Developing a multimodal web application

A Caffrey[1] and R McCrindle[2]

[1]Department of Cybernetics, University of Reading,
Whiteknights, Reading, RG6 6AY, UK

[2]Department of Computer Science, University of Reading,
Whiteknights, Reading, RG6 6AY, UK

aidancaffrey@hotmail.com, *r.j.mccrindle@reading.ac.uk*

[1]*www.geocities.com/aidancaffrey,* [2]*www.cs.rdg.ac.uk*

ABSTRACT

This paper describes the creation of a multi-modal website that incorporates both haptics and speech recognition. The purpose of the work is to provide a new and improved method of internet navigation for visually impaired users. The rationale for implementing haptic devices and speech recognition software within websites is described, together with the benefits that can accrue from using them in combination. A test site has been developed which demonstrates, to visually impaired users, several different types of web application that could make use of these technologies. It has also been demonstrated that websites incorporating haptics and speech recognition can still adhere to standard usability guidelines such as Bobby. Several tests were devised and undertaken to gauge the effectiveness of the completed web site. The data obtained has been analysed and provides strong evidence that haptics and speech recognition can improve internet navigation for visually impaired users.

1. INTRODUCTION

Many inventions have had major effects on people's lives – changing the way they perceive the world around them. Perhaps the most recent example is the internet; since the development of HTML by Tim Berners-Lee in 1990 the usage of the internet has grown exponentially, with corresponding effects on peoples lives. A large percentage of internet traffic contains digital images and with the increase of internet speed, linked to the uptake of broadband connections the growth in transmission of digital images will continue. With more visually impaired users needing to access web products with high levels of graphic imagery, there is a clear need to address the difficulty they may encounter in order to ensure that they are not disadvantaged in a work or social context.

To date, the main methods of web navigation for visually impaired users have been screen readers or a Braille display. However, both these technologies currently fail to address many of the problems faced by visually impaired internet users. For example, web navigation by these technologies is frequently slow and often some of the information on a web page remains unavailable to them. There are a large number of visually impaired people worldwide - over two million in the UK alone. Thus, there is a great demand for improved technology to assist visually impaired internet users with web navigation and web based tasks.

There are several areas of research underway into how to improve internet usage for visually impaired people. One of the most prominent is speech recognition; originally this technology was excellent when computers only had a textual display and a small set of commands to deal with. However, as computer interfaces and applications increased in complexity, the speech models failed to keep up with this rapid pace of development and consequently speech recognition systems do not provide the same degree of internet usage satisfaction to visually impaired users as is available to sighted users browsing via more traditional means.

Another very new way to represent information is through haptics. Information can be conveyed to a user through the sense of touch. This technology in general, appears to have substantial potential (as shown by Wall and Brewster, 2003; Yu et al, 2003b), however integrating haptics into web pages is a novel concept and requires further research into its potential. Lahav and Mioduser (2002) showed that combining haptics

 ISBN 07 049 11 44 2

and speech recognition could give a better overall transmission of information than either on its own. However caution must be applied to this idea - Gladstone et al (2002) showed that excessive or conflicting information (via haptics and speech recognition) can reduce the information transfer process.

A generic problem associated with being visually impaired is the inability of the user to fully mentally map spaces (Lahav and Mioduser, 2002). From this it follows that visually impaired users may not be able to map movements of a mouse to curser movements on a screen – causing problems with targeting (e.g. hyperlinks). Keates et al (2000) invented *gravity wells* – which are a novel method of helping motion-impaired users with targeting tasks.

The main devices for input to a computer are still keyboard and mouse. There are other (haptic) devices that can be used such as tactile mice, trackballs, Gamepads, joysticks and wheels. All these devices have slightly different properties and are good at carrying out different tasks; to try to improve web navigation it makes sense to use a mouse. Currently there is an excellent device available, the Logitech Wingman Force Feedback mouse (shown in Figure.1), that offers full force feedback (which unlike some other devices means it can actually move the device rather than just impede its movement in a direction) and only costs around £50 – so it is affordable by the majority of users.

