DSAI 2010

Proceedings of the 3rd International Conference on Software Development for Enhancing Accessibility and Fighting Info-Exclusion
Proceedings of the 3rd International Conference on Software Development for Enhancing Accessibility and Fighting Info-Exclusion

Oxford, United Kingdom, November 25-26, 2010

Edited by:
Paul Gnanayutham | Hugo Paredes | Ioannis T. Rekanos

Published by:
UTAD - Universidade de Trás-os-Montes e Alto Douro

Portuguese National Library Number: 319749/10
Preface

With the advent of Information and Communication Technologies (ICT) many new opportunities for citizens and consumers have opened up. It is imperative that people with disabilities are entitled equal access to education and employment, equal rights to parenthood, property ownership, political rights, and legal representation. People with disabilities include those who have long-term physical, mental, intellectual or sensory impairments, which, in interaction with various barriers, may hinder their full and effective participation in society on an equal basis with others. Computers, adaptive technology and the internet can help disabled people take advantage of education and employment opportunities leading to sustainable long term empowerment opportunities. In fact, ICT offer capacity building and combating poverty among the disabled in their communities, offering individuals the ability to access knowledge by adapting digital media, and to enhance their social and economic integration in communities. ICTs means new ‘digital opportunities’, ‘new job opportunities’, and of course ‘new hope’. The information society has the potential to contribute more in the form of knowledge resources. The resources must be developed to meet the requirements of all disabled people by overcoming the traditional barriers to mobility and geographic distance.


The main topics of DSAI 2010 include: software and Web accessibility (e.g., Web accessibility tools and evaluation, e-learning and accessibility, design for all); interfaces and interaction (e.g., ergonomics, accessibility and usability, adaptive interfaces, novel and multi-modal user interfaces); improvement of digital TV and electronic communications access (e.g., human-computer interfaces, non-invasive brain computer interaction, gesture interfaces); technologies (e.g., mobile Technology, virtual environments, augmented reality, assistive technology).

Keynote speeches will be given by Professor Gilbert Cockton, School of Design, Northumbria University, on “Inclusion Requires Inclusiveness”, and Professor Morten Tollefsen, Media Lunde Tollefsen AS, Norway, on “Developing an accessible user interface: creativity and the art of stealing”.

This book contains the proceedings of the DSAI 2010, organized by the SAE Institute, United Kingdom, in cooperation with the University of Trás-os-Montes e Alto Douro, Portugal and the University of Portsmouth, United Kingdom.

I would like to express my warm thanks to the co-organizers, reviewers, authors, speakers, participants, and to all that, in different ways, helped in the organization of DSAI 2010.

I hope to meet you again next year for the 4th DSAI International Conference edition.

Steering Committee,

Leontios Hadjileontiadis - Aristotle University of Thessaloniki, Greece
Committees

Conference Chair
Paul Gnanayutham - University of Portsmouth, United Kingdom

Conference Co-Chairs
Hugo Paredes - University of Trás-os-Montes and Alto Douro, Portugal
Ioannis T. Rekanos - Aristotle University of Thessaloniki, Greece

Organization Committee Chair
Jennifer George - SAE Institute, United Kingdom

Organization committee
Itai Bercovier - SAE Institute, United Kingdom
Hugo Fernandes - University of Trás-os-Montes e Alto Douro, Portugal
Steve Hartley - SAE Institute, United Kingdom
Paulo Martins - University of Trás-os-Montes e Alto Douro, Portugal

Steering Committee
João Barroso - University of Trás-os-Montes e Alto Douro, Portugal
Leontios Hadjileontiadis - Aristotle University of Thessaloniki, Greece
Francisco Godinho - University of Trás-os-Montes e Alto Douro, Portugal
Vitor Santos - Instituto Piaget, Portugal

Scientific committee
Adérito Marques - Universidade Aberta, Portugal
Antonio Coelho - Faculdade de Engenharia da Universidade do Porto, Portugal
António Teixeira - DETI/IEETA, Universidade de Aveiro, Portugal
Benjamim Fonseca - University of Trás-os-Montes e Alto Douro, Portugal
Carla Vieira Faria - Ajudas.com, Portugal
Cecilia Sik Lanyi - University of Pannonia, Hungary
Claude Moulin - Université de Technologie de Compiègne, France
Daniel Gonçalves - INESC-ID / IST, Portugal
Daniela Braga - Microsoft Language Development Center, Portugal
David Duce - Oxford Brookes University, United Kingdom
Diamantino Freitas - Faculdade de Engenharia da Universidade do Porto, Portugal
Enrico Pontelli - New Mexico State University, United States of America
Frode Eika Sandnes - Oslo University College, Norway
Franz Puehretmair - KL-I, Linz, Austria
Gabriel Pires - Institute for Systems and Robotics - Coimbra, Portugal
Gaël Harry Dias - Universidade da Beira Interior, Portugal
Gerhard Nussbaum - KL-I, Linz, Austria
Grigori Evreinov - University of Tampere, Finland
Gustavo Rossi - Universidad Nacional de La Plata, Argentina
João Rodrigues - Universidade do Algarve, Portugal
João Varajão - University of Trás-os-Montes e Alto Douro, Portugal
Joaquim A. Jorge - Instituto Superior Técnico, Portugal
John Gill - John Gill Technology Ltd
José Antonio Borges - NCE/UFRJ, Brazil
Jose Luis Sevillano - University of Seville, Spain
Katsuhito Yamaguchi - Nihon University, Japan
Lenka Lhotska - Czech Technical University, Czech Republic
Leonel Morgado - University of Trás-os-Montes e Alto Douro, Portugal
Ling-fu Meng - Department of Occupational Therapy, Taiwan
Luis Garcia - Polytechnic Institute of Beja, Portugal
Manuel Cota - Universidad de Vigo, Spain
Manuel J. Fonseca - INESC-ID/IST/TU Lisbon, Portugal
Miguel Sales Dias - Microsoft, Portugal
Maria Teresa Segarra - IT / Telecom Bretagne, France
Morten Tollefsen - Media Lunde Tollefsen AS, Norway
Paula Morais - Universidade Portucalense, Portugal
Pier Luigi Emiliani - Italian National Research Council, Italy
Ramiro Gonçalves - University of Trás-os-Montes e Alto Douro, Portugal
Ronald Vargas - Aalborg University, Denmark
Roland Wagner - University of Linz, Austria
Santiago Aguilera - Universidad Politécnica de Madrid, Spain
Simon Ball - JISC TechDis, United Kingdom
Teresa Chambel - Faculdade de Ciências da Universidade de Lisboa, Portugal
Vitor Filipe - University of Trás-os-Montes e Alto Douro, Portugal
Willibald Kremser - Austrian Computer Society, Austria
Wolfram Wöß - Johannes Kepler University, Austria
Zeljko Obrenovic - TU/e, Netherlands
Keynote Speakers
Inclusion Requires Inclusiveness
By Professor Gilbert Cockton, School of Design, Northumbria University

Interactive Technologies are useless unless they are used. This applies even more to assistive technologies, since poor uptake and diffusion here is a moral as well as design failure. Good intentions are not enough. Designs for inclusive technologies must be demonstrably successful with respect to personal and social criteria, and not just technical criteria. To achieve this, it is not enough to decide on who a technology is being designed to assist. As well as making choices about technologies and beneficiaries, design teams must make choices about evaluations and design purpose, and co-ordinate these with all other choices. Inclusiveness thus extends beyond choices of beneficiaries to the design process itself. Design processes must be comprehensive enough to maximise the changes of design success.

In my keynote presentation, I will present a high level framework for designing that can be specialised for Inclusive Design. The framework combines six meta-principle for designing with four types of design choice and their interconnections. Examples from past and current accessibility research will provide concrete examples of the impact of inclusive design processes on the achievement of inclusive design.

Bio

Gilbert Cockton is a Professor in Northumbria University's School of Design in Newcastle upon Tyne, UK, where he leads research on human-integrative approaches to design and evaluation, with a focus on balancing creative, technical, scientific and ethical inputs to design decisions. From 1997-2009, he was Research Chair in HCI at the University of Sunderland.

A Fellow of the Royal Society for the Arts, he has published extensively since 1985, with over 190 papers, chapters, books, articles and edited proceedings, and almost 200 invited presentations, on usability, user-experience and accessibility, the uses of empirical grounding, the theoretical bases for worthwhile design purposes, and notations and architectures for interactive software.

His career has blended education, academic research, childcare, design, consultancy, work for and within business and public sectors, directing large regional economic development projects, and professional service. He has secured funding for research and knowledge transfer projects and infrastructure with a value of almost £7M. He is currently scientific co-ordinator for the COST IC0940 TwinTide, a 25 country European network on inter-sector transferability of software design and evaluation approaches.

He has served in many roles within the international HCI community, including Vice-Chair of IFIP TC13 (2004-06), Chair of British HCI Group (2001-2004), Chair of ACM CHI 2003 and BCS HCI 2000 Conferences, and Secretary of IFIP WG2.7 on user interface engineering (1993-99). He is Editor Emeritus of the journal Interacting with Computers, a member of the editorial board of the Journal of Usability Studies, and an advisor to national projects in Japan, Finland and Poland.
Developing an accessible user interface: creativity and the art of stealing

By Professor Morten Tollefsen - Media Lunde Tollefsen AS, Norway

Physiotherapy has been, and still is, a common occupation for visually impaired. Administrative software is used to set up time sheets, for billing purposes, for writing journals etc. In Norway commercial software used by physiotherapists has not been accessible for blind and visually impaired. Funds were therefore granted for the development of an appropriate product / user interface for these users.

In this lecture I want to show how badly a standard product could work with assistive devices. Furthermore, I will show how much potential there is for accessibility with today's technology and development tools.

The program for administration of physiotherapy includes innovative solutions for many of the main challenges for visually impaired, for example efficient use of forms, tables, and use of date fields / calendars. Are these solutions only smart for the visually impaired, or are they smart solutions for everyone?

Bio

Morten Tollefsen took a masters degree in computing in 1991. He worked at the University of Oslo and Oslo municipality until 1999, when he started the company MediaLT with Magne Lunde. Both Lunde and Tollefsen are blind.

They wanted to start a company with focus on three key areas of ICT for the disabled: research and innovation, training, and adaptation. MediaLT currently has 15 employees and has become a leading R & D community in ICT for disabled people in Norway.

Tollefsen has led several national research projects, and participated in a number of EU projects. Tollefsen is particularly concerned with human-machine interaction, and since the introduction of www, has worked with accessibility of web-based user interfaces.
Invited Papers
Inclusion Requires Inclusiveness

Gilbert Cockton
School of Design, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
gilbert.cockton@northumbria.ac.uk

Abstract—Interactive Technologies are useless unless they are used. This applies even more to assistive technologies, since poor uptake and diffusion here is a moral as well as a design failure. Good intentions are not enough. Designs for inclusive technologies must be demonstrably successful with respect to personal and social criteria, and not just technical criteria. Success is more likely if emerging positions and approaches from Interaction Design are adopted. In this keynote presentation, I will present a high level framework for designing that aims at balance, integration and generosity. The framework is inclusive with respect to both intellectual traditions and also the needs of design work. I argue that inclusiveness within the design process is required to achieve fully human accessibility.

I. INTRODUCTION

In 2001, the Assembly of the World Health Organisation (WHO) endorsed The International Classification of Functioning, Disability and Health (ICF [17]). The new classification adopted a balanced and integrative approach to disability. It abandoned an exclusively biomedical approach to disability. Instead, it took what it called, somewhat clumsily, a biopsychosocial approach to ‘disability’, which became an umbrella term for impairment, activity limitation and participation restriction. Impairments, due to limitations in body functions and structures, limit activities and restrict participation, but only in the context of environmental factors and personal factors.

The ICF moved away from a biomedical position that disability solely results from bodily impairments, towards a systems view of disability as resulting from interacting biological, psychological and social factors.

We should expect research on accessibility and assistive technologies to be compatible with this biopsychosocial account of disability. In my keynote, I explore what is required to achieve such compatibility through design research. I make extensive use of very familiar examples of assistive technology, spectacles and contact lenses, to motivate key points.

II. THREE APPROACHES TO ACCESSIBILITY

Accessibility research could logically consider only one, or some combination of, biological, psychological and sociological factors when designing and evaluating assistive technologies (ATs).

A. Biomedical Approaches to Accessibility

A biomedical approach to accessibility would treat bodily impairments as the design problem, and would explore design solutions to overcome them. Design problems here are essentially engineering ones. Bodies that are deficient with respect to some function can be augmented by a well designed innovative AT. Such innovations would succeed when someone could function as if their impairment were not there.

Such engineering approaches have been effective for centuries. For example, the invention of eye glasses in Italy in the thirteenth century was followed by centuries of improvements to lenses as first, scientific understandings of optical physics developed, and then scientific understandings of ophthalmology. At some point, people with vision corrected by prescription glasses could see as well as people with normal vision, e.g., when performing a Snellen test from a fixed distance.

In the UK after 1945, the National Health Service (NHS) provided free spectacles for children, but with a limited range of frames. The most common were round plastic coated wire (pink or blue) that were instantly recognizable as free glasses worn by children whose parents could not afford to buy more individual styles. In terms of performance on a Snellen test, free NHS glasses and purchased ones would be equivalent. However, many children felt stigmatized by NHS glasses, and memories persist decades later. Catherine Hayward, fashion director at Esquire magazine, was recently quoted in the BBC news magazine [8]:

In the 70s when my sister had to wear glasses at 10 or 11, we were all devastated. She was crying. ‘Oh no, being subject to wearing glasses for the rest of your life.’ Those ugly National Health ones. The only one in the class wearing them.

The objective fact (that children with NHS spectacles could see as well as children with normal sight, or with privately purchased spectacles) is a very narrow a basis for judging the success or failure of spectacles. Acceptable accessibility requires more than engineered augmentations of human bodies. Bodies contain people, who have feelings, motives, experiences, social relationships and everything else that makes us human.

B. Biopsycho Approaches to Accessibility

We could move one step closer to the ICF position on disability by considering both physical and psychological factors in accessibility.

To continue with our eyesight correction example, standard issue free spectacles, privately purchased spectacles and contact lenses can be technically equivalent, with the exception of some conditions where contact lenses cannot be worn. However, the
reasons for preferring contact lenses to spectacles are largely psychological, and relate to people’s preferences about their appearance. Given this, it is hard to consider eye correction as being solely an engineering design problem, where a solution can be synthesized, based on a technical specification [15]. People’s feelings about their appearance are not amenable to precise specification, and thus the design problem here cannot be specified in a way that allows solutions to be synthesized on the basis of a formal analysis of the problem. Nevertheless, a short glance at any retail display of spectacle frames, from ‘designer’ models to stylish budget versions, demonstrates that a lack of technical specifications has been no obstacle to the design of eyewear that responds to the needs and wants of diverse populations. As a result, we rarely think of spectacles or contact lenses as prostheses or assistive technologies. They have become lifestyle choices, which we should take as a critical indicator of success for accessibility.

Psychological considerations are additional to biomedical ones. Defects in vision can be objectively measured, as can the quality of optical solutions, so for part of the problem of eye correction, an engineering design approach is absolutely necessary. However, it is not sufficient.

Turning attention to accessible digital technologies, here too it is not enough to design assistive technologies that are compatible with some objective measure of individual capability. One of my former PhD students, Brendan Cassidy, developed the idea of acceptable accessibility [2] to extend our focus on accessibility beyond objective biomedical limitations to body function and structure. In one of his studies, a participant with Fibromyalgia was able to use function keys on ATMs (Auto Teller Machines), but preferred a novel interaction device, because she did not have to remove her crutches to use it. However, as her condition varied, she did not use crutches every day. Her range of objective capability and related preferred challenge thus varied, invalidating any biomedical approach to design based on fixed specifications of individual capability.

Acceptable accessibility covers a span of capabilities. Effective capability can be reduced by comfort factors, where an individual would rather not operate at the ‘official’ limit of their capability. However, impaired individuals are also often highly motivated to avoid limitations to their activities and restrictions on their social participation. Here, an individual’s aspirations may take them beyond some medically assessed limit of their capability. For example, individuals may make strenuous efforts to be independent, and be motivated to work well beyond their comfort zone, and well beyond any capabilities as established in a formal medical assessment.

Variations in individual motivation thus limit the effectiveness of biomedical approaches to accessibility, and the engineering design approaches that are associated with them. Motives, feelings and experiences are more the domain of applied arts design disciplines, such as product, fashion or graphics design. Here, problems and solutions co-evolve [7], so difficulties associated with specifying problems relating to motives, feelings or experiences are far less of an issue.

Engineering design requires a clear tractable problem specification that allows focused systematic analysis and synthesis of solutions that can be validated or verified using quantitative measures. Applied arts design uses solution generation to drive problem definition. A key distinguishing feature of the wicked problems [16] that characterize such applied arts design disciplines is that typically problems are not fully understood until after an acceptable solution has been created.

Accessibility thus requires design approaches that embrace the challenges of wicked problems. As with eyewear, we are beginning to see some assistive technologies moving from crude unattractive engineering solutions to attractive well-considered designs, for example, mobility scooters, wheelchairs, and a range of walking aids such as crutches that come in a range of form factors and styles.

C. Biopsychosocial Approaches to Accessibility

Returning to our example of correcting defective eyesight, contact lenses are a good solution for people who don’t want to expose their defects in vision in public, and/or those who are not happy with their appearance when wearing spectacles. However, contact lenses can be unsuitable when flying, or for prolonged computer use. They cannot be worn without tight fitting goggles when swimming. For some activities, problems arise with both spectacles and contact lenses, especially in environments where lenses will mist up, or glasses or lenses can be physically dislodged. Here, social participation may be restricted, even though it is technically possible to correct eyesight. However, participation does become possible with specialist eyewear such as goggles or masks with prescription lenses. The key point here is that different physical or social environments may require different assistive technologies. A single AT may reduce activity limitations and enable participation in one context, but be disabling in another.

