A Note from the Editor

Dear SIGACCESS member:

Welcome to the new look of the online edition of the SIGACCESS Newsletter – with new layout, the use of sans-serif and larger font throughout, left-justification, and the inclusion of authors’ short biographies and photographs (so that you can say hi when you meet them in meetings and conference).

This issue mainly reports the research that the doctoral students participating in the ASSETS 2008 Doctoral Consortium (DC) are working on. As always, the students are very enthusiastic about their experience at the DC. Some of their comments are included in p. 3 of this Newsletter. Finally, this issue includes a trip report of ASSETS 2008 in Halifax.

Sri Kurniawan
Newsletter editor
SIGACCESS Officers and Information

Chairperson
Vicki Hanson
IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10532, USA.
+1-914-784-6603 (work)
chair_SIGACCESS@acm.org

Vice-chairperson
Andrew Sears
Information Systems Dept.
UMBC
1000 Hilltop Circle
Baltimore, MD 21250, USA.
+1-410-455-3883 (work)
vc_SIGACCESS@acm.org

Secretary/Treasurer
Noelle Carbonell
LORIA,
Campus Scientifique
F54506 Vandoeuvre-lès-Nancy
Cedex
France
treasurer_SIGACCESS@acm.org

Newsletter Editor
Sri Kurniawan
Dept. of Computer Engineering
University of California Santa Cruz
SOE-3, 1156 High Street
Santa Cruz CA 95064, USA
+1 831 459 1037
editors_SIGACCESS@acm.org

Who we are
SIGACCESS is a special interest group of ACM. The SIGACCESS Newsletter is a regular online publication of SIGACCESS. We encourage a wide variety of contributions, such as: letters to the editor, technical papers, short reports, reviews of papers of products, abstracts, book reviews, conference reports and/or announcements, interesting web page URLs, local activity reports, etc. Actually, we solicit almost anything of interest to our readers.

Material may be reproduced from the Newsletter for non-commercial use with credit to the author and SIGACCESS. Deadlines will be announced through relevant mailing lists one month before publication dates.

We encourage submissions as word-processor files, text files, or e-mail. Postscript or PDF files may be used if layout is important. Ask the editor if in doubt.

Finally, you may publish your work here before submitting it elsewhere. We are a very informal forum for sharing ideas with others who have common interests.

Anyone interested in editing a special issue on an appropriate topic should contact the editor.
Comments from Doctoral Consortium Participants

I found the time spent at the Doctoral Consortium to be very rewarding. The Doctoral Consortium gave me an opportunity to interact with experienced professionals in my field, as well as peers. The useful feedback received from the team regarding my research project has greatly helped to shape my work. I strongly recommend the doctoral consortium to PhD students who are just starting out - Grace Mbipom

Doctoral Consortium of ASSETS 2008 was a very good opportunity to share ideas and to get valuable feedback from experienced research colleagues for the rest of my project. It let us know the state of the art in this area, related projects and colleagues working on similar topics. This knowledge will help us to fine-tune our research directions and to find collaboration opportunities with other researchers in this field - Cagatay Goncu

I feel so grateful to have attended the ASSETS 2008 DC. It has provided me with useful feedback from a panel of established researchers and students. I’ve also managed to establish good contacts with experienced researchers and fellow students - Hasni Hassan

The presentation of my beginning PhD at the Doctoral Consortium, and the questions I have been asked, gave me relevant feedbacks on how to improve my methodology, what phenomena should I have to look for, and what basis do I need to consolidate. We have had very useful discussions, with senior researchers or PhD students, about the research of each one, which gave me new ideas for mine. The poster session made me establish new contacts, meet people of which I had read articles before, and present my work to researchers from several different fields of knowledge. I think this Doctoral Consortium, and the poster session at the ASSETS main conference, were valuable experiences – Jérémie Segouat
Automatic Assessment and Adaptation to Real World Pointing Performance

Amy Hurst
Human Computer Interaction Institute, Carnegie Mellon
akhurst@cs.cmu.edu

Abstract
There is a growing population of individuals who are motivated to use a computer but find it physically difficult to do so, particularly when using a pointing device such as a mouse. Common pointing problems include inability to select small targets, difficulty moving a pointing device in a straight line, or difficulty controlling the pointer’s buttons. There are many software adaptations that can improve pointing performance, however the key to these solutions is identifying when to deploy them. In my thesis, I am working to build software that can automatically assess pointing problems and their severity during real world computer use, and deploy appropriate assistive adaptations. In order to evaluate real world interaction with GUIs, I am developing both data capture and analysis tools that are application independent. This software-based approach to improving computer access is powerful because it will be able to adapt to an individual’s changing needs and work across applications.

Introduction
One of the main reasons computers are inaccessible is that they treat all users the same, and usually do not accommodate a user’s changing needs. Currently the best solution to match ability to technology is expensive human-in-the-loop technology assessment. Unfortunately, the high cost of these assessments prevents them from being performed frequently enough to capture an individual’s changing ability. As a result, it is not uncommon to find individuals with pointing problems using either no computer access technology or technology that is not configured for their needs.

I am working to build software that can automatically assess a user’s pointing performance during their computer use and adapt to their current needs. By assessing performance during real world use, assessments can be made more frequently and can better model a user’s changing needs.

My thesis works towards this goal by focusing on three aspects of this problem: real world data collection, assessment and adaptation. In order to maximize the potential impact, I am working to make all pieces application independent and, when possible, operating system independent. I am collaborating with individuals with a wide range of motor and cognitive abilities at United Cerebral Palsy Pittsburgh to observe real world pointing problems and collect both laboratory and real world pointing data.

In this article I will briefly review related work on pointing performance and adaptations designed to accommodate pointing problems. Then I will discuss my progress collecting a dataset of real world pointing use from individuals with and without motor impairments, building learned statistical models to assess pointing performance and deploying pointing adaptations to support user needs.
Related Work

Pointing Performance

Many researchers have studied pointing performance trends of individuals with a wide range of abilities in laboratory settings including [1,3,9,10,11]. These studies have characterized pointing performance in terms of movement trajectories, accuracy, clicking behaviors and speed. I draw from these studies to better understand pointing performance, common pointing problems, and how to assess them.

While these studies have answered many questions about input devices, interface designs and human ability, they have answered many fewer questions about what happens outside the lab, and if these findings transfer to real world use. My work strives to learn more about the characteristics of real world pointing use, compare it to laboratory data by collecting real world pointing data, and analyze it with the same techniques used in laboratory studies.

Interface Adaptations

There is a large body of work exploring how to optimize interfaces for performance and make pointing tasks easier. Examples include changing the way the cursor interacts with targets to make them easier to access [12], increasing or decreasing pointing gain [10], and pausing the pointing device during a click to prevent it from slipping off a target [11]. These adaptations are all designed to accommodate specific pointing problems, and are able to improve performance for individuals who have those problems. However, the key to a large-scale successful deployment of these adaptations is in knowing when to deploy these adaptations and the optimal settings for the current user.

An alternative to changing how the pointing device interacts is to change the onscreen interactors to meet the needs of the user or the input device. One solution to doing this includes collecting both preference information and baseline performance data from a user and automatically designing an optimized GUI for their needs [4]. Another solution is to automatically tailor the interactors in an application to the input devices available to the user. Configurations include redesigning an interface for interaction with a keyboard and mouse, only a keyboard, only a mouse, or even a single switch button [2]. While these approaches represent a valuable way to improve accessibility, these techniques are limited in the applications and toolkits that they can support.

Assessment and Adaptation

Data Collection and Analysis

The best way to build models that will generalize is to use large datasets of pointing actions that represent a wide range of problems. Ideally these examples are representative of problems that happen infrequently. One relatively underexplored area of pointing performance is the study of real world pointing performance. I am exploring this in a study to collect real world pointing performance from individuals without pointing problems. This dataset will be used as a baseline of real world performance data to compare to data from individuals with pointing problems. I am currently conducting a long-term study collecting data from individuals with cognitive and motor impairments at United Cerebral Palsy Pittsburgh.
I have conducted an initial analysis comparing laboratory trials to real world pointing data that found high variance within individuals over time [7]. Figure 1 illustrates the high variance in distance slipped while clicking a target across different login sessions of real world use (sessions 1 through 9) and compared to a baseline laboratory pointing task (session 0). This finding helps motivate the goals of this thesis because one of the best ways to support high variance is through frequent assessment.

One of the largest challenges in analyzing real world pointing data is automatically determining the sizes of the targets interacted with. While accessibility APIs such as the Microsoft Active Accessibility API (MSAA) provide information about some of the targets, they leave many unsupported. I have developed a technique to augment information from Accessibility APIs with computer vision to improve automatic target identification in GUIs. In a dataset of 438 images from 6 applications, my hybrid target identification technique improved overall recognition accuracy from 18.5% (with the MSAA alone) to 73.5% (with MSAA and computer vision). This technique finds targets by analyzing the visual difference before and after a button press, or by using template matching of common target edges to identify a target. Figure 2 illustrates how the MSAA API was unable to identify the target, but my technique found the target successfully. With this dataset I will also investigate the distribution of pointing types and characteristics of frequently used targets.

Figure 1: Real world pointing data from 3 users, illustrating their variance in pixel distance slipped during a laboratory evaluation (session 0) and across real world use (sessions 1 and above).

One of the largest challenges in analyzing real world pointing data is automatically determining the sizes of the targets interacted with. While accessibility APIs such as the Microsoft Active Accessibility API (MSAA) provide information about some of the targets, they leave many unsupported. I have developed a technique to augment information from Accessibility APIs with computer vision to improve automatic target identification in GUIs. In a dataset of 438 images from 6 applications, my hybrid target identification technique improved overall recognition accuracy from 18.5% (with the MSAA alone) to 73.5% (with MSAA and computer vision). This technique finds targets by analyzing the visual difference before and after a button press, or by using template matching of common target edges to identify a target. Figure 2 illustrates how the MSAA API was unable to identify the target, but my technique found the target successfully. With this dataset I will also investigate the distribution of pointing types and characteristics of frequently used targets.

Figure 2: Target selection technique identifying targets in maps.google.com A) Dashed rectangle indicates smallest target MSAA API can detect. B) Bold square indicates target user pressed, which was found with the template matching part of my technique.
targets. Specifically, I am interested in how well real world movements actually fit preexisting performance models such as Fitts' Law [3] and the Steering Law [1]. I will also explore what categories of motion our users perform which cannot be described by these models. This dataset will enable me to investigate the distribution of target sizes and the distances between them, and see how representative they are of the targets and distances used in laboratory studies. Most importantly, I will use this dataset as the training set to build learned statistical models that are able to assess real world pointing.

Assessing Pointing Performance

I have built learned statistical models of pointing performance to automatically detect characteristic movement from individuals with a wide range of abilities. These models distinguish between different groups of use with performance metrics, or features, that describe a pointing interaction. To date, I have built models assessing pointing performance from laboratory studies evaluating menu use and 2D target acquisition.

To better assess performance while using menus, I collected a dataset of novice and skilled menu selections made while using an image editing program. The interaction between the pointing device and menus was analyzed and scored to develop a set of performance metrics, or features, for each menu selection which were used to build a learned statistical model. This model was able to distinguish between novice and skilled menu use with 91% accuracy, without using a task model of the user’s actions [5].

To better assess performance of 2D target acquisition, I built learned statistical models with 3 datasets from prior work in this area to distinguish between performance differences due to age and motor impairment [6]. Specifically, data from the Koester dataset [10] was used to distinguish between the pointing performance of individuals with and without known pointing problems. A dataset from Keates [9] was used to distinguish between pointing performance based on age groups or the diagnosis of Parkinson’s Disease. Finally, a dataset from Trewin [11] was used to predict how well a learned statistical model could predict if the Steady Clicks adaptation would improve performance of individuals who slipped off a target when they tried to click. All 3 models performed with high accuracy results (above 90%), and are illustrated in Figure 3.

Adapt to Pointing Performance

The learned statistical models described in the previous system can be incorporated into applications so they can continuously assess a user’s performance and act accordingly. To adapt to pointing performance, an application should deploy an adaptation designed for a specific performance change once it detects that such a change has occurred. Depending on the specific pointing problem, this adaptation might change the way the pointing device interacts with onscreen elements, or the way it responds to user movement.

I am currently working to build a suite of tools to log user actions and plug-and-play pre-existing adaptations. I have developed one solution to help users acquire frequently used targets by observing a user’s click and drag history [8]. With this history, my software applies a simulated gravity field around the collection of previous interactions to help the user acquire targets. This technique is application independent and operating system independent. While this technique provides information about regions that contain targets, it is limited to adaptations that are agnostic of target size because it does not
know their precise size or exact location. I have developed the hybrid computer vision and MSAA API technique to more accurately identify target size and location so more adaptations can be supported.

![Figure 3: Performance of three learned statistical models built from three datasets from other researchers. These models distinguish pointing performance to predict presence of pointing problem, age differences, and if an individual's performance would improve with Steady Clicks adaptation.](image)

**Status of Research**

**Collect and Analyze Real World Data**

I am working to collect a large dataset of pointing performance from individuals with a wide range of physical abilities performing real world tasks. This real world dataset will be unique and will be analyzed to learn more characteristics of real world pointing including how technique, accuracy, and pointing problems differ over time and across users. In order to perform this analysis, I will employ my techniques to identify onscreen targets using the MSAA API and computer vision. Upon completion of my thesis, I plan on making this database available to other researchers interested in studying real world GUI interactions.

**Assess**

I plan to use this real world dataset to build learned statistical models to identify computer users who would benefit from accessibility tools and make adaptation recommendations. Based on the performance of these models, I will be able to design an appropriate system to translate the output of a classifier into specific adaptations and changes that need to be made.

**Adapt**

Once I have built models based on real world pointing behavior, I will evaluate the performance of these models and automatically deploy adaptations to the interface that are tailored to the user's current ability. Participants will be interviewed about their perceived performance when the adaptations were deployed and how the adaptations affected their pointing performance. These subjective results will be compared to their performance with and without the deployed pointing adaptations.
Conclusion

I am working to make computers more accessible by automatically assessing pointing performance and adapting interfaces to address pointing problems. I am collecting a database of real world pointing behavior, and have had success analyzing real world pointing data, building learned statistical models to identify pointing problems in a laboratory study, and building adaptive systems. With this research, I hope to make computers more accessible for individuals with pointing problems.

Acknowledgements

This work was supported in part by the National Science Foundation under grants EEC-0540865, IIS-0713509, IIS-0325351, Amy Hurst’s NSF Graduate Student Research Fellowship; IBM Research; and the Pennsylvania Infrastructure Technology Alliance.

References

About the author:

Amy Hurst is a PhD student at the Human Computer Interaction Institute at Carnegie Mellon in Pittsburgh, Pennsylvania where she is advised by Scott E. Hudson and Jennifer Mankoff. Her research interests include Assistive Technology and Context Aware Computing. She graduated from Georgia Institute of Technology in 2003 with a BS in Computer Science and a Certificate in Film Studies. She is the recipient of a National Science Foundation Graduate Research Fellowship (2005) and an Anita Borg Scholar (2008).
Improving Computer Interaction for Older Adults
Nic Hollinworth
School of Systems Engineering, University of Reading, UK
n.d.p.hollinworth@reading.ac.uk

Abstract
A point and click interface can present difficulties for older adults (particularly those with physical impairments) who may not easily be able to operate a standard input device such as a mouse or trackball. The use of gestural commands, via a multitouch touchscreen device, is an alternative and direct method of interacting with an application. This research investigates whether gestural commands on a multitouch touchscreen device can make interfaces easier to use, learn and remember for older adults.

Introduction
It is estimated that by the year 2010, 17.6% of the population in the European Union will be over the age of 65, many of whom will undertake computer-based tasks for social, recreational, continuing employment or health-related purposes [9]. However, age-related studies [e.g. 6, 24, 25] have shown that interacting with a computer using traditional computer input devices, such as a keyboard and mouse, can cause problems for many older adults, in particular novices in this age group and those with age-related disabilities such as Parkinson’s disease, arthritis and essential tremor. For example, difficulties can arise when attempting point and click interactions, such as drag and drop, which require manual dexterity and fine motor control. This could be improved significantly with touchscreen interaction, as it has the potential to offer more intuitive methods of interaction using existing skills, operating directly with objects, and requires less visuo-spatial coordination [24].

Existing research into touchscreen interaction [e.g. 15, 23] has relied mostly upon touchscreens that could detect only a single point of contact, which means that objects on the screen were manipulated directly using a single finger, and commands were invoked using single finger gestures, such as rubbing the finger from side to side to erase content or undo the last action. More recent developments in touchscreen technology offers the ability to detect multiple points of contact on the screen (e.g. all 10 fingers), allowing for the possibility of two-handed gestural interaction, such as simultaneous rotation and scaling of images, as an alternative to using a mouse.

This research aims to investigate how to design everyday computer activities for older adults, such as sending an email, by taking advantage of multitouch technologies. This will involve identifying the types of tasks and subtasks that are needed to undertake the given activity, and investigate how the tasks might be mapped to the touchscreen gestures. Previous research exercises [8, 23] have shown that there is variation in how people conceive of gestures for specific tasks. Hence, as a starting point the programme of work begins with an initial study (a ‘paper and pencil’ exercise) to gain a better understanding of what tasks are required for the given activity, the types of (two-handed) gestures used and how these differ between the younger and older age groups.
Related work

Background

Many studies have looked at the effects of ageing on computer interaction [e.g. 14, 19, 24, 25], and some have devised novel hardware and software techniques that could help improve interaction for both users with age-related declines and also disabled users [1, 21, 25, 28]. For example, techniques have been developed that help to steady the mouse pointer when making a selection [e.g. 21, 25] or make double-click actions easier to perform [e.g. 28]. However, these essentially aim to improve point and click interaction, using familiar (indirect) pointing devices, such as the computer mouse, which manipulate an on-screen pointer.

An alternative is to interact directly with on-screen objects using a touchscreen. For example, Potter et al [17] investigated three different methods of touchscreen interaction and compared the speed, accuracy and user satisfaction of each. They found that the method favoured by participants was to have the cursor located above the finger and offset from the finger location, rather than beneath the finger. Touchscreen accuracy was also studied by Benko et al [4] who developed a number of techniques to refine the accuracy of interaction, using a multitouch screen and two fingers, such as using the contact area of the finger tip to determine pressure. Other studies have compared the performance of touch screens with various other devices, and for a variety of applications. For instance, Karat [12] compared user performance and attitudes for menu selection using a keyboard, mouse and touch screen and found that keyboard and mouse performance were about equal, but less than the touchscreen.

Jin et al [11] investigated usability of icons by considering the optimal button size and spacing between buttons for touchscreen user interfaces used by older adults. Their experiments measured reaction time, accuracy and user preferences. Results showed that larger button sizes produced shorter reaction times, and was consistent with other studies such as Sun [20] who investigated button size and spacing for touchscreens used by firefighters and found that there is a trade-off between speed and accuracy that depends upon the spacing between buttons – larger spacing resulted in fewer errors but increased the reaction time.

