ISSUE 89, SEPTEMBER 2007

Accessibility and Computing

ASSETS 2007 Doctoral Consortium

A regular publication of the ACM Special Interest Group on Accessible Computing

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 2007 Doctoral Consortium (DC) are working on. We had delayed this September issue to provide an opportunity for the DC participants to incorporate the feedback and suggestions they received at the DC session in their articles.

Sri Kurniawan

Newsletter editor

Inside This Issue

Association for Computing Machinery

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.

Accessmonkey: Enabling and Sharing End User Accessibility Improvements

Jeffrey P. Bigham

Department of Computer Science and Engineering,

University of Washington, Seattle, WA 98195 USA

jbigham@cs.washington.edu

Abstract

This paper proposes the Accessmonkey Framework for collaborative web accessibility improvement. The goal of the system is to provide a common platform on which end users and developers can create, share and use scripts that specify accessibility improvements. We propose browser plugins that will automatically retrieve user scripts from a common repository and apply helpful transformations to pages visited by users. End user interfaces will allow anyone to create and share new improvements and developer interfaces will allow content creators to edit and save changes scripts make to web content. Accessmonkey seeks to enable those most impacted by web accessibility challenges to directly and independently improve them. This work is being pursued with close consultation and participation by blind web users.

Introduction

Efficient web access for blind users can be challenging. When accessed using screen reader, information encoded visually is inaccessible, functionality requiring the use of the mouse is unavailable, and complex or lengthy content lacking semantic markup is inefficient to browse. We've shown that blind users browse less efficiently as a result and avoid content that is inaccessible [2]. Extensive research has explored the benefits of transforming content to make it more accessible. These tools generally operate either fully automatically or as part of improved developer tools. Comparatively little work has enabled end users to independently improve content. Existing user and developer tools could also benefit by enabling users to share new improvement strategies, and by offering a convenient mechanism by which third-party developers could extend the diversity of existing user and developer tools. The Accessmonkey Framework seeks to address these shortcomings.

Research goals

The goals of the Accessmonkey Framework are threefold:

1. **End User Improvement** - Enable independent improvement of web content by those without programming experience using an accessible interface designed for end users.
2. **Collaboration** - Facilitate sharing of improvements among both users and web developers. Provide a forum for users to help one another.
3. **Common Platform** - Provide a common platform for accessibility improvement that will help improvements reach end users. Technology for improving web accessibility often languishes because no convenient, widely-used platform exists for releasing it.

Proposed solution

The Accessmonkey Framework will provide 1) a common platform for accessibility improvement, 2) end user interfaces that enable users to find, make and share improvements to pages, 3) developer interfaces that enable developers to find, edit and incorporate accessibility improvements into their pages, 4) browser plugins that perform the specified accessibility improvements, and 5) a repository where users can post and retrieve accessibility improvement scripts. The illustration below shows a diagram of the proposed system.

The Accessmonkey Framework will build on existing end user programming tools, such as Greasemonkey [10], Koala [7] and Keyword Commands [8], with added functionality targeted at improving accessibility. Our system will address two shortcomings of these tools: First, writing scripts to improve accessibility with existing tools either requires programming knowledge or requires the use of interfaces that are not accessible. Second, sharing improvements is inconvenient because each tool requires users to visit a separate site and manually install scripts.

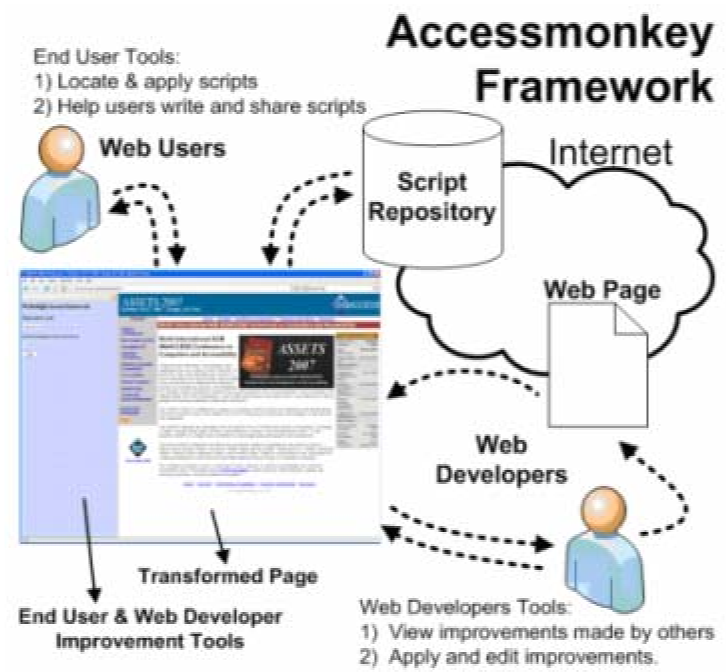

Figure 1: The Accessmonkey Framework enables web users to locate, apply, create and share scripts for accessibility improvement and enables web developers to incorporate those changes in the original pages.

End user improvement

We plan to develop interfaces that will enable blind web users who are not programmers to improve the accessibility of the web pages that they visit. Platypus [12] enables users to create Greasemonkey scripts by visually selecting items and then change their style, moving them elsewhere on the page or deleting them entirely. Keyword Commands [8] allows commands to be entered in pseudo-natural language and Koala [7] further relaxes this syntax and operates in the web domain. These extensions have limited non-visual support and lack features specific to improving accessibility.

Our interface will allow end users to improve content without programming. We are currently collecting many examples of the accessibility problems faced by blind users in order to isolate the specific transformations that should be supported by our tool. Users could, for instance, select text (using either the keyboard or mouse or through a keyword-based interface) and then specify that it should be a heading of a particular level. As other examples, users could add skip links or add list tags to information displayed as such. Future functionality could allow users to assign semantic roles to dynamic content from the ARIA [11] ontology or provide roles to arbitrary content from an ontology designed to improve accessibility [1, 13]. To develop our end user interface, we will use a user-centered design methodology.

Finding scripts

Our system will be designed to efficiently locate, install and apply user-created scripts. Current user scripting tools [10, 5, 7] do not automatically locate and apply scripts because the security of such scripts cannot be guaranteed. Scripts that send private information to a remote site or make other malicious changes are problems for social networking sites [6], and previous versions of Greasemonkey were shown to be susceptible to similar exploits [10]. Instead, existing systems require users to manually find and install scripts, offering a modest level of protection. User scripts need to be powerful yet appropriately constrained to enable arbitrary users to contribute improvements.

We will limit and rigorously secure the operations supported by our end user interfaces, and create a community rating system to help users find the best scripts. More advanced scripts that require more of the Javascript functionality will be thoroughly vetted by the community before being applied automatically.

Preliminary results

We have introduced the common platform component of Accessmonkey [4]. It extends Greasemonkey in two important ways: 1) Changes that are made to pages can be saved by developers in order to improve the original web page. 2) The system is available on multiple browsers and platforms.

We also demonstrated how Accessmonkey scripts can improve accessibility. Site- or page-specific scripts identify a common template or page element and alter it to be more accessible. For example, content is rearranged so that important information is read first and a dynamic menu is made accessible using only the keyboard. General scripts apply to all web pages. Our WebInSight script adds alternative text to web images [3], our context-focused browsing script emulates the Hearsay browser's CSurf [9] and our headings script automatically adds HTML heading tags.

Future and ongoing work

We are currently conducting focus groups with blind web users in order to study how to best create interfaces for selecting and annotating content in a way that will attract interest from the blind community, which will determine the success of this approach. We will soon begin designing, building and testing these interfaces. To identify the most useful improvements that Accessmonkey should enable blind users to provide, we are collecting examples of the accessibility challenges confronted during the everyday browsing of blind web users and isolating common problems. Finally, we are preparing to release versions of Accessmonkey for the Mozilla Firefox and Internet Explorer web browsers.

Further information can be found on our web page:
http://webinsight.cs.washington.edu/accessmonkey/

References

1. S. Bechhofer, S. Harper, and D. Lunn. Sadie: Semantic annotation for accessibility. In Proceedings of the International Semantic Web Conference (ISWC '06), 2006.
2. J. P. Bigham, A. C. Cavender, J. T. Brudvik, J. O. Wobbrock, and R. E. Ladner. WebinSitu: A Comparative Analysis of Blind and Sighted Browsing Behavior. In Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '07), 2007.
3. J. P. Bigham, R. S. Kaminsky, R. E. Ladner, O. M. Danielsson, and G. L. Hempton. WebInSight: Making Web Images Accessible. In Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '06), 2006.
4. J. P. Bigham and R. E. Ladner. Accessmonkey: A collaborative scripting framework for web users and developers. In Proceedings of the International Cross-Disciplinary Conference on Web Accessibility (W4A '07), 2007.
5. M. Bolin, M. Webber, P. Rha, T. Wilson, and R. C. Miller. Automation and customization of rendered web pages. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '05), 2005.
6. T. Jim, N. Swamy, and M. Hicks. Defeating script injection attacks with browser-enforced embedded policies. In Proceedings of the 16th International Conference on World Wide Web (WWW '07), pp. 601–610, 2007.
7. G. Little, T. A. Lau, A. Cypher, J. Lin, E. M. Haber, and E. Kandogan. Koala: capture, share, automate, personalize business processes on the web. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07), pages 943–946, 2007.
8. G. Little and R. C. Miller. Translating keyword commands into executable code. In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '06), pages 135–144, New York, NY, USA, 2006.
9. J. Mahmud, Y. Borodin, and I. Ramakrishnan. Csurf: A context-driven non-visual web-browser. In Proceedings of the International Conference on the World Wide Web (WWW '07), pp. 31–40, 2007.
10. M. Pilgrim. Greasemonkey Hacks: Tips & Tools for Remixing the Web with Firefox. O'Reilly Media, 2005.
11. Roadmap for Accessible Rich Internet Applications (WAI-ARIA Roadmap). World Wide Web Consortium, http://www.w3.org/TR/WCAG20/. 2007.
12. S. R. Turner. Platypus Firefox Extension. http://platypus.mozdev.org/. 2007.
13. Y. Yesilada, S. Harper, C. Goble, and R. Stevens. Screen readers cannot see (ontology based semantic annotation for visually impaired web travellers). In Proceedings of the 4th International Conference on Web Engineering (ICWE '04), pp. 445–458, 2004.

About the author

Jeffrey P. Bigham is a Ph.D. candidate in the University of Washington Computer Science & Engineering Department in Seattle, WA, USA. His interests center around improving the human interfaces used to access web content with a focus on improving the interfaces used by blind web users because of how inefficient access is for this group. Along with his advisor Richard E. Ladner, he started the WebInSight group at the University of Washington, which pursues research targeted at understanding challenges for web access by screen reader users and develops innovative solutions to address them. Prior to his work in web accessibility, Jeffrey pursued research in natural language processing and extracting factual information from large, unstructured resources like the web. http://www.cs.washington.edu/homes/jbigham/

Simulating HCI for Special Needs

Pradipta Biswas

Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD

University of Cambridge, England

pb400@ cam.ac.uk

Abstract

Computers offer valuable assistance to people with physical disabilities. However designing human-computer interfaces for these users is complicated. The range of abilities is more diverse than for able-bodied users, which makes analytical modelling harder. Practical user trials are also difficult and time consuming. We are developing a simulator to help with the evaluation of assistive interfaces. It can predict the likely interaction patterns when undertaking a task using a variety of input devices, and estimate the time to complete the task in the presence of different disabilities and for different levels of skill.

Introduction

Computers offer valuable opportunities to physically challenged people as it help them to engage more fully with the world. However designing and evaluating human-computer interfaces for these users is more complicated than that for able-bodied persons, since the range of abilities is more diverse. Their patterns of interaction are also significantly different from those of able-bodied users. Therefore, existing HCI models are hardly applicable to assistive interfaces. Assistive interfaces are generally evaluated by analysing log files after a user trial. However it is often difficult to find participants with specific disabilities. Petrie et. al. [6] take the approach of remote evaluation but still need to find disabled participants. In this context, a modelling tool that can simulate HCI of users with disabilities relieves the designer from searching for disabled participants to run a conventional user trial. However, research on assistive interfaces and HCI modelling do not overlap. Very few of the existing HCI models have considered users with disability. Researchers on assistive interfaces have concentrated on designing assistive interfaces for a particular application (e.g. Web Browser, Augmentative and Alternative Communication aid etc.), developing new interaction techniques (e.g. different scanning techniques) or developing novel hardware interfaces (head mounted switches, eye-gaze trackers, brain-computer interfaces etc.). They have not looked at designing a systematic modelling tool for assistive interfaces. I am developing a simulator to model HCI of disabled users. It can predict the likely interaction patterns of users when undertaking a task using a variety of input devices, and estimate the time to complete a task in the presence of different disabilities and for different levels of skill. The simulator can be used to compare several existing assistive interfaces and to evaluate new alternatives. Besides disability, I shall also address the shortcomings of existing HCI models and hope to develop a system that will be easier to use than the existing models and support both able-bodied and disabled users.

Related works

The GOMS family of HCI models (e.g. KLM, CMN-GOMS, CPM-GOMS) is mainly suitable for modelling the optimal behaviour (skilled behaviour) of users [3]. On the other hand, models developed using cognitive architectures consider the uncertainty of human behaviour in detail but have not been widely adopted. For example, developing a sequence of production rules for Soar or a set of constraints for CORE [9] is difficult. Usability issues for cognitive architectures are also supported by the X-PRT system [9] for the CORE architecture. Additionally, Kieras has shown that a high fidelity model cannot always outperform a low fidelity one though it is expected to do so [5]. Researchers have already attempted to combine these two forms of model to develop more usable and accurate models. Salvucci and Lee [7] have developed the ACT-Simple model by translating basic GOMS operations into ACT-R production rules. The model works well to predict expert performance but does not work for novices. Blandford et. al. [2] implement the Programmable User Model (PUM) by using the Soar architecture. They developed a program, STILE (Soar Translation from Instruction Language made Easy), to convert the PUM Instruction Language into Soar productions. However, this approach also demands good knowledge of Soar from an interface designer. The second problem of existing modelling approaches comes from the issues with disability. There is not much reported work on systematic modelling of assistive interfaces. The AVANTI project [8] models an assistive interface for a web browser based on some static and dynamic characteristics of users. However, this model does not address the basic perceptual, cognitive and motor behaviour of users and so it is hard to generalize to other applications. My user model [1] breaks down the task of user modelling into several steps that includes clustering users based on their physical and cognitive ability, customizing interfaces based on user characteristics and logging user interactions to update the model itself. However the objective of this model is to design adaptable interfaces and not to simulate users' performance. Keates et. al. [4] measured the difference between able-bodied and motor-impaired users with respect to the Model Human Processor (MHP) and motor-impaired users were found to have a greater motor action time than their able-bodied counterparts. The finding is obviously important, but the KLM model itself is too primitive to use.

Objective

Based on the previous discussion, Figure 1 plots the existing general-purpose HCI models in a space defined by the skill and physical ability of users. To cover most of the blank spaces in the plot, we set our objectives as follows:

Developing a model that can

1. Simulate HCI of both able-bodied and disabled users.
2. Work for users with different levels of skill.
3. Be easy to use for an interface designer.

Present status

We are now developing a simulator that takes a task definition and locations of different objects in an interface as input. Then it predicts the cursor trace and completion time, for different input device configurations (e.g. mouse or single switch scanning) and undertaken by persons with different levels of skill and physical disabilities. The architecture of the simulator is shown in Figure 2 and it consists of the following three components:

The Application model models the task currently undertaken by the user by breaking it up into a set of simple atomic tasks.