Figure 1. *Logitech Wingman Force Feedback mouse.*

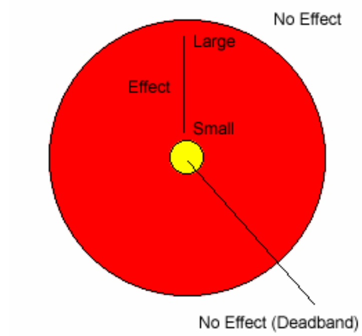

Figure 2. *Where haptic effects are felt in and around an image*.

The novel aspects of the work described in this paper is the implementation of the gravity well idea such that it become much easier for visually impaired people to use a mouse. The particular application chosen to demonstrate this is web navigation – thereby solving two problems simultaneously. Additionally, we use speech recognition and haptics in combination, in order to compensate or circumvent any weaknesses inherent in each technology. The web site created shows several applications of how speech recognition and haptics may be of use to visually impaired users.

2. IMPLEMENTATION OF HAPTIC EFFECTS

2.1 Gravity Wells

The idea of gravity wells originates from the effect found around black holes, whereby all matter around a black hole becomes drawn to its singularity (centre). In computer terms: the curser is the matter, the black hole is an object on the screen and the singularity is the centre of the object. The curser is drawn to the centre only when it has entered the object.

In their original implementation, gravity wells assisted motion-impaired users with their targeting of tasks by applying a haptic effect to move a device to the centre of the target when the curser reached a certain proximity to that target. One of our main tasks is to adapt this approach so that it can be useful for visually impaired users.

2.2 Creating the Gravity Well Effect

The chosen shape for a gravity well is circular, so that the effects experienced are consistent regardless of where the curser enters the image of a gravity well. This representation however is inconsistent with the frequently encountered rectangular representation of images. This creates a mapping problem. To feel effects for only a circle portion of that image, an ellipse mask may be used. This masks the part of the image in which effects should be experienced from the rest of the image - when the mask is entered or exited, effects can be started or stopped respectively as per general gravity well implementation.

Either a spring or a slope may be used to simulate the gravity well effect. Firstly, considering the spring, when the curser is at the edge of the gravity well, the spring is at full extension, so the pull to the centre (and the springs resting position) is the greatest. In and around the centre of the gravity well is a dead band where no effects are felt - this is to allow the curser (and device) to settle in the middle and avoid problems of oscillation and over-usage of motors. The strength of the spring is proportional to the distance from the centre, up to the edge of the gravity well where the strength is the maximum possible. Figure.2 illustrates these concepts.

Secondly, a slope can be used in the place of a spring. The effects are similar, however, instead of the spring extending, the slopes steepness increases as the edge of the gravity well is approached. Again there is a small dead band around the centre. The slope can be customised more precisely than the spring – so has the potential to provide the better effect; unfortunately it is not yet supported by the Immersion Web Plug-in and hence the spring was selected for use in this project.

2.3 Coding the Effects

There are several programs provided with the Wingman mouse such as Immersion Studio 4.1.1 which can be used to choose effects, and their parameters and store them in Immersion Force Resource (IFR) files. However, due to differences in Macromedia Dreamweaver MX version to previous versions, using these files is not yet feasible. This means that the haptic effects need to be coded in HTML sending commands to the Immersion Web Plug-in directly. All the effects need a function to create and initialise them, a function that calculates the position of the centre of the gravity well and starts the effect, and a stopping function.

2.4 Haptic Problems and Solutions

The main problems faced were as follows: there is a lack of primitives that can be represented well as effects, the haptic interfaces cannot simulate the effects accurately enough and sometimes haptic effects can confuse and unnecessarily complicate situations – making the transfer of information worse.

The first problem is dealt with by selecting primitives (e.g. spring) that are well represented as effects. The Wingman mouse is an excellent haptic device and can simulate the required effects well thus solving the second problem. The third problem is combated by carefully selecting when to use effects – keeping them to a minimum and not conflicting with other outputs.

3. IMPLEMENTATION OF SPEECH RECOGNITION

3.1 System Setup

For useful speech recognition the following is needed: text to be read out from a web page, constrained commands to be received and processed from a user, ability to navigate between web pages, and text-to-speech to inform the user what will happen when clicking gravity wells.