However, the majority of work has been concerned with single point of contact devices, i.e. those limited to single touch interaction, often using just the index finger. More recent developments [e.g. 10, 22] can detect multiple points of contact using one or two hands, and allow for much richer interaction through gestures [e.g. 8, 14, 18]. Studies have looked at developing gesture sets and how they naturally map to computing tasks [8, 23], but these have so far been mostly for single handed interaction. Studies have yet to be conducted which consider two-handed multitouch gestural interaction and gestural interaction involving older users.

Interacting with Gestures

The operation of a device such as a desktop computer requires a rich set of gestures, since it needs to be able to activate a (potentially) large number of commands. Many studies have been undertaken that have investigated gestural input for computer interaction using devices such as touchpads and touch surfaces. For instance, gestural communication involves more muscles than keyboard or speech, so gestural commands
can be rather tiring to carry out in practice and therefore must be concise and quick, and not require high precision over a long period of time. Wu et al [26] considered this problem and developed a set of design principles for constructing multi-hand gestures that could be ‘relaxed’ after the initial contact with the surface, once the command had been recognized by the system. It included gestural reuse to reduce the number of gesture primitives that a user must learn.

Wobbrock [23] constructed a user defined gesture set, which reflected users’ behaviour and was based upon the agreement of the participants involved in the study. The results suggested that the users didn't care about the number of fingers used in the gesture, preferred one hand over two, were strongly influenced by the desktop paradigm, but for some of the commands there was little agreement over what gesture would be appropriate.

A similar study was undertaken by Epps [8] who investigated user preferences for tabletop gestural interaction. They found that there was a high degree of consistency amongst participants for a small number of gestures and the most commonly used gesture was the index finger on the preferred hand, followed by spread hand and also a flat hand. An interesting observation was that the subjects raised their hands above the surface frequently when responding to tasks suggesting that the surface provides important feedback to the user, and acts as a convenient resting place for the hand.

Yuan et al [27] carried out a study with a group of users who had various disabilities that affected hand and finger movements and so they designed a set of command-like gestures to address this problem. Since many existing gestures for touchscreens rely upon full hand function, they may not be accessible to people with a physical disability. For instance, a person with contracted fingers, perhaps due to paralysis, may only be able to achieve a clenched hand and there may be limited movement of wrist. Thus, the possible contact shape of the hand on the surface will be very different from an able user. So, instead of using finger-based gestures, they used a clasped hand (a fist) and based gestural interaction on the trajectory and angle of the hand.

These studies suggest that the use of gestures may afford an alternative and more natural interaction method, since users can apply existing motor skills and interact directly with virtual objects, rather than having to develop the necessary skills to manipulate a pointing device. In addition, the ability to use both hands offers potential for a more efficient interaction method, and perhaps one that readily transfers existing manual skills. However, little work has been undertaken so far with respect to one and two-handed touchscreen interaction using gestural input for older users and perhaps those with age-related impairments and disabilities.

**Bimanual Touchscreen Interaction**

Gestural interaction using a multitouch screen offers potential for an improved user experience, and for two-handed interaction. But there is a need to study how people use their hands in computer interaction using gestures, and how tasks are divided and assigned to each hand. Although there has been extensive research in two-handed input [e.g. 5, 13] this has yet to be extended to multitouch interaction.

With multitouch gestural interaction commands may be easier to learn and remember if they are designed to reflect the manipulations of real world physical objects, since this is an existing skill that people already possess. Interacting with gestures can also be more
entertaining and may encourage people to play and explore an interface. In this sense, gestural interfaces should be made ‘discoverable’ so that people can find out for themselves how they work [18]. The iPhone [2] is a good example due to its simple and intuitive interface, with controls that are designed to help users understand how to operate them, without the need for a manual, and invite further exploration.

There are also many challenges that need to be addressed for older adult users, since with increasing age also comes many declines in physiological functions. The likelihood of developing a disability tends to increase with age, with many older people having at least one chronic disability such as arthritis, or a hearing or vision impairment [7]. Most touchscreens, and in particular virtual touchscreen keyboards, lack tactile response, so for users with visual impairments some other modality (such as sound) may need to be employed to provide adequate feedback in response to an action. The design of applications will need to be researched too, since existing interfaces operated through point and click often have screen objects which are too small or too close together to be operated easily using fingers, so are not well suited to gestural interaction.

**Proposed Research**

The aim of the research is to learn how older adults can make effective use of a multitouch touchscreen computer through two-handed gestural commands and perform an everyday computer task, such as sending an email. The same tasks will be undertaken by younger users and then the results will be compared. This is an emerging technology with only a handful of devices currently available, and a largely unexplored area for older computer users. The main objectives are the following:

1. To understand more about how older users can undertake computer-based tasks with a touchscreen using both hands.
2. To examine preferences for gestural interaction, and how gestures naturally map to tasks. This has already been partially explored by a few researchers, but not for older users, nor users with any (age-related) disabilities.
3. To study the difficulties encountered when using a multitouch touchscreen computer. For instance, the ability to form gestures may not be possible for adults with certain conditions such as Parkinson’s disease or arthritis, and in this case it would be interesting to find out if alternative gestures could be employed for those gestures which are too difficult.
4. If particular gestures are not possible or not appropriate, then what are the characteristics of these gestures?
5. Which tasks are most suited to multitouch touchscreen interaction? For example, the direct manipulation of images when editing may be easier than using point and click with a traditional mouse. Conversely, selecting a single pixel when image editing may be simpler with a mouse.
6. To investigate novel techniques for gestural interaction when devising new applications specifically for multitouch touchscreen.
7. To evaluate possible performance gains, ease of use, learnability, and retention. For instance, does direct interaction using a touchscreen actually promote learnability and retention?
The research will involve three linked phases starting with a paper and pencil exercise that will investigate and compare how older and younger users undertake a physical task. The first exercise consists of writing a simple letter, addressing it, enclosing a photograph and posting it. It is designed to find out more about how the hands are used in a simple everyday task, and how tasks are naturally assigned to each hand by the participant. The results of this will be used as the basis for the second study, which looks at a similar activity (apart from writing the letter), but displaying the objects as (static) images on a screen. Using a similar method as Wobbrock [23] participants will be provided with the effect of a gesture and asked to perform its cause. The outcome will be a set of gestures used in the performance of the actions.

Once some understanding of how users produce two-handed gestures in response to given stimuli has been gained, the results will inform the development of a multitouch email application. Rather than building a standard point and click interface, an application will be built that responds to gestural commands and the direct manipulation of screen objects.

The research questions include:

- What tasks would the users find most difficult to accomplish when using a real application and what tasks did not map to gestures easily or at all?
- Is the application easy to learn and remember?
- Are there gestures which the users could not carry out because of physical impairments? For instance, perhaps a user with arthritis could not manipulate an object because of the necessary gesture? What are the characteristics of the gesture that makes it difficult or impossible to use?

This research will provide a better understanding of how older adults use gestures to interact with a multitouch computing device, and provide some insights into the preferences of users, and limitations of two-handed touchscreen interaction due to physiological constraints. In addition, we will gain a better understanding of how gestural interaction differs between older and younger users.

Acknowledgements
Thanks to my supervisor Faustina Hwang, and to the staff and volunteers at Age Concern UK, and the University of Reading. This work is funded by School of Systems Engineering, University of Reading.

References


About the author:

Nic Hollinworth is a PhD student in the Interactive Systems Research Group, School of Systems Engineering, at the University of Reading. His research area is in improving computer interaction for older adults, with a particular interest in multitouch touchscreen interaraction.
Common Input Devices for Malaysian Computer Users with Motor Impairments

Hasni Hassan

School of Computer Science, University of Manchester, U.K
Hasni.Hassan@postgrad.manchester.ac.uk

Abstract

This research aims to investigate whether there are common input devices that are more effective than a standard mouse for Malaysian computer users with motor impairment. The tasks range from very controlled Simple Reaction Time (SRT) tasks to Fitts’ Law and browsing tasks. There are several potential contributions of the research beyond finding the input device that is more effective than a mouse: 1) The Fitts’ Law model and information architecture for people with motor impairment and 2) the relationship between subjective and objective measures in highly controlled as well as naturalistic tasks.

Introduction

Various studies with people with motor impairment had strongly suggested that access to computers and the Internet could provide means of freedom and better opportunities in education, employment, and their social lives in general. The use of a suitable input device plays a significant role in ensuring optimal access to computer. Undoubtedly, people with disabilities would benefit from the use of assistive input devices and technologies. However, in many low-income and middle-income countries, in which Malaysia still belongs to, only 5%-15% of people who require assistive devices and technologies have access to them [20]. Most of the times, in situation when access might be possible, costs are prohibitively high.

Taking these factors into consideration and in view of the current scenario of people with disabilities in Malaysia, a study to investigate the use of other common input devices by people with motor impairment was conducted in order to understand whether there is any device that can perform more effectively than a standard mouse.

In this study, the choice of common input devices was based on previous study on input device usage by older person [9], hence user performance was compared using mouse as the baseline and notebook’s touchpad, touch screen and tablet with stylus. Although there are various representative tasks for pointing such as target acquisition, steering and many others [16], the focus of the study was on three types of ‘motor’ tasks, ranging from a very simple to a natural one.

The three tasks used were the Simple Reaction Time (SRT) tasks, experimental Fitts’ Law tasks and browsing tasks. The SRT task are essentially aiming and clicking tasks that was used to measure how accurately people with motor impairment can aim and click on a target when the location of the target remained constant. In contrast; in Fitts’ Law tasks, the location of the next target was always random. On the other hand, whilst Fitts’ Law tasks provided an objective comparison of performance differences by devices, the browsing tasks allows us to get an insight into tasks that are closer to real-life activities such as browsing the Internet.
Related Work
Numerous research involving people with disabilities and input devices have been conducted especially in the West such as in the USA and UK. Some studies that focused on people with motor disabilities include the evaluation of joysticks for people with tremor [12], mouse and keyboard manipulation difficulties [17, 19], and comparison on the use of input devices by older adults [2]. Some focused on various means of providing assistance to improve computer access for people with disabilities while some were related to the use of a mouse either by providing means to assist cursor movement or to improve mouse operation [3, 4-6, 9, 18].

The growth of the Internet, computing access and its associated facilities provide significant advantages to everyone including people with disabilities. It was reported that the people with disabilities are among heavy participants of public services and potentially have most to gain from convenient, customer-focused channels of electronic deliver [10]. However, in addition to the prohibitive cost of current computer systems, a lack of suitability and adaptability were the major barriers to easy access of computers by people with disabilities. Unfortunately, no research has been found that investigated the use of computers among people with motor impairment in Malaysia, especially in the context of computer input devices.

Overview of the Research
The research goals are to:
1. Find out whether there is common low-cost input device that is more effective and preferred than a standard mouse.
2. Investigate the performance using common input devices in controlled and naturalistic tasks (SRT, Fitts’ Law and browsing) through subjective and objective measures.
3. Fit the empirical data to the various Fitts’ Law models proposed by MacKenzie and Buxton [8] and to investigate the best fitting model.
4. Investigate the effects of individual differences in reading speed and comprehension on browsing performance.
5. Investigate the effect of information architecture (depth vs. breadth) on browsing performance.

Research Procedures
Taking into consideration the effect of fatigue and conditions of potential participants, it was decided that the study would be conducted in two modules. The first module involves a pilot and main study to meet the first four objectives while the second module would focus on the last objective of the study.

Pilot Study
In this study, 3 people with motor impairment were recruited (P1 has Multiple Sclerosis and uses a wheelchair, P2 is tetraplegic and uses a wheelchair, P3 has Cerebral Palsy). A 12” Twinhead notebook was used while devices used were JNC Optical mouse, the notebook’s touchpad, 4” x 3” Tablet with Stylus (Cordless Natural Pen Device) and MagicTouch Add On Touchscreen. The participants signed consent forms and filled in
demographics, computer, input devices and Internet experiences. The participants then performed the SRT test [16].

The controlled experiment consists of three tasks (the order of the tasks was balanced across participants). The first task was discrete pointing using the devices using the Generalized Fitts' Law Model Builder (GFLMB) [15]. The second task was to measure participants' reading and comprehension skills.

A Reading test set was constructed using 8 articles of four different complexities based on the Flesch-Kincaid (FK) grade level [11]: FK12, 14, 16 and 18 (since potential participants for the experiment would be adults with minimum education of 12 years). The length of each article was between 240 and 250 words. Each article has five questions to be answered. Two articles were used for each level to improve reliability.

The third task was a browsing task where the participants browse through online information to answer questions. The information was presented in an 8x8 web hierarchy (this was indicated as the most effective information architecture on browsing studies with older persons). The final leaf pages contain articles of 140-150 words. The participants' browsing time was defined as the moment they clicked ready to the moment that they clicked the 'answer found' button. This was when the information disappeared and they answered verbally.

At the end of the experiment, participants answered a post-task questionnaire on opinion on the use of each device. They were also asked to give their overall opinion of the experiment. The post task questions were adapted from Perceived Usefulness and Ease of Use (PUEU) questionnaire [1] while questions on the usability of the device and the overall opinion were adapted from Computer System Usability Questionnaire (CSUQ) and After Scenario Questionnaire (ASQ) [7]. To anticipate fatigue caused by lengthy experiment, the participants were asked to perform the SRT test using only the mouse just to test the SRT protocol. Their SRT scores varied from 0.28 to 0.63 seconds.

Results from Pilot Study

Based on a pilot experiment done with students, out of the 5 models proposed by MacKenzie and Buxton [8], the Sum-Of and Smaller-Of models are the best fitting models (they do not differ significantly). Therefore, the pilot data was fitted to these two models. Figure 1 and 2 present the Fitts' Law data as modelled using the Sum and SMALLER-OF models.

It can be observed that for both models, the fit was the highest in term of R2 when a mouse was used. The Smaller-Of model has significantly better fit across all devices compared to the Sum model. Therefore, it was decided that from here on the data will only be analyzed using the Smaller-Of model.

Results from the reading tasks show that as the level of difficulty increases, the time taken to read the articles increased. However, the participants' comprehension scores did not correlate with FK levels.

The results of browsing tasks show that using a mouse and a tablet with stylus yielded a similar browsing time (77s – mouse, 78s – tablet) and using touchpad and touch screen also yielded a similar browsing time (both at 112s, which was worse than when a mouse or a tablet with stylus was used). In addition, feedback from the post task questionnaire
indicated that there was no consensus as to the device was the preferred one or the easiest.

![IDSum vs Movement Time](image1.png)

**Figure 1:** The Sum model

![IDSmall vs Movement Time](image2.png)

**Figure 2:** SMALLER-OF model
Implications from Pilot Study

The pilot study carries two implications. First, participants took between 3 and 5 hours with breaks to finish the study even though they are regular computer users (computer use per week ranges from 60 to 65 hours). Since some of the time were spent doing the reading tasks, it was decided to drop articles at FK18.

Second, there was no consensus yet as to which device was the best. From the Fitts’ Law tasks, the tablet with stylus was by far the worst device (reinforced by users’ complaints when performing the Fitts’ Law tasks). However, the median times of the browsing task seemed to indicate that tablet with stylus was not the worst device. However, as we had to shorten the time the participants take, we decided to drop the tablet with stylus as it was the device with the worst performance in the Fitts’ Law tasks and the most complained one.

Main Study

The real study was conducted with 18 participants with various types of motor impairments ranging from cerebral palsy, spina bifida, quadriplegia, tremor and arthritis. The apparatuses and methods were the same as those of the Pilot Study, except the tablet with stylus was no longer tested. Most participants were familiar with computer and mouse; some were familiar with touchpad. Most of them were not familiar with touch screen.

Results from Main Study

Results from the SRT test suggests that the participants performed the fastest when using a mouse, followed by a touchpad and then the touch screen. The means are significantly different for those three devices.

Figure 3 shows Fitts’ Law model for the three devices. Using Zar’s method [21] for comparing the three regression lines show that even though the slopes are not significantly different from each other, the intercepts are, and therefore these are still not the same lines. In other words, the mouse was significantly faster than the touchpad, which in turn faster than the touch screen, confirming the results of the SRT. The ANOVA of the R2 also reveals that the fits are significantly different. Interestingly, the mouse has the lowest fit of the three devices, indicating that there is a large variation in the participants’ performances when using the mouse compared to the performances when using the touch pad or the touch screen.

Results for reading time and comprehension were similar to the Pilot Study. Nevertheless, it was discovered that valid predictions of text comprehension could possibly be hindered by surface characteristics of the text, reader’s cognitive aptitudes and cohesion and coherence of the text [13].

For the browsing data, the average browsing times for the participants were 63.3s (mouse), 71s (touch screen) and 74.2s (touchpad). One-way ANOVA shows that the average browsing times for the participants are significantly different across the three devices. The LSD posthoc analysis also reveals that each device’s browsing time average is different from one another. The result is that, in line with the data from the SRT and Fitts’ Law that mouse is the best device. However, the device with the worst performance this time is the touchpad rather than the touch screen.
For the post-task questionnaires, the means differences were analyzed using ANOVA. Table 1 shows the significant questions with $p<0.05$. An LSD posthoc analysis was performed for those that were significantly different. The posthoc pair wise comparisons were done on mouse vs. touchpad, mouse vs. touch screen, and touch screen vs. touchpad subsequently.

When asked to rank the devices in terms of ease of operation and preference, the ranking was identical. Ten people ranked mouse, 5 ranked touchpad and 3 ranked touch screen first.

<table>
<thead>
<tr>
<th>Question</th>
<th>Posthoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using the device in my job would enable me to accomplish tasks more quickly</td>
<td>NS, ***, *</td>
</tr>
<tr>
<td>I found it easy to get the device to do what I want it to do</td>
<td>NS, ***, ***</td>
</tr>
<tr>
<td>I found the device to be flexible to interact with</td>
<td>***, ***, NS</td>
</tr>
<tr>
<td>I can accurately complete my tasks using this device</td>
<td>NS, ***, ***</td>
</tr>
<tr>
<td>I feel comfortable using this device</td>
<td>NS, **, NS</td>
</tr>
<tr>
<td>Overall, I am satisfied with the ease of completing the tasks</td>
<td>NS, ***, ***</td>
</tr>
<tr>
<td>Overall, I am satisfied with the amount of time it took to complete the tasks</td>
<td>NS, ***, ***</td>
</tr>
</tbody>
</table>

**Current Status**

The second phase of the study was designed to meet objective (5). In essence, apart from investigating the effect of information architecture on browsing performance, it would also...
provide evaluation on users’ performance when using the three devices. Results that will be obtained could be compared with the results in main Study.

There are inherently several limitations of the study. Given the large variability in motor impairment found in our participant pool, there is always a possibility that any non-significant results are a result of this large variance. However, since all the objective measures showed significance, we could probably assume that the non-significant results in the subjective ratings (which are less affected by motor impairment as they are measures of opinions) are due to lack of noticeable differences in the aspects investigated.