The Interface Model decides the type of interfaces to be used by a particular user and sets parameters for an interface.

The User Model simulates the interaction patterns of users for undertaking a task analysed by the task model under the configuration set by the interface model. It uses the sequence of phases defined by *Model Human Processor*. The perception model takes a list of keyboard and mouse events and a sequence of bitmap images of an interface as input and produces a sequence of eye-movements and the visual search time as output. I am developing the cognitive model by using two interacting Markov processes following the concept of dual space model. The motor behaviour model predicts the completion time and possible interaction patterns for performing an action. It will be developed by statistical analysis of cursor traces of disabled users.

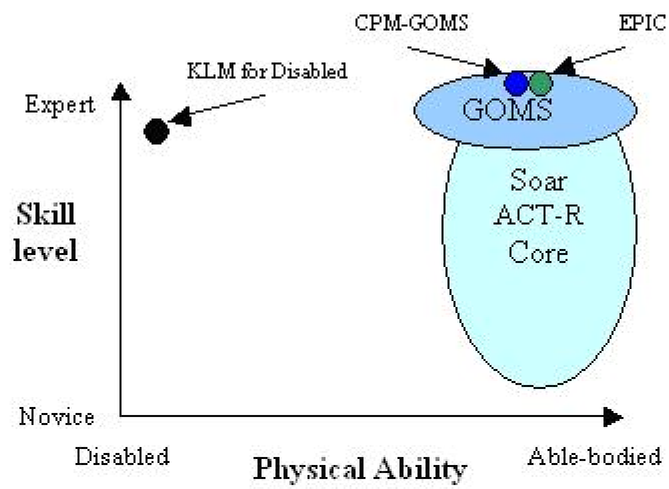

Figure 1. Existing HCI models w.r.t. skill and physical-ability of users

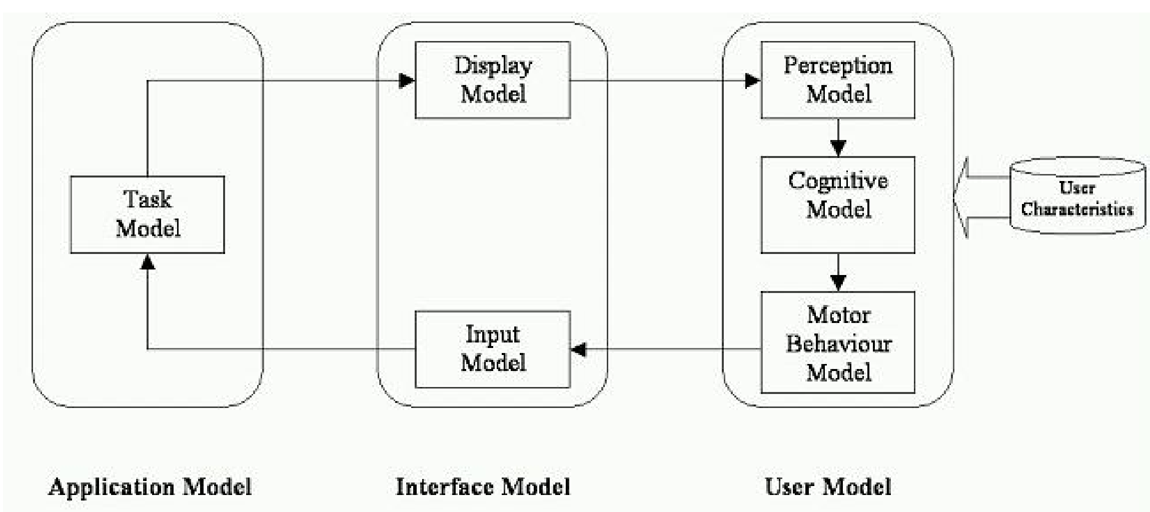

Figure 2. Architecture of the Simulator

We have confirmed the correctness of the model for novice users by an experiment with able-bodied persons. Our next step is to populate the remaining components of the models with more details and to validate them with an experiment with physically challenged people.

References

1. Biswas P. et. al., User Model To Design Adaptable Interfaces For Motor-Impaired Users, In *Proceedings of the Tencon '05 – IEEE Region 10 Conferences*, Australia, 2005, 1801-1806
2. Blandford A., Butterworthb R., Curzonb P., Models of interactive systems: a case study on programmable user modelling, *International Journal of Human-Computer Studies*, vol. 60 (2004), 149–200,
3. John, B. E. and Kieras , B. E., The GOMS family of user interface analysis techniques: Comparison and Contrast. *ACM Transactions on Computer Human Interaction*, Vol. 3 (1996), 320-351
4. Keates S., Clarkson J., Robinson P., Investigating the Applicability of User Models for Motion Impaired Users, In *Proceedings ASSETS 2000, ACM/SIGACCESS Conference on Computers and Accessibility*, November 13-15, 2000
5. Kieras, D. E.. Fidelity Issues In Cognitive Architectures For HCI Modelling: Be Careful What You Wish For. In *Proceedings of 11th International Conference On Human Computer Interaction (HCII 2005)*. Las Vegas, July, 2005
6. Petrie H. et. al., Remote usability Evaluations with disabled people. . In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '06) ,Montreal, Canada, April 22-27, 2006. ACM Press, New York, NY, 2006, 1133- 1141
7. Salvucci, D.D., Lee, F.J., Simple cognitive Modelling in a complex cognitive architecture, In *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems*, Fort Lauderdale, FL, 2003, 265–272
8. Stephanidis C., et. al., Adaptable and Adaptive User Interfaces for Disabled Users in the AVANTI Project, *Intelligence in Services and Networks, LNCS-1430*, Springer-Verlag 1998, 153-166
9. Tollinger I. et. al., Supporting Efficient Development of Cognitive Models At Multiple Skill Levels: Exploring Recent Advances In Constraint-Based Modeling, In *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems*, Portland, Oregon, USA,2005, 411 – 420

About the author

Pradipta Biswas is a research student at the University of Cambridge Computer Laboratory under the supervision of Prof. Peter Robinson. His work concerns about modelling Human-Computer Interaction of physically-challenged people so as to simulate the performance of different interfaces and interaction techniques. He is a recipient of Gates-Cambridge scholarship. He has completed his M. Tech. in Information Technology in 2006 from Indian Institute of Technology, Kharagpur, India. Previously he was a student of University of Kalyani, India and completed his B. Tech. in Information Technology in 2004.

Improving the Efficacy of Automated Sign Language Practice Tools

Helene Brashear

Georgia Institute of Technology, GVU Center, College of Computing, Atlanta, Georgia, USA

brashear@cc.gatech.edu

Abstract:

CopyCat is an America Sign Language (ASL) game, which uses gesture recognition technology to help young Deaf children practice ASL skills. Our database of signing samples was collected from user studies of Deaf children playing a Wizard of Oz version of the game at the Atlanta Area School for the Deaf. We have created an automatic sign language recognition system for the game. We believe that we can improve the accuracy of this system by characterizing and modeling disfluencies found in the children's signing.

Introduction

Since early childhood is a critical period for language acquisition, early exposure to ASL is key for deaf children's linguistic development [14,15]. Ninety percent of deaf children are born to hearing parents. Most of these parents do not know or are not fluent in sign language [7]. Often a child's first exposure to signing is at school.

CopyCat is a research prototype combining an interactive computer game with sign language recognition technology. CopyCat aims to assist young deaf children's language acquisition by interactive tutoring and real-time evaluation [12]. The goal is to encourage the linguistic transition from single, isolated utterances to phrase level signing. CopyCat is designed to support ASL-based communication between the user and characters in the computer game. The child is asked to sit in front of the computer which is equipped with a video camera for computer vision recognition system. He or she wears colored gloves with wrist-mounted accelerometers to assist the recognition. While playing the game, the child communicates with an animated character through ASL. This game is both mentally and physically engaging and allows the child to practice ASL.

Related Work

Sign languages are used around the world by the deaf and speech impaired as a means of communication. These sign languages use hand, body and face gestures as well as spatial structures to communicate information. Automatic sign language recognition (ASR) is the process of using sensors to collect data from a user's signing and use computers to recognize the signs. Sign languages are rich, multi-faceted languages, and their full complexity is beyond current gesture recognition technologies. As a result of this, most researchers focus on a subset or aspect of a sign language to reduce the complexity of their task. Some of these subsets are finger spelling, hand-based gestures, facial gestures, body gestures, inflections, and spatial use.

Research on sign language recognition has been done around the world, using many sign languages, including American Sign Language, Korean Sign Language, Auslan, Taiwanese Sign Language, Chinese Sign Language, Japanese Sign Language, and German Sign

Language. Many sign language recognition systems use Hidden Markov Models (HMMs) for their abilities to train useful models from limited and potentially noisy sensor data. Sensor choices vary from data gloves and other tracker systems to computer vision techniques using a single camera, multiple cameras, and motion capture systems to hand crafted sensor networks. For a more detailed summary of work in the field of ASR see the work of Ten Holt *et al* [10] and Loeding *et al* [13].

Holt *et al* [10, 9] provide a comprehensive summary of t problems of ASR research. Holt describes several significant problems specific to ASR:

- Distinguishing gestures from signs
- Context dependency (directional verbs, inflections, etc)
- Basic unit of modeling (what are the phonemes? how do we describe them?)
- Transitions between signs (movement epenthesis vs. coarticulation effects)
- Repetition (cycles of movement may vary in length)

Transitional Movements

Without modeling of the basic structure of signs (which is debated), it is problematic to use traditional phonemic methods of modeling. Transitions between signs include both coarticulation effects and movement epentheses. Coarticulation effects refer to the changing of signs when they overlap[11]. Movement epenthesis is the actual movement between signs when there is a difference between a signs ending location and the starting location of another sign [17]. Distinguishing gestures from sign is a fundamental problem, since many of these non-sign gestures occur during conversation and act as disfluencies.

Speech recognition typically trains co-articulation models in context by creating biphones and triphones. Gao et al used this approach in modeling Chinese Sign Language with HMMs [8]. Vogler and Metaxas modeled each transition individually as a phoneme using parallel HMMs [18]. Rule based methods have also been used; these methods use heuristics such as acceleration and velocity to detect borders and transitions [16].

Data Sets

Most data sets for automatic sign language recognition are scripted data sets collected in the laboratory by the researchers [9]. Though scripted datasets provide a good testing bed for developing research systems, the field of speech recognition has found that they are limited in their representation of how language is used and lack examples of common conversational artifacts such as accents (since they are often over-enunciated), disfluencies, and inflections [11]. Additionally conversational signing may contain register variation which may result in more or less formal signing depending on the signer's environment. This register variation can affect how signs are performed, what vocabulary is chosen and what grammar is used [17]. Datasets collected in formal, scripted settings may lack many of the important language facets that are needed to fully model the language for use in live recognition systems.

The dataset from the CopyCat project [1] is particularly interesting because it was designed to capture more conversational signing. The dataset was collected from multiple children in a school environment. The children signed to characters in the game as conversational partners in order to complete tasks within the game story line.

Research

The CopyCat dataset was collected from 9 children ages 8-11 at the Atlanta Area School for the Deaf. Children played the game for 5 sessions each, over a period of two weeks. There were 30 game phrases per a session and a game vocabulary of 20 different hand-based signs. The grammar was limited to `Subject+Preposition+Object` format and includes sentences of four, five and six sign lengths. The system used cameras, colored cotton gloves, and accelerometers as sensors.

The dataset was collected using a Wizard of Oz approach. The Wizard of Oz technique is an evaluation method which uses a human ``wizard'' to simulate the functionality that will eventually be provided by the computer. The Wizard is situated out of sight of the subject, receives the subject's input, and controls the system manually, thus emulating the missing functionality of the system [3]. During game play, the children interact normally with the computer game while a human Wizard simulates the computer recognizer and evaluates the correctness of the player's signing. The child is not aware of the Wizard's presence and believes that he/she is using a fully functioning system.

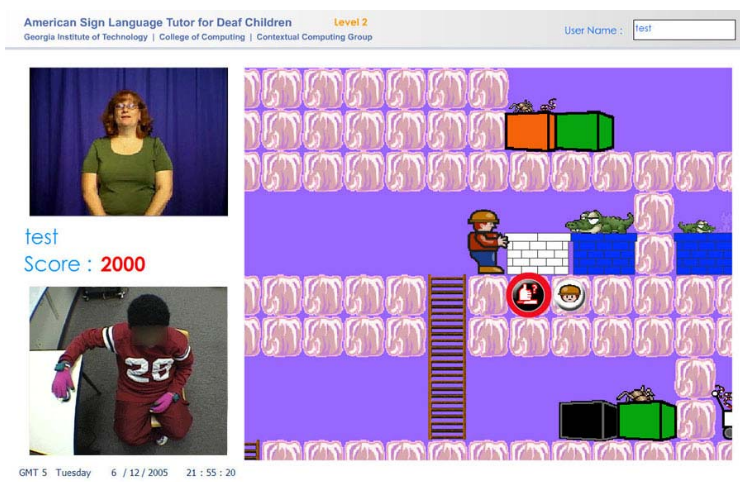

Figure 1: CopyCat screenshot: The child directs Abe (A), the hero, to defeat the villain (B). If the child is unclear on the game scenario, the he/she presses the help button (E) and a tutor video (C) of the properly signed phase plays. The child then presses the start button (F) and signs to the character Abe, making sure that Abe can ``see'' him or her in the camera view (D).

Thesis

My sign language recognition research is focused on creating user-independent models for recognition of hand-based American Sign Language gestures[1]. Our data set will be a collection of samples of the children interacting with the game via a Wizard of Oz setup at the Atlanta Area School for the Deaf (AASD). These models will be used to create our live, interactive game. Since most of the sign language sets (including our own past work) used for machine learning have been collected in a controlled, laboratory environment, these sets do not fully explore common disfluencies in sign. Our data set provides many samples of children signing naturally as they interact with the online characters. It has many examples of non-signing activities such as scratching and fidgeting and also includes false starts, hesitations, and pauses.

Thesis Statement: I will show that modeling the disfluencies that occur in a natural signing context can improve the accuracy of a sign recognition system for an ASL practice tool.

The transition from lab collected signing samples to real-world datasets for sign language recognition necessitates expanding models to include more diverse linguistic information. Just as the speech recognition community found that there is more to speech recognition than well-enunciated speech signals, there is more to sign language than perfectly performed signing. The linguistic scope of sign recognition is still largely limited by technology, with most groups focusing on hand gesture recognition.

Contributions: I will **identify significant gestures in our dataset**, including game vocabulary, communications directed towards game characters, and disfluencies in sign. We will enumerate and model the disfluencies that are relevant to our data set. I will **create an ontology for disfluency classification** for recognition purposes. This ontology will help group disfluencies by their structure and analyze their impact on the recognition system. I can **improve our system by modeling select classes of gestures from the ontology**. These models will be added to the recognition engine in order to improve the recognition rates.

Table 1: Categorization of sign examples on two axes: Clear vs. Unclear and Game Correct vs. Game Incorrect.

	Machine Learning	
	Clear	**Unclear**
Game Correct	Clear, accurate signing which follow the game grammar; these are the ``textbook'' correct samples	Signing which follows the game grammar, but may contain disfluencies; contains signing which is intelligible and grammatically correct, but may contain unusual linguistic artifacts
Game Incorrect	Clear, accurate signing which does not reflect game grammar; may contain good signs without any grammatical structure	Signing which does not follow game grammar and is incorrect; may include gibberish, wrong signs and disfluencies

Modeling Signs

The CopyCat dataset can be divided into four categories, as shown in Table 1. The traditional approach to pattern recognition datasets would be to choose the good examples (the intersection of ``Game Correct'' and ``Clear'' on Table 1). These samples would then be used to both build models and to evaluate the accuracy of these models. We have shown results from this approach for a dataset collected from an earlier phase of data collection for CopyCat [1].