Social considerations further complicate treating accessibility solely as an engineering design problem. For example, with training, an impaired individual and their regular carers may be able to use an assistive device, but other individuals who are unfamiliar with a device may not be able to interact with someone using it, e.g., a communication device for young children with severe motor impairments. Dawe’s studies of families with children with cognitive impairments [5] demonstrated the need to consider the multiple individuals and stages involved in the technology adoption process. It is not enough to demonstrate that
impaired individuals can use a device. All members of an individual’s care circle must be able to use it, directly or indirectly. Just as a psychological perspective on accessibility extends our focus to acceptable accessibility, so too does a social perspective further extend our focus to socially appropriate accessibility.

Turning to digital accessibility, accessibility standards sometimes have less to do with someone’s condition than with the ATs that they use. Most web accessibility standards for the visually impaired have no relation at all to an individual’s physical condition, but instead relate to current limitations of screen reading technologies, such as a longstanding inability to deal with dynamic web capabilities such as JavaScript. Here, disability is clearly a result of an individual’s technological environment, where it is the interaction of websites and screen readers that limits activities and restricts participation, and not simply a person’s body. Activity limitations and participation restrictions need not be solely the concern of website developers. Indeed, it would be far more effective to improve everyone’s screen readers, perhaps as a public good, rather than stick with inadequate commercial offerings. Whereas standard cheap spectacle frames caused embarrassment to socially disadvantaged British children in past decades, a highly capable standard screen reader could greatly improve web accessibility for the visually impaired. ATs are now components in digital product-service ecosystems, and thus cannot be considered in isolation. From this perspective, the next version of the WHO’s ICF needs to take a biopsychosociotechnical approach!

D. Fully Human Accessibility

The above raises the important question as to whether current design research settings are adequate for meeting the challenges posed by ICF for Accessibility. We need design settings that combine appropriate elements of both engineering and applied arts design, which could focus on all of the following:

- Technical options for augmenting human movement, perception and cognition
- Technical options for integrating and co-ordinating across assistive and non-assistive devices, without loss of capability in any integrated technology
- Creative options for designing for human aspirations and comfort
- Creative options for designing for human social relationships

Such design settings must include approaches from both engineering and applied arts design traditions. Neither design tradition is adequate on its own. Engineering design approaches are needed to implement systems that satisfy specified technical requirements. Applied arts approaches are needed to design for the wide range of human values, concerns and issues that are not amenable to precise appropriate technical specifications. Here, solutions depend on the creative simultaneous refinement of problem and solution spaces. While engineering design approaches tend to dominate research on accessibility and inclusion (e.g., rehabilitation engineering, computer systems design), applied arts approaches have been successfully applied. For example, Hutchinson and collaborators developed an award winning creative multimedia piece that focused on terminal illness [13]. There has also been a over a decade of creative work on technology support for art therapy for people with cognitive and emotional difficulties (e.g., [4,11]). For fully human accessibility we need both engineering and creative approaches to design: one or the other in isolation is not enough.

III. NEW DESIGN SITUATIONS FOR THE ICF CHALLENGE

We require new design settings that span approaches from both engineering and applied arts design traditions. As with any design setting, these new ones can be characterized by the range of design choices that have to be made in them. We can thus expect types of design choices from both engineering and applied arts design to be combined or merged within design settings that can fully meet the challenges of the ICF.

The pivotal type of design choice is the choice of design purpose. In engineering design, such choices can be made too early, before a design problem is truly understood. In applied arts design, such a choice may never be explicitly made. We need approaches to design purpose that let the problem specifications of engineering design be extended to embrace loosely characterized human concerns. Unlike engineering design, where an early precise verifiable problem specification is preferred, a looser sense of design purpose may steadily evolve alongside explorations of the solution space. The ICF demands a broad view of design purpose for accessibility, extending purpose beyond simple compensation for impaired body functions through reduction of activity limitations to increasing participation in society. Any true reduction of information and other forms of exclusion requires more than a narrow engineering solution that simply compensates for an impaired bodily function.

The first type of design choice has been introduced as a choice of design purpose. In its engineering form as a problem or requirements specification, it is well understood as the foundation for success. However, too often these are not presented as a matter of choice, but of inevitability. Customer and user research is often presented as a set of activities that will inevitably converge on a set of requirements that are correct and complete. The result is the problem specification rather than a problem specification that is the result of a series of choices on how to conduct requirements elicitation.
and how to interpret the collected evidence. It is vital to see all design decisions as choices, where there are always alternatives, and where the choice of one option over another always requires judgements that favour one set of evaluative criteria (i.e., values) over others. Where there is time pressure to get problem specification over and done with as soon as possible, the poor choices of purpose are bound to result.

Difficulties in establishing valid stable clearly specified problem statements early in the development process can be managed via an iterative approach to design. To drive each iteration, it is important to carry out design evaluations that can indicate what needs to be done next. This exposes a second type of design choice, i.e., the choice of evaluation. As with design purpose, this is chosen and should never wholly be the result of some fixed requirements. For example, a design team may decide to test a prototype with able bodied or impaired users depending on the maturity of the prototype. They may additionally choose to conduct a laboratory based evaluation, with each participant session lasting no more than a few hours. Alternatively, they may choose to leave a device with a few participants for several weeks. For the first option, all interaction may be closely observed and recorded, for the second, participants and/or their carers may be relied upon to collect and/or recall data.

Different types of design choice need to be co-ordinated. For our first two types of design choice, evaluation criteria should be derived from design purpose, rather than be related to measures that are easy to make, but may have no bearing on design success, such as time on task, error counts, frequency of help requests and similar easily collected measures.

If we had based our analysis of design choices on popular media on design, we would have seen design choices as wholly focused on artefacts. The role of the designer would be solely to make choices of materials, features and qualities. Design purpose and evaluation are rarely regarded as types of design choice in popular accounts of design. However, they are clearly positions and activities that can only result from human choices.

**Choices about artefacts** are the most obvious type of design choice. They are typically made deliberately and explicitly in much engineering design, but may be tacit in applied arts practices. Similarly, choices of evaluation will also typically be made deliberately and explicitly in much engineering design, but in applied arts design, evaluation may be hidden away within the reflective practice of designers. The same applies to design purpose, which will be explicit in engineering design (albeit in a form of a problem specification that can be too narrow in scope), but may never be articulated at all in the applied arts, especially as a design is evolving through reflective practice.

There is a fourth type of design choice which can be also be narrowly constructed in both engineering and applied arts design. We design for people, and thus we should know who we are designing for, as well as what we are designing (artefact choices), why we are designing (design purpose) and whether we have succeeded (choices of evaluations). In making choices about who we are designing for, we are choosing who will benefit. **Choices of beneficiary** are thus a fourth type of design choice. To meet the challenges of the ICF, we need to widen the scope of beneficiaries from individuals with specific impairments to a wide circle of social roles associated with impaired individuals.

To have an adequate span of concerns to be able to address the challenges of the ICF, a design setting has to have an abstract core that can make four types of design choice:

- **Purpose** – why is an AT being designed?
- **Artefact** – what specific AT is being designed?
- **Beneficiaries** – who is a specific assistive technology being designed for?
- **Evaluations** – whether a specific AT delivers the intended benefits to each included beneficiary?

Engineering design approaches have little, if any, support for broad choices of beneficiaries. Nor do they provide broad support for evaluation of a wide range of benefits. Human-centred design approaches, by contrast, are strong on developing understandings of beneficiaries and evaluating benefits. However, human-centred approaches can assume that people’s needs and wants can be easily accessed in ways that make design purpose obvious. In contrast, good engineering designers expect to expend considerable initial effort on defining and specifying design problems, resulting in a focused design purpose that is less common in human-centred approaches from applied arts design.

Concrete design settings can thus be stripped bare to an abstract core, comprised initially from different types of design choice. This abstract core can be called an **abstract design situation** (ADS). The level of abstraction here however is not at the expense of practical utility. ADSs can reshape the ways in which we think about design. As the number of choice types increase, it becomes harder to see design as having any single centre. For example, design purpose cannot be wholly user-centred, since choices here inevitably involve judgments by the design team, which may rightly be strongly influenced by the type of design solution that they have in mind. Design purpose is not solely a consequence of choices of beneficiaries. Instead, it results from how design teams’ understandings of technical and creative opportunities interact with their understandings of beneficiaries.

More complex ADSs (and thus their corresponding concrete design settings) cannot be organized around some single **central** type of design choice. Instead, the required design practices are balanced and integrative [1]. Furthermore, given that solutions can often overshoot design purpose, such design practices have the potential to be generous, delivering unintended benefits, perhaps to unintended beneficiaries.
Big challenges (such as ICF) require big solutions. By requiring explicit choices of purpose at some point in the design process, we create design settings that are broader than simpler user-centred approaches. By requiring deep understandings of potential beneficiaries and an imaginative range of evaluation options, we create design settings that are broader than tightly defined engineering design. Rather than being biased towards either technology or people, design must integrate across different choice types in a balanced manner. In the process, opportunities often emerge for delivering unforeseen benefits. The result is a design setting that is Balanced, Integrative and Generous, i.e., one that is BIG enough to meet the ICF’s challenges.

IV. DESIGN WORK FOR NEW DESIGN SITUATIONS

ADSs are useful for exposing contrasts between different design traditions, such as engineering and applied arts, and within these traditions (e.g., artefact-centred, designer-led, user-centred, technology-push, user-pull). ADSs are a simple auditing tool that can expose the current commitments within design settings. Some design settings may commit to all four identified types of design choice (or perhaps more, the above may not be exhaustive). Other design settings may only commit to two (e.g., artefact and evaluation in designer-maker craft settings) or three (e.g., technically-led design with only purpose, artefact and evaluation). Furthermore, such an audit can be extended to consider types of connections between different types of design choice.

As ADSs become more complex (i.e., have more design choice types), the potential connections between choices become more complex too. No concrete design setting would ever attempt to explicitly consider every type of connection, since three-way, four-way and more complex connections are possible, and there are even connections to connections. For example, a scenario description is a narrative of usage that can connect the features and capabilities of a proposed design to its intended benefits (scenario outcome/s). The scenario connects an (envisaged) artefact to its benefits (i.e., design purpose). However, if we evaluate this scenario, we would be evaluating a connection, and thus connecting from the evaluation criteria to a connection between an artefact and design purpose. This example of a connection to a connection indicates the potential complexity of connections within even an ADS, never mind concrete design settings.

In theory, as connection types become ever more complex, an infinite number of connections are possible. This requires design teams to have a clear view of not only the types of design choice that will be explicitly made, but also the types of connections that will be forged between them. Once extended to include the possible types of connections between types of design choices, the ADS concept extends the scope of audits of concrete design settings. A design team can establish their commitment to a range of design practices through identifying their explicit design choice types and their explicit connection types between them. At this point, the ADS concept becomes something that can support design work, albeit at a very abstract strategic level.

Committedness is our first example of a meta-principle for designing. Meta-principles are a further form of abstraction that indicate the work that design approaches can support. The ADS concept is a simple design management tool that supports committedness by scoping out complete design processes. Other metaprinicples focus on specific design work.

Six meta-principles for designing have been derived and motivated [3]. These extend ADSs by scoping out a range of design work that supports balanced, integrative, generous design, with its fusion of engineering and applied arts practices.

Three simple meta-principles relate to forming menus of design options from which choices can be made. They are simple because they can apply to only a single design choice or connection type. They focus attention on support for making choices within design, with each meta-principle focusing attention on distinct aspects of design work, i.e., collecting, communicating and checking design options.

The second meta-principle, receptiveness, emphasizes the need to be open to a wide range of ways to collect options. Design teams need to be receptive to possibilities for collecting options all types of design choice and inter-connections. Brainstorming, sketching, personal experience, state of the art surveys, co-design, literature surveys, design patterns and guidelines, historical surveys, theoretical sensitivities, creativity methods, trends analysis, and human science research methods are examples of design approaches that can improve receptiveness.

The third meta-principle, expressivity, emphasizes the need to effectively communicate within design settings. Design teams need to communicate options for all choice and connection types to all stakeholders. Sketches, prototypes, research reports, evaluation plans, requirements specifications, mood boards, value propositions, product visions, scenarios and personas are examples of design approaches that can improve expressivity.

The fourth meta-principle, credibility, emphasizes the need to be realistic about genuine options. Design teams need to test the credibility of possible options for all choice and connection types. Formal modelling, feasibility studies, pilot testing, data analysis methods, predictive models, literature surveys and expert reviews are examples of design approaches that can improve credibility of design options in isolation.

Two further complex meta-principles relate to the work that can be achieved as a result of balancing and integrating a range of design choices. They are complex because they must apply to more than one
design choice and connections between them. The fifth meta-principle of improvability focuses attention on the extent to which design choices can be systematically evaluated, understood in relation to the overall combination of choices, and iterated to address problems. Evaluation methods cannot provide sufficient support for improvability. It is not enough to identify the strengths and weaknesses of a design (evaluability). Design approaches are also needed that can expose the cause of strengths and weaknesses (understandability) and that can generate appropriate responses on the basis of understandings (responsiveness). While evaluation method are well developed, more broad iteration methods that can translate evaluation results into effective responses are still not well developed for universal design.

The sixth meta-principle of inclusiveness focuses attention on the extent to which all design choices and their co-ordinating connections can deliver benefits (at acceptable costs) to a wide range of beneficiaries. Within information systems design, stakeholder analysis has been the main approach to improving inclusiveness. However, most existing stakeholder analysis depends on formal organisational structures. For ATs, the care circle of an impaired individual always spans organisational, community and family boundaries. Innovative approaches such as worth webs [9] are currently under development. Advances here can draw on existing research on supporting parents of neonates in intensive care [14].

Extending abstract design situations with meta-principles for designing creates a more extensive basis for meeting the ICF’s challenges. Committedness to a broad enough set of design choice types is necessary, but not sufficient, for fully human accessibility. Once a broad enough set of design choice types has been committed to, as well as a sufficient set of interconnections to allow effective co-ordination of these choices, a design team needs to ensure that sufficient high quality options can be collected, checked and communicated for choices and connections. In addition, choices need to be co-ordinated to maintain the improvability and inclusiveness of the overall design process.

Hirsh and Johnson’s Comprehensive Assistive Technology (CAT) Model [12] provides an alternative approach to responding to the ICF’s challenges. The CAT model visualises a hierarchy of prompts that can guide AT design and evaluation. For example, the CAT model’s ‘Person’ and ‘Context’ attributes can be used to identify a target audience of beneficiaries. Also, its ‘Activities’ attribute can be used to identify accessibility challenges that could form the basis for a problem specification approach to design purpose. The CAT model can also be used to design a potential communication solution (artefact) by analysing a disabled individual, to evaluate specific assistive devices, and to evaluate choices of AT for a specific individual. The CAT model is clearly versatile and comprehensive. However, there are no clear ways to extend it or strengthen it, and it makes limited use of the wide range of design and evaluation approaches that have been developed in relevant areas such as product design, interaction design, occupational therapy, rehabilitation engineering, or universal design. By starting with high level concepts such as an ADS extended with meta-principles for designing, we can create a more open framework that can incorporate a design team’s existing approaches, while drawing their attention to gaps in their repertoire and ways to fill these. Table 1 illustrates how existing approaches from interaction design can be associated with design choice types and meta-principles.

**Table 1: Scoping methods by Choice Type and Meta-principle**

<table>
<thead>
<tr>
<th></th>
<th>Receptiveness</th>
<th>Expressivity</th>
<th>Credibility</th>
<th>Inclusiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artefacts</td>
<td>Literature review</td>
<td>Sketches</td>
<td>Feasibility</td>
<td>Sustained ability</td>
</tr>
<tr>
<td>Beneficiaries</td>
<td>Probes</td>
<td>Personas</td>
<td>Ethnography</td>
<td>Worth webs</td>
</tr>
<tr>
<td>Evaluations</td>
<td>Human Science, Arts</td>
<td>Reports, Presentations</td>
<td>Human Science</td>
<td>Value-sensitive design</td>
</tr>
<tr>
<td>Purpose</td>
<td>Theory</td>
<td>Product Vision</td>
<td>Human Science</td>
<td>Theory</td>
</tr>
</tbody>
</table>

Design and evaluation approaches can be shown to support one or more meta-principles, typically for a single type of design choice or connection. A key benefit of meta-principles for designing is that, when combined with ADSs, they support the scoping of design and evaluation approaches. Gaps in existing design practices can be identified, and new approaches can be developed to fill these gaps.

Few existing concrete design settings have sufficient methodological power to be able to fully meet the challenge of the WHO’s ICF. This is particularly the case when designing, not to simply reduce limitations for simple isolated activities, but instead to demonstrably remove restrictions on participation in society for impaired individuals.

To meet the challenge of the ICF, we need to develop frameworks of design and evaluation approaches that support balanced, integrative and generous design. To achieve this, for any ADS, with its committed to design choice types and connections, we need approaches that:

1. Promote receptiveness to a wide range of design options for each committed to choice and connection type
2. Establish the credibility of design options through analyses of validity, feasibility, etc.
3. Promote expressivity of design options to all
stakeholders
4. Support **improvability** for a current concrete design situation
5. Balance **inclusiveness** for a current concrete design situation

Most current combinations of design and evaluation approaches do not fully support all relevant meta-principles for all committed to design choice and connection types. It is easy to highlight gaps that will typically prevent design teams from developing highly inclusive designs. Once identified, we can begin to explore ways to fill such gaps, such as receptiveness to care circle structure as an approach to identifying a wide range of potential stakeholder beneficiaries [9].

In my keynote talk, I will present examples of work by former and current PhD students that illustrates the principles of balanced, integrative and generous design by co-ordinating different types of design choice.

