for complex websites outweighs the errors resulting from mispronunciations and recognition errors for link name referencing. This implies that the usability of web browsing for the elderly can significantly be improved through speech recognition, efficient referencing styles and adequate feedback methods. The project investigation was completed within seven weeks with the guidance from the project supervisor and efficient software tools. Testing was the longest aspect of the project but was one of the fundamental components. The results indicate various aspects of improvement of the usability of applications for the elderly.

12. IMPROVEMENTS

Improvements to the project include the enhancement of currently implemented aspects of referencing styles and layout components of the application GUI. Aspects include the development an algorithm to automatically select text to be spoken for link name referencing. Layout components include experimentation with different annotation methods for the GUI to improve the visual appearance of the application for elderly users. These concepts and additional recommendations for future work are described in appendix I.

13. CONCLUSION

The application was decomposed into iterations to ensure all application specifications were met. The specifications were used to formulate requirements for each iteration. Upon examining the results it was concluded that for simple web pages users prefer numerical referencing. However, this technique recorded the highest amount of errors in comparison to link name referencing which performed significantly better. For complex web pages there is no apparent preference between referencing styles. However, numerical referencing performs substantially better. In conclusion, for optimal performance and accuracy link name referencing must be applied to simple web pages and numerical referencing to complex web pages. Numeric identifiers are short and easy to pronounce which significantly reduces the possibility of errors and mispronunciations. Link highlighting and verbal feedback were considered adequate feedback methods for users. Characteristics of the referencing techniques can further be improved through the use of efficient algorithms and variations in web page annotations. In addition these techniques can be applied to different types of websites and systems to ascertain which techniques or combinations improve the usability of these systems for the elderly.

REFERENCES

[1] Anderson S, Liberman N, Bernstein E, Foster S, Cate E, Levin B. *Recognition of elderly speech and voice-driven document retrieval.* Dragon Systems, Inc, 1999 IEEE, pp 1.

[2] Getting B., Practical ecommerce. Basic Definitions: Web 1.0, Web 2.0, Web 3.0. http://www.practicalecommerce.com/articles/464-Basic-Definitions-Web-1-0-Web-2-0-Web-3-0, Last accessed 27 October 2011.

[3] Conn N, McTear M. *Speech Technology: A Solution for People with Disabilities*. Faculty of Informatics, University of Ulster at Jordanstown, United Kingdom, 2000 IEEE, pp 1.

[4] Baker M J, Pinto F D. *Optimal and suboptimal training strategies for automatic speech recognition in noise, and the effects of adaptation on performance*. Dragon Systems, Inc, 1986 IEEE, pp1.

[5] Wilton-Jones M., HowToCreate. What do HTML, CSS and JavaScript do? http://www.howtocreate.co.uk/, Last accessed 27 October 2011.

[6] Speechapi.com. Online Speech Recognition API. http://www.speechapi.com/, Last accessed 27 October 2011.

[7] Noble C., *An investigation into voice controlled web browsing for the elderly: Voice controlled interface*, ELEN4012, Laboratory Final Report, School of Electrical and Information Engineering, University of the Witwatersrand, South Africa, 2011.

[8] Vliet H V. *Software Engineering Principles and Practice*. John Wiley & Sons, Ltd, England, third edition, 2008, pp 62-64.

[9] Git – Fast Version Control System. git. http://git-scm.com/, Last accessed 27 October 2011.

[10] 24.com. News 24. http://www.news24.com/, Last accessed 27 October 2011.

[11] Association for Computer Machinery, ACM Code of Ethics and Professional Conduct.

[12] Institute of Electrical and Electronics Engineers, *IEEE Code of Ethics.*