We have previously used a single hidden Markov model for each sign along with a strong grammar designed to account for coarticulation effects, movement epentheses, and some pauses. These results showed user-dependent models with testing word accuracies from 90.80% to 95.65%. User- independent models were tested and showed word accuracies ranging from 76.90% to 92.62%. The signing skills of the participants varied, but the word accuracies show that their signing could be modeled and recognized with some success, even with such a small number of children to sample from.

The next step for expanding the functionality of the system is to move beyond these traditional techniques and recognize the signs using more detailed language models. The children have a variety of disfluencies in their signing both as a result of conversational mannerisms, as well as a result of their concentration and effort in the game. It is important to explore the kinds of disfluencies the children exhibit, as well as further analyze the structure of these disfluencies production in order to improve our modeling techniques. A live version of the game would need ASR that can differentiate between a badly formed sign and disfluencies such as head scratch or a long pause. Though there has been limited investigation by linguists, the work of Eklund [4] provides a short overview of some perspectives on disfluency research in sign languages.

Preliminary Results

A preliminary analysis of the data evaluated 16 of the 45 total sessions in order to further refine the groupings of sign. This evaluation included between one and three sessions from each child spread across various sessions. Of 514 total phrases, only 117 of them were from the ``Game Correct'' and ``Clear'' category from 1. The majority of the phrases (354) were in the ``Game Correct'' and ``Unclear'' category. These samples included a wide range of variations in signing that were conversationally correct, but not perfect. ``Game Incorrect'' phrases accounted for 43 of the samples, but only seven of those contained bad signs. The rest contained ``Clear'' or "Unclear" signing which was understandable, but wrong for the game (for example referring to the wrong color).

During game play of Castle Quest, the children interacted with Abe as a natural conversational partner. This interaction resulted in instances of ``out-of-band'' comments to the character. These instructions to Abe were not part of the game story line, but instead, were directly related to game play. Self-corrections accounted for a large part of the ``out-of-band'' comments and some of the disfluencies. Detecting self-corrections may be aided by some ``out-of-band'' comments such as `wrong` and `start again` or some non-sign gestures such as a hand waving ``erase'' gesture that several children exhibited. Studies have shown that adult signers appear to normally edit their sign during conversation (similarly to spoken language) and produce self corrections such as `wrong` or a head shake[2,6], however there have not been systematic studies of conversational repairs in ASL[5].

There is a wide variation in how pauses occur in the library of sign examples. Some pauses occur between signs at various stages: post-completion of the first sign, during the movement epenthesis, and immediately prior to the second sign. Hesitations and false starts are commonly interjected into parts of the phrase where children have to concentrate on the subject matter to successfully complete a game goal. These hesitations often occur mid-sign as the child attempts to think of the next sign in the phrase.

False starts are also sometimes symptomatic of extra concentration and self-corrections. In particular this seems to manifest itself as starting a sign with the wrong hand shape and then changing hand shapes, because the intended sign has changed. For example, the child starts to sign with the closed hand shape for the sign `spider`, then realized that they need to sign `alligator` and switches to the claw hand shape to begin the sign for `alligator`.

While there may be some consistency between non-sign gestures for the children, it is not clear whether this generalizes outside of this particular dataset. Fidgeting and wiggling account for some active variation in signing and that will probably generalize fairly well since it only causes slight permutations in the sensor data.

The children were inconsistent in their performance of signs which contain cycles. The actual repetition of cycles was sometimes more or less than required for the sign. Modeling this repetition will involve changing the structure of the models at a sign level.

Next Steps

It has become clear that it will be important to expand sign labeling schemes to include more linguistic information than the traditional ``gesture label'' by adding tags for variations in the sign that are a result of signer preference, accent, or regional variation. While some of these variations can be accounted for mathematically in the models, it is clear that a more detailed labeling scheme could provide more accurate models. The next step will be to begin to catalog these variations and develop approaches to modeling them.

Conclusion

In order to create live applications that use automatic sign language recognition, language models much be expanded from tradition gesture recognition based approaches to include more detailed linguistic information. Most automatic sign language recognition systems are built on scripted sessions collected in the laboratory. The CopyCat dataset provides an interesting alternative to these scripted systems because it was collected using Wizard of Oz methods. These methods allowed researchers to collected data of children signing conversationally to characters in the game. Further analysis reveals a variety of interesting variations and disfluencies in the children's sign which serve to inform new sign language modeling techniques and bring live automatic sign language recognition closer to a reality.

Acknowledgments

This work is supported by the NSF, Grants # 0093291, # 0511900, and the RERC on Mobile Wireless Technologies for Persons with Disabilities, which is funded by the NIDRR of the US Dept. of Education (DoE), Grant # H133E010804. The opinions and conclusions of this publication are those of the grantee and do not necessarily reflect those of the NSF or US DoE.

Special thanks to students and staff at the Atlanta Area School for the Deaf for their generous help with this project.

References

1. H. Brashear, K.-H. Park, S. Lee, V. Henderson, H. Hamilton, and T. Starner. American Sign Language Recognition in Game Development for Deaf Children. In Assets '06: Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, New York, NY USA, 2006. ACM Press.
2. V. L. Dively. Conversational Repairs in ASL. In C. Lucas, editor, Pinky extension and eye gaze: Language use in Deaf communities, pages 137-169, Washington DC, 1998. Gallaudet University Press.
3. A. Dix, J. Finlay, G. Abowd, and R. Beale. Human-Computer Interaction, chapter 6.4 Iterative Design and Prototyping. Prentice Hall, 2004.
4. R. Eklund. Disfluency in Swedish Human-Human and Human-Machine Travel Booking Dialogues: Dissertation No. 882. PhD thesis, Linköping Studies in Science and Technology, Linköping University, Sweden, 2004.

5. K. Emmorey. Language, Cognition, and the Brain: Insights From Sign Language Research. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2002.
6. K. Emmorey and B. Falgier. Talking about space with space: Describing environments in ASL. In E. A. Winston, editor, Story telling and conversations: Discourse in deaf communities, pages 3-26, Washington DC, 1999. Gallaudet University Press.
7. Gallaudet. Gallaudet University. Regional and National Summary Report of Data from the 1999-2000 Annual Survey of Deaf and Hard of Hearing Children and Youth. Washington, D. C., 2001.
8. W. Gao, J. Ma, J. Wu, and C. Wang. Sign language recognition based on hmm / ann / dp. In International Journal of Pattern Recognition and Artificial Intelligence, volume 5, pages 587-602, 2000.
9. G. T. Holt. The Eye of the Beholder: Automatic Regonition of Dutch Sign Language. Master's thesis, University of Groningen, Netherlands, 2004.
10. G. T. Holt, P. Hendriks, and T. Andringa. Why Don't You See What I Mean? Prospects and Limitations of Current Automatic Sign Recognition Research. Sign Language Studies, 6(4), Summer 2006.
11. D. Jurafsky and J. H. Martin. Speech and Language Processing. Prentice Hall, Upper Saddle River, New Jersey, 2000.
12. S. Lee, V. Henderson, H. Hamilton, T. Starner, H. Brashear, and S. Hamilton. A Gesture-based American Sign Language Game for Deaf Children. In Proceedings of CHI, pages 1589-1592, Portland, Oregon, 2005.
13. B. L. Loeding, S. Sarkar, A. Parashar, and A. I. Karshmer. Progress in Automated Computer Recognition of Sign Language. In Proceedings of International Conference on Computers Helping People with Special Needs, pages 1079-1087. Springer-Verlag Berlin Heidelberg, 2004. Lecture Notes in Computer Science.
14. R. I. Mayberry and E. B. Eichen. The Long-Lasting Advantage of Learning Sign Language in Childhood: Another Look at the Critical Period for Language Acquisition. Journal of Memory and Language, 30:486-498, 1991.
15. E. L. Newport. Maturational Constraints on Language Learning. Cognitive Science, 14:11-28, 1990.
16. H. Sagawa and M. Takeuchi. A method for recognizing a sequence of sign language words represented in a Japanese Sign Language sentence. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pages 434-439, Grenoble, France, March 2000.
17. C. Valli and C. Lucas. Linguistics of American Sign Language: An Introduction. Galludet University Press, Washington DC, 1998.
18. C. Vogler and D. Metaxas. Toward scalability in asl recognition: Breaking down signs into phonemes. In Proceedings of the Gesture Workshop (Gif-sur-Yvette, France), volume 1739, pages 211-224. March 1999. Springer Lecture Notes in Artificial Intelligence.

About the author

Helene Brashear is currently a PhD student at the College of Computing at the Georgia Institute of Technology. Her research focus is in using pattern recognition to create intelligent systems for assistive technology. She is working on her dissertation "Improving the Efficacy of Automated Sign Language Practice Tools" with her advisor, Dr. Thad Starner.

Using Networked Multimedia to Improve Educational Access for Deaf and Hard of Hearing Students

Anna C. Cavender

Computer Science and Engineering, University of Washington, Seattle, Washington

cavender@cs.washington.edu

Educational technology has the potential to better include deaf and hard of hearing students in the academic mainstream. This research involves development and testing of a classroom platform for deaf and hard of hearing students to access remote interpreters and captioners, avoid visual dispersion, and facilitate interaction in the classroom.

Introduction

Entering mainstream universities (at all levels) involves extra challenges for people who are deaf and hard of hearing: skilled sign language interpreters and captioners with advanced domain knowledge can be difficult to find; multiple visual channels of information in the classroom can be difficult to juggle; and collaboration inside and outside the classroom is often strained due to language barriers.

Classroom technology research is currently improving educational experiences for all students and this creates opportunities to better include deaf and hard of hearing students. Wireless networks, data projectors, and portable computing devices can allow remote interpreters, support sharing and capture of instructional materials, and provide additional communication channels for all. A more digital academic environment creates an opportunity for customization to better suit the needs of individual students.

This research will investigate and develop technology to help manage the many academic tasks required of deaf and hard of hearing students. Development will parallel other educational technologies (namely Classroom Presenter [1] and ConferenceXP [2]) so that technology for deaf students will be similar to those used by all students.

Background and Motivation

Students currently utilize an array of accommodations in academic settings including: interpreters, real-time captioners, hearing aids, FM systems, and note takers. However, several issues currently create extra obstacles for deaf and hard and hearing students in university-level classes:

Isolation

As more deaf students enter mainstream universities, there is a growing need for skilled sign language interpreters and real-time captioners with specialized knowledge in advanced courses. Finding an appropriate interpreter considering the variety of topics available can be a challenge, especially for complex courses such as Complexity Theory. The best interpreter for a given student may be located at another university in another state or country.

Dispersion

Because deaf students receive nearly all classroom information visually, they must juggle their visual attention between instructor, slides, interpreter and/or captioner, and personal notes or handouts. Due to this juggling, information can easily be missed. In fact, many deaf students request a student note-taker in order to eliminate at least one of their many visual tasks. While this helps ease visual burdens during class, the student may miss out on the value of taking and studying personal notes.

Exclusion

In spite of the plethora of possible accommodations, communication and participation within the classroom can be strained. Access to information is not equivalent, especially in terms of delivery time, and this makes call-and-response or question-and-answer techniques unfair. Furthermore, involvement outside the classroom (such as project groups meeting and impromptu study groups) can inadvertently exclude deaf or hard of hearing students.

Proposed Solution

Technology has potential to alleviate these problems, encourage participation, and enhance learning for all students.

Networking

Collaboration between universities through the existing multimedia cyber-infrastructure (software, hardware and other technologies including human expertise) would allow better access to skilled interpreters familiar with specialized, university-level topics, creating more opportunity for matching the best interpreter with specific content. This approach will also allow different types of students to receive differing accommodations based on preference. For example, one student may prefer a remote sign language interpreter while another student prefers real-time captioning.

Consolidation

Consolidating visual content into one device may prevent missed information due to the visual juggling act that many deaf students experience. Miller et. al. proposed using transparent video and overlaid digital ink to reduce the visual distance from the interpreter (video) and the student's notes (digital ink) [5]. Taking this concept one step further, imagine a Tablet PC application that allows students to see the instructor, presentation, and accommodation of choice all on one screen and all within visual distance of personal annotations.

Collaboration

Technology has been shown to enhance education in the classroom and these "digital" environments open up new possibilities for leveling the academic playing field. The University of Washington's Classroom Presenter uses a system of networked Tablet PCs allowing students to electronically submit their work and/or questions to the instructor who can then choose to display submissions and digital ink on lecture slides [1]. ConferenceXP, developed at Microsoft Research, provides the infrastructure for networking the Tablet PCs, and is also used for audio and video distance learning and classroom capture [2]. These two projects will provide a backbone for accessible classrooms for deaf students.

Given the scenario where all students are equipped with a networked Tablet PC, an additional opportunity exists for student collaboration. Kam et. al.'s LiveNotes uses digital ink over lecture slides to encourage group conversations and cooperative note-taking during lectures [4]. This idea could be used to bridge the cultural and language gap between hearing and deaf students and encourage group work.

As academic environments become more digital, capture and retrieval introduce interesting areas to improve content accessibility. Synchronization of video feeds, digital ink, and presentation materials could result in better preservation and easier post-class access, much like eClass [3] and other classroom capture techniques [6].

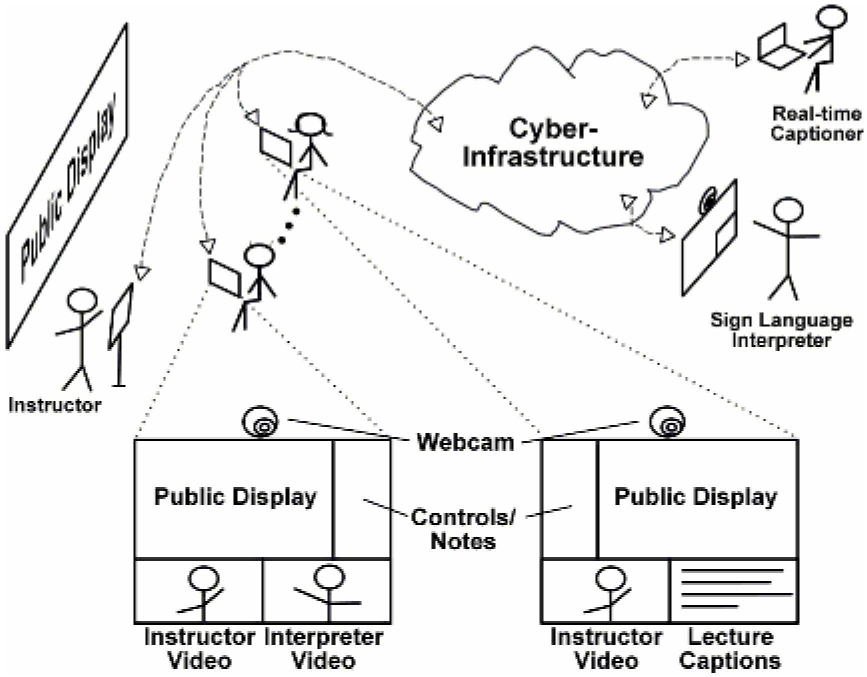

Figure 1: Here, the cyberinfrastructure brings remote interpreters and captioners into the classroom. The instructor uses a microphone, earpiece, and laptop camera to relay audio, video, and presentation materials to the remote interpreter. Students have access to presentation, instructor, accommodation of choice, and their own notes. Students' webcams relay questions and discussions through the interpreter to the rest of the class.