### Table 2a: Major Choices by Type for four accessibility PhDs

<table>
<thead>
<tr>
<th>Eamon Doherty</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Broad range of initial purposes (e.g., recreation, feeding, appliance control), eventually focused on communication.</td>
</tr>
<tr>
<td><strong>Artifact</strong></td>
<td>Cyberlink brain-body interface device with a range of interface styles and configurations with <strong>direct</strong> cursor control</td>
</tr>
<tr>
<td><strong>Beneficiaries</strong></td>
<td>Medical staff, occupational therapists and family members</td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
<td>Formal experiments and longitudinal field studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paul Gnanayutham</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Broad range of initial purposes maintained.</td>
</tr>
<tr>
<td><strong>Artifact</strong></td>
<td>Cyberlink brain-body interface device with an adaptive tiled interface style with <strong>indirect</strong> cursor control</td>
</tr>
<tr>
<td><strong>Beneficiaries</strong></td>
<td>Primarily family members supporting use with no researcher present</td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
<td>Formal experiments (including able bodied participants) and longitudinal field studies</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Brendan Cassidy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Increased/total independence for deaf-blind/visually impaired users of public terminals (e.g., ATMs).</td>
</tr>
<tr>
<td><strong>Artifact</strong></td>
<td>Flexible prototype rig with novel input device (move-select), and vibrational direction and status indicators</td>
</tr>
<tr>
<td><strong>Beneficiaries</strong></td>
<td>Deaf-blind/visually impaired users of public terminals, manufacturers and providers of public terminals</td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
<td>Formal experiments (including ones with able bodied participants)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Jenni George</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Improving choice and usage of AT devices for children with extensive motor impairments.</td>
</tr>
<tr>
<td><strong>Artifact</strong></td>
<td>Social network</td>
</tr>
<tr>
<td><strong>Beneficiaries</strong></td>
<td>Care circles of children with extensive motor impairments</td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
<td>Under development</td>
</tr>
</tbody>
</table>

I have supervised four PhDs with a focus on accessibility. Eamon Doherty (currently at Fairleigh Dickinson University) developed an interaction design approach to the limitations of current brain-body interfaces [6]. Paul Gnanayutham (currently at Portsmouth University) developed algorithmic improvements to Eamon’s work that enabled independent usage of a brain-body interface by individuals and their carers [10]. Brendan Cassidy [2] (currently at UCLAN) developed patented applications of vibrational haptics for visually and aurally impaired users of public terminals. A current PhD student, Jenni George, is currently developing a social networking approach to supporting the care circles of children with extensive motor impairments [9].

The underlying structure of the concrete design settings for these four PhDs is summarised in Tables 2a and 2b. In my keynote talk, I will relate design and evaluation approaches used in these PhDs to meta-principles for designing, and will indicate where improvements could be made to balance, integration and generosity within their design processes.

### Table 2b: Major Choices by Type for four accessibility PhDs

V. CONCLUSIONS

The ICF challenges accessibility research to reach beyond simple compensations for impaired body functions and structures to the demonstrable removal of restrictions on an impaired individual’s ability to participate in society. Neither engineering design, applied arts design, nor user-centered approaches are
adequate in isolation. The ICF challenges us to combine existing design paradigms in order to develop fully human approaches to accessibility.

This paper has presented an abstract framework for reasoning about the range of choices and connections that have to be made within design settings, and the work that design and evaluation approaches must support to enable balanced, integrative and generous approaches to inclusive interaction design. The framework is ambitious, and it will take several years to develop and test new concrete approaches to filling the gaps in existing concrete design settings for accessibility.

This paper’s title is *Inclusion Requires Inclusiveness*. This requirement extends beyond the meta-principle of inclusiveness to the breadth of the design process itself. To achieve genuine inclusiveness, the design process itself must be inclusive, i.e., it must support a wide range of design choice types and their co-ordinating connections, and both choice and connection types in turn must be given support to deliver on a broad range of meta-principles for designing.

Whenever well-intentioned attempts at universal design fail, it is often because the underlying design process has been narrowly restricted to engineering, applied arts or user-centred paradigms. The familiar example of vision correction devices makes it clear that the existence of the current range of prosthetic eyewear depends on a broad design perspectives. We cannot design for a more inclusive world until we have created a design process that is genuinely inclusive with respect to the types of choice that must be made and co-ordinated, and the types of work that design and evaluation approaches must perform. Even in the abstract, the requirements here are more extensive than we have generally realised. There are significant challenges ahead, but I am confident that they can be overcome.

REFERENCES


Phantom Limb Pain Management Using Facial Expression Analysis, Biofeedback and Augmented Reality Interfacing

D. Tzionas, K. Vrenas, S. Eleftheriadis, S. Georgoulis, P. C. Petrantonakis, and L. J. Hadjileontiadis
Dept. of Electrical & Computer Engineering
Aristotle University of Thessaloniki
GR 54124 Thessaloniki, Greece
leontios@auth.gr

Abstract—Post-amputation sensation often translates to the feeling of severe pain in the missing limb, referred to as phantom limb pain (PLP). A clear and rational treatment regimen is difficult to establish, as long as the underlying pathophysiology is not fully known. In this work, an innovative PLP management system is presented, as a module of an holistic computer-mediated pain management environment, namely Epione. The proposed Epione-PLP scheme is structured upon advanced facial expression analysis, used to form a dynamic pain meter, which, in turn, is used to trigger biofeedback and augmented reality-based PLP distraction scenarios. The latter incorporate a model of the missing limb for its visualization, in an effort to provide to the amputee the feeling of its existence and control, and, thus, maximize his/her PLP relief. The novel Epione-PLP management approach integrates edge-technology within the context of personalized health and it could be used to facilitate easing of PLP patients’ suffering, provide efficient progress monitoring and contribute to the increase in their quality of life.

I. INTRODUCTION

Pain, although it is the oldest medical problem and the universal physical affliction of mankind, it has been little understood in physiology until very recently. Nowadays, pain is generally agreed-upon as an experience which involves more than just physical sensations. The International Association for the Study of Pain (IASP)'s defines pain as ‘an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage’ [1].

The most common types of pain are [1] acute pain (results from disease, inflammation, or injury to tissues) and chronic pain (is widely believed to represent disease itself), which both remain a huge problem. One of the peculiar type of pain is the one that relates with the post-amputation sensation. In fact, many amputees are frequently aware of severe phantom limb pain (PLP) in the absent limb, since the brain cells affected by amputation do not simply die off, but neurons in the brain remain dynamic and excitable. PLP is real and is often accompanied by other health problems, such as depression.

As pain is directly connected to the patient’s everyday life, its management is of great importance to elevated his/her quality of life. The goal of pain management is to easing the suffering and improving the quality of life of those living with pain, enabling them to work, attend school, or participate in other day-to-day activities. Treatment approaches to long term pain include pharmacologic measures, interventional procedures, physical therapy, physical exercise, application of ice and/or heat, and psychological measures, such as biofeedback and cognitive behavioral therapy, like relaxation and the use of imagery as a distraction that provides relief [2]. Actually, pain management sets a challenging field, where edge technology could be incorporated within pain management approaches to create an effective environment that could contribute to better pain handling; hence, improving the quality of life at a global level.

In this work, a PLP management system is presented. This is, actually, a module of an holistic computer-mediated pain management environment, namely Epione[3]. The proposed Epione-PLP scheme uses advanced facial expression analysis, as a means to form a dynamic pain meter, which is then used to trigger biofeedback and augmented reality-based PLP distraction scenarios. The latter are structured upon a model of the missing limb for its visualization, in an effort to create to the amputee the feeling of its existence and control, and, thus, maximize his/her PLP relief. The proposed system introduces a novel environment of PLP management that could assist patients and physicians during the process of easing of suffering, contributing to the new culture of Personalized Healthcare2.

II. BACKGROUND

A. Phantom Limb Pain (PLP)

Ambrose Paré (1510-1590), a French military surgeon, has given the first medical description of post-amputation sensation, who noticed that patients may complain of severe pain in the missing limb following amputation. Paré characterized the post-amputation syndrome in his ‘Haquebusses and other guns’ and

1 Epione (Ηπιόνη) was the goddess of the soothing of pain. She was the wife of the medicine god Asklepios, and the mother of a number of minor healing gods, including Hygeia (Good Health), Panakeia (All-Cures) and Iaso (Healing).
2 Personalized Healthcare is an important new approach to medicine, providing great benefits to patients, by improving treatment, enabling people to monitor their own health, increasing compliance and providing inputs to computer systems and associated communications networks and embedded devices that will make it possible to organize, interpret and interrogate this information and apply it to tailoring medical treatment. Personalized Healthcare is at the core of the European Community Initiative ‘European Research Area 2030: Preparing Europe for a New Renaissance-A Strategic View of the European Research Area’ [4].
proposed different models to explain the pain [5]. Subsequent studies by Charles Bell (1830), Magendie (1833), Rhone (1842), Guéniot (1861) and others gave detailed descriptions of the phenomenon and, in 1871, Mitchell introduced the term 'phantom limb’ [6], [7].

Nowadays, it is common knowledge that virtually all amputees experience phantom sensations after limb amputation; the phantom becomes the site of severe pain, which may be exceedingly difficult to treat. A clear and rational treatment regimen is quite difficult to establish, since the underlying pathophysiology is not fully known. Nevertheless, nerve injury is followed by a series of changes in the peripheral and the central nervous system; hence, these changes play a role in the induction and maintenance of chronic phantom pain [8]. Following a nerve cut, formation of neuromas are seen universally, which show spontaneous and abnormal evoked activity following mechanical or chemical stimulation [9]. The ectopic and increased spontaneous and evoked activity from the periphery is assumed to be the result of a novel expression or up-regulation of sodium channels [10]. The sympathetic nervous system may also play an important role in generating and, in particular, in maintaining PLP [11]. Studies based on electrophysiology have documented the existence of nociceptive specific neurons and wide dynamic range neurons in the cerebral cortex. Following limb amputation and deafferentation of adult monkeys, there is a reorganization of the primary somatosensory cortex, sub-cortex and thalamus [12]. A similar reorganization has been observed in humans based on the information from the magnetoencephalogram. Interestingly, this cerebral reorganization was seen mostly in patients with PLP and there was a linear relationship between pain and degree of reorganization [13].

From the aforementioned, it is likely that the first PLP events occur in the periphery, which subsequently generates a cascade of events that sweep more centrally and also recruit cortical brain structures [8]. The occurrence of PLP seems to be independent of age in adults, gender and level, or side of amputation [14]. Moreover, the incidence of PLP is similar following civilian or military accidents [15]. Several studies have shown that 75% of patients develop pain within the first few days after amputation [16]. However, phantom pain may be delayed for months or years. PLP is described as shooting, stabbing, boring, squeezing, throbbing, and burning [16]. Moreover, PLP is primarily localized in distal parts of the missing limb (fingers and palms in upper limb amputees and toes, instep, top of the foot and ankle in lower limb amputees) [17].

B. PLP Management Strategies

Treatment of PLP can be classified as medical, non-medical and surgical [8]. In general, treatment should be based on non-invasive techniques as surgical procedures carry a risk of further deafferentation resulting in even more pain. In this vein, some PLP management strategies have been proposed, based on medication [18]. However, some others try to provide the amputee with alternative ways to handle the PLP suffering, with the 'mirror box'-based being the most well-known therapy, introduced by Dr. Ramachandran [19] at the University of California San Diego.

'Mirror-box'-based therapy is a drug free treatment and has been described in medical literature to be of benefit to 80% of users (some even report numbers as high as 95%) and that rehabilitation can be dramatically improved by integrating physical and mental practice. This is achieved by utilizing mirrors to trick patients' brains into thinking that they were moving their missing limb. The use of visual feedback with the 'mirror box' as a technique accelerates the recovery from PLP. In fact, the patient places the affected limb inside the 'mirror box' and his/her unaffected limb in front of the mirror. Seeing the reflection of the unaffected limb, the patient thus receives visual feedback from a virtual image of his/her affected limb appearing as if it is normal (Fig. 1) [20].

There is, however, great variability in the experienced authenticity of the 'mirror box' illusion and its ability to alleviate PLP. The effect of this positive visual feedback to the amputee can be very therapeutic. The latter translates into reposition of phantom limbs, which were perceived to be held in painful or awkward positions, into non-painful postures, giving temporary relief from PLP. Amputees' perceptions of their phantom limbs, however, often differ greatly from their original limbs [21]. In such cases, the phantom limb can be shorter, or longer, vary in thickness, have gaps or be continuous, in comparison to the original limb. This is one cause for the inability of the 'mirror box' illusion to provide any therapeutic value in some cases, as the reflected image bears no resemblance to the phantom limb as it is perceived by the amputee. These irregularly shaped phantoms cannot be viewed in the mirror, as it necessarily reflects the image of the intact arm.

To circumvent the aforementioned disadvantages, Augmented Reality (AR)-based 'mirror box' illusion simulation has been proposed [22]. There, a 3-dimensional (3D) graphical representation of an arm on a flat screen that is controlled by a wireless data glove is implemented. The latter is worn on the intact arm, whilst the phantom appears on the flat screen in place
Transcutaneous electrical nerve stimulation (TENS) is the popularized name for electrical stimulation produced by a portable stimulator and used as a biofeedback system for the pain treatment [23].

3 Transcutaneous electrical nerve stimulation (TENS) is the popularized name for electrical stimulation produced by a portable stimulator and used as a biofeedback system for the pain treatment [23].
expression analysis, the biofeedback system and the AR-based PL modeling and interfacing.

A. Advanced Facial Expression Analysis

The aim of this part is pain quantification. In general, pain is typically assessed by patient self-report. Self-reported pain, however, is difficult to interpret and may be impaired or in some circumstances (i.e., young children and the severely ill) not even possible [24], [25]. To circumvent these problems, behavioral scientists have identified reliable and valid facial indicators of pain. Hitherto, these methods have required manual measurement by highly skilled human observers.

In the Epione, an approach for automatically recognizing acute pain without the need for human observers was implemented; hence, pain is automatically detected. The patient’s pain status identification is realized following the steps of the block-diagram illustrated in Fig. 3 (up). In particular, a statistical approach (Active Appearance Model-AAM [26]) is adopted, in which a model is built from analyzing the appearance of a set of labeled image examples where structures vary in shape or texture, it is possible to learn what are plausible variations and what are not. A new image can be interpreted by finding the best plausible match of the model to the image data.

Frame-level ground truth was calculated from presence/absence and intensity of facial actions previously associated with pain. Active appearance models (AAM) (Fig. 3 (bottom-left)) were used to decouple shape and appearance in the digitized face images. Support vector machines (SVM) [27] were further employed using information from the change of the activated action units (AUs), defined by the Facial Action Coding System (FACS) [28], to form pain level classification. The latter is formed by the following equation:

\[
Pain = AU4 + (AU6||AU7) + (AU9||AU10) + AU43, \quad (1)
\]

that is, the sum of AU4, AU6 or AU7 (whichever is higher) AU9 or AU10 (whichever is higher) and AU43 to yield a 16-point scale of pain [AUs are scored on a 6-point intensity scale that ranges from 0 (absent) to 5 (maximum intensity). Eye closing (AU43) binary (0 = absent, 1 = present)].

B. Biofeedback System

This part aims at providing a biofeedback to the patient using the regulation of the TENS activity [23] (Fig. 4), according to the pain level estimated by (1). Basic science studies show that high and low frequency TENS produce their effects by activation of opioid receptors in the central nervous system [29].

In particular, high frequency TENS activates delta-opioid receptors both in the spinal cord and supraspinally (in the medulla) while low frequency TENS activates mu-opioid receptors both in the spinal cord and supraspinally. Further high frequency TENS reduces excitation of central neurons that transmit nociceptive information, reduces release of excitatory neurotransmitters (glutamate) and increases the release of inhibitory neurotransmitters (GABA) in the spinal cord, and activates muscarinic receptors centrally to produce analgesia (in effect, temporarily blocking the pain gate). Low frequency TENS also releases...
serotonin and activates serotonin receptors in the spinal cord, releases GABA, and activates muscarinic receptors to reduce excitability of nociceptive neurons in the spinal cord.

The biofeedback system is realized as a wearable device, involving TENS, which cover the complete range of transcutaneously applied currents used for nerve excitation [23]. TENS are connected to the skin using two or more electrodes. The battery-operated TENS unit is able to modulate pulse width, frequency and intensity according to the Epione CC handling at high frequency (>50 Hz) with an intensity below motor contraction (sensory intensity) or low frequency (<10 Hz) with an intensity that produces motor contraction. Data are handled by a local microprocessor and transmitted to the Epione CC via Bluetooth link.

C. AR-based PL Modeling and Interfacing

The role of this part is to implement an AR-based PL modeling and interfacing, as a means for not only creating the 'mirror box' illusion, but, in an extended perspective, to facilitate the functionality of the recreated missing limb into real-life scenarios. A missing limb 3D model (e.g., hand model, Fig. 5-up) is used to provide various degrees of freedom. As to establish completely independence between the existing and the AR-reconstructed limb, a wireless accelerometer (Fig. 5-down-left), mounted on the missing limb site, is used to capture the movement of the remaining part of the amputee limb, driving the movements of the AR-reconstructed limb. To this end, by acquiring information from patient’s gestures/voice commands activity, using the wireless accelerometer (Fig. 5-down-left) and speech recognition, and transmitting it via Bluetooth to the Epione Vault, multimodal interaction and task performance (e.g., gesture-based navigation throughout the Microsoft Virtual Earth using the reconstructed missing limb) is feasible using AR-based glasses (Fig. 5-down-right). In particular, the latter provide stereoscopic preview of the Epione Vault to the amputee (Fig. 6), in order to enhance his/her illusion about the reconstructed missing limb. The amputee can then interact with the Epione Vault, either following specific tasks with varied difficulty (e.g., trying to touch specific colored squares with the AR-based limb-Fig. 6 (up)), and/or using free gesture interaction (e.g., handling Epione Vault facilities-Fig. 6 (bottom)). In addition, the use of AR-based interfacing enables the superposition of the missing limb reconstruction to the real world (via its embedded camera), thus, clearly enhancing the illusion of the limb replacement, independently from the intact one. An example is given in Fig. 7, where the amputee chats with his friends through the Epione Vault handling the communication with the AR-based hand through an AR-based interfacing that augments the reality with this embedded functionality. This allows for free-gesturing in a real-life setting, with the facility to use his/her reconstructed limb as s/he uses the intact one, and enhances imagery interaction (hence PLP relief) in a more experiential way.