The speech engine used is the Microsoft English (U.S) v6.0 Recogniser - this allows speech recognition and text-to-speech. Also needed to allow speech recognition to work in web pages is the Microsoft Internet Explorer (I.E.) Speech Add-In. Finally a standard microphone and set of speakers are required.

3.2 Voice Web Studio

Only since 2002, when Speech Application Language Tags (SALT) was released, has speech recognition in web pages been feasible. SALT extends other languages to provide speech capabilities, in this case the language extended was HTML. An Extension to Dreamweaver MX called Voice Web Studio (VWS) was used to include SALT into HTML. VWS offers four main types of speech elements: speech prompts, speech listens, dialogs and play dialogs.

‘Speech prompts’ are text-to-speech where either text on the screen is read out or pre-defined hidden text is read out. ‘Speech listens’ are elements that listen for user speech input, then compare the input to a list of pre-defined commands and process the result. ‘Dialogs’ are speech prompt and speech listens running in conjunction, to form a simple dialog between the user and the computer. For example a speech prompt may ask the user a question, to which they answer and then receive another speech prompt in response. Finally ‘Play dialogs’ are special speech elements that can be bound to objects in a web page, such as images. They can be used to play a speech prompt to the user in response to events including OnMouseOver.

To make speech recognition work in a web page that page must first be ‘speech activated.’ This means selecting which prompt/listen/dialog will play when the page loads and choosing the browser type. It is standard in this work for a dialog to start whenever a page is loaded to identify the page and allow the user to instantly navigate around the page/site.

The speech navigation model used in the web site is able to access, at any point, a set of standard commands and at some points be able to access some specialist commands. The standard commands will allow the user to move to any page within the site, access help or turn off speech. The specialist commands will only work on certain pages and wherever possible their use will not block subsequent use of the standard commands. This is achieved by predominantly using dialogs so speech listens are active most of the time, allowing fast and easy speech navigation for new and experienced users.

3.3 Speech recognition Problems

Using speech recognition demands a substantial CPU resource. This can cause older machines to lag, especially when the CPU is performing other tasks, listen elements can appear to not work under conditions of lag. This problem can be solved, most of the time by not running unnecessary applications and using a highly specified PC. This problem is likely to persist however, as newer and more complex speech models are continually being developed to utilise the growing speed capacity of newer PCs.

Another major problem is mis-recognitions. This is when valid commands are not recognised. This can be caused by varying accents, user age and sex. There is no standard user voice; the software used is designed for a US voice – the people tested are all from the UK. This can be reasonably dealt with by training the speech recognition software to a users' voice; VWS had event handlers for events such as mis-recognitions that can limit the impact of the problem.

4. WEB PAGE CREATION

4.1 Page Layout

The web pages have been constructed with consistency in mind. The difficult mental mapping task need only be carried out once (so when a user has learnt to use one page, they should be able to use the entire site). As per standard there is a navigation frame on the left of the page (which will contain the gravity wells); the information frame is on the right. Users will always enter into a dialog when a page loads and will start a play dialog whenever they enter a gravity well (giving advice on what clicking the gravity well will do).

The gravity wells are laid out in a vertical stack of three (in the navigation bar) with spacing between them to allow easy distinction between them. There are three because then the most common user requests in a web site are covered: loading the site index, loading the next page and re-loading the current page. Also this design is extendable to enable more layers to be added to the site – instead of loading the site index, a sub-index could be loaded and so forth.

Finally there is an animated gif composed of a series of slightly differing fractals as the background of the navigation bar. The fractals were made in Ultrafractal 3.0.3 and bound together in Macromedia Fireworks MX. The last frame is similar to the first to allow for a smooth changeover. The main reason for this is so that the navigation bar is very striking from the rest of the page – thus making the gravity wells easier to locate by partially sighted users. A additional reason is to make the page(s) aesthetically pleasing to sighted for users, for reasons of recommendations etc.

4.2 Gravity Well Design

The gravity wells are designed so that they can be easily located – this is done by keeping them in a consistent place on the page and increasing their size sufficiently. Also they have been designed such that the users need not necessarily click in the centre of the gravity well, any point inside it will suffice, thereby reducing the time needed for navigation.