The participants were all computer users who are mostly familiar with the mouse. This could definitely sway the results toward the mouse. A follow-up study with non mouse users could definitely provide a more objective result. Despite of these limitations, this study provides some insight into the types of common input devices that Malaysian computer users can potentially use.

Conclusions
This study investigates whether there is a better common input device than a standard mouse for Malaysian computer users with motor impairment. The study found that overall the mouse is still the device of choice, although familiarity might have affected the results (given that almost all of the participants were familiar with the mouse). However, the other devices did not lack behind much from the mouse, opening opportunities for training using devices in which touch pad or touch screen is the standard input devices rather than mouse (such as in laptops).

References


About the author:

Hasni Hassan is a PhD student in Computer Science School at the University of Manchester, working under the supervision of Dr Simon Harper. Her research interests are in the area of human computer interaction with a focus on the people with disabilities. She holds a M.Sc. in Information Technology from the MARA University of Technology Malaysia (UITM) in 2000 and a B. Eng in Electrical Engineering from the University of South Australia in 1997. Currently attached with Universiti Darul Iman Malaysia (UDM), she is awarded with full time study leave, sponsored by the Government of Malaysia.
Critically Analyzing Workplace Discourse to Inform AAC Device Design

Carrie Bruce
Georgia Institute of Technology
carrie.bruce@coa.gatech.edu

Abstract
The purpose of this paper is to introduce a study on workplace discourse that will be conducted to inform design of AAC vocabulary to better support workers’ communication needs. Although some work has been done by AAC researchers and mainstream discourse researchers, there is limited evidence that can be used in AAC device prediction systems that is related to workplace and, more specifically, task-oriented discourse. The limitations in our understanding of workers’ conversational needs warrant investigation into their discourse and the associated physical, temporal, and social environment factors. This study will collect and critically analyze spoken workplace conversation from 20 workers using discourse analysis and corpus linguistic techniques.

Introduction
For over 3.5 million individuals with severe communication impairments [6], full inclusion into the workplace is not guaranteed due to difficulties meeting the spoken and written demands of the job. Consequently, working-age adults with severe communication impairments may use augmentative and alternative communication (AAC) strategies and technologies such as communication boards and speech generating devices to minimize or eliminate communication barriers.

While AAC systems can potentially provide access to predictable and unpredictable workplace vocabulary, one of the most commonly cited challenges is slow communication rate compared to natural speech [5, 9, 20, 22]. Furthermore, limitations in the vocabulary available to and spoken by the user of the AAC system impact the mechanics of the communication exchange, resulting in an imbalance in conversational equity (e.g., lower initiation rates and/or fewer conversational turns) [23]. Moreover, speed and message content have been shown to negatively effect a partner’s attitude towards the person who uses an AAC system [5, 13]. Consequently, an individual who uses an AAC system is likely at a disadvantage when collaborating with colleagues, interacting with customers, and conversing with superiors and subordinates due to the significantly slower communication rate and the limitations associated with available vocabulary, as well as the potentially negative attitudes and conversational inequality that result from rate and vocabulary issues.

Research on rate enhancement strategies for AAC systems, specifically in speech generating devices, has been ongoing for over 25 years. A substantial amount of effort has concentrated on automated prediction of single words, phrases, utterances, and topics [for example, 1, 4, 12, 18, 19, 21]. Language modeling and natural language processing (NLP) have played a significant role in these prediction systems in an effort to improve prediction success rates and therefore, produce utterances that more closely
resemble natural speech in speed and message content. In a majority of word and phrase prediction systems, language models are trained on databases of spoken or written text called corpora. However, only a small number of formalized corpora, such as the Cambridge and Nottingham Business English Corpus (CANBEC) and the Enron emails, relate to work settings. In this limited pool of work-related corpora, none specifically include people with disabilities, have been analyzed to inform AAC devices, or have been included in AAC prediction development efforts. Consequently, corpora used in AAC prediction systems have not included work-specific language that could improve communication efficiency and effectiveness in workplaces.

Despite the absence of workplace data in prediction corpora, there are a few researchers who have investigated workplace conversations to identify potential vocabulary needs of individuals who use AAC [2, 3, 11, 20]. Balandin and Iacono [2, 3] and Tönsing and Alant [20] examined social conversations among workers without disabilities and found that they discussed fairly similar topics during meal breaks. However, they did not investigate conversations during work tasks, including task-oriented discussions. In contrast, File and colleagues [11] collected conversational samples during work tasks. This project yielded a corpus with examples of task-oriented conversation, but further analysis would be required to identify detailed discourse information (e.g., pragmatic functions and discourse markers). Consequently, there is a limited evidence base within AAC research for understanding what workplace and task-oriented conversational data is missing from existing devices.

Literature outside of AAC can potentially provide insight into workplace conversation as spoken and written discourse has been more extensively studied. These studies have characterized and analyzed work-related and social discourse that occurs through face-to-face, phone, or computer mediated communication (CMC) (e.g., instant messaging, email, and chat) interactions [7, 8, 10, 14-17]. The results of these studies indicate that there are conversational differences depending on various factors, including the interaction medium (e.g., face-to-face, email, etc.) and the goal of the interaction (e.g., small talk, gossip, etc.). However, similar to the AAC literature, much of the research in workplace discourse has not specifically looked at discourse during work tasks.

**Purpose**

Based on the stated gaps in our understanding of task-oriented workplace conversations for informing AAC vocabulary and the potentially useful research in mainstream workplace discourse, a multi-phase formal study of workplace discourse across various communication channels including face-to-face, email, and instant messaging is being undertaken. This study will initially investigate face-to-face workplace discourse, focusing on task-oriented conversation. Additionally, contextual factors, such as time, location (e.g., a colleague's office versus the copy room), and conversational partner (e.g., a supervisor and an employee versus two coworkers) will be attributed to the discourse to determine any potential associations. It is hypothesized that this critical analysis of workplace discourse based on temporal, physical, and social environmental information (i.e., when, where, and with whom one is conversing) could better guide NLP techniques used in AAC device prediction systems, consequently enhancing communication rate and message content [11]. This paper discusses the methodology for this initial phase of the study.
Methodology

Participants
To obtain a representative sample of office discourse, 20 employees without communication impairments whose primary job tasks require verbal communication (e.g., public speaking, phone use, or working on group projects) will be recruited from at least 10 different office settings.

Data Collection
Discourse samples will be gathered over the course of 5 consecutive work days to ensure that a wide range of routine and novel topics are included. Participants will meet with the researcher to discuss the procedures of the study, map out the workplace, participate in a brief discussion about conversation habits at work, and review operation of the recording device. Conversations will be recorded via wearable voice activated audio recorders. Participants will be asked to maintain a journal of conversation-related information such as location, partner, and tasks. Conversations will be time-stamped to provide a record of when they took place. Location and conversational partner will be identified through participant reporting in the journal and prompted recall (using samples or description of audio recording to retrospectively prompt identification). Locational references will include space (e.g., copy room, personal office, etc.) and/or products (e.g., copy machine, telephone, etc.). While the participant may provide a name code for the partner, the researcher will ask the participant to characterize the person according to role (e.g., immediate supervisor, in-office authority above supervisor, etc.) and level of social closeness (e.g., stranger, acquaintance, etc.) during a prompted recall session that will occur within one week of the recording dates.

Data Analysis
Although participants' conversations will be continuously recorded, specific time intervals will be selected for transcription and analysis to ensure that conversational content is task-oriented. Transcription and coding will be performed on each sample using common discourse analysis and corpus linguistics techniques. The results of this phase of the study on spoken discourse will be compared to other forms of discourse (e.g., email and instant messaging) that will be collected and analyzed in later phases. The compilation of discourse vocabulary will be incorporated into a context-aware AAC device that will use location, conversational partner, and time to predict work-relevant vocabulary. The new device will improve communication rate and facilitate access to work-relevant vocabulary, thus enabling greater workforce participation by individuals who use AAC.

References


About the author:

Carrie Bruce is a research scientist at the Center for Assistive Technology and Environmental Access (CATEA) at Georgia Tech and an investigator for the RERC on Workplace Accommodations. She is distinguished for her work in examining environmental design issues that impact participation. Her recent work has focused on descriptive analyses of assessment instruments with a person-environment fit focus, development of a workplace accommodations assessment instrument, and investigation of universally designed exhibit interpretation. Additionally, Ms. Bruce is a PhD student in the Human-Centered Computing program at Georgia Tech where she is concentrating on developing a context-relevant communication system for workers with speech and language impairments.
A Study of Sign Language Coarticulation

Jérémie Segouat
LIMSI-CNRS (Orsay, France) / Websourd (Toulouse, France)
jeremie.segouat@limsi.fr

Abstract
This paper presents a coarticulation-modeling project in French Sign Language (LSF). We briefly introduce what sign language is about and why we study a specific part of it: to coarticulate isolated signs in order to create parameterized utterances. Then we explain what coarticulation means and how it has been studied in several research fields. We describe the methodology we have set up to reach our goal: creation of sign corpora, annotations of these corpora, analysis of these annotations, and finally design of a coarticulation model. The evaluation process will be made by different kind of people in various pieces of software.

Introduction
Sign Languages (SLs) are visual-gestural languages, made of signs performed by a person. Signs are constituted of two main types of features: manual (hand movement, configuration, spatial location) and non-manual (eye gaze, facial expression, movement of the torso, of the head, etc.). SLs are natural languages, thus there are linguistic rules (syntactical, grammatical) that have to be taken into [1].

SL generation consists of using 3D avatars to perform SL via several media (television, mobile phone, WWW, etc.). These 3D avatars expressing SL utterances are called “signing avatars”. Animations are created using motion capture, animation description language, or more recently, rotoscopy technique. This technique consists of animating an avatar by recopying key frames from a video. It allows designing realistic 3D animation including manual and non-manual features. SL generation with rotoscopy makes it possible to create complete utterances or parameterized utterances (where signs can be replaced). These utterances can be for example real time information in public transport. As the context of our study is to provide information in French Sign Language (LSF) to deaf travelers in railway station (delay of a train, warnings, and all the daily information hearing persons have) we use rotoscopy. It is co-directed by the LIMSI-CNRS laboratory (linked to the public ministry of research) and the Websourd Company1 for a project funded by the national railway society (SNCF).

We plan to achieve our goal by using pre-build animations and to concatenate them as parameterized utterances. Because we want to provide a real and natural language, a simple concatenation is not enough. Transitions between signs must be taken into account. Signs are slightly different when one performs them isolated or in utterances: the beginning of a sign is modified depending on the previous sign, and the end of the same sign is modified depending on the following sign. This phenomenon is called “coarticulation” and is the main issue addressed in this paper.

1 http://www.websourd.org
In section 2 we present a state-of-the-art related to coarticulation from several field of research, then we propose our methodology in section 3. Section 4 describes what has been done and what remains to do, and section 5 explains how our contribution allows deaf people to get informed.

Coarticulation
We have to study how signers perform this coarticulation, and how we can model it in order to use it in SL generation software.

Definition
We have presented a state-of-the-art of how “coarticulation” is defined in several research fields (spoken and signed languages both in recognition and in synthesis / generation) in [2]. Manual and non-manual features of SLs are used at the same time by the signer, but are not always correlated. That means that coarticulation must take into account that some times both manual and non-manual features has to be modified at the same time (but not the same way), but other times it has to be done separately [3]. Signs are slightly different when they are performed isolated or in utterances. The beginning and the end of a sign S, performed in utterance in comparison as when it is performed isolated, is modified respectively depending on the previous sign, and the following sign. Modifications, either it is addition or deletion of features, can even occur within signs. We call this modifications process: “coarticulation”. We have established that there are various definitions of the phenomenon that is why we have decided to keep a generic definition, to be able to refine it later.

Related Work
Coarticulation for SLs is a recent field of research in SL generation and recognition. Several project in the past ten years aim to produce a signing avatar (Figure 1): ViSiCAST [4], eSign [5], etc. and more recently in Huenerfauth’s dissertation [6].

![Figure 1: Signing avatars: Visia by ViSiCAST (a), vGuido by eSign (b), Huenerfauth’s (c)](image)

These projects focus on various goals, but they all implement a simple concatenation of signs, with various interpolation techniques. Some projects related to SL recognition focus on segmentation: [7] use the bi and triphones approach in modeling Chinese Sign Language with HMMs; [8] model each transition individually as a phoneme using parallel HMMs: rule based methods have also been used: these methods use heuristics such as acceleration and velocity to detect borders and transitions [9]. Even if there are some progress in what coarticulation is and how to separate signs from “transition” between signs, there is no existing reliable model of coarticulation either in generation or recognition of SL. Furthermore these models don’t allow spatial representation of signs, which is part of the SL grammar. In vocal coarticulation research field there are some interesting studies, for instance about the influence of phonemes (up to six before and after [10]). Co-verbal coarticulation [11] has also to be taken into account because of the gestures made while
talking. It is not the same kind of gestures as we have in SL, but we can notice that sometimes co-verbal gestures are not correlated with voice along time, and we have the same phenomenon in SL between manual and non-manual features. Here we are interested in media (voice/gesture vs. manual/non-manual) comparison, and not linguistic comparison.

Because those studies relies only on manual gestures, and because our definition for coarticulation implies others major aspects, we have to design a new methodology in order to take into account the others features that are essential in our point of view.

**Methodology: From Corpus Analysis to Synthesis**

We want to set up a coarticulation model to animate a signing avatar, to produce real Sign Language, and not “signed language”. It is obvious that deaf people need to get information in their language, not a mixed one that they won’t understand (or, worse, misunderstand). In order to reach our goal, we have decided a several step methodology detailed in [12].

Our model will be both a computer one (to animate a 3D avatar), and a linguistic one (to perform real SL). We will analyze coarticulation under the scope of Johnson’s Posture-Transition model (to represent signs structure [13]). The goal of this model is to “devise a system of phonetic notation (…) in order to account for variability in sign forms as they appear in real utterances”. Signs and utterances are constituted of segments defined as temporal alignments of articulator behaviors, and transitions between those temporal alignments. These alignments are constituted of several articulator features that represent hands configuration, locations, fingers flexions, torso movement, etc.

We have to analyze corpora to determine what a sign is and what is not, in order to fix what is due to coarticulation in sign modification. Following our methodology, we have recorded a first video corpus. We have chosen a group of deaf people who work on LSF linguistics in relation with a linguistic research group, to be our LSF linguistic experts. We have chosen a deaf person who works as translator in the WebSourd Company, to be our LSF expert to be filmed. We have begun with a corpus that do not contain all signs needed for the application, because we first want to completely evaluate all the steps of the methodology, before running it on all the signs. This corpus has been designed with two aims, implying specific technical and linguistic conditions. On one hand, we want to study coarticulation, and on another hand we want to build 3D animations using rotoscopy, in order to animate a signing avatar expressing SL with parameterized utterances. To build the avatar by rotoscopy technique we need a video model of each isolated sign, and to study the coarticulation phenomenon we need both signs isolated and in the context of an utterance. Thus our corpus is constituted of isolated signs and utterances. We have used two cameras, one in front, and the other on the side of the LSF expert. We have recorded both isolated signs (Figure 3a), and, complete and varying utterances including these signs (Figure 3b).

In our examples, we can notice that there are slightly different facial expression and location of the right hand (relatively to the head location), depending on whether the sign is performed isolated or within an utterance. For instance, we have recorded the following utterance: “Platform 1, stay away from the edge of the platform. Beware of a passing train through the station. Platform 1, stay away from the edge of the platform.” (as shown in
Figure 3b), and the isolated signs and utterances “Platform 1” (as shown in Figure 3a), “stay away from the edge of the platform”, “Beware of a passing train through the station”. We also have recorded the same utterance with modifications in the reason why people have to stay away from the edge of the platform (because the train is going to move, for instance), and in the platform identification (either a number or a letter). We then compare utterances between themselves and with isolated signs in order to highlight modifications.

By using annotation software (Anvil [14]) we have begun to temporally segment several manual and non-manual features, and to label those segments. For our study, we need to segment these features in order to fix which modifications of signs are due to coarticulation and which are not. At the time we are writing this paper, we are annotating our corpus (Figure 4).

Using Anvil, we first annotate isolated signs, then utterances, and finally display all these annotations simultaneously with a composition of the corresponding videos. The video of the utterance (A) is on the left side, and alternatively on the right side at the top (B) and at the bottom (C) we have the display of the several isolated signs used in the utterance. This video composition and the colored annotations allow us to create hypothesis based on relevant visual information, about how signs are modified in utterance in comparison as when they are performed isolated. In our example, we have displayed all the annotations we have made, both on the complete utterance A, and on the utterances B and E ("platform 1"), C and F ("stay away from the edge of the platform"), and D ("beware of a passing train through the station"). That means that at the same time the utterance is playing, we have the video display of the first isolated utterance B at the top right side, then the display of the utterance C at the bottom right side. They are not displayed at the same location because isolated utterances have a different duration as in complete utterances, and we exactly want to see what modifications occur within an isolated utterance and at the border of this utterance. Thus we provide an alternative display of each isolated utterance B, C, D, E, and F, at the top and the bottom right side.

Yet, we do not have annotated signs with the Johnson’s description model. We have instead chosen to describe in the whole what occurs, and then we will focus on some
phenomena and depict them with the model. For instance, we have annotated non-manual features like eye gaze (direction and opening) and mouthing, and manual features like hand configuration. These annotations provide us some clue about what to look after as modifications at the beginning and at the end of an isolated utterance or sign. We have to perform these annotations on the entire corpus to be able to propose first hypothesis about what modifications the coarticulation process implies.

Figure 4. Annotation (bottom) on the video corpus (top) with ANVIL

Then, statistical analyses (partially based on our visual hypothesis) will allow us to accurately compare isolated and in utterances signs. We will be able to determine what the modifications are and how they occur depending on the context (the signs before and after the focused one in the utterance).
Afterwards we will propose a first model of coarticulation, to be evaluated in the information display software in railway stations. The evaluation will be user-based, with questionnaires and interviews in LSF on linguistic, ergonomic and technical aspects. The panel will be composed of ICT and LSF deaf experts, and anonymous deaf users, with different experience in using LSF.

We will improve our proposal thanks to users’ feedbacks, and then create a second corpus to build a refined model based on all utterances and signs, to be evaluated again, following the same rules used for the first one. The evaluation process will be based on Huenerfauth’s protocol of evaluation used in [5], on some ideas from the ECAs’ (Embodied Conversational Agent) research field [14], and on vocal information system evaluation [15]. In our laboratory, we will display information in 3D with and without coarticulation, and ask for several comparisons: between 3D utterances, between video and 3D utterances, and between written and 3D utterances. In the railway station we will ask for free opinions. The questions will focus on understandability, relevance of information, etc.

We have already built some complete 3D utterances that have been performed by our signing avatar in a railway station (Figure 5). These realizations gave us feedbacks about the 3D creation methodology that is now improved.