V. TECHNOLOGIES USED

The technologies used for the development of Epione-PLP were: (i) Servers: IIS 7.0; SQL Server 2008 R2, (ii) Development Tools: Visual Studio 2010 Professional, Visual Studio 2008 Professional SP1 - Other Technologies: .NET Framework 4.0 (C#, WPF, WCF, ASP.NET); Windows 7 Professional; Windows Mobile 7.0; XML Web Services; MS DirectX 11, Microsoft Expression Suite 3, Microsoft Speech API (SAPI) 5.3, MATLAB R2010a, Autodesk 3D Studio Max 2010, Facebook API (Facebook Developer Toolkit), Twitter API (twitterizer), Skype API, Bing Maps 3D, OpenCV (emguCV), (iii) Hardware: Vuzix iWear VR920, Arduino Lilypad, Sparkfun Witilt v3.0.
VI. DISCUSSION

Epione-PLP scheme is a revolutionary PLP management that provides to PLP sufferers an integrated pain management environment. It is the first pain management tool, which translates the medical needs to technology-based procedures that take into account patient's response in an adaptive way that fulfills each patient's specific therapeutic needs and provide to the physician a plethora of treatment scenarios.

The innovations behind Epione-PLP consist of a broader approach to the problem examined. Most pain management centers gather information from the patient in a static way, simply using questionnaires in order to set up different therapeutic tasks. In this case, the physician has to consider a variety of parameters combined with patient's pain behavior in order to construct the optimum course for each one. On the contrary, the proposed system is designed to circumvent the above drawbacks considering the user’s needs. In particular, based on computer vision inference algorithms and social behavior markers, Epione combines all the necessary information to automatically construct the appropriate computer-based pain management environment (Epione Vault) for each patient. This turns the PLP management to a socially supported activity that combines both the personalized sensation of PLP relief with the support of the community environment.

Moreover, it incorporates biofeedback stimulation (TENS) during the therapeutic session, adjusted, accordingly, to the activation of pain alarms. The involvement of transparent monitoring of the patient’s
interaction with the Epione Vault, the adaptivity of the latter to increase distraction, imagery relaxation and variety in therapeutic scenarios (according to the pain type and/or patient's categorization), along with the AR-based solution for the PLP relief makes Epione an integrated solution for the pain management problem towards normalization of the pain management procedures to each patient's needs (personalized healthcare).

Until now, pain management software refers to questionnaire analysis and patient's history data gathering. Using advanced signal processing and cutting-edge technology Epione-PLP provides an integrated PLP management environment, which is closer to the patient's needs and, therefore, realistic and more efficient. The AR-based experiential feeling of the reconstructed limb extends further the concept of the 'mirror box' illusion, setting it to a more realistic environment, which could support the reconstruction of the body image to the amputee, in a way that his/her body would efficiently handle a series of experiences. As explained in [20], phantom limb experience depends on integrating experiences from at least five different sources: (i) from the stump neuromas; (ii) from remapping, e.g., the spontaneous activity from the face is ascribed to the phantom; (iii) the monitoring of corollary discharge from motor commands to the limb; (iv) a primordial, genetically determined, internal ‘image’ of one’s body; and (v) vivid somatic memories of painful sensations or posture of the original limb being ‘carried’ over into the phantom.

In the case of PLP e.g., due to amputated hand, messages from the motor cortex in the front part of the brain continue to be sent toward the muscles in the hand, even though the hand is missing. In fact, the part of the brain that controls movement does not ‘know’ that the hand is missing [20]. It is likely that these movement commands are simultaneously monitored by the parietal lobes, which are concerned with body image. In a normal person, messages from the frontal lobe are sent either directly, or via the cerebellum, to the parietal lobes, which monitor the commands and simultaneously receive feedback from the arm about its position and velocity of movement. There is, of course, no feedback from a phantom arm, but the monitoring of motor commands might continue to occur in the parietal lobes, and thus the patient vividly feels movements in the phantom. The visualization of the missing limb through the AR-based interfacing could help the amputee to better organize his sense of his body image and create the appropriate feedback to stop painful movements, such as ‘clenching spasms’ in the phantom hand.

The Epione-PLP scheme could foresee various updates and extensions. For example, further increase in the variety of PLP tasks, including various 3D models of different missing limbs and further customization of pain therapy tasks to different pain sources and patient's age (kids, elders) could make it more effective and adaptive. In addition, from a technological perspective, further optimization of the equipment (e.g., more ‘transparent’ AR-based glasses

Fig. 7. The interaction with the Epione Vault using the AR-based reconstructed right hand, for a task of chatting with the community of amputee's friends, in an augmented reality user-scenario (being in his room at his home).
and even smaller interfaces (e.g., accelerometers), and portability, via the implementation of the Epione-PLP interface on a PDA, would maximize its added value. Finally, its large-scale experimental use by a variety of patients could contribute to its optimization and user-acceptability, as its resulting refinement would provide the user with an efficient PLP management environment.

VII. CONCLUSIONS

In this work, an innovative phantom limb pain management environment, namely Epione-PLP scheme, has been presented. Epione-PLP (as a module of an integrated pain management environment) uses advanced signal processing techniques to analyze the user’s facial expressions in a real-time context, resulting in a quantitative pain meter index. The latter is used to form the level of biofeedback to the user and adaptively adjust Epione’s 3D AR-based interacting environment. The AR-based interfacing extends the ‘mirror box’ illusion and sets the experiential sensation of the body image to a more realistic setting. In this way, a more complete information handling strategy is initiated that eases down the user’s PLP feeling. The design and development of Epione-PLP allows for flexible expansion to different user-scenarios, providing personalized PLP relief and management, contributing to a better quality of life.

REFERENCES


# Table of Contents

## Session 1 - ICT Accessibility and Virtual Environments

Usability of 3D controllers in Second Life

*Patrícia Fernandes, Célia Ferreira, António Cunha, Leonel Morgado*

Web Accessibility 2.0: Web Accessibility Status of the Banks With Activities in Portugal

*José Martins, Ramiro Gonçalves*

Avatar Based Computer Assisted Translation from Italian to Italian Sign Language of Deaf People

*Davide Barberis, Nicola Garazzino, Paolo Prinetto, Gabriele Tiotto*

233 - Modeling a Game-Based Adaptive Unit of Learning to Support Adults with ADHD in Virtual Learning Environments

*Laura Mancera, Silvia Baldiris, Ramón Fabregat*

Ethical Considerations When Using Video Games as Therapeutic Tools

*Jason Colman, Paul Gnanayutham*

## Session 2 - Assistive Technology

Enhancing the Academic Achievement of Disabled Students in Greek Universities through Accessible Technology Services

*Anastasia N. Gkouvatzi*

Farfalla: a step toward an inclusive Web

*Andrea Mangiatordi, Riccardo Dondi*

Interaction between families and family school relationship: the impact of information and communication technologies on special education needs

*Carla Silveira Maia, Ana Margarida Pisco Almeida*

Elderly Speech Collection for Speech Recognition Based on Crowd Sourcing

*Ana Júdice, João Freitas, Daniela Braga, António Calado, Miguel Dias, António Teixeira, Catarina Oliveira*

The Companion – Home Shopping From a General Purpose Assistive Technology Platform

*Mary Breeze, Justin A. T. Halls, Edward Varney, Heinz S. Wolff*
Session 3 - Multimodal Human-Computer Interfaces

User Interface Modeling for Accessible Web Applications with the Unified Modeling Language
Helmut Vieritz, Daniel Schilberg, Sabina Jeschke
Omnipresent Care: Voice Commander
Reis L, Marcelino I, Fernandes H, Rijo R, Loureiro P, Pereira A.

“One-Button” Brain-Computer Interfaces
Jason Colman, Paul Gnanayutham

Running Graphical Desktop Applications on Tactile Graphics Displays Made Easy
Christiane Taras, Michael Raschke, Thomas Schlegel, Thomas Ertl

Enhanced Autonomous Control in Interactive Multi-Sensory Environments
Kenneth Brown, Phil Ellis

Session 4 - Computer Vision for Accessibility

Fiducials Marks detection to Assist Visually Impaired people Navigate
Paulo Costa, Hugo Fernandes, João Barroso, Paulo Coelho, Leontios J. Hadjileontiadis

The SmartVision Navigation Prototype for the Blind

A vision system for detecting paths and moving obstacles for the blind
J. José, M. Farrajota, J.M.F. Rodrigues, J.M.H. du Buf

A Non-Visual Photo Collection Browser based on Automatically Generated Text Descriptions
Frode Eika Sandnes

DSAI 2010: Usability and Accessibility in Video Games for the Colorblind
Jose Manuel Bidegain Hanania
Papers
Chapter 1

ICT Accessibility and Virtual Environments
Abstract – The Second Life platform unites various technologies: 3D virtual environments, virtual reality, Internet connection between participants and the possibility of customization of the three-dimensional content. It is a virtual collaborative environment that allows users to explore digital worlds and interact with others in real time. Increasingly, new peripherals to interact with virtual environments are emerging in the market, and good examples of this are 3Dconnexion’s 3D controllers. These are designed to be easy to use, allowing users to zoom, move around in space and rotate 3D objects with a simple hand tap on the device. With this project we wanted to identify usability problems among 3Dconnexion's devices in Second Life 3D virtual environment. We created a usability test and implemented it. With its data we identified usability features of these 3D controllers.

Keywords – Usability, Second Life, virtual environments, virtual worlds, 3D devices, 3Dconnexion

I. INTRODUCTION

According to Arevolo [4], virtual environments like Second Life (SL) are promising platforms for different business areas, such as education, trade and customer support. By understanding their applicability and their limitations, companies can find real opportunities for business in virtual environments. Despite this promising scenario, there aren’t many widely-available hardware interfaces to simplify or optimize the use of virtual environments such as Second Life. Three-dimensional (3D) interaction devices have existed for several years now, but were typically developed with a focus on specialized applications, such as air traffic control or product engineering/design [7, 10], and usability studies on their efficiency and ease of use have followed this narrow focus [21]. Regarding users that are new to 3D interaction, as is the case of Second Life users, three-dimensional interaction is still a field which is little explored, with little knowledge of relevant facts for usability [9].

We wanted to determine whether using 3D controllers could simplify users’ interaction with virtual worlds, as a first step towards a more detailed assessment of their feasibility in support of accessibility to these platforms by citizens with special needs. To do so, we decided to first conduct tests with non-special needs adults, as a first approach to this problem. The focus was on a specific class of interaction devices, that of 3D controllers developed by 3Dconnexion, with users that are unacquainted with 3D modeling techniques, in their use of the Second Life virtual world. For this, we employed the SpacePilot [1] and SpaceTraveler [2] controllers. These 3D controllers or 3D mice, as they are also known, were developed with the aim of minimizing the need for mouse-keyboard coordination to position models or objects on the screen in 3D applications. They can be used in combination with traditional mice (albeit that combination was not the focus of our study): with a touch of the device, users can zoom and rotate 3D objects in a fluid motion with one hand, while simultaneously selecting, creating or editing with the regular mouse on
the other hand[3]. In this context we created and applied a usability test for these devices, based on a literature review on usability concepts and tests.

II. USABILITY

Simplicity, efficiency, optimize, facilitate, enhance, accelerate are verbs circulating around the concept of usability. Usability is synonymous with greater flexibility and more interaction; it is also a feature of what is usable. For Smith and Mayes [19], usability considers three aspects such as ease of learning, ease of use and satisfaction in using the system by the user. According to the standard ISO 9241-11, usability is the “extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [13]. In the same document, “effectiveness”, “efficiency”, and “satisfaction” are further defined as:

- effectiveness: accuracy and completeness with which users achieve specified goals;
- efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals;
- satisfaction: freedom from discomfort, and positive attitudes towards the use of the product.

Usability is directly related to the interface [5, 11, 12, 15, 18, 20], which together with the user and the interactive computer system constitute the 3 main components of the process of Human-Computer Interaction. Software can be well designed in terms of functionality, but if its usability is not good, the user may reject it. There are several positions around the concept of usability. E.g., Shackell [17] considers 4 parameters to measure usability: efficiency, learning, flexibility and attitude of the user. For Hix & Hartson [12], usability is related to the effectiveness and efficiency of the interface and the reaction of the user concerning it. Nielsen [15] lists 5 parameters that he considers as traditionally accepted to measure usability: easy to learn (the user can quickly interact with the system, learning the options and functionality of the navigation buttons); efficient to use (once the user has learnt how it works, the user can find the information he or she needs); easy to remember (even for a user that uses the system occasionally, the user doesn’t need to re-learn how it works, the user can remember it); little subject to errors (users don’t commit many errors while using the system, or if they’re committed the user should be able to fully recover, catastrophic errors must not occur); and pleasant to use (the users feel satisfied with the system, they like to interact with it).

III. THE VIRTUAL ENVIRONMENT OF SECOND LIFE

Usability studies are intended to indicate flaws in the design of a product so that they can be resolved in time to provide a quality product for the user. To better understand the usability of the Second Life virtual environment, when interfaced by the 3D mice mentioned above, we aimed to identify factors hindering or facilitating users’ interaction in this virtual world. The user can interact with the system to: navigate the virtual space, communicate with others, and manipulate/create objects, using the Internet as a connectivity channel. Second Life is also a collaborative virtual environment: users’ actions with other users and with the common virtual environment are visible to all users in real time (i.e., a change caused by one user is visible to all other users, almost instantly). While some may mistake Second Life as a game, Second Life does not involve scores, competition, levels, winners, losers or indeed a purpose: it is more akin to an on-line 3D Web, in that users may not only navigate it, but also hold "real estate", i.e., pieces of virtual land where they can build and store their own content (houses, shops, offices, etc.), either privately or publicly accessible – as they would with Web server space and Web pages. The use of Second Life is thus complex and layered. Regarding the 3D component, there are 2 major elements: navigating the virtual space and creating or editing 3D objects. For this initial study, we focused on navigation. This involves controlling the avatar as it walks on virtual surfaces, but also as it flies around (a preferred – faster – method of movement). Further, navigation also involves camera controls: in Second Life, a user does not have to exactly position the avatar in a place where a sign or element is visible: rather, the user can activate camera controls in order to rapidly focus and zoom in on the intended element.

IV. METHODS AND TECHNIQUES FOR EVALUATION OF USABILITY TESTS
A usability test is characterized by the completion of certain tasks by a set of users. The purpose of a usability test is to collect data about the tasks, typical and critical of an under study "object", to be executed by the participant. For a usability test to return as much detail on a prototype or system (“object of study”) we should apply or combine different evaluation techniques. These techniques are adopted from known procedures in other areas such as ergonomics, where we can find methods such as:

- Heuristic evaluation;
- Ergonomic criteria;
- Inspection based on standards, style guides or recommendation guides;
- Inspection by checklists;
- Cognitive route (or inspection);
- Empirical test with users;
- Interviews and questionnaires.

Some evaluation techniques for usability tests may include a list of methods that directs user efforts to perform a variety of tasks on a prototype or system. While the user performs the stipulated tasks, the user is observed by observers who collect data on user interaction processes, including mistakes made by the user, when and where they are confused or disappointed, the speed with which the user performs the task, if success is achieved in doing the task and user satisfaction with the experience.

Usability tests involving real users in interaction procedures become a painful and complex procedure. The use of heuristics, for example, allows one to identify more serious and difficult to be identified errors. However studies show that the joint use of both processes, the application of heuristics and usability tests, is the best approach to usability research (e.g., Kantner [14]). Usability tests are intended to indicate flaws in the design of the product so that they can be resolved in time to provide a quality product for the user. Thus usability should take into account the characteristics of measurement of quality by providing two types of analysis:

- Performance: measurements or empirical observations of user behavior, emphasizing the performance of the task and quantifying the performance of a specific task.
- Attitude: measurements or subjective observations of the participant’s opinion while the user performs the activities on the system, quantifying the user’s satisfaction in using the system.

The set of attributes represented in usability testing shows the effort required to use a piece of software. The test can be used for different purposes that involve types of tasks, performance measures and arrangement of scales, interviews or inspections to be implemented, finding usability problems and make recommendations to eliminate problems and improve the usability of the product or in order to compare two or more products. With usability testing, one can register best results obtained for future events, leading to cost reduction of services for user support, sales growth and anticipate the launch of products with fewer usability problems and more competitive.

V. PLANNING THE USABILITY TEST FOR 3D DEVICES IN SECOND LIFE

For Dumas [8] (acc. [9]), good usability planning is absolutely necessary, because simply joining people just to see how they react to a product is a waste of time. At least one must bear in mind which aspects of the product must be addressed, how to select participants that are representative of the intended audience, and what tasks to be performed, even in a short period of time. In this context defining targets and concerns facilitates the rest of the planning, because all future decisions will stem from these definitions. For Dumas (ibid.) a goal or objective is generally a declarative sentence where as a concern is generally defined as a question. In 2008, a Brazilian Master student conducted usability tests of the Second Life environment, with the default access methods (keyboard and mouse), and defined a goal and a concern which we find adequate as a starting point for our own study, since it had the same focus, but different input devices [9]. Based on his approach, we adapted his general goal and concern thus:

- Users should be able to enter Second Life and, with ease, conduct basic tasks involving movement on the ground and in the air, circumventing objects, and adjust the viewpoint of the camera.
• Users should be able to perform basic tasks easily, without losing interest or lose the sense of presence in the virtual world.

VII. TEST PARTICIPANTS

According to Barnum [6], experience and motivation are more important in understanding how people use a product in comparison to demographic factors such as age, gender or social level. Some factors suggested by the authors to take into consideration when developing a profile of participants are:
• Work experience (function, time on task, accomplished tasks, etc.);
• General experience with computers (time of experience, software used, etc.);
• Specific experience with computers (type, duration and frequency of experience);
• Experience with the product to be tested (basic or advanced features);
• Experience with similar products.