Gravity wells also have a visual element to them; when the curser is not in one they have a very bold design (helps partially sighted users with targeting their location). When the curser enters the gravity well a short animated gif plays; it visually mimics the feeling of being drawn to the middle of the gravity well by the haptic interface. This adds to the immersion for sighted and partially sighted users. The last frame is very different to the first, this sudden 'jump' between frames keeps attention on them.

4.3 Pages Created

Several different types of web pages were created during this project in order to demonstrate the different potential areas of use of this technology to visually impaired people. Initially when the site loads, a page that is almost text only (so it will work fully with previous technologies) details the software/hardware needed to fully operate the site. Also the choice of a normal version or a plain background version of the site is offered. Also for greater accessibility purposes a small collection of text-only pages containing explanations of the features in the pages was created.

The first main page is the site index, this page introduces the site (and the project behind it), offering a more detailed background if required. The index page is shown in Figure. 4. From this point onwards speech and gravity well navigation is enabled. The second page is the training page, this offers the user an opportunity to understand how to use speech recognition and find and use gravity wells. This is a very important page because ideally it removes the need for prior training and users can teach themselves how to use the site.

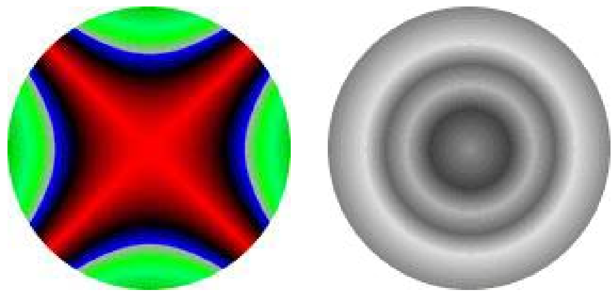

Figure 3. *A Gravity well: when (left) the curser is not in it and (right) when the cursor enters the well.*

Figure 4. *Index page of gravity web.*

The third page is news page, on this page users may choose articles to be read out and skip between paragraphs on the articles. This page is designed to replicate and improve on the function of a screen reader – showing off the new speech software now available for use on the internet.

The fourth page has links to two animations that contain haptic effects within them. This first of these is a flash animation of a picture moving around the screen, the Wingman mouse follows the movement of the picture. The second is of a flash button, haptic effects are felt when clicking the button. Both the animations were made in Macromedia Flash MX and are designed to show the potential for haptic effects in animations.

The fifth page is a comments page, on this page text-fields can be filled via voice commands and then emailed to the webmaster. This page is designed to mimic the process of filling in forms online, such as online shopping or emailing and show it can be done by speech recognition. The sixth page is a publications page where relevant works can be found. The purpose of this page is to demonstrate the potential for loading other websites within the sites frame structure and also to demonstrate how it would be possible to linked back from these external sites to increase awareness and publicity of the site.

4.4 Achieving Accessibility

For a project such as this achieving some form of accessibility was very relevant. Two different free online testing services were used: Bobby and UsableNet – Bobby was the service predominantly used. A URL can be passed to Bobby and a list of accessibility errors are generated in response. Many of the errors were easily solvable such as stating the language of the page, providing alternative content for images and ensuring that fonts are set to relative rather than exact sizes. Solving these problems allows previous technologies such has screen readers to use the pages better.

There are different levels of accessibility award available from Bobby; the standard is Section 508 approval and the highest is Bobby AAA approval. The vast proportion of the site we developed achieved Bobby AAA approval and the entire site achieved at least Section 508 approval.

As haptics and speech recognition on the internet are both very new it is not surprising that errors were generated. For example, one of the main problems found is the use of the 'object' tag (which is essential) because this tag is not supported by older browsers or PDAs. Of course older browsers do not support haptics or speech recognition anyway, but nevertheless descriptive tags of what would happen if the tag was supported are needed.

Another accessibility issue considered was MarkUp Validation. The problem with HTML browsers is that where code is not clear, the browser will try to guess what was meant. Not only may the guess be wrong, it may also be different from browser to browser. The validator used is the WC3 MarkUp Validation Service; as with Bobby a URL can be given and a list of errors is generated. The same problem occurred – as the software used the newest tags available old versions of HTML were shown not to be able to support them. However the process generated a number of errors that were solvable, thus further increasing the sites accessibility.