![Figure 5: Signing avatar displayed in some French railway stations (expressing “Ladies, Gentlemen, we remind you that smoking is forbidden in the station” in LSF)](image)

**Work Done and Expected**

The study has begun in January 2008. We have spent the first year to establish a state-of-the-art: coarticulation in both vocal and sign languages, either in recognition or generation. We also looked for papers from co-verbal studies. We have set up our corpus creation methodology, and begun to create the video corpus. We have built some complete 3D utterance that has been signed by our signing avatar in a railway station. This realization gave us feedbacks about our 3D corpus creation methodology, which is now improved.

We are now annotating our first corpus and we should soon be able to propose a first coarticulation model for SL: we should be able to propose it by the end of the first semester of 2009. After evaluating this first attempt, we will improve it thanks to the users’ feedbacks. Meanwhile we will annotate the whole corpus, and propose an enhanced model based upon the corrected first one and the whole corpus annotations. We plan to propose a reliable coarticulation model for FSL by the end of 2010. This model will be used in the display information system in railway stations in France.
Contributions

The French law of February the 11th 2005 makes it a rule to provide accessibility for disabled people: all public buildings should inform deaf signing people in LSF. This is for the "official part" of our contribution, but in a more practical way, this study will allow deaf people to get pieces of information. Nowadays, when available information is displayed in written form; but 70% of French deaf people are illiterate so they can’t get it. Displaying SL and not signed language will also contribute to the recognition of SL as a real language and not a simple non-vocal code.

Our model will be used in railway station to display information at first, but we have in mind that it could be used for airport, underground, etc. to provide a global accessibility in LSF. Moreover, we could use it for different SL, as SLs differs mostly in lexicon but very slightly in syntax and grammar.

Conclusion

Our goal is to set up a coarticulation model of French sign language to be used in a display information system. That system will provide information by the mean of a signing avatar, in a real natural sign language thanks to our model. We have described the methodology and the evaluation process. We also have defined what coarticulation is. We will soon achieve all steps of our methodology and could evaluate our model to improve it. Our research has direct and concrete applications and corresponds to deaf people’s needs. It should have social and cultural impacts.

References


About the author:

Jérémie Segouat is currently a French PhD student in both the ILES team at the LIMSI-CNRS laboratory and the WebSourd Company. His research focus is in sign language synthesis to provide information in sign language to deaf people. He is designing a coarticulation model of French Sign Language to be used in railway stations to provide information in sign language, by the mean of a 3D virtual signer. He is working on his dissertation “Synthesis of sign language by the mean of a virtual signer: signs’ coarticulation modeling” with his advisor, Dr. Annelies Braffort. He is a graduate sign language interpreter.
Context-Enhanced Interaction Techniques for More Accessible Mobile Phones

Shaun K. Kane
The Information School, DUB Group, University of Washington
skane@u.washington.edu

Abstract
Modern mobile phones enable users to access a wide variety of information and communication services anytime and anywhere. In order for people with disabilities to benefit from these services, mobile devices must provide accessible user interfaces. Currently, users with motor and visual impairments often have difficulty using mobile devices due to issues such as screen readability and small controls. These problems may be exacerbated when attempting to use such devices in crowded, noisy or otherwise distracting environments. We are developing accessible mobile interaction techniques that adapt the mobile device interface to the user by leveraging information about the user’s current location, activity, and level of ability.

Introduction
While relatively uncommon even 10 years ago, today mobile phones are used daily by people of all ages, occupations and abilities. As mobile phones have become more ubiquitous, they have also gained features such as e-mail, web browsing, and global positioning system (GPS) capabilities, thereby providing access to additional information and services while on the go. This access is beneficial to all users, but may have special significance for some people with disabilities. For these users, mobile phones offer a constant connection to friends, family, and caregivers, allowing them to act independently in the world while remaining connected to essential support and services [1].

Despite the ubiquity of mobile phones, many commercially available phones remain inaccessible to people with disabilities. Users with visual or motor impairments may encounter difficulties using these devices due to issues such as small screens, small controls that are difficult to manipulate, and a lack of multimodal feedback [9]. Furthermore, because mobile phones may be used in a wide variety of contexts, such as when walking down the street or riding public transportation, environmental factors such as ambient noise, unsteady surfaces, and inclement weather may further impair a user’s ability to interact with the device [9].

Although mobile phone accessibility issues may be exacerbated by environmental factors, understanding the effects of these factors makes it possible to develop user interfaces that can adapt to these limiting situations, thereby improving overall accessibility. We are currently investigating mobile interaction techniques that use information about a user’s context, location, and activity to adapt to the user’s abilities and increase accessibility.

Related work
Modern mobile phones often feature on-device sensors such as accelerometers, light sensors, and GPS devices. Prior studies have explored using these sensors to detect a user’s context and change the user interface accordingly. Hinckley et al. [4] explored interaction
scenarios that used on-device sensors to augment a user interface, but focused on novel interactions rather than increasing accessibility. For example, their prototype rotated its on-screen display when the device itself was rotated, and began audio recording when the device was held near the user’s face. Dey and Mankoff [2] developed a prototype augmentative and alternative communication (AAC) device that used information about the user’s location to populate a word menu. This was intended to improve usability by populating the menu with likely candidates, but a pilot evaluation of the prototype did not show significant benefit over existing AAC systems. Our research extends this prior work by exploring new adaptive techniques that use contextual information to increase accessibility.

This research also extends our prior work in developing accessible and adaptive mobile user interfaces. Barrier Pointing [3] introduced an alternative target selection method that increased pointing accuracy on touch screen-based mobile devices for users with some types of motor impairments. Slide Rule [5] introduced gestural input and audio output techniques that enabled visually impaired people to use a touch screen-based mobile device. These interfaces improved accessibility for users, but did not adapt to the user’s context or level of ability. In another study, we developed a prototype walking user interface (WUI) that increased the size of touch screen targets while the user was in motion. Our current research extends this prior work to explore how accessible interaction techniques can be combined with context-driven interface adaptation to provide accessibility support when it is needed most.

**Situational factors in mobile accessibility**

While many mobile phones are difficult to use for people with disabilities, these people may encounter further accessibility issues when attempting to use a device in a busy environment or while moving. A mobile phone button that is difficult to press while sitting at a desk will likely be even more difficult to press when the user is on a moving train or is walking down the street. This negative effect on performance caused by environmental factors can be thought of as a situational impairment [8], since it effectively impairs the user’s ability to interact with the device. Situational impairments may affect mobile device users with and without disabilities. Situational impairments can be caused by a range of factors, including user movement, inclement weather, impeding clothing, or user fatigue [6]. These impairments affect user performance in various ways, including reducing reading speed [7] and target acquisition speed [6].

User interfaces that can adapt to the user’s context may reduce the effects of these situational impairments. These situational accommodations [6] may overcome the effects of situational impairments by adapting the user interface so that it is easier to use in a given situation. For example, a user interface might make text larger while the user is moving in order to make it easier to read, or might increase text contrast when the lighting conditions are poor. These accommodations may be triggered automatically when needed, thereby reducing the effort required to interact with the device in a distracting or impairing situation.

**Proposed solution**

We are developing a system of interaction techniques that leverage contextual information about the user to reduce accessibility issues during mobile use. These
techniques draw on multiple types of contextual information, including the user’s location, movement, and current activity. This information can be detected by sensors on the mobile phone itself. The system can also monitor the user’s current performance, and can take action if the user’s performance is significantly lower than normal. For example, the system might detect that the user is pressing the BACKSPACE key significantly more than usual and infer that the user is having difficulty typing. The user may also set explicit preferences for adapting the user interface. Once a potential accessibility issue is detected, the user interface can adapt to allow for more accessible interactions. The proposed system is illustrated in Figure 1.

The proposed system consists of three primary components. First, the system must determine the user’s context using sensors on the mobile device. Prior research has shown that it is possible to determine a user’s location [2] and movement [4] using on-device sensors. Data can be combined from multiple sensors to create a model of the user’s current activity and capabilities.

Second, the system must provide a set of interface adaptations that can address potential accessibility issues as they arise. These adaptations may take a number of forms, including: (1) rescaling or reordering a user interface to accommodate interaction difficulties, (2) automatically performing contextually appropriate actions to reduce required effort, (3) activating an alternative interface that is optimized for a given context, or (4) activating typing or pointing correction to reduce errors. Examples of some typical mobile use situations, potential situational impairments, and adaptations to address these impairments are listed in Table 1.

Finally, because each user may have different needs, the system must allow users to customize accessibility features. Users should be able to choose which accessibility features

Figure 1. Accessible mobile interaction techniques leverage contextual information about the user, such as their location, movement, and error rate, to adapt an interface to the user’s current needs and abilities.
they want to use, and determine when they will be activated. Users may also further customize the adaptations, for example by choosing minimum and maximum sizes for on-screen targets. These customization features must be easy to use and accessible to people with a range of abilities.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Impairing effects</th>
<th>Interface adaptations</th>
</tr>
</thead>
<tbody>
<tr>
<td>User is in motion</td>
<td>Reading ability is reduced</td>
<td>Text size increased; text-to-speech activated</td>
</tr>
<tr>
<td>User is in a busy bus station</td>
<td>Attention is reduced; crowded space impairs movement</td>
<td>Bus schedule application automatically launched</td>
</tr>
<tr>
<td>User is riding on a bumpy bus</td>
<td>On-screen targets are difficult to hit while moving</td>
<td>On-screen target size increased; keyboard error correction activated</td>
</tr>
<tr>
<td>Phone is in pocket</td>
<td>User cannot see the screen</td>
<td>Voice output activated; touch screen gestures enabled</td>
</tr>
</tbody>
</table>

Table 1. Mobile usage situations, potential impairing effects, and interface adaptations that address these effects.

**Future work**

We are currently conducting a qualitative investigation of mobile device accessibility, with an emphasis on accessibility issues that occur while moving around in the world. We will interview people with disabilities about the mobile devices that they use and about the accessibility issues that they encounter when using these devices. This investigation will reveal what accessibility problems are experienced by mobile device users with disabilities, how environmental factors can affect these problems, and how users adapt to these problems.

Following this investigation, we will develop an initial set of accessible, adaptive interaction techniques for mobile phones. These techniques will be refined through participatory design and iterative prototyping involving users with visual and motor impairments.

Finally, we will conduct a field study to evaluate the effects of these interaction techniques on mobile phone usability and accessibility. When possible, we will deploy prototypes to the mobile devices that the participants are already using, rather than requiring them to adopt new devices for the experiment.

The expected contributions of this research include further understanding about mobile accessibility issues in the field, a set of new accessible mobile interaction techniques, and a framework for integrating and managing these interaction techniques. These contributions may significantly increase the accessibility of current mobile phones, and will help ensure that future mobile devices are accessible to users with disabilities.

**References**


About the author:
Shaun K. Kane is a Ph.D. candidate at The Information School at the University of Washington. He is a member of the Accessibility, Interaction, and Mobility (AIM) Research Group and is advised by Jacob O. Wobbrock. He previously received a B.S. and M.S. in Computer Science from the University of Massachusetts. Shaun’s research focuses on developing accessible mobile devices for people with diverse abilities.
ICT-related skills and needs of blind and visually impaired people

Carolina van Puffelen

University of Twente, Dpt. of Technical & Professional Communication, The Netherlands
m.c.j.vanpuffelen@utwente.nl

Abstract:
This study focuses on the relationship between the ICT-related training offered to blind and visually impaired people and their actual, self-reported and demonstrated, competencies for online activities and information processing. The findings of the study can shed light on how people with severe visual disabilities are prepared to access the web for educational, institutional and social participation. The study also gives insight in the validity of instruments to measure ICT-linked skills for the target group and creates an empirical foundation for improvements of ICT-related training. The first phase of the study investigated how blind and visually impaired people perceive their participation in society through ICT. An extensive interview showed how this audience perceives the frequency and quality of their Internet use (or absence thereof) and how they acquired these skills.

Introduction
In the Netherlands blind or visually impaired people can obtain a formal ICT training from one of the three Dutch support organizations for people with a visual impairment. The trainings usually consist of tailor-made instructions fitted to the individual supporting that person in participating in society through ICT-use. The organizations for which these trainings are an integral part of their rehabilitation support are interested in an assessment of the effects. They want to know whether they are of good quality and whether they fit the needs of their audience which consists of young and elderly people with a severe visual disability.

Related Work
The study is related to other studies on access and ICT-skills conducted in the Netherlands [1, 2]. These studies examine ICT-linked participation and skills assessments for Dutch people without a specific visual disability. In this sense the study forms part of a larger effort to chart the Dutch population on these aspects.

Motivation
There is a considerable body of research on visually impaired people that concentrates on access technologies and web designs. This study starts ‘from the other side’ by looking mainly into the question what this audience needs to make good use of these facilities. We depart from a situation in which a visually disabled person has a computer with supportive devices and online access only to discover that he or she has or lacks adequate skills to navigate the Web and process online information. With this study we hope to contribute to an integrated view on accessibility from which we can fruitfully reduce the barriers that people with a visual disability face everyday.
Background to the study
In many studies it has been demonstrated that ICT is not only reducing barriers of access, but at the same time creating barriers [3, 4, 5, 6]. People with visual disabilities face special barriers in using the Internet, aside from those related to material access and computer-related trainings. Dobransky and Hargittai [7] mention technical accessibility problems as one of the extra barriers that people with a visual disability need to tackle. Bayer and Pappas [8] found that especially navigation and screen reading posed problems for blind internet users.

For a large majority of people with visual disabilities ICT-related training forms an integral part of their rehabilitation effort. The purpose of the present study is to investigate whether these ICT trainings empower them to overcome the mental and technical ICT-barriers and to participate online in educational, social and institutional activities.

One of the specific issues addressed in this research is the measurement of ICT-skills. Many studies on digital literacy measure the skills involved with self-report measures such as questionnaires. Hargittai [9, 10] has shown that there can be a severe mismatch between people’s perceived and actual, demonstrated ICT-skills leading to over-estimates as well as under-estimates. The present study includes self-reports as well as observations of actual behaviour. Thus, it seeks to contribute to developing valid quick-scan assessment of people’s ICT competence. The findings should also form a foundation for further improvement of the ICT-related training for young and elderly people with visual disabilities.

Research questions
The target groups in our study are young (10-14 years) and elderly (55+) people with a vision of <0.3. Young visually disabled need ICT skills for communication as well as for entering the labor market and for expanding their educational opportunities. For elderly people with visual disabilities ICT-skills are important to get and stay in touch with peer groups, with government institutions, and with services related to their disabilities. Using Van Dijk’s [1] framework for analyzing digital inequality, the study focuses on: (1) Educational participation, (2) Social participation, and (3) Institutional participation. In these areas of participation we concentrate on the following activities that our audience is likely to engage in, or might want to do so: emailing and chatting, information seeking and conducting transactions online. We seek to find an answer to the following research questions:

- How do people with a visual disability assess their skill levels in educational, social and institutional areas of participation?
- Which audience characteristics are related to the skill levels?
- How have people with a visual disability acquired their skills?
- Which types of ICT-enabled tasks do visually disabled people engage in, or aspire to engage in?
- Which barriers do visually disabled people encounter when performing key ICT-supported tasks in the three areas of participation?
Research instrument

The key skills that we examine in this study are: (a) synchronous and a-synchronous communication skills (use of e-mail and chat), (b) Information search and retrieval in websites (use of static web pages), and (c) Conducting transactions in websites (use of forms in dynamic web pages and web services). A list of key skills components in each area of participation is developed. This list forms the basis for conducting a telephone interview with people with visual impairments. The pre-structured skills interview designed for this study is based on the types of access and digital skills described by van Dijk [1].

He distinguished four successive kinds of access to computers and internet connections: motivational access, material or physical access, skills access and usage access.

![Figure 1. ICT skills access questions](image)

Respondents

Participants are recruited through the three national support organizations. The inclusion criteria are: a vision of < 0.3 and having received ICT training from these organizations in the past three years. All participant information will be treated in accordance with the professional code of conduct for social researchers. Data have been gathered through telephone interviews with 142 respondents. These data have been analyzed to answer research questions 1 and 2.

Preliminary results

We analyzed data collected from 142 telephone interviews with 69 elderly respondents whose ages ranged from 55 to 87 and 73 young respondents with ages between 10 and 18 years [11]. We found remarkable differences between age groups in use, skills confidence and skills acquisition methods. The WWW is used by 97.3% of the young people against 71% of the elderly and the younger respondents perceived their skills, except for the use of email, higher than the elderly. The majority of the elderly respondents (71.4%) learned to go online and using email from receiving computer training whereas 80.3% of the younger people learned these internet activities themselves.
Ongoing and future work
At the moment we are completing the second phase of the study in which course material was analyzed, training sessions observed and trainers and trainees interviewed.

In the next phase an inventory will be conducted of tasks that participants need to master if they want to engage in educational, social and institutional areas of participation through ICT.

From the interviews and observations a list of specific ICT tasks will be developed based on the framework of Van Deursen and Van Dijk [2]. They distinguish four types of digital skills:

Operational skills
- Operate an Internet browser
- Operate an online search engines
- Complete an online form

Formal skills
- Navigate on the Internet by recognising and using hyperlinks
- Maintaining a sense of location while navigating

Information skills
- Locate, select and process information
- Evaluate the information source

Strategic skills
- Taking advantage of the Internet by goal oriented action and making decisions to gain personal benefits

A selection of respondents from the first phase will perform these tasks on their computer at home in presence of the researcher. Data resulting from the last phase can be linked with the findings from earlier phases. The results from this study will offer an insight in the way visually impaired people use their digital skills to participate in society. Furthermore it will form a foundation to help the support organizations improving and expanding the ICT-related training they offer the young and elderly people with visual disabilities.

References

About the author:

Carolina van Puffelen graduated with a Bachelor of Science in Education, Design, Multimedia and Management from the University of Twente in 2004. She was awarded a Master in Educational Science and Technology from the same university in 2005. She started her research project in 2007 at the Department of Technical and Professional Communication which specializes in web accessibility. Her research interests are in the area of human and computer interaction and inclusive technology.
Generation of Accessible Diagrams by Semantics Preserving Adaption

Cagatay Goncu
Clayton School of IT, Monash University, Australia
cagatay.goncu@infotech.monash.edu.au

Introduction

Diagrams are important components of educational materials. They can be found in various media. Printed media, such as text books, have traditionally been the most commonly used media for educational material, however online electronic resources, such as web pages and e-books, are becoming increasingly popular sources of educational content. Online materials include static diagrams, like text books, but may also include dynamic content such as animations and components that can interact with the user.

Unfortunately, these educational diagrams are not easily accessible to blind people. Currently, an accessible version must be generated from print-based material, and requires a human to interpret the original visual diagrams and generate the accessible version, most commonly in tactile and/or audio formats.

In Australia, Vision Australia is the main organisation that generates accessible versions of diagrams. People, often teachers and students, submit their interest in a diagram and send either the diagram or a reference to it to the Vision Australia Transcribers Department. This request is then assigned to a transcriber who analyses the diagram and creates a tactile version via drawing tools and using specific guidelines [1] for generating effective tactile diagrams. The tactile diagram is then printed using an embosser and checked by proofreaders who are blind and experienced in reading tactile diagrams. If there is a need, the diagram is updated by the transcribers again. This proofreading and updating process continues until the tactile diagram is finalised.