Research Goals and Status

Our primary research goal is to find ways to increase involvement of deaf and hard of hearing students in university academics. Improving access to resources and providing additional communication channels will involve participatory design and iterative development. Solutions will be viable for traditional classroom environments as well as for lab sessions, study groups, and project meetings.

We are currently collaborating with Rochester Institute of Technology (RIT), home of the National Technical Institute for the Deaf (NTID) supporting over 400 deaf students in the academic mainstream, over 120 sign language interpreters, over 50 captioners. NTID has conducted extensive research on interpreting and captioning in tertiary education. Collaboration will jump start the design of a “cyber-community” to enable deaf and hard of hearing students to advance academically through technological access. Participants will include deaf students, interpreters, captioners, sign language linguistic researchers, educational technology researchers, and cyber-infrastructure experts.

An excellent opportunity for evaluation and feedback of the proposed research is the Summer Academy for Deaf and Hard of Hearing Students hosted each summer at the University of Washington. The top ten college freshmen or sophomore applicants join the program to take college courses focused on computer science and related fields. Because the academy involves mainstream, 9-week courses, it presents an ideal test-bed.

References

1. R. Anderson, J. Chen, L. Jie, N. Li, N. Linnell, V. Razmov, and F. Videon. Supporting an interactive classroom environment in a cross-cultural course. In ASEE/IEEE Frontiers in Education, *to appear*., 2007.
2. J. Beavers, T. Chou, R. Hinrichs, C. Moffatt, M. Pahud, and J. V. Eaton. The learning experience project: Enabling collaborative learning with ConferenceXP. Microsoft Research Technical Report MSR-TR-2004-42, April 2004.
3. J. A. Brotherton and G. D. Abowd. Lessons learned from eclass: Assessing automated capture and access in the classroom. In CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, 121–155, New York, NY, USA, 2004. ACM Press.
4. M. Kam, J. Wang, A. Iles, E. Tse, J. Chiu, D. Glaser, O. Tarshish, and J. Canny. Livenotes: a system for cooperative and augmented note-taking in lectures. In CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems, 531–540, New York, NY, USA, 2005. ACM Press.
5. D. Miller, J. Culp, and D. Stotts. Facetop tablet :: note-taking assistance for deaf persons. In Assets '06: Proceedings of the ACM SIGACCESS conference on Computers and accessibility, 247–248, New York, NY, USA, 2006. ACM Press.
6. S. Mukhopadhyay and B. Smith. Passive capture and structuring of lectures. In MULTIMEDIA '99: Proceedings of the ACM international conference on Multimedia (Part 1), 477–487, New York, NY, USA, 1999. ACM Press.

About the author

Anna Cavender is pursuing a Ph.D. in Computer Science and Engineering at the University of Washington. She graduated with a B.S. in Computer and Information Science from the University of Oregon in 2004. Her research interests lie in Human-Computer Interaction, specifically assistive technology for people with disabilities. She was the recipient of the Computing Research Association (CRA) Outstanding Undergraduate Award in 2004 and is a National Science Foundation (NSF) Fellow.

Enhancing accessibility through correction of speech recognition errors

John-Mark Bell

Learning Societies Lab, School of Electronics and Computer Science, University of Southampton, Southampton, UK.

jmb@ecs.soton.ac.uk

Abstract

The learning within lectures of hearing-impaired students can be hindered by errors in captions generated by speech recognition. My research intends to address this problem by investigating ways of correcting these captions. I summarise approaches to automatic error correction and describe the preliminary studies that have been conducted. These studies show that human editors set a tough benchmark for automatic correction to meet and indicate that automatic correction is feasible. Finally, I summarise my intention to develop a correction framework that will permit quantitative and qualitative testing of correction methods.

Introduction & motivation

University lectures are not an accessible environment for students with hearing difficulties. In lectures, information is mainly imparted through the words that the lecturer speaks. Although visual media (e.g. Powerpoint slides) are often used, these are usually a complement to what the lecturer says and not a complete representation of all that is taught in the lecture. Currently, a number of approaches are used to tackle this problem. These include employing a sign-language interpreter, a stenographer, or third-party note-taker to capture the information content of a lecture. Unfortunately, sign-language interpreters and stenographers are a limited resource. Thus it can be difficult to find suitably qualified interpreters and stenographers are often employed in more financially rewarding environments (such as court reporting). Third-party note-takers are often students who are taking the same course. The quality of the notes is variable and there is no guarantee that they capture the full information content of the lecture.

Existing research has investigated the use of real-time, automatic speech recognition (ASR) based, captioning of lectures [1]. It shows that ASR captioning can, in theory, make lectures accessible for hearing-impaired students by providing access to the words the lecturer speaks. The captions can either be displayed upon a screen at the front of the lecture theatre for all students to view or on a personal display client running on a student's laptop or PDA. Bain et al [1] also present anecdotal evidence that ASR transcription is of use to those without hearing impairment (e.g. those who have difficulty taking in-class notes). ASR-generated captions also pave the way to automatically generated lecture transcripts that can be placed online after the lecture for revision purposes.

However, the results of the research show that "nearly 40 percent of faculty participants reached the benchmark of at least 85 percent accuracy" (where "accuracy" is defined as the number of correct words in the ASR output divided by the total number of words spoken) and the mean accuracy rate across all participants was 77% with a standard deviation of

9.58. From this, it is clear that over 60% of lecturers did not reach the required accuracy rate in their lectures and that improvement is required.

Errors within lecture transcripts can be particularly problematic as they may distort the lecture content, either rendering it meaningless or, worse, imparting incorrect information. For example, it is not uncommon for ASR-generated captions to insert or remove "not" from a sentence, which clearly distorts the meaning of what was said.

The focus of my research is to improve the experience of hearing-impaired students within university lectures. As it is unlikely that significant improvements in the accuracy of ASR will be achieved in the short term, this will be approached through investigation into methods of correcting ASR output. The intention is to determine whether automatic post-processing of ASR output offers any significant improvement over the raw ASR output within the educational setting.

Related work

The impact of errors upon ASR usage is application dependent. For example, an ASR-based airline ticket booking system might be more susceptible to certain types of recognition error than a lecture transcription system. This results in a number of error handling strategies. These can be divided into two broad categories; strategies for detecting the location of errors and mechanisms for correcting errors. It is important to note that this categorisation is entirely conceptual – in practice, the separation may be implicit within an error handling technique.

The initial problem is to detect reliably the identity of errors. Sarma et al [6] propose using contextual analysis to identify likely errors. This exploits the relationship between words in a given context, usually over a larger range than afforded by the trigram models used by an ASR engine. By generating statistics for the likelihood of a word occurring given the words around it, it becomes possible to predict whether a given word is in error.

Another approach to error detection is to apply Naive Bayes classifiers to features of the ASR output. One possible feature set would be the confidence scores associated with the output words, the alternate hypotheses for a given word and the difference in scores between output words, as discussed by Zhou et al [9]. However, as these approaches are entirely statistical, there is no guarantee that they will successfully identify all errors in the output. Nor do they avoid false positives – i.e. not all words flagged as being in error are truly errors.

After detecting the potential presence of errors, it becomes necessary to correct the errors (or mitigate their effects). One way of solving the error correcting problem is to employ humans to edit ASR output. This was investigated by Wald et al [8]. Their study shows that, on average, a human editor may correct 23.9% of the errors in ASR output in real time. To correct an error, the user has to detect, select, and then edit it.

Alternatively, a technological approach may be used. For example, previous research into error correction within dialogue systems (such as the airline booking system mentioned previously) has shown that statistical post-processing of ASR output can have a beneficial impact upon accuracy [5]. This is achieved through use of a fertility channel model to "translate" between the ASR output and what was actually said. Dialogue systems can also enter a corrective sub-dialogue where the human participant performs the correction through prompting by the dialogue system (e.g. [2]). In situations where keywords are of paramount importance, utterance verification may be used to detect and selectively correct errors [7].

Other correction approaches include using probabilistic techniques or knowledge bases to reorder candidate lists [4].

Solution & methodology

Existing ASR engines generally have some form of facility that permits editing and correction of the output by the end user. This is designed for use during dictation or in off–line editing. Therefore, it is not well suited to the requirements of use in a lecture. Usually, when corrections are made using this facility, the ASR engine's statistical model is updated to reflect the correction. This allows the ASR engine to learn words and phrases that are commonly used by the speaker. Additionally, the impact of a single correction upon the ASR engine's statistical model is likely to be small, due to the engine's model being the result of a large corpus of training data.

The approaches identified could lead to an automatic correction process operating on the output of an ASR engine (i.e. as a post-processing step). This would be more suited to a lecture scenario and could interface with the manual editing system described by Wald et al [8].

Domain-specific processing

Generalised ASR engines are trained using a large corpus of data, which is non-domain-specific in nature. Due to this non-specificity, it is likely that the engine will make mistakes when encountering domain-specific content, even if all words are in the engine's vocabulary. The intention is to produce word-based statistical models that would be able to compensate for the ASR engine's genericity by post-processing the output. These models would contain information about the kind of language found in lectures and specific knowledge about the subject matter. For example, techniques from the field of machine translation may be used to produce a model that "translates" between ASR output and what was actually said.

Linguistic analysis

ASR engines' linguistic knowledge is based upon statistical information generated from a large corpus of training data. An approach using linguistic analysis would use more explicit knowledge of language constructs to detect grammatical errors. This is likely to be particularly useful on a higher level than the statistical word sequence model used by ASR engines. As ASR engines' models are very narrow (limited to two or three words), higher-level concepts such as grammatical correctness are not tested for. The use of linguistic analysis is intended to introduce this higher-level consideration of the output.

Candidate list analysis

As a side effect of the statistical approach used by ASR engines, a list of alternative hypotheses is commonly generated. This list contains words or phrases that are considered to be the most likely alternatives for the output word. Therefore it is the ideal place to look for potential correct output. For example, if a system has access to content-related information (such as Powerpoint slides or lecture notes), it could use this to hypothesise the correct word in the candidate list.

Uses of phonemic information

Phoneme information may be used to mitigate the effect of segmentation errors within the ASR engine's acoustic modelling stage. Anecdotal evidence suggests that, in a reasonably large number of cases, mis-segmentation is the likely cause of a misrecognition. Therefore, if the phonemic information is available on output, a phoneme-based re-segmentation technique might be used to attempt to correct this situation. Phonemic information may also be used to counteract out-of-vocabulary (OOV) errors through the use of a phoneme-sequence to word dictionary.

Current work

So far, some preliminary investigations have been undertaken into the impact of multiple, independent, human editors upon the output and into the feasibility of utilising candidate list information to automatically correct ASR output. A further study has investigated the feasibility of using a machine-translation style model to correct the ASR output automatically.

Effectiveness of multiple human editors

Raw data collected by Wald et al [8] were re-evaluated to gain some insight into the effect of multiple human editors working independently, thus giving a benchmark against which automatic correction techniques may be compared. In the original study, five users attempted correction of errors in sixteen separate test cases, using a prototype tool. Assuming the existence of a system capable of collating the output of their editing, I wished to determine what impact the addition of editors would have.

To do this, each combination of editors was examined and their output merged, to find the combination that produced least errors for each of 1, 2, 3, 4 and 5 editors. On average, a single editor was capable of correcting 24% of the errors in the transcript. In the best case, using more than two editors gains little extra benefit (two editors corrected 44% of the errors, on average). It is likely that the majority of editors are correcting the same (or very similar) errors, thus the remaining errors are less likely to be corrected by adding further editors.

Candidate list evaluation

An experiment was carried out into the improvement possible when considering the candidate list generated by an ASR engine. It was assumed possible to detect erroneous output reliably. A metric was used to obtain a theoretical upper bound on the achievable level of improvement to the output transcript's error rate. The metric treated as correct all erroneous words for which the correct transcription could be found in the candidate list, no matter where in the list the correct transcription occurred. The results are promising and indicate that, for a mean initial word error rate of 22%, an absolute reduction in error rate of 7% may be achieved with a standard deviation of 1.67. This corresponds to 32% of the errors in the ASR output being corrected.

Machine translation model

A preliminary investigation was conducted into the feasibility of using a machine translation model to correct errors in ASR output automatically. This was based upon the larger-scale experiments (from the natural language translation field) described by Casacuberta [3]. These previous experiments were designed to determine the better of two model training techniques. Essentially, it involved creating "extended symbols", which mapped from the input word to the output word sequence, then using these to create a finite-state transducer.

My experimentation was upon a reasonably small corpus, which was divided into four blocks. In the experimentation, three types of model were created – bigram, trigram and four-gram, and two types of extended symbol were used (type-i and type-ii). This resulted in 6 models being built and tested, with the aim being to discover:

1. The usefulness of machine translation models in this situation
2. The percentage of errors in the ASR output that could be corrected
3. The most successful model

A cross-validation experiment was then carried out in which three blocks were used for training each model and the remaining block was used for testing. This was repeated four times (once for each permutation of the blocks). All words were in vocabulary to eliminate the potential for out-of-vocabulary errors.

The results of this investigation are promising. They show that the machine translation approach could result in a significant reduction in the number of errors in the ASR output. The type-ii extended symbol was clearly better than the type-i extended symbol, regardless of whether a bigram, trigram or four-gram model was built. No difference in improvement was observed between trigram and four-gram models, which is likely to be the result of the small size of the corpus. Both trigram and four-gram models performed better than the bigram models (for their respective symbol type). Overall, the best model (type-ii, tri/four-gram) was capable of correcting 27.7% of the errors in the raw ASR output.

Future Work

Future work will focus upon finding reliable methods for identifying, isolating and correcting erroneous sections of ASR output. This will consider the use of linguistic features (such as parts of speech) and confidence scores for error detection. It will also consider the use of domain-specific knowledge and statistical modelling to perform error correction. This will be followed by the development of a framework into which implementations of error detection and correction techniques may be plugged. The framework will permit quantitative testing of individual and combined techniques; thus allowing comparison of them. Further to this, qualitative testing involving users of the system will be conducted with the aim being to demonstrate its usefulness and highlight areas where extra improvement would be beneficial.

References

1. K. Bain, S. Basson, A. Faisman, and D. Kanevsky. Accessibility, transcription and access everywhere. *IBM Systems Journal*, 44(3):589–603, July 2005.
2. C. Bousquet-Vernhettes, R. Privat, and N. Vigouroux. Error handling in spoken dialogue systems: toward corrective dialogue. In *Proceedings of Workshop on Error Handling in Spoken Dialogue Systems*, 2003.
3. F. Casacuberta. Inference of finite-state transducers by using regular grammars and morphisms. In *Grammatical Inference: Algorithms and Applications*, volume 1891 of *Lecture Notes in Computer Science*, pages 1–14. Springer-Verlag, 2000.
4. H. Lieberman, A. Faaborg, W. Daher, and J. Espinosa. How to wreck a nice beach you sing calm incense. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, pages 278–280, New York, NY, 2005.
5. E. K. Ringger and J. F. Allen. Error correction via a post-processor for continuous speech recognition. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 427–430, Atlanta, GA, 1996.
6. A. Sarma and D. D. Palmer. Context-based speech recognition error detection and correction. In *Proceedings of the HLT-NAACL Conference: Short Papers*, pages 85–88, Boston, MA, 2004.

7. A. Setlur, R. Sukkar, and J. Jacob. Correcting recognition errors via discriminative utterance verification. In *Proceedings., Fourth International Conference on Spoken Language*, volume 2, pages 602–605, October 1996.
8. M. Wald, J.-M. Bell, P. Boulain, K. Doody, and J. Gerrard. Correcting automatic speech recognition errors in real time. *International Journal of Speech Technology*, In Press.
9. L. Zhou, J. Feng, A. Sears, and Y. Shi. Applying the naive bayes classifier to assist users in detecting speech recognition errors. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 183b–183b, 2005.