5. TESTING

5.1 Gravity Well Testing

The original implementation of gravity wells (Keates et al, 2000) was shown to aid targeting times. Our implementation has changed so much from this original application that a short test to determine whether the new implementation is also effective at aiding targeting was devised. The test involved asking a set of users to move the curser to a corner of the screen and then move the curser into one of the gravity wells. This was then carried out for each of the remaining gravity wells and then all repeated from each of the corners. The test is carried out twice by each user, once with haptic effects and once without - so the times can be compared. All the usual steps to ensure fair testing such as half the users starting with haptics and the other half without were taken.

Over the twelve targeting tasks carried out by each user, every user experienced a saving in time with haptic effects, with the average time saved per target just over 0.2s and in total 2.5s. The slowest user took 17s to complete the task (without haptics), making a saving of 15%, up to the fastest user of 10.5s, making a saving of 24%. Unsurprisingly the time taken to reach a target was related to the distance to it, barring a few anomalies. Figure 5 shows the cumulative targeting times for all the users.

5.2 Web Site Testing

In this test speech recognition and haptics are working together, this test is important because it replicates the conditions users exploring the site on the internet would experience. The purpose of these tests were to gauge the effectiveness of incorporating speech recognition in the site (commands were tested three times each), the ability of users to find gravity wells with increasing experience and a series of qualitative questions to obtain information about the effectiveness of the site in general and possible improvements to all aspects of it. Yu and Brewster (2002) suggest that there is no significant difference in test performances between blindfolded users' test results and blind users' test results. Working on this assumption, our tests were carried out with a group of students from The University of Reading, for some parts of the tests they were blindfolded to simulate blindness.

Looking first at the speech recognition results, 86% of commands were correctly recognised. No significant difference between the genders was found. Apart from a few instances all users were able to get any command recognised at least one out of three times. The highest percentage of commands correctly received from a user was 93% moving reasonably linearly to the worst of 79%.

When users were requested to find the gravity wells blindfolded, all users experienced a reduction in the overall time needed to complete the test from their first attempt to their third. The average time to complete a single targeting task fell from 3s at the first attempt to 2.4s at the third (20% reduction). Most users improved at each attempt. The total time all users needed dropped 11% from the first to the second and a further 7% from the second to the third (shown in Figure 6). Also notable was that users experienced a drop in the number of times they entered an incorrect gravity well.

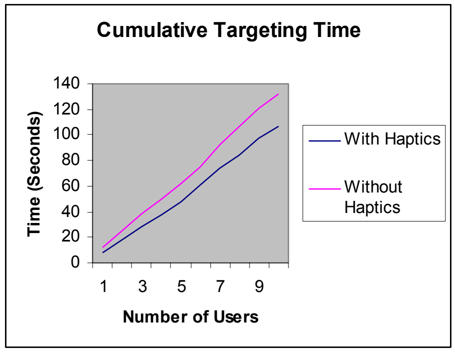

Figure 5. *Cumulative targeting times for users with haptics (bottom line) and without (top line).*

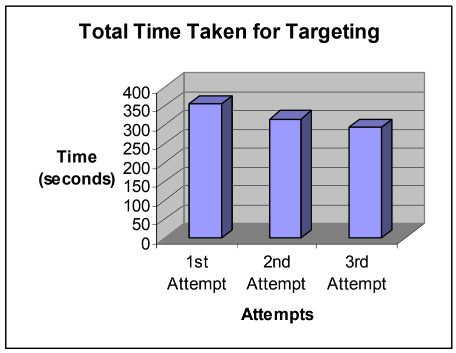

Figure 6. *Total times taken for targeting by all the users at each attempt.*

Finally some of the qualitative information obtained revealed: a general dislike of the computer generated voice, some words were not being pronounced correctly by the speech engine, haptic effects were very popular – especially in animations, the gravity wells were also popular and have a reassuring effect on users.