Generation of accessible diagrams using this kind of process is not scalable. It requires extensive effort and resources and will not scale up to handle the vast number of diagrams that are, and in the future will be, available in on-line educational materials. Thus, there is a need to automate this transcription process. There have been other attempts to automatically generate accessible diagrams, but they are not generic approaches that can be applied to different diagram types. Instead, they are specific to certain types of diagrams, interaction styles, and devices.

This research project “Generation of Accessible Diagrams by Semantics Preserving Adaptation” aims to provide a generic approach for automatic generation of accessible diagrams. This report addresses the motivation and the proposed approach of the project. It is organised as follows: In the next section, the research background is given. In the third section, the current approaches to provide accessible diagrams are summarised. In the fourth section, our proposed approach and the research questions to be answered are presented. The current status and the timeline of the project are also given in this section. Finally, in the last section, future implications of the project are given.
Background

There are various factors that should be considered when generating accessible diagrams. A basic understanding of visual diagram representation and comprehension is necessary to determine how to present an accessible diagram. It is also crucial to understand the human sensory system, especially the relevant subsystems: visual, aural and haptic, to determine the most appropriate ways of presenting information in non-visual modalities. We must also understand the context that the diagram will be used, such as the task and the viewer’s preferences and skills.

Diagram Representation and Comprehension

Diagrams are used to visualise, understand, manipulate and reason about entities and relations among them [40]. They vary enormously [21], and can be categorised based on representation related (1-5) and context-related (6-8) aspects [5] which are:

1. Basic graphic vocabulary: Primitive graphic elements (points, lines, areas) and their properties (size, colour, texture)
2. How they use of pictorial abstractions: Depictions of physical objects in realistic, schematic, and abstract forms, such as a river on a map.
3. Graphic structure: Configurations that determines the organisation of diagram components.
4. Mode of correspondence: Relationship between a representation and its meaning, such as a wine glass may represent a bar, or a fragile item.
5. Represented information: The domain where the diagram is used, and what it represents such as definition of a space, or a time frame.
6. Tasks and interaction activities performed on a diagram
7. Cognitive processes: The issues related to diagram users, such as perception, interpretation and reasoning.
8. Social context: Cultural conventions used to represent and interpret diagrams.

Diagrams are useful for a variety of reason. They extend working memory and the properties of the visual system can be exploited to allow faster inference and search than is possible with an equivalent textual representation [33]. Furthermore, diagrams can store large amount of information in a small space [47] which can make them an extremely effective tool to convey information.

Like natural language, diagram descriptions have syntax, semantics and pragmatics. The syntax of a diagram is given by its graphic elements such as points, lines, rectangles, circles ...etc, and the rules for positioning. They also have visual properties associated to them such as size, colour, and texture ...etc. Thus, graphic elements, their visual properties and the placement of them on the layout form the syntax of a diagram [20].
Semantics is another aspect of a diagram. It is the abstract meaning of a diagram that includes the entities, their types and properties, and relations among them [20]. Unlike the sentences of a sentential language construction is done by putting a syntactic element next to the previous one, diagrams are constructed by placing the graphic elements on a layout with different relations such as left-of, touches, and inside. This construction affects the potential diagrams that can be generated by that syntax and includes semantic information. In the classical example of Euler’s circles where the premise “All A are B” is represented by the circle A inside the circle B, the syntactic production of putting a circle inside another one has a strong relation between syntax and semantics, which makes it an effective diagram. Thus, syntax and semantics of a diagram are not separated clearly. In a well-formed diagram, semantics can be extracted from the syntactic structure. However, there may be many different syntactic structure with the same semantics, but not all of them are equal.

Pragmatic issues guide the choice of appropriate syntax among these different structures. They cover the interpretation of a diagram by a human, and contribute to the effectiveness of the diagram. The relative size of graphic elements, their similarities to each other, positions on the layout, and emphasis by using visual properties such as colour, and texture that are chosen by a graphic designer are examples of pragmatics. For instance, in a network diagram where each node represents a computer, router, or a hub, and each edge represents a link between the nodes, the placement of the nodes is determined by the common conventions based on experience. An arbitrary placement of nodes will not change the semantics, in this case which node is connected to which one, but leads to an incomprehensible diagram with lots of intersecting edges and overlapping nodes. Hence, both syntactic and semantic components of a diagram are related to pragmatic issues [20]. Semantics determine the entities and relations among them, and pragmatics determine the most effective way of presenting this semantic with syntactic elements. Since pragmatics adds extra information, such as emphasising the importance of some aspects of the diagram, to the visual presentations they effectively extend the core semantics of a diagram.

**Human Sensory System**

The human sensory system has different subsystems which have specialised sensors for receiving and interpreting information from various resources. Among these subsystems, visual, aural, and haptic subsystems are the most relevant to our topic.

The visual subsystem has sensors that receive light and provide visual information such as shape, size, colour, intensity and position [12]. It has a wide area of perception that provides parallel information acquisition which is also a continuous flow [22]. Although, it perceives information from a wide area, it also has a narrow focus of attention frame which can provide detailed information. Moreover, it does not need to have a physical contact with the objects to acquire this information.

The haptic subsystem is specialised to process tactual and kinesthetic stimulus. It has sensors that receive stimulus about touch, temperature, and motion [12]. The cutaneous actuators are on the skin and sense the touch, and temperature stimulus. The kinesthetic actuators are on the muscles and joints of the body and sense the motion stimulus [18]. The haptic subsystem can provide most of the same information as the visual subsystem such as shape, size, texture, and position [22]. However this functionality is not the primary
function of this subsystem and it is not as fast and as effective as the visual system [27]. It can provide more accurate information about substance properties such as texture, and hardness [34]. The haptic subsystem needs to be in direct contact with the objects to acquire stimulus. The extent of the perceptual field is less than the visual field and limited to the hands. Moreover, since the perception of an object requires movement of the fingers and hands, information acquisition is considered less parallel than vision [22].

The aural subsystem has sensors that receive aural information such as audio, and speech [12]. It is more effective in acquiring sequential stimulus [22]. Since aural subsystem provides binaural hearing it can also locate the source of an stimulus [18]. The aural subsystem does not need to have a physical contact with the objects to acquire this information.

Among the visual, aural and tactual sensors, the visual sensors have the most capacity to carry information. Visual sensors carry four times more information than the tactual sensors, and two times more than the aural sensors. The bandwidth limits for the visual, tactual and aural senses are, and bits/s, respectively [28].

For blind people, audio and haptic stimuli are the main sources of information. To present diagrams in these modalities, the information in a diagram should be converted and presented in a way that aural and tactual sensors can receive and interpret effectively.

**Tasks on Diagrams**

A person can perform different tasks on a visual representation of a diagram. Based on a task analysis developed for information visualisation, these are [10]:

- Overview: Recognising the type of the diagram.
- Zoom: Getting a closer view of elements in an area.
- Filter: Removing the irrelevant elements from the focus.
- Details-on-demand: Getting information about a particular element.
- Browse: Moving around a diagram.
- Search: Searching for a particular elements and/or information.

Interaction with a diagram starts with getting the overview of the whole diagram. This overview provides a skeleton of the diagram structure and gives an idea whether the diagram is relevant or not. The browse task is used to get information about the diagram elements [4]. During browsing, the focus frame [11, 51] of the diagram is changed by zooming in and zooming out. Moreover, users can search for the elements in the diagram. Filtering can be applied to the diagrams to focus on the relevant information. This information can then be queried.

Each of these tasks should also be performed on accessible diagrams by a blind user to get the necessary information. Thus, this study will investigate methods of performing these tasks on accessible diagrams in an effective way.

**User Profiles**

World Health Organisation classified vision related problems into two major categories which are visual disturbances and “blindness and low vision”. Visual disturbances include disabilities such as day blindness, night blindness, colour blindness, and temporary vision.
loss. Blindness and low vision are categorised into six visual disability categories according to their “best-corrected visual acuity”2 values [48].

In this study, solutions for people who have no sense of size, form and contour will be investigated. The profile of these people will be used to determine the appropriate devices and presentation styles. Since the capabilities of the human sensory system and its subsystems are different for each person, the user profiles will be used to characterise the capabilities of blind people in terms of their tactual and aural sensitivities.

Tactile literacy is another important point to be considered. It includes both Braille alphabet and tactile diagram reading skills. The Braille alphabet has special characters encoded by six-dot matrices that are used to represent letters and symbols. The words can be constructed from these characters in two different ways. In Grade-1 Braille, the letters of a word are converted one-to-one to their Braille equivalents. Because of the size of these words, Grade-2 Braille uses a set of contractions to make the words shorter such as contracting commonly used words (the, of, ..etc), and word endings (-tion, -ing, ...etc).

Tactile diagram reading requires training and/or experience to distinguish the lines, textures and texts in a tactile diagram. Moreover, it requires graphicacy skills which are the same set of skills required for sighted user to understand diagrams. These include view, format and commentary connection in diagrams [3]. Views provide different perspectives of the same object such as cross-sections, magnified and distant views. Formats define the context of the elements, and the conventions about them. For instance, in a bar chart it is important to understand the conventions that the axis defines the independent and dependent values, and the bars represent the relations between these values. Commentary is another prior knowledge that should be learnt. The elements provide commentary about other elements that make up the graphics should be interpreted separately. For instance, the arrows used in a pulley system to indicate the movement, and the symbols associated to a pulley are the commentary objects of that diagram.

The duration of blindness should also be considered as a factor in the user profile. It is expected that late-blind people can have some advantages over early-blind people due to the previous visual exposure. A study with early- and late-blind people shows that late-blind users perform better in tactile picture identification [23].

**Assistive Devices**

Assistive technologies refer to both hardware and software systems that provide assistance to users with disabilities for maintaining and improving their capabilities. Assistive devices for visual disabilities can be grouped into four main categories based on the modalities of interaction they use: audio, tactile, kinesthetic, and hybrid.

**Audio**

The most widely used audio form is synthesised speech. Screen readers use this form of audio to provide textual information from a screen, a document and a web page. JAWS3 and HAL4 are well-known screen-readers which produce synthesised speech.

---

2 Visual acuity is sensing size, form and contour with one or two eyes for near and distant objects

3 www.freedomscientific.com/

4 http://www.dolphinuk.co.uk/products/hal.htm
There are also devices that uses OCR (optical character recognition) technology to extract the text from an image source and read it in speech. Kurzweil 1000⁵ and VERA⁶ are the products extend screen readers by using OCR technology to read text contained in a printed image.

Sonification is another technique to present information via more complex non-speech audio. It includes encoding the information in terms of audio elements such as tone, pitch, timber and loudness. It can be in 1D, 2D, and 3D space:

- 1D sonification: Only one audio signal is used.
- 2D sonification: Stereo audio signals are used.
- 3D sonification: Ambisonic⁷ systems [41] are used.

Tactile
There are many different displays that facilitates tactile interaction. These displays can be grouped into two main categories: (i) non-refreshable, and (ii) refreshable [50].

Tactile paper⁸ (Figure 1) is a non-refreshable medium which is created by embossers, such as Tiger Braille Embosser⁹.

![Tactile embossed paper](http://www.quantech.com.au/piaf)

Embossers punch the paper with varying height dots to create raised shapes. They are also used to print Braille text. Thermo-form (swell) paper¹⁰ (Figure 2) is another non-refreshable medium that includes thermo capsules that rises based on the applied heat. The more the heat is applied, the higher the capsules rise. These swell papers are created by heat lamps and tactile image makers, such as Piaf¹¹. With heat-lambs, users trace on the swell paper to create the tactile graphics. The height of the raised graphics on a particular point depends on the heat lamp’s duration of stay on that particular point. As for the tactile image

---

5  http://www.kurzweiledu.com
6  http://www.nanopac.com/Reading era.htm
7  A three-dimensional sound production system that models the direction and the distance of audio signals
9  http://www.viewplus.com/products/braille-embossers
10 http://www.nctd.org.uk/images/Brain.jpg
makers, they use images printed on a swell paper and apply heat to produce tactile images. In these devices, the darkness of the ink on the paper determines the height of the raised shapes [52], the darker the lines the higher the raised lines. The second category which includes refreshable displays can be grouped into two: (i) static displays, and (ii) dynamic displays. Static displays are fixed sized devices which have a fixed number of actuators. When the display is activated, the user traces the area to figure out what is on the display. On the other hand, in dynamic displays there is a single line of actuators which dynamically change in time and trace the area automatically.

Figure 2: Swell paper

Apart from these devices, there are also new and upcoming technologies such as electronic dynamic surfaces,12 and band-aid-size tactile displays.13 Electronic dynamic surfaces will have electro-active polymer films that can raise and form tactile graphics by electric signals. As for band-aid-size tactile displays, they also use electro-active polymers that are organised in a grid structure and put onto a flexible tactile display that can be wrapped around any part of the body. In the near future these technologies will hopefully provide the capabilities to present not only static content but also dynamic content, such as animations.

**Kinesthetic**

Kinesthetic devices (force-feedback devices) use the sense of movement. Interaction with these devices is achieved through a pointing device such as a joystick or a stylus. The users use the pointing device to navigate in the diagram. When the user passes over an element of the diagram the device gives either a force-based feedback or an audio feedback depending on the design of the device. These devices have the capability to provide the sense of movement in two and three dimensional space.

PHANTOM14 [Personal Haptic Interface Mechanism] [37] is a well-known example. The user interacts with this tool in a 3D space via a stylus whose movements are tracked by the system. Feedback to the user is given by the reflected forces to the stylus (Figure 3).

---

12 http://www.jhu.edu/~news_info/news/home07/nov07/tactile.html
14 http://www.sensable.com/documents/images/LargePHANTOMOmniImage.jpg
The main disadvantage of these devices is that the focus area is limited to a small portion of the whole diagram layout. Thus, the user cannot get an overview of a diagram when presented by these devices.

![Phantom](Image)

**Figure 3: Phantom**

**Hybrid**

Assistive devices can also use more than one interaction style to improve the effectiveness of the interaction. Using tactile and audio together is the most common approach. An example to this is the TTT (Talking Tactile Tablet) [32] which uses a print tactile graphic on top of a touchpad that can track the position where the user touches the tactile graphic\(^{15}\) (Figure 4). Based on the position the device can give both speech and non-speech audio feedback.

![Tactile and audio device](Image)

**Figure 4: Tactile and audio device**

Another interaction style combination is kinesthetic and audio. SoundView [49] is an example that utilises this type of interaction. It uses a moving pointer device to trace a colour image. Non-speech audio feedback is produced based on the colour on the surface and the trace speed.

\(^{15}\) [http://www.mcb.ac.uk/t3/assets/images/overlay_9.jpg](http://www.mcb.ac.uk/t3/assets/images/overlay_9.jpg)
Current Approaches

There have been many approaches to the generation of accessible diagrams. These can be grouped into four main categories:

- Manual approaches
- Approaches using standard diagram input formats
- Approaches using non-standard diagram input formats
- Semantic based approaches

The following sections summarise the approaches in these categories, and their limitations. The properties of each approach in these categories are classified by using the following dimensions:

- Input format: The diagram format used as an input for the tool.
- Diagram type: Types of diagrams.
- Task: Type of tasks intended to be supported.
- Presentation: Modality of the presentation.

Manual Approaches

These approaches require a human designer to generate the accessible diagram basically from scratch. Table 1 summarises these approaches.

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Task</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDraw [30] Free hand drawing</td>
<td>Diagrams with lines and polygons</td>
<td>Author diagram</td>
</tr>
<tr>
<td>CorelDraw, Inkscape Free hand drawing</td>
<td>Diagrams with lines and polygons</td>
<td>Author diagram</td>
</tr>
</tbody>
</table>

Table 1: Non-automatic approaches for accessible diagram generation

TDraw [30] is an example of diagram authoring tool which is used by a blind person. It uses swell-paper, thermo pen, and a digitiser are used as system components. The swell-paper is put on a digitiser to get the position and the state (pressed/released) of the thermo pen. When the thermo pen traces on the swell paper, the thermo capsules raises because of the heat produced by the thermo pen. The height of the raise depends on the duration of the stay. In the meantime, the digitiser stores the position and the state of the pen to reproduce the diagram visually or tactually.

CorelDraw16 and Inkscape17 are example of drawing tools that are used to generate accessible diagrams from scratch by transcribers such as those at Vision Australia. Designers use these drawing tools to create an accessible diagram from the original visual representation.

The main disadvantage of these approaches is that they are time consuming and cannot scale to handle the rapidly increasing material available on electric environments such as e-books, and web pages.

---

16 www.corel.com
17 www.inkscape.org
Approaches Using Standard Diagram Input Formats

These approaches use standard input formats such as images, and SVG. They use image processing or syntactic pattern recognition algorithms to recognise important features of a diagram such as labels, fill colours and shape outlines and replacing them by Braille text or speech audio, patterns and lines with varying weights, respectively. They are generic and can be applied to any kind of diagram. Table 2 summarises these approaches.

<table>
<thead>
<tr>
<th></th>
<th>Input Type</th>
<th>Task</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>iReader</td>
<td>Image, Diagram w/ text on print media</td>
<td>Read text</td>
<td>Speech Audio</td>
</tr>
<tr>
<td>vOICe</td>
<td>Image, Diagram on print media</td>
<td>Browse diagram</td>
<td>Non-speech Audio</td>
</tr>
<tr>
<td>Optacon</td>
<td>Image, Any diagram</td>
<td>Provide diagram content</td>
<td>Tactile on refreshable display</td>
</tr>
<tr>
<td>AudioGraph</td>
<td>Image, Diagrams w/ components</td>
<td>Provide diagram content</td>
<td>Non-speech Audio</td>
</tr>
<tr>
<td>TGA</td>
<td>Image, Any diagram</td>
<td>Provide diagram content</td>
<td>Tactile</td>
</tr>
<tr>
<td>Krufka</td>
<td>SVG, Any diagram</td>
<td>Provide diagram content</td>
<td>Tactile</td>
</tr>
</tbody>
</table>

Table 2: Generic approaches for accessible diagram generation

iReader [42] is one of the tools using optical character recognition to extract text from the captured images of a book.

The vOICe\(^{18}\) system uses a line scanner which traces the image horizontally at a constant rate. At each step of the trace, it scans the y-axis from beginning to the end indicating this by a increasing frequency of the audio. It changes the loudness of the audio depending on the pixel value of the image at the scanned coordinate.

Optacon [7] system has a photo-transistor array, and a small array of vibrating pins. The photo-transistor array is used to scan the visual material and recognise the marks on the material. These marks are then converted to vibrations based on the densities of the marks.

AudioGraph [29] uses a slightly different approach. It defines a grid system with horizontal, and vertical coordinates and associate different instrument sounds to these axis. The mapping between the horizontal and vertical positions and the audio signals is done by associating a lower pitch to lower coordinate values, and higher pitch value to higher coordinate values. The diagram is then converted to a bitmap image and placed on this grid structure. Later, the components of the diagram are identified and presented to the user by generating audio signals based on the positions on the grid. This presentation is done component by component, so the user can build a mental model of the diagram on the grid.