For this usability test, we looked for participants with a good level of experience and frequency of use of computers in order to avoid difficulties of use not related to Second Life itself. Furthermore, we tried to combine different levels of prior experience with three-dimensional interaction in order to observe the effect of this factor in usability. These levels of experience and frequency were measured through a pre-test questionnaire answered by all participants. Given the lack of literature as reported earlier, we specifically decided to include only non-special needs participants, in order to acquire an initial overview of potential issues and difficulties – thus providing improved know-how for future tests with special needs users.

Another aspect to define is the number of participants in the test. This is a matter of intense discussion in the usability community. According to Virzi [20] 80% of usability problems were detected with four to five participants and 90% were detected with 10 participants. Furthermore, Nielsen suggests the use of five participants in terms of best cost and benefit [9, 15]. Obviously, if the goal of testing is to validate statistical results, a higher number of participants will be required to conduct the appropriate analysis and generalization of a specific population.

Since the objective of this project is to attain a perception of the usability of 3D controllers in Second Life and possibly expose underlying issues, we chose a sample of 5 participants to conduct our test.

VIII. TEST LOCATION

Acc. Barnum [6], a prepared room to apply the usability test provides benefits, but it’s not a requirement for individuals and organizations interested in conducting a usability test. To conduct a test without an equipped room, all that is needed is enough space for the test conductor (observer) to sit next to the participant. For the application of our usability test we used a college classroom. This location allowed us to observe what was happening on the computer screen as well as the participants’ reactions and expressions.

VIII. DEFINITION OF TASKS

A usability test is a sampling process, therefore it is essential to choose the tasks that will be part of the test proper. Dumas [8] suggests some criteria for selection of tasks:
• Tasks with the potential to generate usability problems;
• Tasks suggested from their concerns and previous experiences;
• Tasks that users will perform.

The tasks were designed to reflect what a user needs to do in Second Life to get a basic interaction experience. We chose two Second Life locations, one a replica of the Belém Tower at the simulator Utopia Portugal XXIV, the other a simulator that was mostly empty (at the time), [omitted in blind paper for review], where participants performed the tasks. The former location was chosen because it is a common example of a Second Life location, with large and small
buildings, water areas, pathways, etc; the latter because we could use the free space to build specific situations for testing.

Dumas [8] suggests two important points: tasks must follow a natural flow for the participants and, the most important tasks should be at the beginning of the test since it’s possible for participants to give up before the end of the test. We established four types of basic functionalities: orientation (understand one’s surroundings within the virtual world), moving on ground (reach a location by walking), moving in the air (reach a location by flying), and adjusting the perspective of the camera (attain the desired viewpoint and zoom). These features were evaluated by performing tasks with different difficulty levels, as we detail below.

The tasks chosen under these considerations were as thus:

1. Being able to adapt and get oriented in Second Life (Belém Tower area);
2. Moving along a simple route, with only slight curves;
3. Moving up and down stairs and being able to stay on track, i.e., without falling to the lower level;
4. Moving along a path, avoiding obstacles;
5. Flying and circumventing an object;
6. Fly and landing at a specific location;
7. Flying and landing just in front of another avatar;
8. Controlling the camera in order to focus a panel and be able to read its content.

XI. METRICS TO BE USED IN THIS USABILITY TEST

Dumas [8] indicated that from a usability test it is possible to gather performance measures or subjective measures. Performance measures are quantitative and generally involve counts of actions and visible behavior, while subjective measures are the perceptions, opinions and judgments. On the assumption that the usability evaluation is a process to produce a “value/measure” of the ease of use of a resource, and that this value/measure may be qualitative or quantitative in nature, these tools can determine the degree of user satisfaction [16]. Usability measures the quality of user experience when interacting with a product or system. It includes several factors such as Ease of learning (speed at which a user who has never seen an interface assimilates it well enough to perform basic tasks); Ease of remembering (if a user has never used the interface, whether he/she will be able to remember enough to use it the next time or will have to start learning how to use it again); Efficient use (speed at which tasks are carried out by an experienced user, having learned how to use the system); Frequency and severity of errors (how often a user makes mistakes using the system, what is the severity of errors, and how the user corrects or overcomes these errors); Subjective satisfaction (the degree of user satisfaction in using the system).

Considering this framework, for our test the measurements of performance were defined as:

- Quantity of errors to complete an action;
- Number of requests for assistance in order to complete an action;
- Observation of a frustration attitude;
- Observation of confusion;
- Observation of a satisfaction attitude;
- Time spent to complete the task.

We collected data on these metrics using a scale while observing the participants performing the tasks. In order to obtain immediate reactions of each participant, we also performed a short post-task questionnaire to capture the difficulty level to complete each task. Dumas [8] emphasized the usefulness of these questionnaires: it not only allows a break between tasks but also gives the observer enough time to prepare the next task.

X. ANALYSIS OF THE 3D CONTROLLERS’ USABILITY TEST

This section is dedicated to the analysis of all tests completed by the participants of the usability tests. It also presents some partial conclusions of these tests.
A. Analysis of the pre-test questionnaire

The usability test participants received and completed the pre-test questionnaire, whose objective was to confirm the characteristics previously defined for the user profile. The list of participants and gathered data are as follows in Table 1 (all measures are based on 6-point Likert scales from 0 to 5).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Experience with computer</th>
<th>Frequency of computer use</th>
<th>Experience with 3D environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>Male</td>
<td>23</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Participant 2</td>
<td>Female</td>
<td>23</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Participant 3</td>
<td>Male</td>
<td>28</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Participant 4</td>
<td>Male</td>
<td>26</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Participant 5</td>
<td>Male</td>
<td>22</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

The age span was not large: the youngest participant was 22 years and oldest 28 years. Regarding the level of experience in computer use, it is confirmed through the questionnaire that the participants had a medium to high level, as can be seen in Table 1. Regarding the frequency of computer use, most participants were daily computer users. Finally, on experience with 3D environments, two users had significant experience, one had average experience, one low, and one had no previous experience.

B. Overview of the observation grids completed by observers

Based on the grids that we completed as observers, during the execution of the tasks that comprised the usability test, the following can be depicted (observations are valid for both controllers, except where noted):

- One of the things in which users felt more difficulty was in moving the avatar forward, as they often triggered the zoom function instead.
- Participants showed some frustration when they made a mistake or failed to complete the task.
- Another mistake was running into walls or objects. One of the reasons that may be the causing this error is participant disorientation, but we cannot assert whether that would be due to lack of domain of the Second Life space or lack of experience in the use of the 3D controller.
- With the smaller controller (SpaceTraveler), most of the users used two hands, one to hold the controller in place and the other to perform the task; this made users feel safer and more comfortable in using the controller. With the bigger controller (SpacePilot) this behavior did not occur.
- Generally, in tasks in which participants were successful expressions of satisfaction were demonstrated. A minority of participants did not express any kind of reaction.

C. Post-task questionnaire

Usability test participants received and filled out the post-task questionnaire, with the aim of self-evaluating themselves after the completion, or not, of each task. Table 2 presents the data obtained with this questionnaire, using 6-point Likert scales from 0 to 5. The column groupings reflect the type of activities: Task 1 was about getting bearings on the environment, tasks 2-4 about walking, 5-7 about flying, and Task 8 was on camera control.

<table>
<thead>
<tr>
<th>Task 1</th>
<th>Task 2</th>
<th>Task 3</th>
<th>Task 4</th>
<th>Task 5</th>
<th>Task 6</th>
<th>Task 7</th>
<th>Task 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant 1</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Participant 2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Participant 3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Based on the averages obtained we could say that Task 7 was the most difficult for participants. This task consisted in flying and landing right in front of another avatar. At the other extreme the easiest task for participants could be said to be Task 6, which consisted in getting the avatar to fly and land in another location. However, due to the small number of participants, we must not look at averages, but rather at individual results. For instance, Participant 3 (the one with the least experience in computer use and no experience in 3D environments), found Task 7 extremely difficult, while Participants 4 and 5 found it at the level of most others.

D. Analysis of post-test questionnaires

In the following section there’s a summary of the post-test questionnaire carried out by each participant.

Participant 1
According to the pre-test questionnaire, participant number 1 had used Second Life, but doesn’t use it often. Regarding the main reasons why this participant is not a frequent SL user, he stated that he did not feel comfortable: he had used this software before and did not like the experience because of the difficulty in its use. However, if training was supplied, this user would be receptive.

Before the test, this user was happy and willing to do this usability test. In general, during tasks the user showed to be focused.

This participant did not feel any difficulties concerning the reading and understanding of the tasks.

The main problems identified by the participant are related to the fact that when the user wanted to move forward or back, the full screen camera function (a.k.a. “mouse look”) was activated, making the task a bit harder because the user couldn’t see its surroundings entirely due to this perspective.

For this participant, learning how to use the controller revealed to be difficult, since some movements were complicated due to high sensitivity of the controller. As for the tasks, this participant stated that they weren’t either too easy or too difficult.

Participant 2
The 2nd participant indicated some knowledge in virtual environments, in particular Second Life and the game Counter Strike. The main reasons why this participant is not a frequent Second Life user are: already used it and didn’t like it, and had difficulties in moving and orientating herself in Second Life. Nevertheless, if training was provided this participant would be receptive to use such an environment.

Before the test, the participant was happy and willing to do the test. In general, during the tasks, this participant was focused on the tasks and got a bit confused, which led to requests for help. At the end of this experience, this participant reported that she was happy to have participated in this usability test.

Some difficulty was encountered by this participant in understanding the instructions.

The main problems identified by this participant relate to some difficulties in shifting from walking mode to fly mode. In the first interaction, this participant found the controller too sensitive.

For this participant, learning of how to use the controller proved to be accessible.

As for the tasks, this participant stated that they weren’t either too easy or too difficult: “no major difficulties.”

Participant 3
Participant number 3 had never used a virtual environment. He identifies as the main reasons for not being a user of this environment the following reasons: ignorance of the existence of Second Life, does not know how to use it and has never needed to use it. Before the test, the participant was satisfied to be part of this usability test but with a few
questions about Second Life. In general, during the tasks, this user was focused but somewhat confused. At the end of this experiment the user was happy for having participated.

Little difficulty was encountered by this participant in understanding the instructions.

The main problems experienced by this participant related to the sensitivity of the controller and the zoom control.

This participant found the learning of how to use the controller accessible, considering it easy to learn. As for the tasks, this participant stated that they weren’t either too easy or too difficult.

Participant 4

Participant number 4 had never used virtual environments. This participant identifies as the main reasons for not being a user of this environment the following reasons: did not know about the existence of Second Life, does not feel comfortable using it and never needed to use it before. This participant states that he would use Second Life, if given some type of training.

Before the test the participant had been pleased to carry out this test. In general, during the tasks this participant seemed to be unfocused and somewhat disinterested. At the end of this experiment this user was happy to have performed these tasks.

As for reading and understanding the instructions of the experience this user had little difficulty.

The main problem experienced by this participant relates to the sensitivity of the controller and the zoom control.

This participant found the learning of how to use the controller accessible, however believes that learning how to use the controller is difficult because it’s very sensitive to achieve certain movements. This participant also suggests that it would be easier to learn how to use the controller if it was applied to games.

Participant 5

The 5th participant indicated that he has some knowledge in virtual environments including the game Counter Strike. The main reasons why this user isn’t a frequent Second Life user are: doesn’t know how to use Second Life, does not feel comfortable and until today did not need to use it. However this participant wouldn’t be interested in training if it were provided. Before the experience in Second Life this participant was pleased to perform the experiment, but also concerned about doing something wrong and full of questions about SL. During the experience the user seemed to be concentrated on the tasks and a bit confused.

After completing the tasks the user was happy to have performed the experience. As for reading and understanding the instructions of the experience this user had a bit of difficulty. According to this participant learning how to use the controller is difficult. As for the tasks, they were easy to carry out.

XI. CONCLUSIONS

Data analysis involves a search for trends and surprises in the data collected. Dumas [6] justified this saying that a usability test is an empirical evaluation method and thus needs to be justified from the data collected.

Based on the realization of this usability test we can conclude that during the various tasks, with different difficulty degrees, the interaction of participants with Second Life through 3D controllers has evolved positively over the successive tasks, despite the progressively increasing complexity of these tasks.

In spite of the problems encountered by our usability test, this device has shown to be manageable. Whether it is more so than having to coordinate between keyboard actions and normal mouse actions, is something that only a comparison trial would be able to establish. We have not performed a comparison of both methods of interaction with SL, and this will be a further phase of research, but a critical element for the results of such a test would be the status of the zoom/move forward confusion in 3D controllers. As it stands, we are unable to say just from this test whether this issue will render 3D controllers better or worse than a keyboard/mouse combination.

The biggest problem highlighted by this usability test was related to the zoom functionality, because the functionality to move both forward and back is very similar to the zoom feature. The only difference is the pressure that is used on the controller. If one presses it too much in order to move forward, one accidentally activates the zoom functionality. This should be taken into account to produce a more usable product in the future.
REFERENCES

Web Accessibility 2.0

Web Accessibility Status of the Banks With Activities in Portugal

José Martins, Ramiro Gonçalves
Universidade de Trás-os-Montes e Alto Douro, Vila Real, Portugal
E-mail: jlbandeira@gmail.com, ramiro@utad.pt

Abstract – Web accessibility presents itself as one the most important factors for the equality of societies in the modern days. This document is a simple and clear snapshot of the accessibility status of the websites of the Banks with activity in Portugal.

Economic activities are part of everyday life since always. The banking sector has a large dimension for a long time, and has been doing a major contribution to economic growth in Portugal.

We present an evaluation of the Portuguese banks websites against W3C Web Content Accessibility Guidelines 2.0.

The results that were obtained show a severe lack of accessibility in the evaluated websites, what can be translated in the existence of a great number of barriers to all citizens that present some sort of disability.

I. INTRODUCTION

Since its creation, the Web has been growing both in size and complexity. It is argued that this happens due to its decentralized properties: anyone can contribute to the Web without a central authority dictating what is allowed to be published or not. Consequently, people with insufficient technological skills to produce quality content (e.g., usable), became massive Web publishers. This makes the Web being perceived as a living organism, constantly evolving in different directions regarding its size and quality [1].

As the number of people using the Web is increasing, also the number of minor groups, such as the ones with some sort of disability, become more present.

Everybody accesses the internet in different ways. Individuals needs determine whether they have to change browser settings to view web pages or use assistive technologies such as speech and Braille output, speech input, or screen magnification. People may also use different platforms, ranging from Personal Computers (PCs) to Personal Digital Assistants (PDAs) or mobile devices such as a phone [2].

The growing need for access to online information impels a guarantee of accessibility to this web content. According to the 2001 communication of the European Commission [3], there are 37 million disabled European citizens who need to be granted full access to web content.

The banking sector has a large dimension for a long time, and has been doing a major contribution to economic growth in Portugal. In this new millennium the advent of Internet has had a significant impact on the banking service that is traditionally offered by banks [10]. With help of the Internet, customers can do their banking anytime, anywhere, since Internet access is available. This service is called Electronic Banking (EB), and is in explosive growth in many countries, transforming the traditional banking practices [10]. By offering services with EB, the financial institutions seek to reduce operating costs, improve banking services, retain consumers and expand customers share [11].

II. WEB ACCESSIBILITY CONCEPTS AND CONCERNS

The term accessibility can easily be defined as the possibility for disabled people to interact with a product, resource, service or activity in an equal way as someone without any disability. Concerning ICT, we can define accessibility as the creation of interfaces that are perceived, operable and easy to understand for those with a wide range of disabilities. This includes all disabilities, such as visual impairment, hearing problems, physical, cognitive or neurological limitations. Beyond this, accessibility features also make all products more available to those that do not have any kind of disability [9].

According to Shawn Henry and Jim Thatcher, the web accessibility goal is providing all disabled citizens the ability to perceive, understand, navigate and interact with the Web, regardless of their disability [4].

A. Regulations and World Perspective

According to the European Commission, there are about 37 million European Citizens that present some sort of disability or impairment, that need access to all web content [5]. As an example, in the United Kingdom, there’s an estimate that the disabled citizens
present a market value of about 120 billion pounds [12]
The World Health Organization – WHO claims that about 10% of the world population suffers, from some kind of disability or incapacity. This number clearly shows the existing need for health and rehabilitation services. Due to this, the WHO created an action plan called “Disability and Rehabilitation Action-Plan 2006-2010”, whose mission goes not only, for trying to disseminate and create awareness of this reality throughout the world community, but also to create initiatives that help in the process of recovery and re-integration of disabled people back to society [6].

W3C Accessibility Initiative
Web content accessibility has been a priority for various world entities, such as the W3C consortium which in 1999 created the World Accessibility Initiative – WAI. This initiative was created with the aim of being a parallel organization to the W3C and its mission was to develop guidelines that would be understood as the international standards for web accessibility; as well as to develop support materials for a better understanding and development of web accessibility, and to develop new resources through international cooperation [9].

Portuguese Web Accessibility Regulations and Concerns
In the year 2002, the Portuguese National Institute of Statistics - INE promoted a demographic study named “Censos 2002 – População residente com deficiência segundo o grau de incapacidade e sexo”. According to this study, there were 634000 Portuguese citizens with some kind of disability. This number represents 6% of the entire Portuguese population [7].

III. BANKING CONCEPTS
A. Banking
Although the concept of banking is directly related to the financial area by involving the movement of capital, its definition can lead to several subjectivities. Many sociologists believe that the financial and banking market is a system of social interactions, where banks are the key intermediaries.

B. Portuguese Financial System
The Portuguese financial system has been subject to major structural reforms, which contributed to the increasing of its efficiency, with the privatization of existing banks, the liberalization of markets, in deregulation and alignment of existing regulations in Portugal to other economies, and increasing modernization of the institutions, mainly driven by the adoption of new technologies.

Between 1985 and 2003, while the national GDP has increased, in terms of market prices, more than 6 times, the assets of banks grew more than 12 times, representing twice the growth. Although the Portuguese economy has suffered some progress, not all of its developments have been positive. Through Figure 1 is possible to obtain an approach of Portuguese economic situation in historical perspective. Portugal has recently suffered a slowdown in economic growth relative to richer countries of the EU, leading the national banking institutions development to constraints on their growth [15].