5.3 Analysis and Explanation

The results obtained provide encouraging evidence that not only the method in general but also the exact implementation of gravity well synchronised with speech recognition is a viable and promising method of web navigation. Further, the evidence demonstrates that this is the case for all users – as to varying extents all users need to carry out targeting and navigating tasks.

The speech recognition test showed that all users received high recognition percentages. Caution must be exercised however, as a high percentage is often still not acceptable enough. However, the perceived standard of speech recognition technology by users in general, is often not that high and this view may have contributed to the percentages we achieved being seen as favourable by the users taking part. It should be noted that extensive testing was carried out prior to the user test and any commands that proved hard to recognise were replaced if it was possible to do so. When a command was not recognised it was usually on the first attempt made by the user, on failing they would then speak more clearly and more often than not the command was then recognised. On a few occasions users were simply not able to pronounce a command in such a matter that it would be recognised, so failed on all three attempts made.

There were a few anomalies within the gravity well testing (where haptics appeared to increase targeting time) that can be explained by observation made during the experiment. User targeting was sometimes a little off, leading to them being caught momentarily in the gravity of an adjacent gravity well. Sometimes users moved the Wingman mouse very quickly and moved through the gravity well and out the other side and then had to move back into it.

As with most tasks the users were able to learn how to find the gravity wells and perform it faster. Mental mapping and muscle memory may well be the cause of this. Backing up that claim is the fact that users curser trajectory was shaky and non-linear at the first attempt, becoming a more direct and linear movement at the second and third attempt. Once the trajectory was learned the users increased the speed at which they moved the curser.

6. CONCLUSIONS

This project investigated the feasibility of a new implementation of gravity wells, with new speech recognition software, for improved web navigation for the visually impaired. The results show that this approach has a high feasibility to provide significant improvements on how visually impaired people navigation the web and hence improve their quality of life.

The new speech software performed well, with good event handlers to deal with problems on the occasions when it did fail. The haptic effects developed have shown that haptics can improve the internet for all users, particularly in animations. Several effects were developed and the one selected was received very well.

The Wingman mouse is an excellent device that performed well for this project, simulating the desired effects well. Some users complained of slight discomfort when using it, despite this it should be considered for any haptic project where the complexity of effects needed is not too great.

Performing accessibility and MarkUp checks was very important not only for this project, but for all web sites in general. The problems encountered were due to the checks requiring that legacy software is able to run the HTML code. When using the newest software available this is not always possible to fully achieve; the process of trying to achieve accessibility is very worthwhile as accessibility will increase even if the highest standards are not quite met.

Within this area of research there is a lack of projects being carried out. There are a vast number of disabled (in this case visually impaired) users – so there is a large demand for more extensive research into ways of providing equal access to the internet for these users. Replacing the information lost by not having a visual channel is a difficult task that haptics and speech recognition cannot yet achieve fully. However there is great promise in the future of these technologies and further research is required.

Having successfully proved our initial concepts, it is our intention to carry out further testing and development with the involvement of target groups of visually impaired users.

Acknowledgements: We would like to thank the students at The University of Reading who took part in the testing for free.

7. REFERENCES

K Gladstone, H Graupp and C Avizzano (2002), Assessing the utility of dual finger haptic interaction with 3D virtual environments for blind people, Proc. 4th Intl. Conf. Disability, Virtual Reality and Assoc. Tech, Hungary.

S Keates, P Langdon, J Clarkson and P Robinson (2000), Investigating the use of force feedback for motion-impaired users, CNR-IROE, Florence.

O Lahav and D Mioduser (2002), Multisensory virtual environment for supporting blind persons' acquisition of spatial cognitive mapping, orientation and mobility skills, Proc. 4th Intl. Conf. Disability, Virtual Reality and Assoc. Tech, Hungary.

S Wall and S Brewster (2003), Assessing haptic properties for data representation.

W Yu and S Brewster (2002), Comparing two haptic interfaces for multimodal graph rendering, Proc. 10th Symp. On Haptic Interfaces for Virtual Environments and Teleoperator Systems, USA, Florida, pp. 3-9.

W Yu, K Kangas and S Brewster (2003), Web-based haptic applications for blind people to create virtual graphs, 11th Haptic Symposium, Los Angeles.