About the author

John-Mark Bell is a PhD student within the Learning Societies Lab in the School of Electronics and Computer Science at the University of Southampton. His studentship, which began in October 2006, is funded by an EPSRC grant and is supervised by Dr Mike Wald and Dr Andrew Gravell. His undergraduate study was also at the University of Southampton, being awarded an MEng in Computer Science in 2006. John-Mark's research interests include accessibility and universal design, with a specific focus upon hearing impairment. His PhD is focussed upon improving the accessibility of university lectures for hearing-impaired students, especially through improvement of captioning

Increasing the accessibility of pen-based technology: An investigation of age-related target acquisition difficulties

Karyn Moffatt

Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

kmoffatt@cs.ubc.ca

Abstract

This paper describes the author's dissertation research on improving the accessibility of pen-based technology. The first step for this research was to gather information on the underlying causes of target acquisition difficulty. To meet this goal, a controlled laboratory study was conducted, which uncovered three sources of difficulty: slipping, drifting, and missing just below. The remaining work will be to address these difficulties by implementing new interaction techniques and experimentally evaluating their effectiveness.

Introduction

The average age of the world's population is increasing. As of the 2000 Census, 12.4% of the U.S. population (approximately 35 million individuals) was aged 65 or over. Moreover, it is projected that by 2050 this proportion will have risen to over 20% [4]. Many older adults prefer to live independently [9]; however, the prevalence of cognitive and sensory age-related impairments can make this challenging.

Technology is increasingly being promoted as a means of addressing these age-related impairments and enabling individuals to live more independently (e.g., [2, 6, 7, 10] to name a few). Because they are small, mobile and powerful, handheld technologies such as Personal Digital Assistants (PDAs) and Tablet PCs are appealing platforms for these endeavors. However, for these technologies to be viable for this purpose, it is essential that older users be able to perform basic interactions with them (such as selecting an icon or menu item).

Age-related impairments rarely occur in isolation, and as such many individuals have associated impairments; often these impede their ability to interact with small devices. For example, in our own work designing mobile technology for cognitively impaired individuals [6], we informally observed many participants struggling with target acquisition using a stylus (e.g., selecting an icon or a menu item). This has motivated us to gain a better understanding of pen targeting difficulties and to ascertain the extent to which age is a factor.

The high-level objective of this thesis is to increase the accessibility of pen-based technology (such as Tablet PCs), by investigating mechanisms for assisting individuals, and in particular older individuals, to select more easily using pen technology.

Background

There has been considerable research aimed at developing improved target acquisition techniques. However, despite these advances, point and tap (i.e., selection by (i) tapping down, (ii) possibly moving the pen, and (iii) tapping up, with selection determined based on the location of the tap up) remains the de facto standard. Unfortunately, as noted above, many users find this action difficult to perform accurately and efficiently.

We have identified three characteristics that span the majority of previous work and that we believe have limited its ability to address the basic acquisition needs of a wide range of users. Firstly, there has been a narrow focus on young-healthy adults, who can more easily adapt to different techniques. There are many parameters, including a user's sensory and motor ability, that are likely to affect target acquisition and manipulation skill. Thus, a broader perspective can be gained by examining a range of users and abilities. Secondly, there has been a focus on evaluation with a single, typically highly constrained task. It is important to include multiple tasks to capture both concrete comparative measures and complex interaction. Finally, much of the focus has been on designing and evaluating novel techniques over developing a deeper understanding of how users manage basic tapping. Focusing on developing new techniques and evaluating them against the status quo (point and tap) has led research towards gross measures of overall speed and accuracy. While these measures provide comparative data about which technique is superior, they do not reveal underlying limitations or tell us where innovation is still needed.

In light of this, our approach is to try to fill this niche by adopting the following three-phase approach: (1) gather information on the underlying causes of target acquisition difficulties across a range of ages and tasks, (2) develop new interaction techniques to better support pen-based interaction, and (3) experimentally evaluate these new techniques against standard point and tap. In the following sections we elaborate on these phases. To date, we have completed the first phase. The second and third phases are currently underway.

Understanding the Sources of Difficulty

In the first phase, we conducted a controlled laboratory experiment to examine target acquisition difficulties across the adult lifespan. Specifically, our main goal for this study was to perform a detailed analysis of the types of difficulties users encounter while tapping to acquire targets. Additionally, we were interested in determining whether these difficulties vary in terms of their nature and severity with age or task situation. Full details of this study can be found in [5]; here we provide a summary, highlighting the key findings.

To address the goals outlined above, we performed an evaluation of two selection tasks (multi-dimensional discrete tapping and menu selection) across three adult age groups (young: 18–54, pre-old: 55–69, and old: 70–85). Participants completed both tasks, which were counterbalanced. We included two tasks to gain a better understanding of how task might affect targeting ability, especially in terms of accuracy. A tapping task was selected because, it is the gold standard for evaluating input techniques, and provides well understood measures of speed and accuracy. A menu task was selected because it provides a greater degree of realism than the tapping task, and may require slightly more cognitive effort.

Our results revealed three primary sources of target acquisition difficulty: missing just below, drifting and slipping. Slipping was specific to older users and common to both tasks, while the other two errors, drifting and missing just below, were specific to the menu task, but affected users of all ages.

Missing just below occurs when a user's tap distribution is downwardly shifted, resulting in frequent erroneous selection of the top edge (i.e., the top 10% or 2 pixels) of the item below the target item, and infrequent selection of the corresponding top edge region of the target item itself (see Figure 1). Our results indicated a general trend towards missing just below: none of our participants made notable use of the top edge of the target item, while a substantial subset made use of the edge of the item below. Specifically, our data suggested that a selection along the top edge of an item is 11 times more likely to be intended for the item above it.

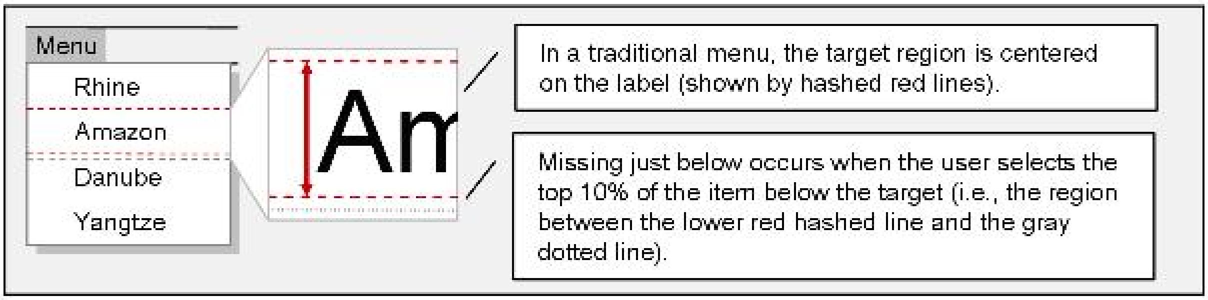

Figure 1. Missing just below.

Drifting, though not one of our planned measures, was a dominant pattern observed during the menu task sessions, occurring when the user accidentally enters the hover region of an adjacent menu. As with a mouse, this action causes the open menu to switch. The problem is that with a Tablet PC hand occlusion often results in users lifting their hand up and away to see the menu contents. Depending on the distance lifted and the angle of this action, the pen may accidentally "drift" to the next menu, resulting in the target menu unexpectedly closing. Thirty-five out of 36 participants drifted at least once, and this behavior did not improve over the course of the study; that is, participants did not get used to eh designed interaction. Moreover, drifting impeded performance: trial time almost doubled when drifting occurred.

Figure 2. Drifting occurs when the user accidentally hovers over an adjacent menu (e.g., Edit in this example), causing it to open, and the desired menu (e.g., File) to close.

Finally, a slip error is the result of landing on the desired target, but slipping off before lifting (as shown in Figure 3). For the old age group, it accounted for approximately half of the errors in each task. In addition, while missing remained relatively constant across age, slipping clearly increased. Although slips were a major source of errors, they were relatively short; for the tapping and menu tasks, the average lengths were 12 and 10 pixels, respectively.

An additional finding of this study was that the behavior of older participants enabled us to uncover difficulties common across the lifespan. The most prominent example of this was drifting, although it also applies to missing just below. Drifting was not a behavior we predicted; rather our observations of the older users during the experimental sessions prompted us to investigate it in detail. It was only upon closer examination of the data that we discovered that all participants were impacted by drifting. Because the older participants moved more slowly overall, it was easier to follow their actions and catch inefficiencies. Also, they were more overt about their interactions. Younger individuals, on the other hand, recovered more quickly and were considerably less verbal about their experience; thus we were not able to detect the effect for them from our observations alone.

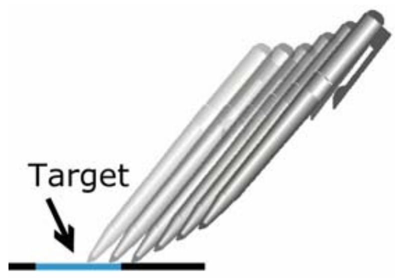

Figure 3. A slip error occurs when the user correctly lands on the target (show in blue) but slips off before lifting the pen.

Addressing the Difficulties Uncovered

The next two phases of this research focus on addressing the difficulties uncovered in the first phase and evaluating the proposed solutions. These two phases are currently underway. Specifically, we have developed and evaluated solutions for missing just below, and we are in the early stages of exploring potential solutions to drifting and slipping.

Missing Just Below

For the first difficulty, missing just below, we designed and developed two approaches: (1) reassigning selections along the top edge, and (2) deactivating the top edge. Figure 4 demonstrates both of these approaches relative to a standard menu. In the *reassigned edge* approach, the top edge of each menu item was reassigned such that taps in that region resulted in selection of the item above. This approach effectively shifted the target region of each menu item down (in motor space), while leaving the visual appearance unchanged. In the *deactivated edge* approach, the top edge of each item was deactivated such that taps in that region were ignored. This approach effectively shrunk the height of each item (in motor space), and added an invisible menu separator between items. It also left the visual appearance unchanged.

The existence of a downward shift in the tap distribution (as reported in [5]) implies a disparity between where users are aiming and the center of the menu item. Thus, the idea behind the reassigned edge approach is to reduce this inconsistency by matching the target bounds to the user's actions. We suggest that the main advantage of the reassigned edge approach is that it changes the current interaction the least. We predict that most users would not notice the small shift, but would simply benefit from fewer errors. Its disadvantage is that it turns a small number of correct selections into errors (i.e., those on the top edge of the target item itself).

Selecting the wrong menu item can have a high cost; for example, selecting the wrong program from the Windows Start menu not only requires the user to go back and reselect the correct item, but also requires the user to wait for the undesired program to load, before closing it. As such, the motivation behind the deactivated pixel condition is that it is less costly to tap an inactive region of the menu and have to re-tap, than to tap an incorrect item and have to correct the selection. However, we have previously noted that users typically do not wait to see if their tap registers, but rather move on, subsequently realize they have not actually made a selection, and then have to go back to try again [5]. Thus, although we predict that the deactivated condition will remove the greatest number of erroneous item selections, it may have other costs. Moreover, these costs may particularly affect older users as older adults have previously been shown to be less able to adapt to changing task requirements [3].

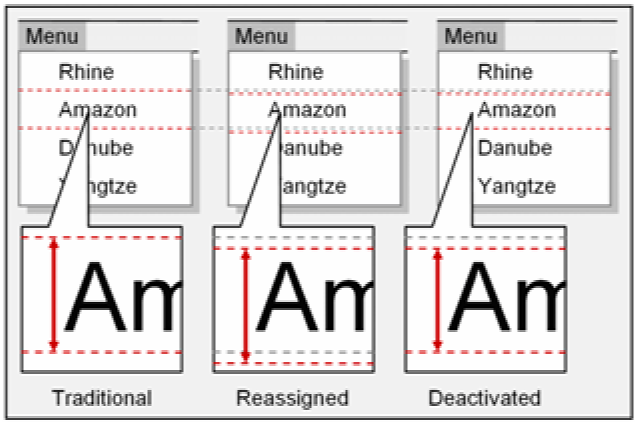

Figure 4. In a traditional menu, the target region (shown by hashed red lines) is centered on the text (left); in a reassigned edge menu, it is shifted down by 10% of its height (center); and in a deactivated edge menu, it is reduced by 10%; taps on the top edge are ignored.

To examine these questions, we ran a controlled laboratory experiment with younger (19–30) and older (66–81) adults to compare these two experimental interfaces, relative to each other and to a traditional edge control condition. In contrast to the first study, we did not see a clearly defined downward shift in the tap distributions, but rather saw two diametric distributions: one that was downwardly shifted (the low hitters) and one that was upwardly shifted (the high hitters). Our performance results reflected these two opposing categories. When we considered our participants as a whole, only the deactivated edge approach showed any benefit over the control condition. However, when we consider the tap distribution groups individually, we see that low hitters did benefit from both the reassigned and deactivated edge approaches, while the high hitters were negatively impacted by the reassigned edge approach, and relatively unaffected (in terms of top edge selection errors) by the deactivated edge approach. Nonetheless, the deactivated edge approach was unpopular with the high hitters as it increased the number of taps required to make a selection, without providing them with any benefit. Moreover, the variability between users suggests that additional research is necessary to explore the practical implications of deploying these techniques.

Drifting

To address drifting, we note that in the first study, none of the participants intentionally used hovering to switch menus. Thus the simplest way to prevent drifting may be to turn off that feature and require a tap to switch between menus. This may not be the best approach, however. In our first study, participants were all novices to pen-based interaction and were prompted to the correct menu. We suspect that for more expert users, or when the task requires browsing through menus to find the correct item, being able to switch menus without touching the screen may prove more useful. An alternative approach would be to introduce some form of delay to the switch, either by time, distance, or a combination of the two. However, it is unclear whether or not browsing and drifting behavior can be distinguished by these measures. We are in the early stages of exploring these questions, and plan to conduct a user experiment to evaluate the effectiveness of these techniques.

Slipping

Finally, the slipping difficulty was a problem for the older users, a result that is consistent with research on the mouse [8]. However, with the mouse, slipping has generally been attributed to an inability to hold the mouse steady while clicking. As tap selection does not have an analogous button clicking action, it is surprising it was also a problem here.

One approach to preventing pen-based slip errors would be to adapt Steady Clicks [8], a slip assistance technique designed for the mouse that assists the user by freezing the cursor at the mouse down position. However, like many mouse-based interaction techniques, Steady Clicks alters the ratio between mouse and cursor movement. The direct mapping between the cursor and the tip of the pen makes this less ideal. One possibility is to handle the freezing internally, and not manipulate the cursor. The drawback is that some users may find this lack of visual feedback confusing.

Another approach would be to combine freezing with area cursors [1]. With an area cursor, it is not the tip of the cursor that defines the object selected, but rather a larger selection area centered on the tip of the cursor. We believe this small degree of separation may provide the flexibility needed to allow a natural form of freezing. On pen down, the area cursor would freeze. Freezing would break, if the pen crosses the edge of the cursor.

Each of these approaches has inherent benefits and drawbacks. The first approach is cognitively simpler, but the lack of visual feedback may be a disadvantage for some users. In contrast, the combined approach provides strong visual feedback, and requires less positioning precision, but adds complexity to the interaction. Furthermore, even if both approaches are viable solutions, it may be that there are situations in which, or user groups for whom, one solution is better than the other. To evaluate these tradeoffs, we propose conducting a laboratory experiment comparing these approaches to each other and to standard tap and point across a range of adult ages.