C. E-Banking
For centuries, many financial transactions required human presence. This peculiarity has been amended with the emancipation of ICT. Currently, a customer can make several financial activities without entering a bank branch. This phenomenon is known as EB [16], offering a range of services and activities by digital means, usually provided through a bank.

D. Appearance of EB
It is estimated that EB via Internet began in 1995, indicating the use of Internet as a channel for delivery to remote banking services [17], however, some banks have had informational Web sites before using phone as EB channel. In Portugal a first step in EB was given by BCP, with the release of several multimedia kiosks providing the customer with bank information [16].

E. Factors that influenced the adoption of the EB
Several converging areas of reference and theories suggest many influences on the adoption of EB, including theories of customer behavior in the choice and use of the media, diffusion of innovation, acceptance of the technology, service and transaction costs, among other [18].

F. Adoption and use of EB
Since EB’s appearance, there have been some constraints in its usage and adoption. In Europe, differences between countries appeared to be largely explained by differences in the availability of Internet access. Recently, EB users are relatively satisfied with the quality of banking online (as can be seen through
In Portugal, according to data from the study Netpanel and Marktest in May 2008, 1072 thousand accesses were made on the banking websites, representing a growth of 3.2% when compared to April 2007, and 4.3% from the same month of 2007.

IV. WEB ACCESSIBILITY EVALUATION

A. Web Accessibility Evaluation Proceedings

According to the W3C Web Accessibility Initiative, the accessibility evaluation of a website is a process achieved by the following steps: definition of the scope of the evaluation, definition of the evaluation tools, definition of the proceedings for the manual evaluation and definition of which reports will result from the evaluation process [13].

For the definition of the scope of the evaluation, we had to identify the criteria to be used for the accessibility evaluation and who would be part of the target group. In order to achieve good results, we decided to use the “AAA” accessibility level announced by W3C as the evaluation criteria [8].

For the evaluation tool, we’ve chosen Sortsite [13] because it could automatically evaluate an entire website against WCAG 2.0, and because it delivered an evaluation report that allowed us to analyze and compare these same results.

For this project we decided that the manual evaluation of the entire target group websites was something that we wouldn’t be able to achieve in the period of time that we had for the resolution of this project. As a result of this situation we decided not to manually evaluate the referred websites.

Concerning the kind of reports that would result from the evaluation process, we decided to do a group of simple statistical studies (average, standard deviation, maximum and minimum), that would represent the reality of the web accessibility levels presented by the Portuguese enterprises.

B. Evaluation Target Group

For the definition of the target group we’ve chosen the main financial companies with activities in Portugal. These 31 companies are all registered in the Portugal banking sector regulator – Bank of Portugal [14].

C. Evaluation Results

The evaluation starting point was the target group analysis. With this initial analysis we aimed to check all the banking companies that belong to our target group and that had an available website. The results of this analysis concluded that 90% of the initial 31 banking companies had a website that could be evaluated. The remaining 10% of websites were divided into two groups. One that was composed by the websites that were incompatible with the used tool (7%) and another one that was composed by those banks with a website in maintenance or without a website (3%).

After this initial process we’ve done the second evaluation stage. By evaluating the remaining 26 websites, we’ve concluded that none of the evaluated websites presents itself as accessible to all citizens.
high, the number of errors can still be considered valid indicators of an existent reality.

**Target Group Evaluation Results**

<table>
<thead>
<tr>
<th></th>
<th>Priority 1</th>
<th>Priority 2</th>
<th>Priority 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>419.62</td>
<td>208.79</td>
<td>459.76</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>6655.39</td>
<td>320.87</td>
<td>700.86</td>
</tr>
<tr>
<td>Median</td>
<td>340</td>
<td>71</td>
<td>197</td>
</tr>
<tr>
<td>Min.</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Max.</td>
<td>3010</td>
<td>1300</td>
<td>2865</td>
</tr>
</tbody>
</table>

Figure 3 - Statistics of the evaluation process

Analyzing the Web accessibility levels of the evaluated banking companies, we can see that these companies websites are not, in a general manner, accessible to all citizens (including those with some sort of disability). This claim is clearly perceived by analyzing all the achieved results.

A more detailed look on the obtained results indicate that 76% of the evaluated websites present a number of priority 1 errors below average. In the priority 2 errors, the numbers are not so different. 72% of the evaluated websites have a number of errors below average. 83% of the evaluated websites present a number of priority 3 errors below average.

**V. CONCLUSIONS**

With this work we managed to achieve our initial goal which was, delivering indicators on the actual accessibility levels presented by the banks with activities in Portugal.

As the presented results demonstrate, a considerable number of accessibility errors was detected on all the websites belonging to the target group. This fact indicates that the accessibility levels presented by the websites of the target group is indeed low.

With this work we could also conclude that 3% of the initial target group don’t have an available website. This indicates that not all banking companies present a concern in offering a constant EB service.

**REFERENCES**

3. EU, Delivering eAccessibility. 2002, Comission of the European Communities.
17. Gopalakrishnan, et al. (2003). A multilevel analysis of factors influencing the adoption

Avatar Based Computer Assisted Translation from Italian to Italian Sign Language of Deaf People

Davide Barberis, Nicola Garazzino, Paolo Prinetto and Gabriele Tiotto
Politecnico di Torino, Dipartimento di Automatica e Informatica, Torino, Italy
e-mail: {davide.barberis, nicola.garazzino, paolo.prinetto, gabriele.tiotto}@polito.it

Abstract — This paper presents a system for the assisted translation from Italian written Language to Italian Sign Language of Deaf people with particular focus on the sign language assisted translation process. The objective is to create a platform for the creation of contents that can be visualized through a signing virtual avatar. The system provides features that help the user during the translation process. The system is under development in the ATLAS project that targets the development of a platform for the automatic translation from Italian written text to Italian Sign Language.

I. INTRODUCTION

Research in Machine Translation (MT) aims to develop systems able to translate from a source language to a target language. Automatic MT systems are designed to translate text or speech without human intervention. Systems that perform computer assisted translation (CAT) are designed to include human assistance in order to correct any errors that may occur. This is due to the fact that present MT technology is not able to provide fully automatic high quality translations [1]. Thus, in CAT systems, translations are performed by the system in collaboration with human translators within a CAT framework [2].

CAT takes effort of a series of tools aimed at providing assistance for the translator with regards to both coherence and consistency of his work and speed. Research in CAT technology brought important contributions from the early 80s with significant improvements in interactivity [3] [4] [5] [6]. Important contributions are brought by the TransType project [7] that embeds data driven MT techniques with the interactive translation environment.

Following these ideas researches evolved into more complex approaches that use a complete MT system to produce full or partial target sentence hypotheses which can be accepted or amended by a human translator.

Usually the assisted translation applications offer some functionalities, such as:

• searching lemmas in the text,
• base dictionary creation,
• consultation of the meaning of the lemmas,
• insertion of new lemmas in the dictionary at runtime,
• checking for the quality of the translation,
• checking if the sentence already exists in the memory of the system,
• storage of the phrase translation in the system.

The tools used for assisted translation, have some modules designed to simplify the work of the user such as semi-automatic algorithms for the simplification of complex structures. In this case, if the original sentence has a complex structure (i.e. with relative clause), the automatic tool can extract it and organize the lemmas splitting the source in two sentences.

By means of this operation, is easier to find the grammatical features of the sentence and thus can be analyzed and translated in the destination language with more precision. Moreover sign languages have poor use of relatives and sentences structure is usually very simple. This makes sentence simplification of particular interest for our application.

Another issue to be solved is when a lemma in the original language has two or more meanings in the destination language. This problem is called ‘disambiguation’. To solve this problem usually the tool has to convert the lemma in the singular form (if noun) or infinitive form (if verb). By performing this transformation, it can be easier to take the correct lemma in the destination language. In general, systems have been designed to solve any type of ambiguities (lexical, syntactic and semantic) [8]. To our knowledge no research targeted sign language assisted translation. In this paper we present a platform for the Italian to Italian Sign Language (LIS henceforth) CAT. In section 2 we present an overview on sign languages, in section 3 we discuss the requirements of sign language assisted translation and in section 4 and 5 we describe our system and the translation process on which it is based. The paper concludes with section 6 and outline some future perspectives.

II. SIGN LANGUAGES

There are hundreds of sign languages. Wherever there are communities of deaf people, sign languages
with unique vocabulary and grammar. Even within a single country, regional variations and dialects may be encountered, like in any spoken language. Deaf people from different regions may communicate the same concept in different ways. However, there is no direct correlation between SLs and spoken languages. Users of SLs communicate through concepts, not words. While it is possible to interpret SL into a spoken language such as English, such an interpretation would not be a direct translation. SLs are visual languages that incorporate gestures, facial expressions, head movements, body language and even the space around the speaker. Hand signs are the foundation of these languages. Many signs are iconic, meaning the sign uses a visual image that resembles the concept it represents. Actions are often expressed through hand signals that mimic the action being communicated.

SLs have a grammar of their own; the existence of grammar rules is one of the most important and distinctive elements that distinguish SLs from other gestural forms of communication, such as gestures and pantomime. The grammar is expressed principally through specific alterations of the sign’s location and of some movement features, such as direction, duration, intensity, or width.

In the following lines will be presented the most representative aspects of SLs, in particular those that are particularly relevant when approaching their synthesis, like linguistic features, timing and spatial constructions.

Prior to William Stokoe’s seminal work in American Sign Language (ASL) [9], it was assumed that linguistics was exclusive to the study of human speech. Sign language was regarded by linguists as a series of pictorial gestures with no linguistic structure. Stokoe demonstrated that signing was indeed a rich, linguistically complex language. A fundamental contribution of his work was redefining the basic unit of a sign to units he termed cheremes as opposed to the sign as a whole, these units are analogous to speech phonemes: minimally contrastive patterns that distinguish the symbolic vocabulary of a language.

Sign Languages simultaneously combine shapes, orientations and movements of the hands, as well as non-manual components, such as facial expressions, to convey meaning. Their spatial nature makes it difficult to write or even transcribe them. Moreover, there is not a one-to-one relation between a sign language and the related verbal language; a word may be represented by more than one sign; likewise, one sign may be translated into more than one word.

According to Valli and Lucas [10], signs are made up of smaller components that are sometimes called parameters. Their parameters include handshape, location, movement, palm orientation and non-manual signs. Common linguistic features of deaf sign languages are extensive use of classifiers, a high degree of inflection, and a topic-comment syntax. Many unique linguistic features emerge from the ability of sign languages to produce meaning in different parts of the visual field simultaneously.

In general two features of sign language should be taken into account when considering sign language translation and assisted translation in particular: Signing Space Model and Timing Model.

A. Signing Space Model

Sign language processing is often performed by processing each individual sign. Such an approach relies on an exhaustive description of the signs and does not take into account the spatial structure of the sentence. Talking about space is an important component of all human languages. Spatial constructs enable us to share a mental representation of a space in conversation with another individual and to find objects and events whose locations are unknown. The ability to share information about the location of objects in space has obvious adaptive as well as practical value. The use of the space in our discourses enables us to extract and represent the positions of objects and people relative to each other out of the infinite number of configurations in which entities can be arranged. Recent works in Sign Language linguistics [11] have shown that the meaning of a sign language production can be improved by considering the construction of the signing space. The signing space is the space surrounding the signer and where the signs are produced.

According to Stokoe [9], the signing space is defined as extending from just above the head to the hip area in the vertical direction and extending close to the extents of the signer’s body in the horizontal direction. Stokoe identifies 12 elemental locations in the signing space which during the sign language production the signer will use to position the entities that are evoked in the sentence and to materialize their semantic relationships, so that the resulting construction can be considered as a representation of the meaning of the discourse.

Lenseigne and Dalle [12] propose a system that integrates along with a good knowledge about the SL grammar and syntax, also a representation of the interpretation of discourse through a modeling of the signing space. The symbolic representation of the signing space consists of a cube surrounding the signer, regularly divided into Sites. Each location may contain a single Entity, each Entity having a Referent. A Referent is a semantic notion that can be found in discourse. Once it has been placed in the signing space, it becomes an Entity and has a role in the sentence. So that, building a representation of a
sign language sentence consists of creating a set of Entities in the Signing Space.

**B. Timing Model**

Phonetic studies of languages consider the formation of linguistic signals on both a simultaneous and sequential basis. At any moment in the signal, a series of features align to create a posture, and a string of postures combine temporally to create the signal.

The simultaneous nature of Sign Languages is given by the fact that discourses contain information about several simultaneous independently articulated channels of performance, like the eye gaze, the head tilt, the shoulder tilt, the facial expression, and the location, palm orientation, handshape of the signers hands and so on. Specifying how to coordinate these channels as they change over time is an important part of the LIS generation if the temporal relationships between information on these channels are not correct, then the meaning of the signed discourse could be affected. For instance, a particular facial expression may need to co-occur with a specific movement of the hands. In order to achieve a correct sign language translation, the implemented translation system must consider the signed discourse as a set of multiple coordinated channels and try to translate them accordingly.

Considering the sequential basis of signs, Kita [12] suggests the following model for dividing signs into general sections, where signs inside curly brackets require at least one repetition, and signs inside square brackets are optional:

\[
\text{Unit} = \{\text{Phrase}\} \\
\text{Phrase} = [\text{Preparation}]\text{Stroke}[\text{Retraction}]
\]

In linguistic terms, Strokes are linguistic units, or signs, that provide semantic content to the discourse; Preparations and Retractions, which together form transitions, are nonlinguistic segments that do not offer “grammaticalized” semantic information. Despite lacking semantic value, transitions in signed languages are necessary and they must be fluid movements that allow the sign stream to continue intelligibly.

**C. Italian Sign Language**

Understanding deaf community needs regarding the translation of the LIS represents the start point of the research aimed at generating the LIS translator.

The main topics of interest in this area are: the type of content the deaf community wishes to access by using an automatic translator, in which situations the human interpreter is mandatory and when it may be replaced by an automatic signing character, etc. It is also of great importance analyzing the different sign languages currently used in Italy, that vary from region to region, in order to define which one can better represent “the” national sign language and to define an appropriate “vocabulary” of signs and language complexity to offer suitable guidelines for the automatic translator phase. In order to do this it is fundamental to have good knowledge of the main aspects and peculiarities of the LIS. Like many signed languages, the LIS is very different from its spoken counterpart, so that it has little in common with spoken Italian but shares some features with non-Indoeuropean oral languages, for example it is verb final, like the Basque language; it has inclusive and exclusive pronominal forms like oceanic languages; interrogative particles are verb final.

A sign variety of spoken Italian also exists, the so-called Signed Italian, “Italiano Segnato” which combines the LIS lexicon with the grammar of spoken Italian: however, this is not the Italian Sign Language.

![Figure 1. One to one translation example](image)

III. REQUIREMENTS FOR ASSISTED TRANSLATION

To solve the problem of Sign Language translation a solution based on a complete automatic translation would open significant scenarios in Deaf people lives. Nevertheless is clear how this task offers critical challenges that are nowadays not solved. A solution based on a one to one translation (see figure 1) between the source and destination lemma. This solution is relatively simple in terms of architectural definition and implementation. More in detail this solution can be based on the use of a dictionary that allows to translate every single lemma from the source to the target language.

By analyzing this kind of implementation with respect of the requisites imposed by the sign language grammar (see section 2) is clear how this solution will result hardly intelligible to Deaf people. The reasons for this are the following:
During the translation from Italian to LIS some morphological elements of the sentence are not present in LIS. To solve this problem we suggest the use of a morphological analyzer that selects the lemma to be deleted according to the LIS grammar.

The order of lemmas in LIS is different from the Italian one. In this case it is necessary to introduce a syntactic analyzer that allows to derive the role of every single lemma within the sentence. It is possible to specify a correct order between lemmas resorting to the syntax of the sentence and defining the entities that take place in the signing space.

Signs suffer modifications within the signing space with respect to the semantics of the sentence. The single sign cannot be translated to its correspondent in LIS without considering such modifiers.

More specifically the visual modality that characterizes sign languages appears through different communication channels, that we consider, for the translation purpose, orthogonal among them. They include for example hands, facial expressions, eye gaze, body movements, etc…

While the spoken language has one dimension that is the spoken voice communication channel, with sign languages the dimension of the communicative act increases as it is expressed through the different communication channels we referred above. As a consequence during the translation phase we need a set of parameters that cover the modifications involving the communication channels and to manage the synchronization among them.

The goal of our system is to perform the visualization of the translated signs by means of a virtual character. This means that the attributes used in the translation process have a strong impact on the way in which the character will sign the translation. The signs need to be fluent and precise to be intelligible. In other words

IV. THE TRANSLATION PROCESS

Our system performs the CAT process (figure 2) from the Italian language to LIS. The process starts from the sentence in written Italian Language. The text is segmented and analyzed by a morphological analyzer. The analysis gives the Italian sentence as a string of lexemes through tokenization. The sentence is then parsed with a Tokenizer and a Chunk-rule based dependency Parser. These features are provided by the Turin Linguistic Environment (TULE) [13]. Resorting to TULE the system provides a rich set of data as it adopts a representation format based on the dependency paradigm centered upon Augmented Relational Structure (ARS), where each relation is implemented as a feature structure that can include values for morpho-syntactic, functional-syntactic, and syntactic-semantic components [14]. After the analysis process we need to consider the LIS

![Figure 2. Sign Language CAT process](image_url)
grammar to provide the language in an intermediate format which conveys both linguistic and visualization information. From the written language domain we are shifting to the visual language domain. The attributes tagging step allows the user to perform basically two operations:

1- Operations on the lemmas, such as shifting, insertions, deletions.
2- Detection of the information that express variations of the signs depending on semantics and tagging of the lemmas with respect of this information.

The first operation allows to create a sequence of lemmas that reflects the LIS sentence structure. The second operation allows to add significant information for the virtual character depending on the semantics of the sentence.