TGA [31] uses image processing algorithms to generate tactile diagrams. This consists of image segmentation and image simplification. In image segmentation, the text in the image are identified, and replaced by the Braille text. In image simplification the visual properties such as colours, shading, and textures are simplified. The image is then scaled to satisfy the required fixed size of the Braille characters.

\(^{18}\)http://www.seeingwithsound.com/javoice.htm
In another approach (by Krufka), SVG\(^{19}\) is used as an input to automatically generate tactile diagrams. In this approach, the hierarchy of the elements in the SVG files are used to extract the important boundaries of the image. The importance of the boundaries are determined by the depth of the element in the hierarchy; the higher the element is on the tree, the more important it is. In the tactile presentation, the more important elements are represented by higher raised lines than the less important elements [29].

The main disadvantage of these approaches is that they tend to preserve the same layout as the original representation without performing any adaptation based on the diagram type, task and the user profile. Also, they do not take into account the semantics of the diagrams which can lead to bad layout for tactile diagrams. Moreover, effective audio versions cannot be generated due to the lack of semantics.

**Approaches Using Non-Standard Diagram Input Formats**

These approaches use tool specific formats which make accessible diagram generation easier by providing both the location of the elements and their semantic information. Thus, they are specialised to certain type of diagrams, and assistive devices. Table 3 summarises these approaches.

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Task</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audiograf [26]</td>
<td>Any diagram designed by the tool</td>
<td>Tactile on touchpad + Non-speech + Speech Audio</td>
</tr>
<tr>
<td>TMAP [39]</td>
<td>Map</td>
<td>Tactile + Non-speech/Speech Audio</td>
</tr>
<tr>
<td>IVEO [17]</td>
<td>Any diagram</td>
<td>Tactile + Non-speech/Speech Audio</td>
</tr>
<tr>
<td>VAR [15]</td>
<td>Any diagram designed by the tool</td>
<td>3D Non-speech Audio</td>
</tr>
</tbody>
</table>

Table 3: Non-generic approaches for accessible diagram generation

Annotated-SVG: SVG file which has extra meta-data information added by the tools.
Audiograf [26] let the user use a touch panel and audio system to interact with diagrams. The user navigates through the diagram by using the touchpad. During the navigation the system sets the focal point to the touched point. The graphic components at the focus area are identified by special sounds while textual information is read in speech by the system.

In TMAP (Tactile Map) [39] the authors combine different technologies and modalities to generate tactile/audio street maps. TMAP is a joint project by The Smith-Kettlewell Eye Research Institute\(^{20}\) and Touch Graphics\(^{21}\) who are responsible for the tactile map generation and associating audio with the tactile map and presenting on a TTT (Talking Tactile Tablet) \[^{32}\] touchpad, respectively. In this study a tactile street map is created by an online TMAP engine based on the user’s input and preferences. The output, which is an SVG file, is then downloaded by the designer. The designer uses software for TTT to

---

19 http://www.w3.org/Graphics/SVG
20 www.ski.org
21 www.touchgraphics.com
associate audio feedback with the tactile elements in the tactile map. The users use this system by first printing the tactile map by an embosser and then putting it onto a TTT touchpad.

IVEO [17] system also uses a touchpad and a tactile printout to present diagrams. Using its design software, a designer associates audio to relevant positions on the diagram. The final version of the diagram is saved as an SVG file. They are then printed by an embosser and put on a touchpad. When the user traces on the tactile printout, he/she can press at any point to get the associated audio information.

The VAR (Virtual Audio Reality) system [15] has three main components. The host computer, the joystick and headphones. It enables the users to get an overview of an environment on the host computer, and perform task on a graphical user interface by using the joystick. The elements on the visual interface on the host computer are associated with earcons [9], which are short audio representations of elements, and placed in a 3D space. The user traces this 3D space by using the joystick. During the tracing, earcons associated to elements are provided via the headphones.

The main disadvantage of these approaches is that they are specialised to a specific kind of device and/or presentation styles. Moreover, they do not use any standard representation of a diagram to generate its accessible version, rather they use their own representations which must be generated by a human.

**Semantic Based Approaches**

These approaches use the diagram’s underlying semantics to generate a accessible version of it, rather than using its visual representation. Table 4 summarises these approaches.

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Task</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elzer [13]</td>
<td>Image</td>
<td>Bar charts</td>
</tr>
<tr>
<td>Kevin [6]</td>
<td>Tool specific format</td>
<td>Data flow diagrams</td>
</tr>
<tr>
<td>IGR [44]</td>
<td>Image</td>
<td>Bar and pie charts</td>
</tr>
<tr>
<td>TeDub [43]</td>
<td>Abstract syntax of diagram</td>
<td>UML diagram</td>
</tr>
<tr>
<td>GrassML [16]</td>
<td>Semantic of diagram</td>
<td>User-defined diagrams by their semantics</td>
</tr>
</tbody>
</table>

Table 4: Semantic based approaches for accessible diagram generation

A probabilistic system (by Elzer et al) [13] which hypothesises the intention of information graphics (bar charts) is one of these approaches. In this system, a visual extraction module analyses the information graphics and creates an XML representation of it. This representation is then used by the caption tagging module, and intension recognition module to exploit the communicative signals by using a Bayesian network which uses pre-defined intension categories. Based on the different visual properties detected by image processing algorithms, a probability is assigned to each intension category. According to the probability values, the output is given by the content planning module, and message organisation module.
Kevin [6] is a system specialised for data flow diagrams. The system converts the data flow diagrams to its equivalent representation, which uses a $N \times N$ matrix to show the flow of data from a source to a target represented by rows and columns, respectively. The user interacts with the system by using speech audio.

The Image and Graphic Reader (IGR) [44] is another tool that uses image processing techniques to process business charts. The image processing algorithms recognise the primal geometric objects of the diagram. These objects are then classified into components of a particular chart such as bar, and pie chart. The classification includes both text and shapes. The content is given to the user through a speech synthesiser.

TeDub (Technical Drawings Understanding for the Blind) [43] is an EU-Funded project which has many partner institutions. The aim of the project is to navigate through the technical diagrams and get information about the elements in the diagram for blind users. It uses a image processing system to classify and extract information from the technical drawings. The diagrams are represented as connected graphs. The structure of the diagram is converted to the TeDub's Node, Node Info, Connections, Connection Info, and Annotation elements. This conversion provides easy navigation through the diagram which is problematic in the text versions of the diagrams. It is stated that the system will use screen reader, tactile overlay (this only includes the generic view of the diagram), force feedback joystick and 3D-sound with varying pitch and direction to provide the information from data storages such as databases, outputs of technical application, and internet [14]. In the current version (TeDub System Version 3.25.0), it uses sound, joystick or keyboard to handle visual data in technical drawings created in UML.

GrassML [16] uses a high level description in which a diagram is created by specifying only its semantics without any visual representation. At the presentation level, it generates the representation of a diagram in different modalities. For instance, in the graphical representation SVG format is used. At the structure level, it uses a specific language called ZineML. This language includes the basic shapes that can be used in a diagram.

MyLanguage is another language defined in GrassML framework to define the semantics of diagrams. It is specific to an application domain, and represents the entities and relationships among them. A particular diagram representation is generated by a set of rules at two transition points: (i) semantic level to structure level, (ii) structure level to presentation level. In the first set of rules, the entities and relations defined in MyLanguage are mapped to the shapes defined in ZineML. In the second set of rules, the shapes in ZineML are mapped to its representation. The representation can be a visual diagram in SVG or speech audio.

The main disadvantage of Elzer, IGR, and TeDub is that they both are specialised to specific diagrams. IGR can be extended to other types of diagrams by introducing new algorithms, but Elzer's approach may not be appropriate to other diagram types where information about individual components are also required as well as the intension of the diagram as a whole, such as flowcharts. As for GrassML, it requires extensive amount of work to create the input rules, and it is unlikely that designers will prefer such a diagram generation process. Thus, in practice, it is not feasible that most diagrams will be authored in such a way so as to explicitly include the semantics.
Project Proposal

Thus we see that, while there has been considerable research into generation of accessible diagrams, there is still no generic technique for automatically generating accessible diagrams from a standard input format such as SVG or image. Developing such a generic technique is the aim of this project. This generation should preserve the same semantics and pragmatics as its visual representation as far as possible, and be appropriate to users, tasks, devices and presentation media.

This leads to the following research questions:

1. How do we extract the syntax, semantics, and pragmatics of a diagram from its original visual form?
2. What are good alternative presentations of an accessible diagram based on diagram type, users, tasks, devices and presentation media?
3. How do we generate accessible diagrams such that they convey the same semantics, and pragmatics as possible as their visual representation?

As seen in the following Figure 5, our proposed architecture takes a generic input of a diagram and creates its intermediate representation which includes its syntax, semantics and pragmatics, and then combine this with the user profile and contextual information (devices, and presentation media) to generate an accessible version.

![Figure 5: The proposed solution](image)

We believe that diagrammatic notation can be specified by a visual language which details the diagram’s graphic components, and relations among them in a two or three dimensional space. The syntax of this language is defined by a grammar that includes primitive components and production rules. These production rules not only define how to create composite components from primitive ones, but also define the required geometric relations among these components. In contrast to text grammars where “immediately precedes” is the only relation, in diagrams there can be various relations, such as touches, next-to, and left-to. The semantics of the diagram can then be constructed by associating meanings to the components and combining them according to the relations [36]. This approach has been successfully used in different application areas such as sketch recognition, image processing, and graphical user interfaces [36]: PDL (Picture Description Language) [46] to specify pictures, G-Lotos [8] to define process specification, toolkits for building user interfaces such as Palette [19], and visual language parsers [24].
Defining a diagram type with a visual language allows us to translate it into its accessible version automatically by using the techniques of programming languages and compiler design. In computer programming, a program is written in a specific programming language such as Java, and C++ that can be understood by humans. This language is defined by a grammar and production rules. The program created by this grammar is used by a compiler to generate a computer executable code which is semantically equivalent but defined by another programming language. The compiler performs this translation by analysing the source language, and synthesising it to the target language. During the analysis, the compiler creates the tokens and syntax tree of the source program. It also creates a symbol tree that stores information about the source program that will be used at synthesis. During the synthesis phase, the target program is constructed from the intermediate representation of the source program and the symbol table [2].

With a similar approach, a diagram defined by a visual language can be transformed into another representation which is its accessible version. The grammar and production rules of the visual language allows us to construct a syntax tree from its definition. This syntax tree will be used for the intermediate representation. We will add extra information (such as pragmatics of a diagram type) to this syntax tree, which we call a decorated syntax tree, and then use it with the user profile and the context information to generate the accessible diagram. This decorated syntax tree can then be automatically translated into the decorated syntax tree for the diagram in another representation. Related methods have been used in direct manipulation interfaces, such as TRIP2, for bidirectional translation between the application data and its visual representation [38](Figure 6).

To illustrate our approach, a grammar is defined for bar charts. In Figure 7 the types and tokens of the grammar are given. Type definitions include the type name and its possible
values. The tokens are the primitive objects that are used to construct a bar chart. Their
definitions include their name and the attributes. In Figure 8, the production rules of the
grammar are given. They are in the form of constraint multiset grammars [35] where the left
hand side symbol of the rule can be rewritten in terms of the right hand side symbols such
that the constraints specified in the “where” clause are satisfied. In addition, the “using”
clause in these rules specifies the attribute value assignments of the left hand side symbol.

<table>
<thead>
<tr>
<th>TYPES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>character</td>
<td>&quot;A&quot;</td>
</tr>
<tr>
<td>digit</td>
<td>&quot;0&quot;</td>
</tr>
<tr>
<td>pattern</td>
<td>&quot;white&quot;</td>
</tr>
<tr>
<td>character</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>digit</td>
<td>&quot;,&quot;</td>
</tr>
<tr>
<td>pattern</td>
<td>&quot;empty&quot;</td>
</tr>
<tr>
<td>number</td>
<td>Any integer or real number</td>
</tr>
<tr>
<td>string</td>
<td>Any sequence of characters and digits</td>
</tr>
<tr>
<td>orientation</td>
<td>&quot;vertical&quot;</td>
</tr>
<tr>
<td>axisType</td>
<td>&quot;independent&quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TOKENS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>coordinate</td>
<td>x: number, y:number</td>
</tr>
<tr>
<td>graphicsObject</td>
<td>position: coordinate, embossedHeight: number</td>
</tr>
<tr>
<td>data</td>
<td>seriesName: string, independentValue: string, dependentValue: string</td>
</tr>
<tr>
<td>barChart</td>
<td>layout: orientation, values: set of data</td>
</tr>
<tr>
<td>title</td>
<td>text: string</td>
</tr>
<tr>
<td>label</td>
<td>text: string, indicatorLine</td>
</tr>
<tr>
<td>axis</td>
<td>chart: barChart, type: axisType, [gridInterval: number]</td>
</tr>
<tr>
<td>legend</td>
<td>chart: barChart</td>
</tr>
<tr>
<td>line</td>
<td>graphicsObject</td>
</tr>
<tr>
<td>rectangle</td>
<td>graphicsObject</td>
</tr>
<tr>
<td>style</td>
<td>style: barChartbars.width: number, barChartbars.fill: colour, axis.lineWidth: number, horizontalAxisIndicatorLength: number, verticalAxisIndicatorLength: number, legendSymbolSize: number</td>
</tr>
<tr>
<td>style</td>
<td>style: barChartbars.width: number, barChartbars.fill: pattern, axis.lineWidth: number, horizontalAxisIndicatorLength: number, verticalAxisIndicatorLength: number, legendSymbolSize: number</td>
</tr>
</tbody>
</table>

Figure 7: Bar Chart Grammar Types and Tokens

The grammar is actually for two different representations: visual and tactile. The definitions
in black are those definitions that are common both to visual and tactile bar charts. The
ones in red are the definitions specific to visual bar charts, whereas the ones in blue are the
definitions specific to tactile bar charts.

In our proposed approach the productions in the grammar for the visual representation are
used “bottom up” to recognise the bar chart diagram and the constraints in the grammar
are used to control the applicability of the production rules. Once the syntax tree has been
created this can be used to automatically generate the corresponding syntax tree for the
tactile representation. The tree is essentially isomorphic using equivalent tokens and
production rules from the grammar. However the attributes for the tokens will not yet be
known. These are constructed by traversing the syntax tree and sending the geometric
constructs off to a constraint solver. This solver determines the geometric attributes of the tokens.

**Figure 8: Bar Chart Grammar Production Rules**

A sample visual bar chart and its tactile representation are given in Figure 9, and Figure 10 respectively. Based on the defined bar chart grammar, the constructed decorated syntax
trees are given in Figure 11, and Figure 12. In these syntax trees, the decompositions of chart components are shown in vertical order. The properties of the components are shown on the right of the nodes, and the constraints on these elements are shown on the left of the nodes.

Research Approach
We believe that in near future diagrams will be generated in Scalable Vector Graphics (SVG) format rather than raster images. Thus, in our approach, we will focus on SVG graphics as the input visual representation. We plan to work on four different diagram
types: (i) bar and pie charts\(^2\) (Figure 13), (ii) node-link diagrams\(^3\) (Figure 14), (iii) flowcharts\(^4\) (Figure 15), and (iv) information maps\(^5\) (Figure 16). These diagrams are commonly used, and they are sufficiently different to provide a good test of the genericity of the proposed approach. While tools such as IGR already handle bar charts, we think that it is useful for us to compare our approach with previous approaches.

![Figure 11: Sample Bar Chart Decorated Syntax Tree](image)

The components of our architecture are listed as follows:

- **Diagram Reader**: Generates visual tokens from a visual representation.
- **Diagram Interpreter**: Generates the high level diagram syntax tree from visual tokens and diagram grammars.
- **Diagram Grammars**: The formation elements and rules for different types of diagrams.
- **Transformation Rules**: The mappings of elements and rules in the grammars between the visual and accessible versions of diagram type.
- **User profiles**
- **Context**: The factors that affect the adaptations on accessible diagrams.
- **Accessible Diagram Generator**: Generates the accessible versions of diagrams.

We will first define diagram grammars which include the equivalent grammar for the definitions of the components for particular diagram types, and their accessible versions. The SVG input will be processed by the Diagram Reader which will analyse the visual representation of the diagram, and will generate the visual tokens in the diagram. The main role of this component is to perform transformations such as rotation so as to generate a SVG description that is faithful to the visual representation. The visual tokens are used to determine the relevant diagram grammar based on the diagram type.
Figure 14: Node-link diagram sample

Figure 15: Flowchart sample
This grammar and the visual tokens are then used by the Diagram Interpreter to generate the decorated syntax tree of a particular diagram which includes its syntax, semantics, and pragmatics. Here we plan to use the Cider Toolkit [25], which is a component based toolkit for creating smart diagram environments. The context information will be collected separately, and will include user profiles (user preferences, duration of blindness, Braille literacy), tasks, devices, and presentation media. The Accessible Diagram Generator will take the decorated syntax tree and the context information as its inputs and produces the equivalent decorated syntax tree in the appropriate accessible version. It will use a constraint solver to determine the attributes of the tokens in this tree and hence the appearance and position of graphic elements in the accessible version.

To date, a broad literature review has been completed. This literature review includes the assistive devices and technologies used in presenting diagrams to blind people. Moreover, tactile diagram generation has been investigated by visiting Vision Australia offices. At Kooyong office, we had meetings with transcribers to understand the process of tactile diagram generation. At Prahran office, we obtained information about the devices that are available to blind users. And, in Sydney office, we have contacted people in TabMap community, whose main interest area is generation of tactile maps, to understand the issues in tactile diagram generation. In addition to these meetings, sample tactile diagrams were created by using drawing tools to better understand how to present tactile diagrams. These diagrams include bar charts, organisation charts, and network diagrams. They were examined by transcribers and proofreaders at Vision Australia, and gave us great insight into the difficulty of transcription.

A prototype Tactile Chart Tool (see Figure 17) has been implemented for generating tactile bar and pie charts (see Figure 18). This tool has been implemented in Java and gets its input as a text file and generates an SVG file based on the parameters specified by the user such as the minimum bar width, the axis line widths, minimum separation between adjacent bars. The parameters are used to modify the elements of the tactile diagram. For the output, an SVG file which includes the tactile diagram elements and the parameters is generated by using the Batik SVG Toolkit. For now, the tool is using ad-hoc algorithms to generate the diagrams. In the next releases, it will include constraint based layout algorithms which make it possible to handle various and numerous constraints of tactile diagram generation. So far, we have had pilot studies with proofreaders and obtained positive feedback about the diagrams generated by this tool.

**Figure 16: Information map sample**
Figure 17: Tactile Chart Tool

Figure 18: Charts generated by the tool
Conclusion
There have been attempts to automatically generate tactile diagrams by using a variety of different technologies. But most of these projects were highly dependent on the original visual presentation. We believe that semantics based diagram generation is a better way to generate accessible diagrams. This requires the development of automatic generic techniques for extracting diagram semantics from a visual representation and then transform this into an accessible version of the diagram.