Conclusion

To summarize, the expected contributions are:

(1) Empirical evidence identifying pen-based target acquisition difficulties, and demonstrating how these difficulties vary across task situations and with age.

(2) New pen-based target acquisition techniques that build upon existing techniques to address the needs of older individuals using a pen-based device.

(3) Controlled user evaluations that: (i) demonstrate the effectiveness of our techniques and show that existing techniques can be combined and modified to meet the needs of older users and pen interaction, and (ii) ascertain that there are trade-offs, revealing that the choice of technique depends on the constraints of the user population.

Mobile technology is increasingly being promoted as a means of addressing cognitive and sensory age-related impairments and enabling individuals to live more independently. However, for this to be a viable approach, the accessibility of these technologies needs to be improved for older adults. By addressing the usability of basic input for older adults, this research will provide a foundation for broader innovations in these other areas.

Acknowledgements

This research is supported by Canada's Natural Sciences and Engineering Research Council (NSERC), and by the Canadian Institute of Health Research (CIHR). The author also thanks her supervisor Dr. Joanna McGrenere, and her committee: Dr. Peter Graf, Dr. Alan Mackworth, and Dr. Shari Trewin.

References

1. Grossman, T., & Balakrishnan, R. (2005). The bubble cursor: Enhancing target acquisition by dynamic resizing of the cursor's activation area. *Proc. CHI '05*, 281–290.
2. Hawkey, K., Inkpen, K.M., Rockwood, K., McAllister, M., & Slonim, J. (2005). Requirements gathering with alzheimer's patients and caregivers. *Proc. ASSETS '05*, 142–149.
3. Heath, M., Roy, E.A., and Weir, P.L. (1999). Visualmotor integration of unexpected sensory events in young and older participants: A kinematic analysis. *Developmental Neuropsychology*, 16(2): 197–211.
4. Kinsella, K. and Velkoff, V. (2001). An Aging World: 2001. U.S. Bureau of the Census.
5. Moffatt, K., & McGrenere, J. (2007). Slipping and drifting: using older users to uncover pen-based target acquisition difficulties. *Proc. ASSETS '07*, 11–18.
6. Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004). The participatory design of a sound and image enhanced daily planner for people with aphasia. *Proc. CHI '04*, 407–414.
7. Mynatt, E. D., Essa, I., and Rogers, W. (2000). Increasing the opportunities for aging in place. *Proc. CUU '00*, 65–71.
8. Trewin, S., Keates, S., & Moffatt, K. (2006). Developing steady clicks: a method of cursor assistance for people with motor impairments. *Proc. ASSETS '06*, 26–33.
9. Willis, S. L. (1996). Everyday problem solving. In J.E. Birren and K. W. Schaie (Eds.), Handbook of the psychology of aging (4th ed., pp. 287-307). San Diego: Academic Press.
10. Wu, M., Baecker, R., & Richards, B. (2007). Designing a Cognitive Aid for and with People who have Anterograde Amnesia. In (ed.) Jonathan Lazar, *Universal Usability*, John Wiley & Sons. 317–356.

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 post-graduate 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.

Online support communities for older people: Investigating network patterns and characteristics of social support

Ulrike Pfeil

Centre for HCI Design, City University, London, U.K.

U.Pfeil-1@city.ac.uk

Abstract

My PhD research aims to identify the components of social support in online support communities for older people. Findings will show how older people support each other in these online communities and how they form a social network based on their communication activities. In addition, my PhD research will reveal how different characteristics of social support evolve and change over time. I will investigate how relationships between older people form, dissolve, and are reconstructed within online support communities, and how individual members are integrated and take on certain roles and responsibilities within the community. Based on the findings, a descriptive model that conceptualises the exchange of social support in online support communities for older people will be developed.

Introduction

Social interactions and collaborations on the internet have become ever more popular in recent years. The rise of social networking sites (e.g. MySpace, Facebook, LinkedIn) and virtual worlds (e.g. SecondLife, ActiveWorlds) illuminates the trend to use the internet to interact with others online. At the same time, traditional ways of online communication (e.g. chats and discussion boards) are becoming more popular as well. A few years ago, people used the internet mainly to retrieve information. Nowadays, typical online activities include communicating with each other in virtual settings in order to socialise [11] and/or to collaborate. Horrigan et al. [3] found in a survey that 84% of internet users participate in online communities. People use online communities to meet other people, develop friendships, play, and exchange experiences and support [11].

In general, researchers describe online communities as settings, where people can meet and communicate with each other online [14]. Preece and Maloney-Krichmar [13] state that online communities are made up from "people who come together for a particular purpose, and who are guided by policies […], and supported by software." In particular, online support communities consist of people who share similar life experiences and build a place of support, compassion and trust [12, 11]. Online support communities do often work as self support groups and display a high level of understanding and emotional support.

Motivation

The degree of internet usage by people aged 65+ has increased by 47% between 2000 and 2004. In Britain, 28% of the older population go online [7]. In the USA the situation is similar, as 22% of older people use the internet and an increase of this percentage is expected [2]. For most of the older people who go online, email is the predominant method of online communication [3], but online communities are also used increasingly by this target group.

Much work has been done by scholars and practitioners to make the internet more accessible. Guidelines that assert accessibility standards have been developed and evaluated (e.g. [5]). This work has primarily focused on the accessibility of information on the internet. However, considering the increasing amount of social activities on the internet, it becomes clear that only ensuring access to information on the internet is no longer sufficient. The new challenges lie in considering the social aspects of activities on the internet. These include taking into account how older people communicate with each other, how they exchange information and support, and how they form relationships and groups in online settings. Thus, it is necessary to study how older people socialise and interact with each other in online communities. In my research, I focus on the exchange of social support in online support communities for older people.

The content that older people share with each other in online support communities has been the subject of several studies [1, 4, 16, 18, 19]. Also, query-based techniques have been applied to investigate how older people perceive social interactions in online settings [6, 17, 18]. However, these studies have been partial in their approach, as only few of these studies investigated social support specifically, and few used multiple methods to conduct their research. An integrated investigation is needed to fully understand how older people exchange social support in online support communities. My PhD research aims to provide a holistic description of the exchange of social support in online support communities for older people.

Research Questions

As elaborated above, the key Research Question of my PhD research is:

"How is social support exchanged in online support communities for older people?"

This can be broken down into the following Sub-Questions:

I. What are the appropriate methods for studying online support communities for older people?
II. What are the characteristics of social support in online support communities for older people?
III. What are the network properties and patterns of online support communities for older people?
IV. What are older people's needs concerning online support communities? How do they experience support in offline settings and how does this influence their participation (and non-participation) in online support communities?
V. How does the content and the network patterns of online support communities for older people develop over time? (e.g. How does the network pattern change over time? Does the relation between content and network pattern change over time?)
VI. How does the behaviour and network position of individual members change over time? (e.g. How do people gain and loose power and influence in the online community? What roles do they take on over a period of time?)

Completed work

Content analysis of an online support community for older people

In order to study the components of social support in online communities for older people, I did a content analysis of a subset of the messages of the discussion board about 'depression' within SeniorNet [15]. I collected and analysed the conversation on the board for a period of 1.5 years between 2000 and 2001. Four hundred (400) messages were exchanged in this time and qualitative content analysis was used to determine how social support is expressed and facilitated in online communication. The findings identified different components of social support, and elicited the different roles that people take on in the online support community. A detailed description of the study can be found in [8].

As a result of this study a code scheme that describes the different aspects of support in the investigated online community was developed. Table 1 lists the seven main categories of the code scheme with the short descriptions and examples based on the analysis of the messages. The categories describe the characteristics and components of social support as exchanged in online support communities for older people.

Table 1: Developed code scheme

Category	Description	Examples
Self-disclosure	Text units in which people post information about themselves. This can be done in different ways (e.g. emotional, narrative, medical)	"I yawn all the time. I want to go to bed. I know you're supposed to get out, but I don't have the energy to do that much."
Community building	The text unit includes people's opinion about the online community and meta-information about communication activities on the discussion board.	"Thank God for this board, as I can sit here and cry and rattle on---you are the only ones who understand."
Deep support	Supporting text units are often emotional and customised towards the unique situation of the target that the message is for.	"Words are so hard right now. So I place my hand gently over yours and let love and sweetness flow through to you."
Light support	The text unit is supportive and uplifting. It is written in a generic way, for another person or the whole community.	"Hang in there" , "I am thinking about you")
Medical facts	These text units include questions and answers about factual information within the topic (e.g. medication).	"So in "both cases" situational depression and bipolar depression they alter chemicals in the brain?"
Technical issues	The text units are concerned with technical problems or suggestions to solve them.	"Read in your browser screen and have Notepad or Wordpad minimised…"
Slightly off	Text units that are about others or about topics that strayed away from the theme of the discussion board.	"Sorry to hear Iowa's weather yesterday. Minnesota is much too cold and damp."

SNA of the communication patterns within the online support community

In order to study the communication patterns within an online support community for older people, I analysed the social network structure of the discussion board about 'depression' within SeniorNet. I applied social network analysis (SNA) to analyse the communication patterns and relationships between members of the discussion board. I investigated who was talking to whom in the online support community and constructed a social network based on the communication activities. In addition to looking at the structure of the exchanged messages within the discussion board as a whole, I also investigated the impact of the communication content on the social network patterns. In particular, I investigated whether conversations in each of the seven identified categories (see Table 1) have an impact on network characteristics (e.g. density of the network, building of cliques, and inclusiveness of the network). Figure 3 shows the sociogram of all investigated communication content of the discussion board about depression within SeniorNet for the period of 1.5 years between August 2000 and February 2001. The red nodes represent the members of the online support community, and the black lines indicate the communication activities between them.

Findings show distinct differences between the social network patterns of empathic (related to support) and non-empathic (not related to support) communications. For example, members are more connected and closer to each other in the social networks that are related to support compared to communication that is not related to support. Also, the difference between seeking and giving support has an impact on the network structure, as messages that seek support are directed to the whole online support community and messages that give support are more commonly targeted to specific members. Additionally my results show that the type of support has an impact on the social network structure within the discussion board. Whereas 'Light support' is freely shared between all members of the online support group equally, 'Deep support' is often exchanged in small sub-groups within the online support community. For a full description of the study and findings, refer to [9].

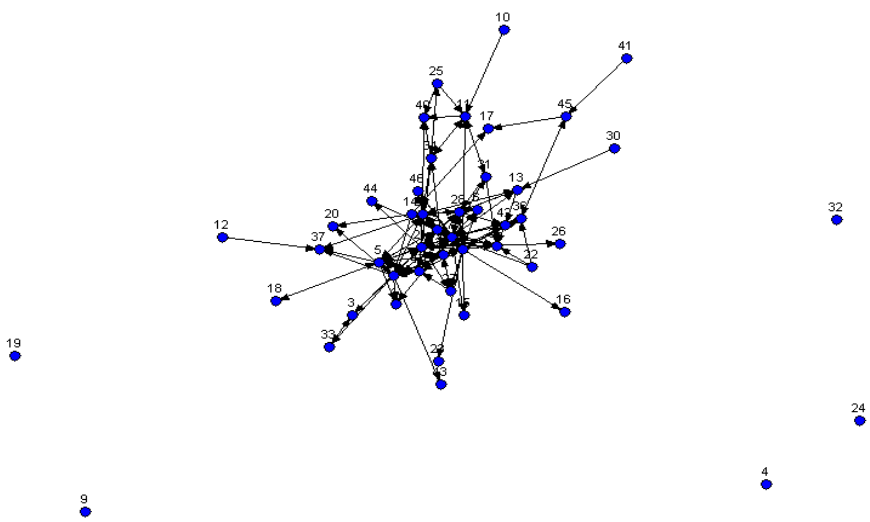

Figure 2: Sociogram of the investigated communication activities within the online support community

Elicitation of user needs and preferences

In order to get a deeper understanding about the needs and preferences of older people regarding online support, I conducted interviews with 31 older people who had different levels of expertise in using the internet and online communication (non-internet users, users who use email, and users who use online support communities). I studied the perception and experiences of older people concerning support in their everyday offline life. In addition, I also investigated their usage of online communication (e.g. emails and online support communities) in order to exchange support. The aim of the interviews was to elicit older people's motivation to exchange support in online settings and the reasons for reluctance to do so. The developed categories (see Table 1) were used as a basis for the interviews and people were asked about their experience of online and offline support relating to these categories. By analyzing the perceptions and experiences of older people, I investigated how the different characteristics of online communication can facilitate or hamper the exchange of support for older people.

Findings show that online support communities for older people do indeed have the opportunity to enhance older people's lives, but they also indicate that in order to facilitate online support for older people, special care has to be taken about the needs and preferences of this target group. A complete description of the study and its findings can be found in [10].

Future work

The remaining work of this PhD includes a longitudinal study to investigate the dynamics and changes of message content and network structure in these communities over a longer period of time. The first step in the data collection phase will be to collect messages of an online support community for older people. The chosen online support community will have to have an archive with a sufficient amount of messages to do longitudinal analysis with.

There will be two approaches to the analysis of the collected data: Firstly, I will investigate the characteristics of the whole community over time. Linking both, the analysis of the content and the network structure over time, this study will give insight into the development of online support communities over time and further inform the phenomenon of social support in online support communities for older people.

Secondly, I will observe individual members of the online support community over time. Findings of this study will inform the exchange of social support in online support communities for older people from an individual perspective.

Based on the findings of this study and of the studies already completed, a model will be developed that shows how social support is exchanged in an online support community for older people. Additionally to the components of social support and network characteristics that emerge of supportive communication, roles and responsibilities of individuals in the process of exchanging social support will be incorporated. Also, dynamic changes of support and network patterns over time will be modelled to allow for a more flexible explanation of the phenomenon.

The developed model is proposed not only to explain the exchange of social support in online support communities for older people, but also as a tool to be used to study online support communities for older people. Thus, the model will be accompanied by methodological guidelines on how to apply it to study online support communities. For example, guidance will be provided on how to use the categories as a basis for content analysis.

Additionally, a variety of online support communities for older people will be consulted with the aim to validate the generalisability of the model. I will take parts of the model as a basis to reflect on its applicability to study online support communities for older people in general.

Anticipated contribution

The results of the studies will shed light on the patterns of social support exchanged among older people in online communities. I believe that the outcome of the PhD will be of use for both researchers and practitioners in the area of HCI/CMC as it investigates an area that is currently understudied.

The proposed model will go beyond existing research and provide a tool for analyzing the characteristics and network patterns of online support communities for older people. Scholars will be able to use the model, both to explain social support in online support communities as well as to apply the model as a methodological tool in their own research activities.

Also, findings from the studies and the description of the model will help better understand how older users interact with each other in online support groups which will also benefit practitioners and designers of online support communities for older people. If we understand the aspects of online social support and how it is exchanged by older people in online communications, we can also find ways to nurture it and design online communities to better facilitate supportive communication.