The output format is called AEWLIS, ATLAS Extended Written LIS and has been defined in the ATLAS project as an annotation schema and as the main output of the Computer Assisted Translator. The translator produces an XML with tags and attributes. Each tag is a communication channel (see section 2) and each attribute is a particular modification that occurs in the sign for that communication channel.

### A. Atlas Written Extended LIS

AEWLIS is a formalism defined within the ATLAS project to properly represent the different sets of information items associated with a given LIS sentence.

In particular, within the ATLAS project, the AEWLIS is used in 3 different use-cases:

- as the output of a translation step, from Italian text to AEWLIS;
- as the output of an annotation step, from a LIS video to AEWLIS;
- In a modification/correction/upgrading phase, from AEWLIS to AEWLIS.

When building the AEWLIS formalism, the following Communication Channels (CCs) (see section 2) were considered to be relevant in LIS: Hands, Direction, Body, Shoulder, Head, Facial, Labial and Gaze. All the previous CCs are assumed to be each other orthogonal and require an independent recording of the actions performed on each of them. It is also assumed that signs can not be signed contemporaneously.

Furthermore, the LIS sentences are assumed to be represented as sequences of Signs. The set of Signs used within the ATLAS project include: the signs associated with “Lemmas” included in the Radutzky dictionary; signs widely used within the various Italian deaf communities, but not included in the Radutzky dictionary; signs newly defined within the ATLAS project, previously not existing, and needed in order to represent peculiar concepts, mainly linked with meteorological events and situations; and signs associated with “LIS classifiers”.

### V. THE SYSTEM ARCHITECTURE

From these considerations emerge that what is specified into the AEWLIS file directly influences the virtual character movements. The figure 3 shows the architecture of the system. The login module allows user profiling and manages useful information about
the human translator such as personal information and professional experience to monitor his dependability. User profiling is performed on the basis on their experience in translation, their skills as interpreters or whether they are deaf or not.

The user is assisted through all the process from text import to attributes tagging (see section 4). After the text import the system checks whether that sentences has been previously translated or not. All the sentences that are translated are stored in the ATLAS Multimedia Archive (AMMA). In the case it finds a sentence, or a portion of, that has been previously translated the system suggests it as a possible translation. AMMA stores the data about translations with the correspondent AEWLIS, complete with linguistic information and attributes.

The system provides a visual feedback by using the ATLAS player currently under development. The user can visualize the results of his translation to check if is intelligible when performed by a virtual character.

A. AMMA

The ATLAS MultiMedia Archive has been previously described in detail in previous works [15]. Here we will give a brief overview of its use. It is the data storage system that supports all the translation process within the ATLAS system. It is designed to store:

- Corpora Databases
- LIS Lexicon
- AEWLIS translations
- Wordnet
- Metadata

It allows to share in a structured way information regarding the automatic translation process from Italian to LIS. The design of the database allowed to deeply investigate relations and dependencies between the major entities taking part into the translation process. Moreover it provides support for all the operations from the content ingestion (i.e. text import and metadata management) to the translation storage.

VI. CONCLUSION AND FUTURE WORK

This paper presented a novel system for the Computer Assisted Translation from Italian written Language to Italian Sign Language of Deaf people the paper focused on the translation process and suggested requirements for such purpose. The system supports the manual translation by providing suggestions and feedback on the main operations required by translation such as:

- Tokenization and language analysis
- conversion of the sequence of lemmas resorting to the source and target languages grammar,
- suggestions on the best possible translation on the basis of the previous translation performed in a learning-by-doing fashion,
- Direct link with the lexicon stored on the AMMA database,
- Automatic metadata management,
- Avatar preview on the translated sentence.

The system is based on the ATLAS automatic translator and on the ATLAS player for avatar visualization, both under development within the ATLAS project.

Future work aims at integrating the ATLAS translator and player and at defining the set of metadata to be stored in the AMMA database. The system will be tested resorting to group of users and translators with different background. They will be both deaf and hearing people.

VII. ACKNOWLEDGEMENTS

The work presented in the present paper has been developed within the ATLAS (Automatic Translation into sign LAnguageS) Project, co-funded by Regione Piemonte within the "Converging Technologies - CIPE 2007" framework (Research Sector: Cognitive Science and ICT).

The authors would like to thank Andreea Barticel for the great work done on LIS documentation.

REFERENCES


Modeling a Game-Based Adaptive Unit of Learning to Support Adults with ADHD in Virtual Learning Environments

Laura Mancer, Silvia Baldiris and Ramón Fabregat
Institute of Informatics and Applications
University of Girona
Girona, Spain
e-mail: {lmancer, baldiris}@eia.udg.edu, ramom.fabregat@udg.edu

Abstract—This paper presents an approach to adapt a videogame executed in a virtual learning environment (VLE) to meet the needs of adults with attention deficit hyperactivity disorder. The videogame is modeled as a Unit of Learning (UoL) using the IMS learning design. The UoL will be able to automatically adapt features such as lighting, activities and sounds to each user profile. This user profile is built considering five user characteristics: behavioural conduct, cognitive performance, emotional state, learning style and game-based performance. Based on this information, the game is adapted and constantly updated for the user in order to better personalize the learning process.

Keywords—Attention Deficit Hyperactivity Disorder, Video Game, User Modeling, Unit of Learning, IMS – Learning Design, Adaptive Hypermedia System, Virtual Learning Environment.

I. INTRODUCTION

Many efforts have been made in e-learning to address the individual user’s needs and provide adaptive learning processes. To accomplish this goal, a convergence of Adaptive Hypermedia Systems (AHSs) and Virtual Learning Environments (VLEs) was proposed [1, 2]. Adaptive VLEs, which aim to provide personalized learning and management processes, are a broad field of research covering a diversity of user characteristics and needs, and employing a diversity of technologies that VLEs support. For the latter, several standards and specifications have emerged to provide interoperability, reusability and efficiency among VLEs [3, 4, 5]. The large variety of VLE users includes people with disabilities, disorders or illnesses, and this technology could help them achieve quality learning.

Attention Deficit Hyperactivity Disorder (ADHD) is a lifespan disorder characterized by inattention, hyperactivity and impulsivity [6]. Most of the research on it has been focused on children and adolescents (between 4 and 18) [7, 8, 9]. Recent studies have found that although symptoms of inattention remain a problem, symptoms of hyperactivity and impulsivity decline with increasing age. This may contribute to an under-diagnosis of ADHD in adults [10, 11]. Adults with this disorder typically have serious complications in life, including poor academic performance, occupational difficulties, impaired peer relationships, traffic accidents and violations of the law [12]. The research in this paper is focused on adults (18-60 year-olds) with ADHD symptoms who are involved in a lifelong learning process, especially in a university context.

Students with attention problems usually have great difficulty concentrating in class. However, they can easily play video games for long periods of time because these games are usually extremely over-stimulating [13]. Proposals to use games in the classroom have been well received by students and teachers, and many successful cases have been documented [14, 15, 16]. These cases include games for people with ADHD and other learning disability disorders [14, 17, 18].

We propose to model an adaptive and intelligent video game based on the needs of students with ADHD that runs in a VLE, also known as a Learning Management System (LMS). For this purpose, we have conceived of an educational video game as an adaptive Unit of Learning (UoL), which is modeled using the IMS Learning Design specification (IMS-LD) [19]. This specification allows the creation of advanced UoLs, which can be used to model adaptive learning strategies with games [20, 21]. Furthermore, the IMS-LD permits those learning designs and their processing to be exchanged and applied across different learning environments.

The video game is adaptive: the information acquired through a user models is used to adapt features of the video game for particular user profiles. In this way, a variety of characteristics, environments, activities and other features are shown to different users. The video game is also intelligent: a system based on agents is used to update the user model according to the performance of the students (players) while they play. Being linked to the user model, the video game is capable of making real-time changes to upgrade adaptations for each particular user. In addition, intelligent techniques, especially machine learning techniques, are being used to build the system.

The user model is created considering five features: behavioural conduct, cognitive performance, emotional state, learning style and user’s game-based performance. It is stored in accordance with the guidelines from the IMS Accessibility for Learner Information Profile specification (IMS-AccLIP) [22]. By providing the user model according to this specification, the user’s information stored in the user model can be used by all LMSs that support the IMS-AccLIP. The <e-Adventure> educational game engine is being used to implement the adaptive video game [23] because this platform facilitates the integration of educational games and game-like simulations in LMSs. Moreover, <e-adventure> provides integration with LMSs through game exportation following educational standards, particularly the IMS-LD specification.
This paper is structured as follows: section II introduces the motivation for and the background to our proposal. Section III outlines how our proposal works. Section IV describes the proposed user model. Section V explains the characteristics considered to design the video game and the process to suitably deliver the video game to the users. Finally, section VI presents some concluding remarks and recommendations for future work.

II. BACKGROUND AND MOTIVATION

The proposal presented in this paper emerges from the BCDS group’s interest in using the characteristics of e-learning users to provide customized learning processes. This means adjusting the delivery of learning resources and activities to the principal actors (teachers and students) in the teaching and learning processes. In this context, as indicated before, a convergence of AHS and LMS was proposed [1, 2]. The BCDS group has carried out a variety of research as a result of this merger [24, 25, 26].

The basic idea of an AHS is to know the characteristics of system users. The AHS adapts itself to the features inferred for each user. The system offers resources, activities, recommendations, instructional designs and other kinds of contents to users according to their profiles [27]. The distinctive elements of an AHS are: a) a user model, which represents information about a particular user that is essential for the adaptation effect of an adaptive system [28]; and b) an adaptation process, which is the mechanism that takes adaptation decisions according to the user model.

An LMS is a web application used frequently to publish and provide tools for the management of educational courses and contents [29]. Generating quality e-learning resources can be a difficult task, so there is great interest in guaranteeing the interoperability, the reusability and the efficiency of the information stored in LMSs [30]. Several e-learning standards and specifications are being adapted to enable interoperability among different platforms, to protect the investment in content development and to exchange local and global content. The specifications written by the IMS Global Learning Consortium [3] are widely accepted and supported by many e-learning platforms. The IMS-AccLIP specification [22] provides a framework to define general user characteristics such as identification, preferences, interests, goals, certifications and licenses, acquired and desired competencies and disabilities. The IMS-LD [19] provides a specification to create complex instructional designs or UoLs in a standardized way. Instead of supporting a specific set of pedagogical approaches, the specification provides a framework to define personalized learning flows consisting of plays, acts, activity structures and environments [31].

The IMS-LD defines three levels of detail, 1) Level A is the main part of the specification that provides the reference for building any UoL with the elements: methods, plays, acts, roles, role-parts, learning activities, support activities and environments; 2) Level B adds features to create more complex lesson plans using properties, conditions, calculations, monitoring services and global elements; and 3) Level C adds notifications. Level B UoLs actually model several classical adaptive methods including the reuse of pedagogical patterns, navigational guidance, collaborative learning, contextualized and mobile distributed learning and stereotype-based adaptation. The properties in IMS-LD can be related to the attributes of the IMS-AccLIP. This permits easier personalization of the course, assessment or game level.

Although there are a large variety of users of LMSs, we focus on the educational needs of adults with ADHD. This is because a) the number of people diagnosed with ADHD has grown during the past few years [32], b) the research conducted on ADHD is mostly focused on children and adolescents, overlooking the adult population [7, 8, 9], c) several studies have demonstrated that individuals with this disorder might suffer difficulties and letdowns, such as school and job failures [12], d) ADHD may be undiagnosed among adults [10, 11], and e) several studies have found that most students with deficits such as those that make up ADHD, and who take online courses, drop them in a few days because they find the courses hard to follow [33].

ADHD is a disorder that, at least in childhood and adolescence, is characterized by inattention, hyperactivity and impulsivity symptoms. The basic types of ADHD, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) of the American Psychiatric Association [34] are: predominantly hyperactive-impulsive, predominantly inattentive and combined hyperactive-impulsive and inattentive, depending on the symptoms a person presents. In this paper we consider the inattentive type and combined type, which includes symptoms of the inattentive type, because symptoms of inattention remain in adults [10, 11]. In neuropsychological studies performed on people with ADHD, some cognitive functions were impaired, especially failures in executive functions (EF) [7, 8, 9, 35, 36, 37, 38], which comprise interrelated higher-order cognitive processes responsible for goal-directed and contextually appropriate behaviour [39].

In 1998, an experiment was performed to study the influence of a computerized cognitive-training system (“Captain’s Log”) on the behaviour and performance capabilities of some children with severe cases of ADHD [40]. Results show that the children who were most successful in the training demonstrated the highest generalizations of those skills that were the focus of the treatment. Etchehareborda and Comas developed a multimedia tool called MCC-94 [41] to monitor cognitive performance. The tool evaluates certain cognitive areas and it is used to diagnose children with ADHD. In this context, cognitive tests are used to measure the performance of EF skills. However, ADHD research has mainly focused on cognitive aspects, and much less on the study of the affective deficiencies accompanying the disorder. Several studies have concluded that these affective deficiencies are also important for the effective diagnosis and treatment of ADHD [37, 42, 43, 44]. Specifically, these studies indicate that children, young people and adults with ADHD show a primary dysfunction in the recognition of emotional stimuli. Furthermore, they also present an important inability to
control their emotions, especially the negative ones: aggressiveness, depression, sadness, anger and frustration.

A recent model, known as the ‘Dual Pathway Model’, considers both cognitive and emotional deficiencies in ADHD and concludes that separate pathways exist for them. An individual with ADHD presents impairments in one or both of these pathways [45]. However, cognitive and motivational deficits are not exclusive of ADHD and evaluating ADHD in adults is a complex process [46]. It generally requires performing a retrospective diagnosis of childhood symptoms, using diagnostic criteria proposed by international experts grouped in the DSM-IV [34] or the International Statistical Classification of Diseases and Related Health Problems (ICD-10) of the World Health Organization (WHO) [47], discarding other illnesses or disorders, performing cognitive tests, interviewing the evaluated persons and their families, analyzing patient behaviour, and including, if necessary, more complex studies such as magnetoencephalography and functional magnetic resonance.

This paper considers the following processes to infer if a user has ADHD symptoms: multimedia cognitive tests to obtain deficits in cognitive areas, emotion recognition techniques to recognize users’ emotions in real time, a self-assessment rating scale to characterize and to quantify relevant ADHD behaviour symptoms and a key question that aims to discard other disorders. Although we are studying deficiencies related to ADHD, we will not present a method for detecting ADHD in a medical context.

III. OVERVIEW OF THE PROPOSAL

The work presented in this paper proposes to model and deliver an adaptive/intelligent video game for students with ADHD symptoms performing learning processes in LMSs. Information related to the ADHD of each user is stored in a user profile. The video game is modeled using the IMS – LD specification level B because this specification relates the information stored in each user profile to conditions that define different features of the video game according to the characteristics of each user. The general framework of our proposal is described in this section and presented in Fig. 1.

In Fig. 1, we have considered the creation of external tools modeled as web services in order to evaluate each characteristic considered in the students’ profiles. These tools can be connected with a variety of LMSs. However, some elements in the LMSs need to be adjusted, or incorporated, to complete the user modeling process. These are: 1) the creation of a data structure where the information related to the profile is stored, 2) the implementation of an environment where the process is explained to the students and where the links to access the evaluation tools are shown and 3) the definition of classification rules, from which the user’ profile is built. Communication between the external evaluation tools and the LMS is achieved through an interface using the XML-RPC protocol. Also in Fig. 1, we show that E-adventure, the platform that will be used to create the video game, is an external tool that needs to be connected to the LMS. This platform already offers the connection functionality with a variety of LMSs.

Furthermore, in Fig. 1 two timelines are presented: design and execution time. Two lines have been defined because some user profile information is acquired before the game starts and other user profile information is obtained in real time when the student is playing the video game. Design time is the portion of time when all the necessary conditions to provide the users the adequate characteristics according to their profile are defined and implemented. For this reason, the time it takes the student to perform the tasks and do the tests to extract the information needed to build the user profile is included in the design time. Execution time is the portion of time when an agent system monitors in real time the performance of the users while they play the game. Although the mechanism of real-time facial expression recognition to extract users’ emotional states begins during the design time, it has been included in the execution time because this process is also performed while the student is playing the game. As the user profile information is constantly updated, the initial adaptation of the video game given to each user can also change.

The general process can be described as follows: the students are asked to take two computer question-based tests and perform some computer cognitive tasks. The tests are used to characterize and quantify relevant ADHD behaviours in students and determine their learning styles. The computer cognitive tasks are used to evaluate the students’ cognitive performance. While the users interact with the system, the mechanism of facial expression recognition is executed to extract information about the emotions expressed. A preliminary user model is built form these characteristics. Based on this preliminary user model and the conditions previously generated to model the video game, the first version of the game is delivered to the user. An agent system is constantly executed to consider the performance of users while they play the game and to update variables in the user model. These updates change the values of the condition parameters in the UoL, activating the analysis of the conditions one more time. By changing these values, the characteristics that are shown to or hidden from the user are also adjusted. The intention is to provide better
customization and personalization as the students play the game.

Moreover, the user model will be able to be stored according to the IMS-AccLIP specification through a service connected to the LMS. The user information is saved in the attributes of the <accessibility> element within the information model of this specification. By providing the user profile with this format, we offer an interoperable user model that can be used by those LMSs that support the IMS-AccLIP specification.

IV. USER MODELING

An effective user model for our proposal must be able to recognize if a user has ADHD symptoms. Therefore, we have evaluated behavioural aspects, cognitive performances and user emotions to infer if a user has ADHD symptoms. Furthermore, we have taken into account the user learning style to construct this user model. This is because several studies have found that the methods or strategies that a student employs to select, process and work with information, i.e. the user’s learning style, is an important part of the learning process of people with ADHD [48]. An explanation on how each feature is obtained is presented below.