We believe that realisation of such a generic system will be an important step towards automatic generation of accessible diagrams, and it will be valuable step in allowing blind people to access the information in both printed and online diagrams.

References
15. C. Frauenberger and M. Noisternig. 3D Audio Interfaces for the Blind. Workshop on Nomadic Data Services and Mobility, Graz, Austria, March., pages 11–12, 2003.
17. J.A. Gardner and V. Bulatov. Scientific Diagrams Made Easy with IVEOã„¢.g.


About the author:

Cagatay Goncu is currently a PhD student at Clayton School of IT, Monash University, Australia. He has been working on accessible diagrams with his supervisors Prof. Kim Marriott, and A.Prof. John Hurst. The aim of his research is to develop automatic generic techniques for extracting diagram semantics from a visual representation and then transform this to an accessible version of the diagram.
Good Visual Aesthetics Equals Good Web Accessibility

Grace Mbipom

Human Centred Web Lab, Information Management Group, School of Computer Science, University of Manchester, England, UK
grace.mbipom@cs.manchester.ac.uk

Abstract
This paper summarises a doctoral research project titled Empirical Investigation of Visual Aesthetics1 and Accessibility (EIVAA). The project proposes to investigate the relationship between visual aesthetics and Web accessibility, with the aim of building a framework to guide the development of aesthetic accessible Web content. As a first step, a formative study was conducted to examine the relationship between visual aesthetics and Web accessibility from a technical perspective. Thirty Web pages that had previously been rated on various aspects of their visual appearance in a study by Michailidou et al. [20] were investigated. The visual aesthetic dimensions considered in their study were “clean”, “clear”, “organised”, “beautiful”, and “interesting”. In our study, all thirty Web pages were tested for accessibility using an automatic evaluation tool, “Cynthia Says”2. Initial results show that Web pages which were previously agreed by participants to be visually clean, clear, and organised had lesser numbers of Web Content Accessibility Guidelines (WCAG) 1.0 failures. This suggests that visually pleasing pages along these design dimensions may not pose a barrier to accessibility. The implications of our study and future research directions are discussed.

Introduction
Visual aesthetics have been shown to enhance positive user experience in the context of the World Wide Web (Web) (e.g [8, 15, 17, 20, 23 and 25]). However, there is still much tension between visual aesthetic considerations for Web pages and their effect on ease of use for people with disabilities [22, 26]. On the one hand, most Web developers feel stifled by Web accessibility requirements and associate accessibility with visually unappealing designs. On the other hand, accessibility advocates assert that Web accessibility does not hinder the Web developer [10]. This apparent tussle has not been helpful to the accessibility community [22]. Moreover, most of the work that has been done in this area is theoretical and anecdotal in nature. Whether it is possible to achieve aesthetic accessible designs in practice still remains a pressing problem [10].

The EIVAA project seeks to investigate the relationship between visual aesthetics and Web accessibility with the aim of guiding aesthetic designs that are also accessible. To achieve this goal, extensive user studies and eye tracking information will be used to understand which Web pages are visually pleasing to sighted users, so that technical and manual (involving users and accessibility experts) accessibility tests can be conducted for such pages. We believe that this may reveal the interplay between visual aesthetics and Web accessibility. Based on our findings, research-based Web

1Visual aesthetics here refers to visual pleasure. For the purpose of experimentation, visual aesthetics was investigated along five design dimensions previously studied by Michailidou et al. [20].

2Cynthia Says - http://www.contentquality.com/
design guidelines will be proposed, and a framework for guiding aesthetic accessible designs will be developed.

Related Work

There is currently work in Human-Computer Interaction that seeks to understand the role of visual aesthetics in interaction design. However, many researchers in the field are yet to understand what visual aesthetics entails in practice [9]. This problem is not peculiar to Human-Computer Interaction domains alone. From antiquity, understanding the subject of visual aesthetics has been a problem. Early philosophers were also divided on the issue of visual aesthetics and how its study should be undertaken [19]. While some adopted an "objectivist" approach, others took an approach which was much more "subjectivist". The former group closely associates visual aesthetics with the object being viewed, and sees beauty as a "measurable construct" that is derived from the physical form of the object [27]. The latter are of the opinion that the person experiencing an object decides whether the object in question is beautiful or not [14]. In the context of the Web, the two approaches are often used together because of their inter-relatedness. Typically, study participants are required to state their visual aesthetic preferences for Web pages along design dimensions such as order, symmetry, clarity, beauty, complexity, ease-of-use and so forth. Some of the afore-mentioned design qualities can also be treated as objective measures, depending on the approach the researcher wishes to adopt [7]. Frohlich [6] points out that for interaction design, these two approaches have their implications. According to him, the objectivist approach would require that interactive system designers focus more on the product being designed, and that they strive to design in a way that conforms to the norm in terms of beauty expectations. Of course, these may vary across communities [15, 28]. The subjectivist approach would require that designers design according to the stipulated best practices in their respective industries and believe that by so doing, the user might just be able to see the beauty that exudes from the finished product and come to appreciate the efforts of the designer. Frohlich recommends the latter approach as the former may be "distracting".

Web Visual Aesthetics

Empirical studies on visual aesthetics for the Web have followed of one of two directions:

- There are those studies that have concentrated on replicating experiments to validate visual aesthetic theories of old. In such studies (e.g [16]), Web pages are used as visual stimuli in place of works of art and geometrical shapes which served as test beds in the early days of experimental aesthetics. One popular aesthetic theory that has been investigated in the context of the Web is Berlyne's [1] arousal theory. Berlyne's theory holds that people love to experience aesthetic pleasure at moderate levels.

- There are also studies that have focused more on investigating the relationship between visual aesthetics and different aspects of user experience. Some aspects explored so far include: usability [8, 17, and 24], credibility [15, 23], desirability/preference [25], complexity [20] and more recently, accessibility [18].

Although these studies have consistently shown that visual aesthetics enhances positive user experience, very little is known about the interactions between visual aesthetics and accessibility. Sighted Web users trust Websites they perceive to have good Web page aesthetics, and would more readily transact with such sites [15]. Also, users perceive
information obtained from a Website with good visual aesthetics to be more credible than that obtained from a Website with poor visual aesthetics, even if the two Websites in question have the same content [23]. Visually pleasing pages are generally perceived to be easier to interact with [8, 17] and less visually complex [20]. Moreover, beauty determines the visitor’s initial impressions, hence their preference [25].

The influence of Web visual aesthetics on accessibility is one area that has not benefited much from empirical research. The increasing trend in Web visual aesthetics is speculated to hinder people with disabilities from accessing Web resources effectively, especially people with visual impairments3 [26]. This belief has had an impact on the way most Web developers perceive the Web Accessibility Initiative (WAI)4. Some Web developers are reluctant to comply with proposed guidelines for building accessible Websites, because they perceive them to be "restrictive" design-wise [10, 22]. Perhaps, the inability to clearly distinguish between visually pleasing and displeasing Web designs also has a part to play in this matter. The question that needs to be asked is: what is the true visual nature of Web designs that manifest inaccessibility?

Web Accessibility: The need to make the Web accessible to people with disabilities cannot be over emphasized. Apart from the fact that interacting with the Web gives users with disabilities a sense of belonging [10], it also makes them feel more independent [21]. With the potential socio-economic benefits accessibility brings to the Web community, one would expect that the accessibility initiative would be well received by all. Unfortunately, this has not been the case. Several actions have been taken, mostly legal and educational to make Web developers and builders of user agents, authoring tools, interactive systems and providers of Information Technology (IT) services, conform to accessibility requirements. For example, in the United States of America (USA) Section 5085 which outlines accessibility standards has been adopted. Some notable guidelines include the Web Content Accessibility Guidelines (WCAG) Version 1.0 [5] and 2.0 [4] recommended by the World Wide Web Consortium (W3C)6 under their Web Accessibility Initiative. Other guidelines specified by the World Wide Web Consortium include the Authoring Tool Accessibility Guidelines (ATAG) [11] and User Agent Accessibility Guidelines (UAAG) [12]. WCAG version 1.0 and 2.0 outline the necessary steps to take in order to produce Web content that is accessible to people with disabilities. ATAG specifies what developers of tools used to create and maintain Web content should do in order to advance the accessibility mission, while UAAG explains how to make technologies used to interact with the Web accessible to people with disabilities, and how this helps in making the Web accessible in general [10].

Accessible Designs – The Overlap: The relationship between Web accessibility and visual aesthetics or good design is an area of research that is still poorly understood. One work that has attempted to examine the overlap between accessibility and design is that of Regan [22]. According to him, the current state of affairs is one that can be described as

---

4 Web Accessibility Initiative (WAI) - http://www.w3.org/WAI
5 Section 508 - http://www.section508.gov/
6 World Wide Web Consortium (W3C) - http://www.w3.org/
“a failure of the imagination”. Accessible Websites have poor visual designs [22]. Henry [10] points out that, theoretically, the myth associated with this issue is false. However, she acknowledges that how to achieve aesthetic accessible designs in practice is still a problem. She further states that the lack of resources such as skills and money on the part of Web developers, and inaccessible Web tools and technologies are responsible for this. To set a good example, Henry led a redesign of the Web Accessibility Initiative Website [10]. Figure 1 shows screenshots of the homepage visual design before and after redesign in 2005. She acknowledges that the new and current design in Figure 1 (b) may not be an epitome of visual beauty because of limited resources, but it represents a demonstration of the fact that visual aesthetics and accessibility can be compatible [10]. One question that needs to be asked, however, is whether the designs we classify as beautiful are truly visually pleasing. One Web user may prefer the new homepage design over the old because of more use of colour and graphics. Another person may prefer the old design for its simplicity. Although we may not be able to design for every taste or satisfy subjectivity as Frohlich [6] points out, there is a need to have an understanding of what visually pleasing designs are, or at least an approximation with the help of models of Web users’ aesthetic perception. Such designs may then be investigated for compatibility with accessibility.

Formative Study

We have been able to examine the accessibility level of thirty Web pages that had been previously rated on various aspects of their visual appearance in a study by Michailidou et al. [20]. The visual aesthetic terms that were considered in their study were “clean”, “clear”, “organised”, “beautiful”, and “interesting”. They were chosen from a pool of visual aesthetic terms commonly used in Web research (see [17]). The same visual aesthetic terms were investigated in our study. The aim of our study was to examine the relationship between visual aesthetics and Web accessibility from a technical perspective.

Materials

The Web pages used for our experiment were taken from [20], and had originally been selected from Alexa7 UK Top 100 Websites as of 18th December, 2007 when their study was first conducted. The Web pages were representative of common genres available online such as entertainment, news, e-commerce, personal, academic and social-networking Websites. The ratings for the visual aesthetics of the Web pages were obtained from an on-line study where 55 participants ranked the thirty pages in a counter-balanced manner.
according to the five visual aesthetic characteristics mentioned earlier (see [20] for more details).

Method
The thirty Web pages were examined against WCAG 1.0 which was the stable version of WCAG available at the time our study was conducted. The number of guidelines failed by the individual Web pages was noted. We were interested in the number of guidelines failed, rather than the checkpoints. If at least one checkpoint was violated under a guideline, then that was counted as one failure for the associated Web page irrespective of the number of checkpoints failed under such a guideline. The automatic accessibility evaluation tool used was Cynthia Says. We chose Cynthia Says because it is a free Web-based tool that presents accessibility evaluation results in a clear manner. The Web pages were tested for all three WCAG 1.0 conformance levels: “A”, “AA”, and “AAA”. An automatic accessibility evaluation method was chosen because of the nature of our study, which was formative. Although this method is presently less reliable than manual accessibility evaluation methods, it is important to note that automatic validation and evaluation tools like the one used in our study still play a significant role in helping Web developers identify potential accessibility bugs [13]. In addition, these tools are often the first port of call for Web developers who wish to evaluate their products against stipulated standards.

Results and Discussion
Significant inverse relationships were observed between “clean”, “clear”, and “organised” visual aesthetic attributes, and the WCAG 1.0 failures of the Web pages for all three conformance levels. However, “beautiful”, and “interesting” showed no significant relationships. So, Web pages that were perceived to be visually “clean”, “clear”, and “organised” had lesser numbers of WCAG 1.0 A, AA and AAA failures, suggesting that visually pleasing pages in this context may be more readily accessible than their visually displeasing counterparts. However, further studies involving users with visual impairments are required on this issue. Figure 2 shows the relationship between visual aesthetics and WCAG 1.0 AAA failures.

The visual aesthetic terms “beautiful” and “interesting” appeared to be the most subjective of the five terms examined in [20]. We speculate that this subjectivity may have influenced our results. In addition, the word “interesting” is ambiguous and may not be directly related to a Web page’s visual appearance. A careful look at the rankings done in [20] reveals that some of the Web pages which were agreed to be visually “cluttered”, “confusing” and “disorganised” across participants by virtue of the low visual aesthetic scores given to such pages, received high ratings for being “interesting”, and therefore “visually pleasing”. In our case, such pages were found to have high numbers of WCAG 1.0 failures. Consequently, the term “interesting” exhibited a random behaviour when its relationship with the number of guidelines failed by the Web pages was examined. The same explanation holds for the term “beautiful”. However, ratings of beauty were less subjective than those of interestingness in their study and this effect came across in our results also. For more information, please see the associated technical report [18].

Future Research Plan

In general, we speculate that there may be several factors responsible for the relationships that have emerged from this initial study. For example, it is not clear whether a Web developer’s technical background may have influenced the observed relationships, as this was not taken into account. We speculate that experienced Web developers more readily incorporate accessibility into their designs compared to amateurs, leading to good visual designs that may have few or no accessibility issues. Consequently, we will investigate the effect of a Web developer’s technical expertise and accessibility awareness. It is envisioned that our findings will help Web accessibility regulators tailor Web access equality awareness programmes, and recommendations to the appropriate target groups.

Aesthetic Perception

Further investigations on the ambiguity of the adjectives used to describe Web visual aesthetics in our research will be carried out. In particular, participants’ understanding of the visual aesthetic terms used in our future studies, and the consistency of the visual aesthetic judgments made, will be investigated via extensive terminology surveys and user studies. Since the aesthetic perceptions of Web users remain difficult to understand completely, alternative evidence gathering methodologies are desirable. Consequently, eye tracking techniques will be used to understand if sighted users have a way of interacting with Web pages that are visually appealing as opposed to unappealing ones. We are also interested in expert versus novice ratings of Web visual aesthetics. As such, the visual aesthetic judgments of professional Web developers will be investigated against those of novice developers and users in general. Also, the cultural factors that come into play during visual aesthetic evaluations will be factored into our studies \[15, 28\]. Based on our findings, a holistic model of Web visual aesthetics will be built.

Accessibility Evaluation

The accessibility component of our work will require that Web pages that have been identified as visually pleasing by Web users and developers from our previous studies on aesthetic perception, be investigated for accessibility conformance using different
approaches. Some of the methods that will be used include technical evaluation: involving the use of various automatic tools; user studies: involving people with a wide range of visual impairments; Hybrid methods: involving accessibility experts and the use of automatic tools [3] and expert agreement: based on accessibility ratings from a group of experts [2].

Based on our studies on visual aesthetic perception and Web accessibility, a support framework for Web developers will be built. It is envisioned that this will help guide the development of aesthetic accessible designs. Figure 3 gives an overview of the EIVAA research project. It highlights the two major components of our research: aesthetic perception and accessibility evaluation studies, the methods that will be employed in each case and our expected outcome. For more information please visit our Website http://hcw.cs.manchester.ac.uk/

Conclusion
The EIVAA project seeks to expand the knowledge of the interactions that exist between two important Web constructs: visual aesthetics and Web accessibility. The goal is to build a framework that guides the development of aesthetic accessible Web content. We envision that our findings will be useful to Web developers, interactive system developers, and the accessibility community, who are all currently faced with the challenge of reconciling these two Web constructs.

Acknowledgements
The author would like to thank the anonymous reviewers, faculty members, and participants at the Doctoral Consortium for their useful feedback. She will also like to thank her Supervisor, Dr. Simon Harper for leading her into an interesting area of research. This work is supported by the Akwa-Ibom State University of Technology, Nigeria, West Africa.

References


About the author:
Grace Mbipom is currently a PhD student at the School of Computer Science, University of Manchester, UK. Before now, she had obtained an MSc in Artificial Intelligence (Natural Language Processing) from the University of Edinburgh, Scotland, UK and a BSc in Computer Science from the University of Calabar, Nigeria. Her research interests are centred around Web accessibility.
Automatic Readability Assessment for People with Intellectual Disabilities

Lijun Feng

City University of New York

lijun7.feng@gmail.com

Abstract

My research goal is to advance our understanding of, and quantify, what makes a text easy or difficult to read, in particular for readers with intellectual disabilities. Previous research in automatic readability assessment has looked at a limited class of lexical and syntactic properties of texts. Moreover, these models are usually not targeted towards any particular group of readers. In my own work, by contrast, I have used sophisticated computational tools to build an automatic readability metric that exploits global semantic (discourse level) properties of a text, in addition to well-studied lexical and syntactic features. Our preliminary results (Feng et al., 2009) confirm the value of discourse attributes. My research is targeted towards understanding the particular difficulties faced by readers with intellectual disabilities. The ultimate goal is not simply to model and understand readability issues, but also to aid in the development of automatic language processing tools that can rewrite texts to be more readable.

Motivation and Background

According to the 2006 American Community Survey (U.S. Census Bureau, 2006), about 5% of the civilian noninstitutionalized population, approximately 13.5 million people age 16 or above in the United States have mental disabilities, with intelligence test scores of 70 or below. Among this group of people, about 85% are in the category of mild mental retardation (IQ range 50–75) (Drew and Hardman, 2004). We will use the term “intellectual disabilities” (ID) or “mild intellectual disabilities” (MID) henceforth. People with ID face many challenges in their daily lives; one of these challenges lies in the area of reading literacy. A study conducted by Jones et al. (2006) assessing the reading comprehension of adults with MID reported that the average reading skills of subjects were below that of average 7-year-old readers without disabilities.

Several factors contribute to the lower literacy skills of adults with ID. Above all, the limitation of their cognitive functioning affects their reading comprehension directly. Research has shown that people with ID are often better at sounding words out rather than comprehending their meaning (Drew and Hardman, 2004). It has also been shown that they are often slow at resolving the identity of proper names and encoding them into a symbolic form in their memory (Hickson-Bilsky, 1985). These individuals often have problems remembering and inferring information from text because of their limited working memory (Fowler, 1998). Consequently, as part of the previously read units are lost from the working memory, they often have difficulty with integrating complex information into a cohesive semantic representation (Hickson-Bilsky, 1985), which likely results in their lower reading comprehension.

It is difficult to find reading materials for individuals with MID that are (1) of interest to them and (2) at the right reading level. Reading materials at lower reading levels are typically
written for children, and texts written for adults without disabilities often require a high level of linguistic skills and sufficient real world knowledge, which these individuals often lack. The lack of appropriate reading materials may also discourage adults with ID from practicing reading, thus diminishing their already low literacy skills.