References

1. Brennan, P. F. Moore, S. M. & Smyth, K. A. ComputerLink: Electronic support for the home caregiver. *Advances in Nursing Science*, 13(4), 1991, 14-27.
2. Fox, S. *Older Americans and the Internet*. Washington, DC: Pew Internet & American Life Project, 2004.
3. Horrigan, J.B., Rainie, L., & Fox, S. *Online communities: Networks that nurture long-distance relationships and local ties*. Washington, DC: Pew Internet & American Life Project, 2001.
4. Kanayama, T. Ethnographic Research on the Experience of Japanese Elderly People Online. *New Media and Society*, 5(2), 2003, 267-288.
5. Kurniawan, S. & Zaphiris, P. Research-derived web design guidelines for older people. In: *Proceedings of 7th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS' 05), 2005. ACM Press, Baltimore.
6. McMellon, C.A. & Schiffman, L.G. Cybersenior empowerment: How some older individuals are taking control of their lives. *Journal of Applied Gerontology*, 21(2), 2002, 157-175.
7. Office of Communication. *Consumers and the communications market: 2006*. Ofcom Consumer Panel, 2006.
8. Pfeil, U. & Zaphiris, P. Patterns of empathy in online communication. In: *Proceedings of CHI 2007 – the ACM Conference on Human Factors in Computing Systems* (San Jose, CA, 28. April - 03. May 2007), CHI '07. ACM Press, New York, NY, 2007, 919-928.
9. Pfeil & Zaphiris (under review). Investigating social network patterns within an empathic online community for older people. Journal of Computer-Mediated Communication.
10. Pfeil, Zaphiris, and Wilson, (under review). Older People's Perceptions and Experiences of Online Social Support. CHI Conference, Florence, 2008.
11. Preece, J. & Ghozati, K. Observations and Explorations of Empathy Online. In. R. R. Rice and J. E. Katz, The Internet and Health Communication: Experience and Expectations (2001). Sage Publications Inc.: Thousand Oaks, 237-260.
12. Preece, J. & Maloney-Krichmar, D. Online Communities: Focusing on Sociability and Usability. In Jacko, J. & Sears, A. (Eds.) The Human-Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Applications (2003). Lawrence Earlbaum Assc. London, UK, 596-620.
13. Preece, J. & Maloney-Krichmar, D. Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, 10(4), article 1, 2005. Retrieved 1st November, 2007 from http://jcmc.indiana.edu/vol10/issue4/preece.html.
14. Preece, J., Maloney-Krichmar, D. & Abras, C. History of Emergence of Online Communities. In: B.Wellman (Ed.) Encyclopedia of Community (2003). London: Sage, 1023–1027.
15. SeniorNet. SeniorNet, 2007. Retrieved 1st November, 2007 from http://www.seniornet.org.
16. White, M.H. and Dorman, S.M. Online support for caregivers: analysis of an Internet Alzheimer mailgroup. *Computers in Nursing*, 18, 2000, 168-176.

17. Wright, K. B. Computer-mediated support groups: An examination of relationships among social support, perceived stress, and coping strategies. *Communication Quarterly*, 47(4), 1999, 402-414
18. Wright, K. B. The communication of social support within an on-line community for older adults: A qualitative analysis of the SeniorNet community. *Qualitative Research Reports in Communication*, 1(2), 2000, 33-43.
19. Xie, B. Getting older adults online: The experiences of SeniorNet (USA) and OldKids (China). In: B. Jaeger (Ed.), Young Technologies in Old Hands - An International View on Senior Citizen's Utilization of ICT (2005). Copenhagen, Denmark: DJOF Publishing, 175-204.

About the author

Ulrike Pfeil started her PhD work in October, 2006, funded by a three-year, full-time doctoral research studentship from City University London. She was awarded an MSc in HCI from the same university in 2006 and a BSc in Information Design from the "Hochschule der Medien", Stuttgart, Germany in 2005. Her research interests include Social Aspects of Computing, especially in Computer-Mediated Communication. She specifically investigates the phenomenon of social support among older people, especially how social support changes in online environments compared to offline ones. Within City University, she is working at the Centre for HCI Design. The research area of the Centre is particularly focused around Requirements Engineering, E-learning, Computer-Supported Collaborative Work, Inclusive Design and Social Aspects of Computing.

Early Diagnosis of Autism through Analysis of Pre-speech Vocalizations

Keshi Dai

College of Computer and Information Science, Northeastern University, Boston, USA

daikeshi@ccs.neu.edu

Introduction

Autism is a complex neurodevelopmental disorder that typically appears before age of three years old. The narrower term of autism always refers to classic autistic disorder in the spectrum of pervasive developmental disorder (PDD). In this paper, "autism" and "autistic" are all used to refer to the entire range of autism spectrum disorders (ASD).

Although autism was first described in 1943 by Kanner, there is still no quantitative method to accurately diagnose it. Currently, autism can be reliably diagnosed by the age of 3. However, studies suggest that many children with autism may be accurately identified at 12 months or even younger [1]. The most widely used diagnostic methods nowadays rely on evaluation by professionals of behavioral characteristics. This kind of diagnosis usually requires a two-stage process that includes developmental screening and comprehensive evaluation by a multidisciplinary team [1]. In the developmental screening, parents need to recall when their children reached each milestone in their development through reviewing video recordings, photos, and so on. In the next step, a screening instrument is used to assess whether children are typically developing based on parent reports and observations. Among screening instruments available today are Checklist of Autism in Toddlers (CHAT) [2], the Screening Tool for Autism in Two-Year-Olds (STAT) [3], and the Social Communication Questionnaire (SCQ) that is for 4 years old children or older [4]. These four are mainly to identify children with severe autism. They may not be able to detect some mild ASD. Other screening instruments like Autism Spectrum Screen Questionnaire (ASSQ) [5] and Childhood Asperger Syndrome Test (CAST) [6] are designed to identify mild ASD like high-functioning or Asperger syndrome. If some behaviors of a child in the screening process are identified as indicators of ASD, this child will need a comprehensive diagnostic evaluation. This evaluation is performed by a multidisciplinary team that includes a psychologist, a neurologist, a psychiatrist, a speech therapist, or other professionals in ASD. Apart from traditional diagnostic methods, new technologies such as movement analysis [7] and eye gaze analysis [8] are being developed as potential clinical tools in the diagnosis of autism.

Early diagnosis of autism via pre-speech vocalization

Our purpose is to develop an objective clinical tool for the early diagnosis of autism. We will use technologies developed in Early Vocalization Analyzer (EVA) and visiBabble system to extract features from vocalizations of infants and toddlers from 6 to 18 months old. By focusing on a population with moderate risk for the development of autism, we hope to be able to derive classification rules, based on these detected speech features, which distinguish those children who are eventually diagnosed with autism from other children in the group. We hope that our classifier can be used, in the future, for early diagnosis of autism and therefore for early intervention.

EVA and VisiBabble

EVA is a computer program that automatically analyzes infant vocalizations and derives a "vocalization age" based on detected features. It is able to clinically distinguish infants, six to fifteen months old, who may be at risk for later communication or other developmental problems from typically developing infants [9, 10]. The visiBabble system relies on EVA to process vocalizations in real-time. It responds to the infant's syllable-like productions with visual feedbacks and records the landmark analysis. This system reinforces the production of syllabic utterances that are associated with later language and cognitive development [11].

Technical Methods

We first find landmarks in the acoustic signal and then use them to extract other features. Landmark detector used in our work is built on the Liu-Stevens landmark detection theory [12]. Essential to this theory are landmarks, pinpointing the abrupt spectral changes in an utterance, which mark perceptual foci and articulatory targets. Listeners often focus on landmarks to obtain acoustic cues necessary for understanding the distinctive features in the speech.

In this work, we focuses on three types of landmarks: glottis (marks the time when the vocal folds transition from freely vibrating to not freely vibrating), sonorant (marks sonorant consonantal closures and releases), and burst (marks stop/affricate bursts and points where aspiration/frication ends due to stop closure) [9, 10]. We used three measurements related to landmarks: landmarks per word or utterance, voice onset time that is the time between when a consonant is released and when the vibration of the vocal folds begins, and landmark rate that is the rate of each landmark type in an utterance. Figure 1 shows landmarks produced by our landmark extractor.

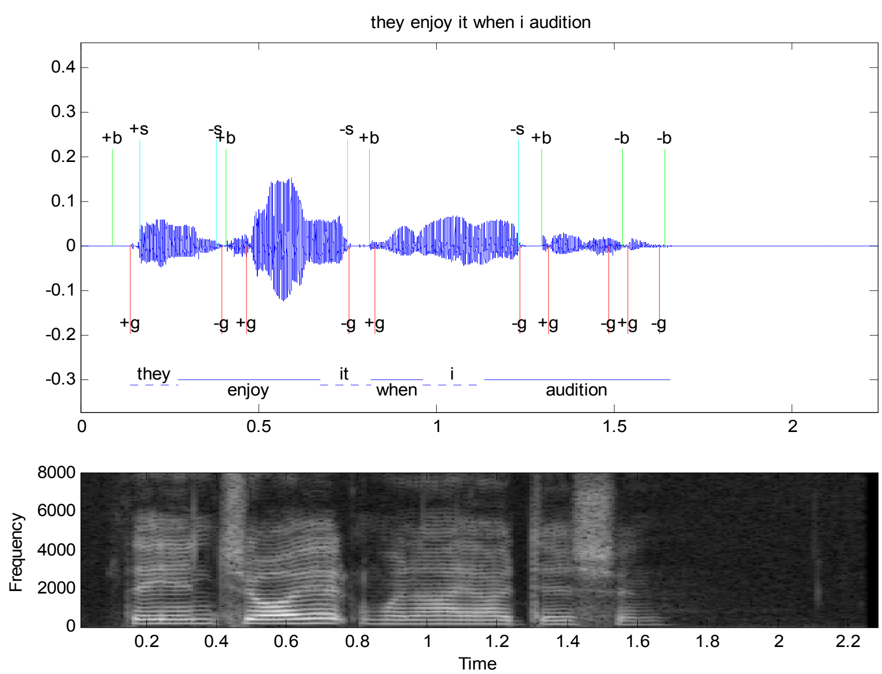

Figure 1: Landmark plot produced by our landmark detector

In Figure 1, we can see that the region between +g and –g is the voiced region. +s/-s landmarks only happen in voiced region, and +b/-b landmarks only appear in unvoiced region. In the spectrogram, the energy of the fundamental frequency in voiced region is the strongest. The +s landmark happens when there is an increase in energy from the Bands 2 (0.8-1.5 kHz) to Bands 5 (3.5-5.0 kHz) and the –s landmark signifies energy decrease in these frequency bands. A +b landmark is detected when a silence interval is followed by a sharp energy increase in high frequency from Bands 3 (1.2-2.0 kHz) to Bands 6 (5.0-8.0 kHz). On the contrary, a –b landmarks signifies a sharp energy decrease in high frequency followed by a silence interval.

Sequences of landmarks are grouped into standard syllable patterns. A syllable is a unit of sound, and is typically made up of a vowel with optional initial and final margins. In our syllable detector, syllable is based on the order and spacing of detected landmarks. It must contain a voiced segment of sufficient length. 38 possible syllables were recognized. 11 syllables begin with +g landmark, 22 begin with +b, and 5 begin with +s. 4 types of syllable features are extracted: syllable rate, syllable number, landmarks per syllable, and syllable duration.

With above-mentioned landmark and syllable features, we can construct new features for early diagnosing autism. Specific patterns of landmarks, particular types of syllable, and anomalous pitch patterns in the utterance may also be important indicators. In addition, features used to estimate vocalization age in [10] may be also useful. Vocalization age delay may be a possible indicator of autism, because autism is also characterized by impaired social interaction and communication.

Planned Study

Subjects

We plan to work with a group at Children's Hospital, Boston, which is studying younger siblings of children with autism. We plan to record these children for 10 minutes every 3 months from one to two years. Features will be extracted from these recordings and we will look for patterns that can distinguish those children who eventually develop autism from those who don't.

Data Processing and Analysis

First we will examine the recordings we collected. We will remove utterances not produced by our subjects, non-speech utterances, and other extraneous noise or voices, and only keep clear babbling. Then, landmark detector will be used to extract landmarks. Syllables and utterances analysis will also be applied to landmarks. We will record landmarks, syllables, and utterances for further analysis. After children are clinically diagnosed as autism (at 2.5 to 3 years), a comprehensive analysis will be conducted based on landmarks, syllables, and utterances collected before. In this step, we hope to find features that distinguish those children who eventually develop autism from those who don't.

Envisioned Contributions

Current research estimates that 2-6 per 1000 children have an ASD in the United States [1]. It is a surprising increase over rates reported in 1980s and 1990s. Our method makes early diagnosis of autism at pre-speech age possible. Early diagnosis allows for early intervention, which is critical, because children with autism who receive early intervention at younger ages make greater improvements than those who receive it at older ages [13, 14].

Early intervention can prevent declines in intellectual development, and improve the communication skills, motor, and cognitive development [15, 16]. Early intervention should commence as soon as possible after diagnosis. It will have a great impact on the lives of children with autism and their families.

References

1. Autism Spectrum Disorder (Pervasive Developmental Disorders). National Inst. of Mental Health, 2004.
2. Baird, G., Charman T., Baron-Cohen, S., Cox, A., Swettenham J., Wheelwright, S., and Drew, A. A screening instrument for autism at 18 months of age: A 6-year follow-up study. Journal of the American Academy of Child and Adolescent Psychiatry, 39 (2000), 694-702.
3. Stone, W. L., Coonrod, E. E., and Ousley, O. Y. Brief report: Screening tool for autism in two-year-olds (STAT): development and preliminary data. Journal of Autism and Developmental Disorders, 30, 6 (2000), 607-612.
4. Berument, S. K., Rutter, M., Lord C., Pickles, A., and Bailey, A. Autism Screening Questionnaire: diagnostic validity. British Journal of Psychiatry, 175 (1999), 444-451.
5. Ehlers, S., Gillberg, C., and Wing, L. A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. Journal of Autism and Developmental Disorders, 29, 2 (1999), 129-141.
6. Scott, F. J., Baron-Cohen, S., Bolton, P., and Brayne, C. The Cast (Childhood Asperger Syndrome Test): preliminary development of a UK screen for mainstream primary-school-age children. Autism, 2, 1 (2002), 9-31.
7. Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., and Maurer, R. G. Movement analysis in infancy may be useful for early diagnosis of autism. Psychology, 95, 23 (Nov. 1998), 13982-13987.
8. Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. General Psychiatry, 59, 9 (Sep. 2002), 809-816.
9. Fell, H. J., MacAuslan, J., Ferrier, L. J., and Chenausky, K., Automatic Babble Recognition for Early Detection of Speech Related Disorders. J. Behaviour & Inf. Tech., 18, 1 (1999), 56-63.
10. Fell, H. J., MacAuslan, J., Ferrier, L. J., Worst, S. G., and Chenausky, K. H., Vocalization Age as a Clinical Tool. In Proceedings of ICSLP '02, (Denver, CO, Sep. 2002), 2345-2348.
11. Fell, H. J., Cress, C., MacAuslan, J., and Ferrier, L. J., visiBabble for reinforcement of early vocalization. In Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility, (Atlanta, GA, Oct. 18-20, 2004), 161-168.
12. Liu, S. Landmark Detection for Distinctive Feature-based Speech Recognition. Ph.D. Thesis. M.I.T. Cambridge, MA, 1995.
13. Harris, S.L., and Weiss, M. J. Right from the start: Behavioral intervention for young children with autism. Woodbine House, Bethesda, MD, 1998
14. Sheinkopf, S. J., and Siegel, B. Home based behavioral treatment of young children with autism. Journal of Autism and Developmental Disorders, 28, 1 (1998), 15-23.
15. Guralnick, M.J. Effectiveness of early intervention for vulnerable children: A developmental perspective. American Journal on Mental Retardation, 102, 4 (1998), 319–345.
16. McGee, G. G., Morier, M. J., and Daly, T. The Walden Early Childhood Programs. In Handleman, J. S., Harris, S. L. Preschool Education Programs for Children with Autism 2nd ed. Pro-Ed, Austin, TX, 2001.

About the author

Keshi Dai is a second-year Ph.D. student in the College of Computer and Information Science of Northeastern University. His research interests include speech processing, early vocalization analysis, and emotion recognition. Currently, he is planning to work with a group at Children's Hospital, Boston to collect utterances of younger siblings of children with autism. He hopes to find features that can distinguish those children who eventually develop autism from those who don't.

Encouraging Speech and Vocalization in Children with Autistic Spectrum Disorder

Joshua Hailpern

Dept. of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA

jhailpe2@cs.uiuc.edu

Abstract

Technology can improve the life of those with Autistic Spectrum Disorder (ASD). Specifically, some children with ASD are not fortunate enough to acquire the ability to communicate with language on their own. With language being an important method of communication, socialization, and interacting with the world, these children need researchers to develop new solutions to help teach vocalization and speech. Without speech, these children will have difficulty communicating their needs, wants, and emotions, as well as being able to function within society at large. This paper examines the existing HCI research on ASD, as well as proposes a new direction for investigation.

Introduction

During the normal development of a child, language is acquired without much explicit effort by parents, practitioners, or the community. However, some children, such as those with Autistic Spectrum Disorder (ASD), are not fortunate enough to acquire this skill on their own. That language is considered "a unique characteristic of human behavior… [that] contributes in a major way to human thought and reasoning"[11], these children need our help to learn to communicate and function in the "real world". Without speech, these children have difficulty expressing their desires, emotions, and communicating on a day-to-day basis. HCI is poised to develop new techniques that use technology to assist practitioners and parents to teach their children with ASD to talk and develop normally. With 1 in 150 children being diagnosed with ASD [4], the need for new solutions is growing.

A Brief Description of Autistic Spectrum Disorder

ASD is a developmental disability impacting social functioning, such as empathy, basic interaction and communication. It is important to note that autism affects each individual differently. These differences cause a wide range, or spectrum, or conditions ranging from individuals who are "high-functioning" or having only slight delays in communication and social functioning, to those who are "low-functioning" having a greater challenge of interacting. Other notable characteristics of many individuals with autism include "insistence on sameness… Preference to being alone… spinning objects [and] obsessive attachments to objects " [1].

Communication Treatments

In the 1960s, Ivar Lovaas began teaching children with autism new behaviors through a technique called "applied behavior analysis", in which a behavior is encouraged or discouraged as it encounters environmental consequences. In short, his technique relies upon using objects, food, and actions as rewards for desired behavior (prompted by a researcher) [11]. Over many trials and sessions, children with autism eventually learn to respond in a predictable fashion by interacting with people in their environment. There are three main drawbacks to this form of treatment:

- It requires many sessions with trained professionals who are in short supply. This can place a financial burden on the family.
- Teaching sessions require intense attention and prolonged contact from a practitioner or parent.
- The child must interact with a human being. One characteristic of ASD is anxious, detached, and "alone" interaction with other individuals [2, 6]. Thus the interaction with a human being, as the primary mode of teaching, might pose some degree of built-in difficulty for the ASD child.

HCI Research on ASD Treatments

Existing work by HCI researchers has approached ASD from three primary directions. Work by Abowd and others have explored the benefits of technology to aid the diagnosis process [5, 8, 9]. This research is crucial, because early detection allows children to begin treatment earlier, allowing them to catch up faster to their non-autistic peers. Further, this work allows us to better understand how to identify autistic characteristics. Although greatly beneficial, this research does not provide a direct method to enhance the education of children with ASD.

Researchers have also explored the effect that technological environments have had on the process of assisting children with autism to learn how to interact with other human beings [7, 14]. This work uses virtual environments, as well as virtual peers, to create situations in which the children with ASD are comfortable. They are then able to learn person-to-person interactions, without the apprehension of having another person in the room. This work, however, primarily has dealt with "high functioning" children, or those who already know how to speak and have a deficit in social interaction. Therefore, it is hoped that principles learned from this body of literature will have the potential to be applied to research targeting children with ASD who have not yet acquired speech.

The third approach seeks to encourage children with ASD to "play," where playing is mediated through technology [10, 12, 13]. By creating technological methods of interaction (visual displays and physical robots), play and comfortable interactions can be garnered from children with autism. There is a feeling of "safety" by having the main form of interaction occur with non-humans. Further, these devices allow the child, rather than a third party, to be in control of the interactions. This research has much potential. To date, however, it has not focused on encouraging more communication-based activities, such as speech and human-to-human interaction.

A Proposal for a New Direction in Research

Through this analysis of existing approaches, we believe there is potential to create a new direction of research, focusing on using technology to encourage meaningful speech in low-functioning children with ASD. Speech would allow these children to express needs, desires and live more normal lives in society. The literature provides strong evidence that interacting with technology often can motivate children with ASD. Further, existing literature shows that real-time visualizations, which act as social mirrors, can influence communication interaction [3]. Therefore, we see the potential of technology to aid teachers in the development of sounds, words, and speech; thereby contributing to what is an exclusively human-to-human interaction. By introducing technology into this form of treatment, we believe we can alleviate a degree of apprehension experienced by the children when interacting with humans, and provide teachers with a new technique to complement and supplement their existing approaches.

Some of the literature has successfully encouraged play by leveraging visualizations that use abstract imagery (shapes that do not directly map to real world objects). Further, research targeting human-human interaction through training on virtual environments shows the potential benefits of training with this technology. We believe that combining aspects of these two approaches could facilitate a new understanding of sound and speech. These results lead us to believe that a similar approach could be used to help children with ASD understand and control their sounds by visualizing their voice. These visualizations could be used as form of reward for making correct sounds, or as a form of feedback to allow them to visually comprehend the sounds they are making. By constructing these visualizations using hand held devices, the learning experience could also be easily extended to day-to-day life, rather than being limited to an office or a school.

Current Research

In order to test whether computer generated stimuli (auditory and/or visual) can be constructed to encourage/reinforce sound production in children with ASD, we have begun the first phase of this research project. Our study attempts to uncover the effect of different permutations of auditory and visual feedback on the vocal production of children in our target population. We have constructed approximately a dozen varying types of visualizations, as well as 4 different metaphors for sound feedback.

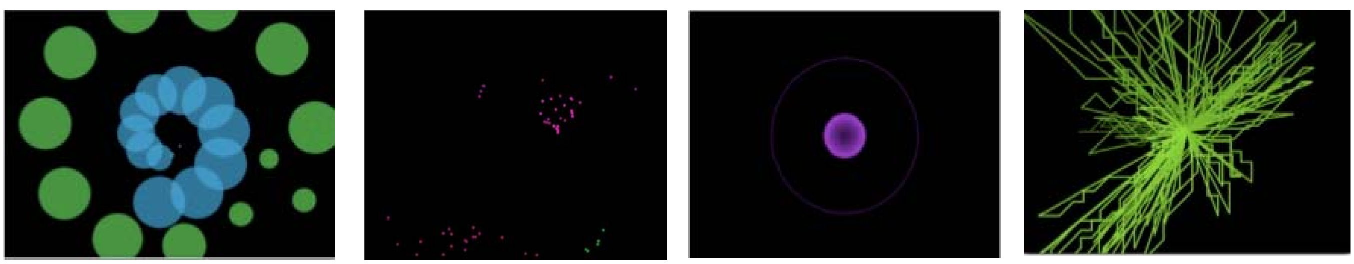

Figure 1. Examples of Visualizations used in Phase 1

Visualizations fall into 4 categories; Falling, Spinning, Found Imagery, Flashy (See Figure 1). We chose to model falling objects and spinning objects in an attempt to replicate stimuli that seems to impact children on the spectrum in the real world. We theorized that because some children enjoy watching objects falling, and spinning objects (as well as themselves) we could leverage that interest by bringing these types of motion into the digital domain.

Children enrolled in this first phase of research attend 6 sessions in which a researcher presents them with approximately 8 different forms of feedback to sound production over about a 40-minute period (Figure 2). Each combination of visual and auditory feedback (trial) is separated by a period of play and relaxation for the child, in order to maintain a positive and stress-free environment. Software, on the systems presenting the feedback, logs critical data points about the child's interaction with the software. With parental consent, sessions are video taped to allow researchers to gather additional data. The configuration of the test room follows one of 3 configurations (Figure 3).

From these sessions, research will attempt to determine not only whether these forms of feedback can encourage sound, but also explore other related behaviors of children with

ASD; is there a form of feedback (auditory or visual) that is more influential, what type and amount of attention will a child pay to the screen on a computer, are children with ASD visual learners, and are all children with ASD the same (in terms of preferences).

Figure 2. A child using a spinning visualization with echoing auditory feedback. His eyes are looking at the screen.

Phase One began in early August 2007, and was completed in November 2007. Initial qualitative results appear promising. Qualitatively, many of the children appear to have forms of feedback for which they will generate far greater number (and frequency) of sounds. In addition, many parents have expressed that their children become excited, if not elated when they realize that they are coming to a session. That could indicate that the visual/auditory stimulus is enjoyable to play with and thus could be used as a form of enjoyable education for children with ASD.

Figure 3. The room for phase 1 was configured based on the needs of the child. If the child had not learned to sit, the room was opened up (left) and a trampoline and/or bean bag chair was used while the visualizations were projected on a screen. Children who learned to sit, but preferred the large screen were placed in a chair surrounded by tables (middle) to help confine their movements and to encourage focusing on the screen. For the remaining children, a large screen computer was used (right), with a chair restricting their focus to the screen. For the first 2 set ups, a speaker was placed behind the projector screen. For the table-top computer, the speaker was placed behind the monitor.

One specific child appears to be reacting over and above the others in terms of sound production. Specifically, he appears to be exhibiting signs of turn taking (with the auditory feedback), as well as mimicking many of the sounds generated (as if playing a call and response game). At the completion of his standard 6 sessions, we are planning to conduct an additional 2 sessions with a new visualization (controlled through a wizard of Oz technique) that requests that the child make specific sounds. Should those sounds be made, visual and auditory reinforcement will be provided. The choice of reinforcement will be based on the qualitative observations of the child's preferences made by researchers. Should the results prove positive, we will have an indication that these systems (with little modification) could be used to teach sound formation to some children with ASD.

Analysis

For the quantitative analysis, we plan to treat Phase One as a single subject analysis across conditions with auditory and video recordings of the following variables: time in chair, percent of intervals with positive emotive state, diversity of phonetic repertoire, frequency of utterances, variation in utterance duration, and exploration of utterance volume. Though we plan to focus primarily on within subject comparisons, we plan to do some across-subject analysis.

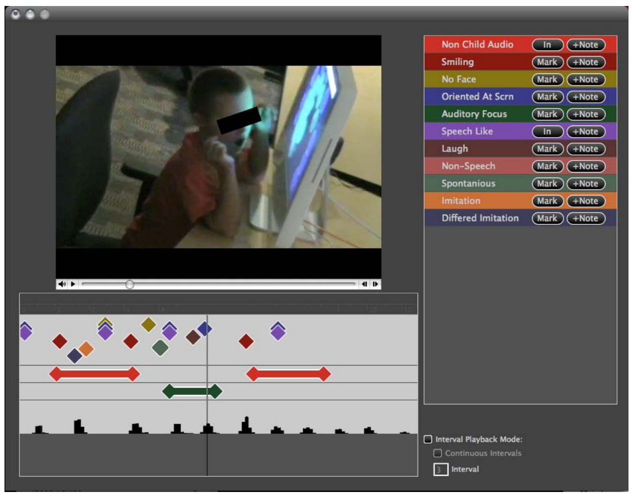

Figure 4. To aid in the video analysis, we are creating a new tool to help facilitate video coding, Vcode. Vcode is designed to facilitate multiple ways to analyzing, and annotating video. Data collected with Vcode can be synched with existing comma delineated data files to create a unified data set. Further, to help determine reliability, Vcode will perform agreement analysis on data collected by multiple coders.

Further studies are planned exploring the effect of visualizations on vocalization frequency, controlling vocalization, instructing children to make specific sounds, and word formation. We also are planning to explore form factors ranging from larger objects to portable hand-held devices (which would allow children to learn outside of an office, in more comfortable environments like the home).

Conclusion

To date, little or no research has been reported on using technology to teach low functioning children with ASD to learn to vocalize or speak. However, the field of augmentative and alternative communication embraces technology primarily as a medium of communication and not as often as a method of instruction. If we can encourage vocalization at the age of 3, a pivotal age for children with ASD, this could lead to an increased communicative ability, which makes not only the child's life easier, but also increases their chances of functioning in the world around them. In addition, if the approach we are recommending proves successful for speech, similar methods could be applied to other behaviors or to other disorders with speech impairments.

References

1. Autism Society of America, ASA. Autism Society of America, 2007.
2. Baskett, C.B. The effect of live interactive video on the communicative behavior in children with autism, University of North Carolina at Chapel Hill, Chapel Hill, 1996.
3. Bergstrom, T. and Karahalios, K., Seeing More: Visualizing Audio Cues. in *INTERACT*, (Rio de Janerio, Brasil, 2007).
4. Center for Disease Control and Prevention, CDC. Autism Information Center, DD, NCBDDD, CDC, Atlanta, April 25, 2007.
5. Hayes, G.R., Abowd, G.D., et al. Designing Capture Applications to Support the Education of Children with Autism in *International Conference on Ubiquitous Computing*, (Nottingham, UK, 2004).
6. Kanner, L. Autistic Disturbances of Affective Contact. in Kanner, L. ed. *Nervous Child 2*, V.H. Winston, 1943, 217-250.
7. Kerr, S.J., Neale, H.R. and Cobb, S.V.G. Virtual environments for social skills training: the importance of scaffolding in practice *Proceedings of the fifth international ACM conference on Assistive technologies*, ACM Press, Edinburgh, Scotland, 2002.
8. Kientz, J.A., Abowd, G.D., et al. Grow and know: understanding record-keeping needs for tracking the development of young children *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press, San Jose, California, USA, 2007.
9. Kientz, J.A., Hayes, G.R., Abowd, G.D. and Grinter, R.E. From the war room to the living room: decision support for home-based therapy teams *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM Press, Banff, Alberta, Canada, 2006.
10. Lehman, J.F. Toward the use of speech and natural language technology in intervention for a language-disordered population *Proceedings of the third international ACM conference on Assistive technologies*, ACM Press, Marina del Rey, California, United States, 1998.
11. Lovaas, I.I. *The Autistic Child*. John Wiley & Sons, Inc, New York, 1977.
12. Michaud, F. and Théberge-Turmel, C. Mobile robotic toys and autism. in Dautenhahn, K. ed. *Socially Intelligent Agents - Creating Relationships with Computers and Robots*, Springer, 2002, 125-132.
13. Parés, N., Carreras, A., et al. Promotion of creative activity in children with severe autism through visuals in an interactive multisensory environment *Proceeding of the 2005 conference on Interaction design and children*, ACM Press, Boulder, Colorado, 2005.
14. Tartaro, A. Storytelling with a virtual peer as an intervention for children with autism. *SIGACCESS Access. Comput.* (84). 42-44.

About the author

Joshua Hailpern is currently a Ph.D. student at the University of Illinois at Urbana-Champaign in the Computer Science Department focusing on Human Computer Interaction. He attended Carnegie Mellon University in Pittsburgh, graduating with a BS in Computer Science and a minor in Music Technology. His dissertation focuses on developing new technology to help teach lower functioning children with autism how to speak and utilize spoken language as a form of communication. He is working with his advisor, Karrie Karahalios as well as Professor James Halle from the Department of Special Education at UIUC and Laura DeThorne from the UIUC department of Speech and Hearing Sciences. Website: http://hailpern.com/joshua/