The need to identify or reformulate texts suitable for lower reading levels is not unique to people with ID. Children, second language learners, and adults with low literacy skills can also benefit from such texts. However, manually adapting written texts is both time and labor intensive. In the past decade, natural language processing (NLP) techniques have been used to develop automatic text simplification systems to assist human readers (Devlin, 1999; Inui et al., 2003). Research has focused mainly on lexical and syntactic simplification. Lexical simplification often uses word frequency or predefined wordlists to identify difficult words and replaces them with simpler synonyms. Syntactic simplification often uses dependency-tree structures and pattern recognition techniques to identify complex syntactic constructs, such as relative clauses, passive voice, and conjoined sentences; transformation rules are then applied to change these constructs into shorter or plainer sentences.

People with MID would certainly benefit from texts simplified in this fashion. However, synonym-replacement and syntax-tree simplification alone cannot fully cover the needs of this group of users, because, in addition to challenges that come from lexical and syntactic factors, they have other difficulties with processing written information as discussed above. Moreover, most earlier text simplification systems process input text one sentence at a time, which inevitably results in increased length of the simplified document, because long and complex sentences are often split into multiple shorter sentences. The resulting increased length of the whole document can pose another challenge to readers with MID because it requires processing and storing more information.

In order to meet the special needs of this group of underrepresented individuals, we are ultimately interested in designing and implementing an automatic text simplification system that modifies a text at the discourse level, in addition to lexical and syntactic simplification. This entails high-level semantic simplification, whereby the most relevant information is retained and less relevant information simplified or completely left out (Feng, 2008).

The following proposed dissertation research is the first step of a longer-term research project into a discourse-level text simplification system. We face several open, foundational questions, which are both self-contained and crucial for further research, and thus form a stand-alone dissertation project. There are two major research questions that are at the center of the design and implementation of such a text simplification system (Inui et al., 2003): (1) How do we identify which portions of a text will pose difficulty for our users? (2) When there are several possible simplification choices, how do we decide which is the optimal one to choose for our users? We need a reliable reading assessment tool to guide the system’s actions during the simplification process and to evaluate the difficulty posed by a given text, whether found, constructed by experts, or simplified automatically.

Relevant Literature and Previous Work

Extensive research has been conducted in the past 80 years to understand what affects the readability of a text and how to assess its reading difficulty. To make it easier for people to judge the reading difficulty of a text, grade levels or number of years of education
required to completely understand a text are commonly used as index for reading difficulty.

Many traditional readability metrics use simple linear functions with two or three shallow language features to model the readability of a given text. For example, the Flesch-Kincaid grade level formula (Flesch, 1979) uses average sentence length and average syllables per word to calculate the grade level of a text. Similarly, the Gunning FOG index (Gunning, 1952) uses average sentence length and the percentage of words with at least three syllable as parameters. These traditional metrics are widely used, especially in educational settings, partly because they are easy to calculate. However, these metrics do not always capture the reading complexity of a text accurately, which has been confirmed by several recent work in the field (Si and Callan, 2001; Feng et al., 2009; Petersen and Ostendorf, 2009). For example, the preliminary results of our research show that, tested on 1433 articles labeled with grade levels ranging from 2 to 5, Flesch-Kincaid metric only predicted 29 (2%) of them with correct grade level, while our metric predicted 881 (61%) correctly.

It is understandable why traditional readability metrics are not reliable, because the complexity of a writing lies in much more complicated factors than just average sentence length or average syllables per word. Moreover, short sentence length does not necessarily indicate easy readability (consider poems, for instance), and there are many infrequently used words with only a few syllables.

Recent work on readability deployed sophisticated natural language processing techniques, such as parsing and statistical language modeling, to capture more complex linguistic features and used statistical machine learning tools to build readability metrics. Si and Callan (2001) used unigram language models to capture content information from scientific web pages. Combined with surface linguistic features, they built a classifier with a linear model to predict the reading difficulty of these web pages with success. Collins-Thompson and Callan (2004) adopted similar approach and used a smoothed unigram model to predict the grade levels of short passages and web documents. Michael J. Heilman and Eskenazi (2007) continued using language modeling to predict readability for first language texts, furthermore, they extracted grammatical features from parsed documents to build a classifier for readability prediction for second language texts. Schwarm and Ostendorf (2005) used support vector machines to combine features from traditional reading level measures, statistical language models and automatic parsers to assess reading levels.

Recent work on readability all confirms the benefits of using statistical language models and/or parsers over traditional measures. It has also be shown that, although some features may outperform the others, when they are used combined, they often help improve the overall performance of the readability metrics. Our preliminary results are consistent with these findings.

**Methodology and Data**

Our research goal is to build and evaluate an automatic readability assessment tool that accurately models the reading difficulty of a text for people with intellectual disabilities.

The problem with building such a metric lies in that text difficulty is not directly observable, we need to do user studies and/or find proxy variables that associate with reading difficulty. Our general methodology relies on the following five proxies. (1) Paired
original/simplified texts. A common assumption is that simplified texts should be easier to read. Paired texts provide valuable clues on how texts with identical subject matter differ. During our modeling, we will use paired texts to analyze and select features that distinguish the simplified texts most from the original ones. (2) Grade levels. They are a commonly accepted index for reading difficulty of a text, especially in educational settings. Reading difficulty increases with grade level. It is an open research question in our study whether grade levels are appropriate readability indicators for our users. Our hypothesis is that text properties that influence reading difficulty for average readers are qualitatively (but perhaps not quantitatively) the same for readers with ID. This hypothesis says that what is hard for average readers is also hard, perhaps to a different degree, for readers with ID. Put yet differently, we are not aware of any text properties that cause problems only for readers with ID. Vice versa, we think a readability metric that is designed for a specific group of users can be useful for the general public as well. Another reason to look at grade levels is because they have been widely used in the existing literature. Novel readability metrics should try to predict grade levels so that they may be compared with existing approaches. (3) Subjective ratings by experts. We will ask experts who specialize in working with adults with ID to rate text difficulty. The motivation behind this is as follows: (a) We rely on their expertise to help us identify factors that may play an important role in affecting reading difficulty for our users. (b) Subjective expert ratings are much more reliable and easier to obtain than from target users. We will evaluate subjective ratings by checking inter-rater agreement, as well as correlation with grade levels and subjective ratings and observations. (4) Subjective (introspective) ratings by users. This will probably be especially problematic in our study, as the users’ subjective judgment may not be fully reliable because of their cognitive impairments. Many research questions remain open as how to design and conduct studies with adults with ID to get effective and valid feedback. However, we believe direct user feedback is valuable in our user-specific study. (5) Objective observations in user studies. We will present our target users with texts at a variety of difficulty levels and record their reading times. Subject will further answer simple comprehension questions after reading, and we will analyze the accuracy of their answers. This will give us the most direct clues about the difficulties faced by our target user group, even though we will need to account for per-subject and other effects. For developing our readability metric, we want to combine all the above observations to get at those underlying text properties that are associated with reading difficulties.

We plan to develop three groups of features based on the analysis of the paired original/simplified texts: (1) traditional shallow features, (2) syntactic features, and (3) novel discourse features. Our shallow and syntactic features (1) are mainly inspired by previous work in the field. Shallow features include those that are often used by traditional readability metrics, such as average sentence length per word, average number of syllables per word, total number of polysyllabic (3) words, percentage of polysyllabic words. Our syntactic features (2) include (a) parse tree features, such as average parse tree height, ratio of terminal and non-terminal nodes, number of relative clauses, noun phrases, verb phrases and prepositional phrases, etc.; (b) cross-domain perplexity features captured by various statistical language models; and (c) part-of-speech features, such as average number of adjective, adverbs, past participles, past tense verbs, present participles, modal verbs, and infinitive markers.

Our selection of discourse features (3) will focus on those that are cognitively motivated by our user characteristics. Our hypothesis is that, aside from infrequently used word and
complex syntactic constructs that even average readers find difficult, the density of the amount of information introduced in a text and whether the information needs to be inferred may pose particular challenges for our users. Named entities, such as people’s names, locations, organizations, together with general nouns, serve as major information carriers in a text. We hypothesize that the more entities there are to be processed, the more they will overwhelm our users’ already slow semantic decoding and limited working memory. Similarly, at a even higher discourse level, factors such as the amount of topics introduced in a text, where each of them starts and ends, where some of them overlap, require excessive working memory to process and keep track of. Moreover, our target users often have problems remembering and inferring information from a text (Fowler, 1998), so information that is not stated directly and needs to be inferred by readers can be challenging for our users. Therefore, our discourse features will include (a) entity density features, such as average number of unique entities (people, location, organization and general nouns) per sentence, total entity mentions per sentence, or the total counts of both in a text; (b) discourse topic features, such as the total number of topics, the average span of a topic, number of overlapping active topics, etc; and (c) coreference features, such as average distance between pronouns and their referents, and the number of nouns and/or pronouns that refer to the same object or person.

We use various existing and novel NLP techniques and toolkits to extract features. To extract syntactic features, we use the well known Charniak parser (Charniak, 2000) for automatic syntactic analysis. We use LingPipe (http://alias-i.com/lingpipe) to extract named entities and solve coreference related issues. We use the lexical chain software developed by Galley and McKeown (2003) to annotate synonyms, hypernyms and hyponyms. The SRI Language Modeling Toolkit (http://www.speech.sri.com/projects/srilm) and the CMU-Cambridge Statistical Language Modeling Toolkit (http://www.speech.cs.cmu.edu/SLM) are used to train various language models.

To investigate the possible impacts of these features on readability and to construct our metric, we use statistical and machine learning techniques, in particular support vector machines (SVMs), to train various classification and regression models on labeled data. Our preliminary experiments use LIBSVM (Chang and Lin, 2001), an integrated software for support vector classification and regression, which supports efficient multi-class classification and provides an automatic tool to perform grid search for optimal parameters, which are essential in improving the performance of our models (C.-W. Hsu, 2003).

An ideal corpus for our research would be a set of original texts paired with their simplified versions adapted just for our target users with MID. Unfortunately, there does not exist such a corpus yet, to the best of our knowledge. In order to develop our methods and test our hypothesis, we collected two comparable corpora from two sources: literacynet.org and Encyclopedia Britannica (Barzilay and Elhadad, 2003). Both corpora consists of paired original/simplified texts. We collected a third corpus from the Weekly Reader (http://www.weeklyreader.com), an online publisher producing magazines for students in grades pre-K to 12. This corpus consists of 1433 articles in total with grade levels labeled ranging from grade 2 to 5. We use this corpus to train various classification and regression models using SVM, and we test the baseline performance of our models on the same data and compare them with existing literature. At the same time we will create our own user-
specific paired corpus for user studies. We will collect texts that are of interest to our users and recruit experts on adults with ID to simplify them.

**Preliminary Research**

Currently, we have collected two unlabeled paired corpora (LiteracyNet and Encyclopedia Britannica) and a labeled corpus (Weekly Reader). We have developed 49 features in total and implemented a baseline library to calculate them. Using these features and the labeled data we have built our models and evaluated them.

Our preliminary results show that our discourse features are competitive to shallow and syntactic features, they help improve the overall performance of our metric when used combined (see Table 1). Compared with previous work, our novel readability metric are very promising in predicting correct grade levels. It outperformed the traditional Flesch-Kincaid readability metric by remarkable margin (see Table 2A), and remains competitive with other methods similar to our approach. Compared with a similar study conducted by Petersen and Ostendorf (2009), our metric outperformed theirs with high margin on articles with lower grade levels (see Table 2B and 2C). For further details, see our forthcoming paper (Feng et al., 2009), accepted for publication at EACL-09.

<table>
<thead>
<tr>
<th>A. Shallow features</th>
<th>B. Syntactic features</th>
<th>C. Discourse features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec</td>
<td>Recall</td>
<td>F-score</td>
</tr>
<tr>
<td>Gr 2</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td>Gr 3</td>
<td>0.49</td>
<td>0.54</td>
</tr>
<tr>
<td>Gr 4</td>
<td>0.46</td>
<td>0.43</td>
</tr>
<tr>
<td>Gr 5</td>
<td>0.64</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 1: Comparison of precision, recall, and F-score for SVM trained with shallow features, syntactic features and discourse features on the Weekly Reader data (10-fold cross-validation).

<table>
<thead>
<tr>
<th>A. Flesch-Kincaid</th>
<th>B. Peterson et al 2009</th>
<th>C. Our metric (10-f CV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prec</td>
<td>Recall</td>
<td>F-score</td>
</tr>
<tr>
<td>Gr 2</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Gr 3</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Gr 4</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td>Gr 5</td>
<td>0.83</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 2: Comparison of Flesch-Kincaid, Peterson et al’s recent work and our metric based on 10-fold cross-validation.

**Future Work**

In the next stage of this research, we will continue development of more discourse features using coreference resolution software and additional complex syntactic features and incorporate these new features into our metric. At the same time we will recruit experts to simplify texts for our users, ask for experts’ subjective ratings, recruit adults with MID as test subjects, design questionnaires for user studies, and conduct user experiments. In Fall 2007, we conducted a pilot study with 14 adults with ID, which provided us with valuable feedback on many aspects, such as what kind of objective measures worked, what still needs to be
improved, and what did not work. We also learned how to interact with our test subjects
during the study to obtain qualitative feedback.

In the end, we will use our metric to predict readability of the texts that are presented to the
experts and the test subjects. We will analyze the correlations of the predictions by our metric
with experts’ subjective ratings and user feedbacks.

References
   Technical report, Department of Computer Science, National Taiwan University.
   of the North American Chapter of the ACL, page 132–139.
   reading difficulty. In Proceeding of HLT/NAACL.
   Sunderland, UK.
   with intellectual disabilities. Merrill, Columbus, OH.
   Consortium.
   readability assessment. In The 12th Conference of the European Chapter of the Association for
10. Anne E. Fowler. 1998. Language in mental retardation. In Handbook of mental retardation and
    chaining. In Proceedings of the 18th International Joint Conference on Artificial Intelligence
    (IJCAI 2003).
    assistance: A project note. In Proceedings of the Second International Workshop on Paraphrasing,
    pages 9–16.
16. Jamie Callan Michael J. Heilman, Kevyn Collins-Thompson and Maxine Eskinazi. 2007. Combining
    lexical and grammatical features to improve readability measures for first and second language
    texts. In Proceedings of the North American Chapter of the Association for Computational Linguistics
    (NAACL’07), pages 460–467.


About the author: Lijun Feng is a PhD candidate in the Computer Science Department at the Graduate Center of the City University of New York (CUNY), where she is working with Prof. Matt Huenerfauth, her thesis advisor. Her research interests include natural language processing (NLP), in particular readability, text simplification, and text comprehension. Her thesis research combines NLP and machine learning techniques to build and evaluate an automatic text readability assessment tool, with special focus on readers with intellectual disabilities. Before entering CUNY, she obtained my masters degree in Computer Science from Brooklyn College in May 2005.
ASSETS’08 Trip Report

Top caliber accessibility research, a small friendly gathering, and a picturesque setting—these were just a few of the highlights of the Tenth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS’08).

ASSETS ’08 took place from October 12–15, 2008 in Halifax, Nova Scotia. Halifax boasts the world’s second largest natural port and is a lively city on the Atlantic coast and home to several breweries (Yum!). The conference was held at the Halifax Marriott Harbourfront Hotel, overlooking the spectacular waterfront, and close to the heart of downtown.

Though most attendees spent Sunday checking into the hotel and visiting the Harbourwalk, for a small number of attendees the conference officially started early that morning. This year 11 students participated in the Doctoral Consortium, a one-day workshop, where PhD students present their research and receive feedback from a panel of senior researchers. The consortium is typically an intense, but exceedingly helpful experience. From what I heard from this year's cohort, 2008 was no exception.

One of my favourite things about ASSETS is its intimate size. Because it is a relatively small conference, it remains friendly and provides ample opportunity to both catch-up with old friends and make new ones. A prime example of this was the welcome event on Sunday night. After gathering in the conference lobby, Simon Harper, our fearless leader (and the 2008 general chair) confidently lead us through downtown in search of local beer and friendly atmosphere. We quickly found a suitable pub, and spent the rest of the evening gathered cozily around a few large tables, meeting each other and enjoying food and drink.

With Monday morning came the start of the main program, and still a little jet-lagged—not to mention a little tired from the night before—we stumbled into the conference. Fortunately, Dr. Ron Baecker roused our sleepy eyes with an invocative opening plenary on the design of technology to aid cognition. I don’t think I was the only attendee a little embarrassed by teary eyes after seeing the video portion of his presentation. His topic was not only captivating, but also timely: there was a marked trend in the program this year towards research on memory and cognition.

In total, 29 papers were presented over the three-day conference covering a wide range of topics. Web-accessibility continued to be a hot topic, but research also ventured in a few new directions including gaming and the use of social networks to improve accessibility. Lunches and coffee breaks provided occasions to mix with other conference attendees and to check out the 34 posters and demonstrations, as well as the posters from the 9 student research competition semi-finalists and the doctoral consortium attendees. There was a lot to see, but luckily a reception Monday night provided even more opportunity to meet, mingle, and catch all the new and exciting research.

ASSETS ’08 was particularly special, as the first ever SIGACCESS Award for Outstanding Contributions to Computing and Accessibility was presented at a banquet Tuesday night. This biennial award recognizes individuals who have made significant and lasting contributions to the development of computing technologies that improve the accessibility of media and services to people with disabilities. This year’s award went to Dr. Jim Thatcher for his pioneering work on screen readers.
Dr. Thatcher graciously accepted the award and humbly joked about his amazing life, as we devoured a scrumptious meal.

In general, the quality of the research presented this year was exceptional, but two papers were noted as standing above the rest. This year the ACM SIGACCESS Best Student Paper Award went to Yevgen Borodin, Jeffrey P. Bigham, Rohit Raman, and I. V. Ramakrishnan for their paper, "What's new?: Making web page updates accessible." And, the ACM SIGACCESS Best Paper Award went to Jinjuan Feng, Jonathan Lazar, Libby Kumin, and Ant Ozok for their paper, "Computer usage by young individuals with down syndrome: An exploratory study."

Mirroring its start, the conference closed with an informal gathering around local beer and friendly conversation. I, unfortunately, missed out on that event as I had to rush off to the airport. But I left full of excitement for next year's conference, and full of ideas for new research. So as much as I longed to stay just a little bit longer, I couldn't wait to get home and to start on those ideas! I hope to see everyone next October (26–28) in Pittsburgh, PA.

About the author: Karyn Moffatt is a PhD student in Computer Science at the University of British Columbia, working under the supervision of Dr Joanna McGrenere. Her research interests are in the area of human computer interaction with a focus on the design of inclusive technology. In 2005, she was awarded a 3-year NSERC postgraduate scholarship to pursue her thesis research on increasing the accessibility of handheld technology for older adults. She also holds a M.Sc. in Computer Science (2004) and a B.A.Sc. in Computer Engineering (2001), both from the University of British Columbia.