A. Behavioural Conduct

In order to characterize and to quantify user behaviours that might be relevant to ADHD symptoms, we used the short version of the Adult ADHD Self-Report Scale – v1.1 (ASRS v1.1) [49]. This scale has been proposed by WHO [47] and validated for Spanish speakers. The questions in the ASRS v1.1 are consistent with the DSM-IV criteria [34] and address the manifestation of ADHD symptoms in adults. In the test, the first four questions are related to attention symptoms and the last two to hyperactivity symptoms. This self-report scale takes less than five minutes to complete.

The information obtained by applying the ASRS is a starting point to identify if an adult might have ADHD symptoms. However, some studies have found that even though this scale provides good convergent validity, sensitivity, specificity and diagnostic capability, there is a high probability of obtaining false positives such as bipolar disorder or schizophrenia. To mitigate this situation and discard other disorders, a key question has been included. The result of the evaluation of behavioural conduct is either positive or negative. For the positive scenario, the result will also determine the ADHD type.

B. Cognitive Performance

Cognitive tasks are used to extract information about a user’s cognitive performance. Even though implementations of these cognitive tasks in a computer format exist, we implemented our own tools considering that most of these implementations are available as proprietary software.

According to the study realized by Marchetta [50], the areas of the EFs that should be evaluated are: working memory, sustained attention, mental flexibility, verbal learning and choice-making capacity. The last one is evaluated while the user plays the game. The results of the evaluation of each area are very low, low, medium, high and very high, calculated according to normalized data. Combining these results, a final cognitive performance is obtained.

Fig. 2 provides an overview of the approach used to obtain a partial user model according to the cognitive performance. When users are asked to present the cognitive evaluation, a message is sent to the service with the user identifier. Then, the service returns a message with the results of the evaluation for this user and the data is stored in the LMS database. The service that evaluates the abovementioned cognitive areas is completely transparent to the student. The capacity to make choices is not evaluated with this service because this cognitive area is inferred while the user is playing the game.

C. Emotions shown by the user

As mentioned before, people with ADHD show an important inability to module their emotions, especially the negative ones. In this work, the emotions shown by the user are considered variables that support the presence of ADHD symptoms in a particular user. The emotions captured are: sadness, anger, frustration, happiness and a neutral state. These emotions are obtained through a real-time facial expression recognition system based on images taken with a webcam while the user is working on the LMS. Fig. 3 provides an overview of the process to capture [51] and to store the emotions in the LMS database.

As shown in Fig. 3, an external service is provided and linked to the LMS to capture the emotions expressed by the user. The tool is executed repeatedly after short periods of
time to record several emotions of each user and define a pattern of the user mood.

D. Learning Style

The behavioural conduct, the cognitive performance and the emotions expressed by the user guide the diagnosis of people with ADHD symptoms. Learning style is also used because a variety of studies have revealed that adapting this characteristic is important for appropriate educational interventions [48].

A learning style is defined as the strengths and preferences in the way people acquire and process information. Felder’s theory [52] classifies different kinds of learning styles along four dimensions, as shown in Table I.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Style</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing</td>
<td>Active</td>
<td>tend to do best when they can do hands-on work and actually conduct experiments or manipulate things manually.</td>
</tr>
<tr>
<td>Perception</td>
<td>Sensitive</td>
<td>gravitate towards concrete facts and figures.</td>
</tr>
<tr>
<td></td>
<td>Intuitive</td>
<td>prefer the conceptual and the theoretical to the concrete.</td>
</tr>
<tr>
<td>Input</td>
<td>Visual</td>
<td>prefer to see what they are learning through graphs, diagrams and pictures.</td>
</tr>
<tr>
<td></td>
<td>Verbal</td>
<td>are most successful when the information is heard or read through words.</td>
</tr>
<tr>
<td>Understanding</td>
<td>Sequential</td>
<td>prefer to have information laid out in a linear and orderly fashion.</td>
</tr>
<tr>
<td></td>
<td>Global</td>
<td>prefer to see a big picture first.</td>
</tr>
</tbody>
</table>

Each characteristic takes one of the two possible style values. Users have to take Felder’s test to define their learning preferences [53] and the learning style of each user is constructed as described in [24].

V. VIDEO GAME MODELING USING LEVEL B IMS – LD

The learning strategy proposed in this paper consists of using a video game adapted to the needs of adults with ADHD. We considered the needs required by the different learning styles and the needs required by the different types of ADHD to design the environments and characteristics involved in the video game. These characteristics include colour level, content and figure quantity, sound balance, animation balance and level of difficulty.

The video game is created so that the decision making – or an evaluation of a user’s ability to make decisions – largely guides the dynamics of the game. Students are given awards to maintain their motivation while they play, since a variety of studies have found that frequent and short-term recognition is important to motivate people with ADHD. The goals and instructions of the video game are given as clearly and unambiguously as possible. The idea is to create a fun and lively, but not extremely over-simulating, vide game.

As stated before, the video game is modeled as a UoL with the level B IMS-LD specification. In the <conditions> element of this level rules can be established to determine if a specific characteristic, activity or a branch of the game flow (a set of activities) is shown or hidden to a specific user.

IF Student have :: (ADHD, type C, level low)
THEN hide activity A1 and show activity A2
OTHERWISE show activity A1 and hide activity A2

Figure 4. Example of rule for adapting the videogame.

In Fig. 4 a rule considering the results from the diagnosis given to a specific user is provided as an example: “If the student evaluation said that a user may have ADHD combined type symptoms and these symptoms are low, then activity A1 is hidden and activity A2 is shown; in any other case, activity A1 is shown and activity A2 is hidden”.

The initial adaptation of the video game presented to users is based on the preliminary data stored in the user model of each student. However, the user model is maintained and updated by a system that is constantly monitoring the users’ performances while they are playing. In this manner, the video game changes as the user profile information does and better decisions can be made about the characteristics to be presented to the users.

The complete process to deliver the adequate features according to each user profile contains five steps. In the first one the system considers the type of ADHD a particular user presents. At that moment, one of the possible paths to start the videogame is selected according to the ADHD type. In the second step the system determines the stage of the game to be played by the user. This is defined by the severity of the ADHD symptoms the user presents: very low, low, medium, high, or very high. This severity is determined by the results of cognitive performance and behavioural conduct. The learning style of each student is considered in the third step. With this feature and the ADHD type, the activities, resources, roles, environments and other characteristics of the video game to be presented to the user are chosen. In the fourth step the adapted video game is delivered to the particular user and in the fifth step the IMS-LD makes adjustments to the properties initially delivered to the user, according to the changes in the user model. This adjustment process is carried out in real time.

The <e-Adventure> educational game engine is used to implement the adaptive video game, because this platform facilitates the integration of educational games and game-like simulations in educational processes, in general, and in LMSs, in particular. Moreover, <e-adventure> provides integration with LMSs through game exportation following educational standards especially supported by IMS-LD.

VI. CONCLUSIONS AND FUTURE WORK

The overall goal of this work is to provide quality learning processes to people with attention problems, more specifically to adults with ADHD who have found in virtual learning an alternative to complete their education. This paper presents an approach to adapt a video game to a VLE, according to the needs of adults with ADHD. However, other strategies can be developed, e.g. practicing relaxation...
techniques, providing recommendations or adapting complete instructional designs. The video game is modeled as an UoL using the IMS LD specification. To deliver an adapted UoL to every user, a user model that considers information related to the ADHD in adults has to be linked to the UoL. With regard to this user information, we have found that in order to provide a suitable learning strategy to people with ADHD, the users’ behavioural conduct, cognitive performance, emotional state, learning style, and game-based performance need to be considered. The strategies for this work must be in accordance with the needs that the ADHD require.

Although this work was initially focused on the necessities of adults with ADHD, we found that other neurological and psychiatric disorders may present the same deficiencies as ADHD. Therefore, a key question was included so as to exclude disorders other than ADHD. Thus, the possibility of obtaining false positive results in the detection of whether a user can have symptoms of ADHD is reduced.

Instead of considering this situation as a problem, we saw it as an opportunity to expand the range of users to which this development can be addressed. On one hand, it can be extended to other disorders by applying specific rating scales for each of them. On the other hand, the project can offer a generalized user model that considers cognitive and emotional dysfunctions.

The expected result is that the adaptations automatically performed by the proposed system can be compared with those suggested by a professional. This will indicate how good the adaptation that the system provides is.

ACKNOWLEDGMENTS
This work is supported in part by the Spanish Science and Education Ministry through the AZUN@ project (ref. MICINN (TIN2008-06862-C04-02/TSI)). The BCDS group (ref. GRCT40) is part of the DURSI consolidated research group COMUNICACIONS I SISTEMES INTELLIGENTS (CSI) (ref. SGR-1202).

REFERENCES


Ethical Considerations When Using Video Games as Therapeutic Tools

Jason Colman¹, Paul Gnanayutham²

¹SAE London, 297 Kingsland Road, London E8 4DD, United Kingdom
²School of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth, PO1 3HE, United Kingdom
{j.colman@qantm.co.uk, paul.gnanayutham@port.ac.uk}

Abstract—Video games have been used in a variety of therapeutic and rehabilitative contexts. However, there are health risks associated with playing video games, including the risk of epileptic seizure. Additionally, video games have been criticised for reasons including their portrayal of women and minorities. For games to be accepted as an ethically valid therapeutic tool, these concerns must be addressed. The authors believe that video games can be used as therapeutic tools when used responsibly.

I. INTRODUCTION

Video games are a new medium, and arguably an art form [20], born in the second half of the twentieth century [23]. The term “console game” refers to games which are played on a dedicated hardware device, called a console, connected to a television. “Handheld games” are portable entertainment devices. The latest generation of handheld games includes the Sony Playstation Portable, and Nintendo DS. Mobile phones also offer sufficient capabilities to be used as portable gaming platforms [16]. Finally, “computer games” are entertainment software applications played on a personal computer. The term “video game” is used here to mean a computer game, handheld game or console game.

Video games have been used in a variety of therapeutic contexts [12], [13]. However, video games have well established health risks, and have also been criticised for other reasons, including the effects of their violent content and portrayal of women.

The authors are interested in new therapeutic uses for games, particularly in the context of people who have suffered a traumatic brain injury and may require a brain-computer interface (BCI) to communicate. This group of users clearly has specific needs – are they more at risk than others from the dangers of video games?

II. ETHICAL ISSUES IN RESEARCH

All research involving human participants today is subject to ethical considerations. Kimmel (p.6) [21] states that “Research ethics comprises principles and standards that, along with underlying values, guide appropriate conduct relevant to research decisions.”

Many ethical issues arise from conflicting sets of values. One example is that naturalistic observation may cause a conflict between methodological validity and participants’ privacy. Many ethical dilemmas are brought to the fore in experiments where the participants are initially deceived by the researcher.

At its heart, the question the researcher must answer is “Should I conduct this study?”

A. Personal, professional and regulatory ethics

There are three main sources of guidance to which the researcher can turn in the ethical decision making process: personal; professional; and regulatory.

A researcher’s personal ethical values are shaped by his or her life experiences, and may lie on a spectrum between means-oriented (“do no harm”) and ends-oriented (“the ends justify the means”) [21].

Three tests for a personal ethical decision are suggested by the Institute of Business Ethics [19] (in the context of a business decision):

1. Transparency: do I mind others knowing what I have decided ?
2. Effect: whom does my decision affect or hurt ?
3. Fairness: would my decision be considered fair by those affected ?

In addition to his or her personal value system, the researcher will be expected to adhere to standards set by their profession. Examples of professional bodies are the American Psychological Association (APA) and British Psychological Association (BPS). These bodies’ standards are relevant to the evaluation of software user interfaces, because the experimental model of psychology is often followed in HCI experiments.

Researchers at hospitals and universities are required to submit their proposed study to an ethical panel or commission, typically called an Institutional Review Board. Kimmel [21] (p.51) notes some reservations about ethics boards. Wueste [41] notes their strengths. The ethical validity of a course of action is judged by several people. If the group of judges all arrive at the same conclusion, the confidence in the decision is
increased; if there is a conflicting decision, this indicates a possible dilemma that has been overlooked.

Finally, legal regulations supersede personal and professional principles. Examples of such government regulations are the US Federal regulations for Human Research; and in the EU, the 2001 European Commission directive (Data Protection Directive 95/46/EC) requiring ethics committees for medical research.

B. Informed consent

Contemporary formal ethical standards can be traced back to the Nuremberg Code [28], arising from the Nuremberg trials of Nazi scientific atrocities. The ten-point Code introduces the concept of voluntary, informed consent, describing it as “absolutely essential” [21].

Saha and Saha [35] discuss informed consent in the context of clinical trials. Informed consent is crucial to allow the participant to decide what risks to take with his or her body. Informed consent protects the human rights of the participant. It is too valuable a principle to be sacrificed for any anticipated research benefit.

C. Working with severely disabled participants

Researchers have a fiduciary responsibility to protect participants, in that an unequal relationship exists, where the more powerful person is entrusted to protect the best interests of the other.

Ethical issues arise when a participant is severely disabled and unable to communicate his or her consent. The APA Ethical code for research with participants from special groups provides the following guidance [21]:

“For persons who are legally incapable of giving informed consent, psychologists [should] nevertheless provide: 1. Appropriate explanation; 2. Seek the individual’s assent; 3. Consider such person’s preferences and best interests; and 4. Obtain appropriate permission from a legally authorised person, if such substitute consent is permitted or required by law.” (Section 3.10b)

The BPS Ethical Principles for conducting research with human participants states that [21]:

“Where real consent cannot be obtained from adults with impairments in understanding or communication, wherever possible the investigator should consult a person well-placed to appreciate the participant’s reaction, such as a member of the person’s family, and must obtain the disinterested approval of the research from independent advisors.” (Section 3.4)

D. Privacy

The right to privacy is enshrined in major human rights codes, e.g. the United Nations’ Universal Declaration of Human Rights; and the European Convention on Human Rights. A basic principle of research ethics is that the privacy and anonymity of participants should be respected.

Issues arise when usage of a system in the field is logged. Collecting data from real users in the field is recommended [25], to gather statistics such as how many features are used; or the rate of errors; to find usability problems which are not apparent during observations, etc. Clear issues of privacy are raised: the user must be made aware if logging is being performed, and must be able to disable it.

Kimmel [21] (pp.122-123) notes that methodological issues may also arise due to privacy issues, as participants may be unwilling to answer questions which threaten their privacy.

E. Participant Debriefing

Participant debriefing is regarded as an integral part of any experiment [18], [21]. The reason is that the subject will have a natural tendency to feel that they, rather than the hypothesis, were tested, and may believe they have failed the test, suffering a damaging blow to their self-esteem [18].

Nielsen [25] makes the same point, in the context of usability studies. In the evaluation stage of interface development, participants will be measured on how well they can use the interface, with attributes such as speed and accuracy being recorded. It must be made clear to the participant that the interface was under test, not the person using it. This can be explained before and during the evaluation, and also reinforced during the de-briefing.

However, concerns have been raised that debriefing may itself cause harm, for example, if participants have been chosen because of some deficit, such as low self-esteem or embarrassing behaviour [21](p. 80).

Experiments involving deception of the participants complicate debriefing. The experimenter cannot lie to participants during debriefing, as this would undermine its purpose and exacerbate ethical problems. However, the participant may believe that the debriefing is also a deception. An explicit debriefing will address this head on and has been found to reduce false beliefs and negative feelings [21].
III. THERAPEUTIC USES OF VIDEO GAMES

Video games have been used as therapy in numerous contexts [12], [13].

A. Physical therapy

Playing video games improves reaction times, hand-eye coordination, and raises the player’s self-esteem [13].

Loftus and Loftus [22] note that sports games, requiring speed, accuracy, strategy, and alertness, are useful in the treatment of problems with eye-hand coordination, visual field, and tracking.

Games have been successfully used in situations where repetitive motion is required of a patient, as physical therapy. Griffiths notes that this success may be due to the motivating nature of games, and their role in distracting attention from discomfort during physical therapy [13]. Griffiths cites examples of games being used as therapy for arm injuries, as a way of increasing hand strength, and improving arm reach for patients with traumatic brain injuries [13].

Burke and colleagues [5] have developed a suite of games designed to aid recuperation after a stroke. This is accomplished by requiring the player to make repetitive arm movements which aid upper limb recovery.

Games can provide entertaining challenge – fun – giving rise to motivation that more conventional forms of therapy may lack. O’Connor and team [30] developed a wheelchair interface to computer games called Game Wheels. This interface motivated spinal cord injury patients to exercise more regularly, by controlling games by driving their wheelchair.

Disorders involving muscles of the eyes have been treated using video games [22]. The monotonous task of visually following a dot on screen can be replaced with the more entertaining task of playing a video game. Such therapy is in use at present, and a study of 60 participants at London’s Great Ormond Street Hospital is planned [37].

B. Children with learning difficulties

Loftus and Loftus [22] (p. 148) report a study of 25 children with learning disabilities, aged between 6 and 13. The children were tested before and after playing a number of video games for 30 minutes, and were found to have improved in motor ability and spatial visualisation.

Demarest [7] describes the benefits of playing video games for her autistic son, aged 7. These are improvements in language ability, basic maths and reading skills, and social skills. Demarest stresses that these benefits occurred as a result of her involvement interacting with, and discussing the games. Demarest found that playing the games improved her son’s self-esteem and made him feel calmer, and has recommended their use to parents of other autistic children.

C. Treatment of behavioural problems

Favelle [10] used the game The Wizard and the Princess (Sierra on line, 1980) as a therapeutic tool to help adolescents with severe psychiatric disorders, in a residential treatment centre. The game allowed for the exploration of alternatives to violence, and the development of problem solving skills. Additionally, Favelle found the game Alter Ego (Activision 1986) effective in individual therapy, the game’s situations providing an opportunity to talk about sensitive issues in a safe environment. Favelle concludes that games can be an effective tool in individual and group therapy, when used in conjunction with skilled counseling.

Spence [38] describes ways in which video games have been used to help children with emotional and behavioural problems. He provides case studies showing ways in which games can bring about changes in the development of relationships, motivation, cooperation, aggression, and self-esteem. In his view, the use of games has effected positive change in the children in his care, subject to some guidelines which he provides.

D. Pain management

Griffiths [12] cites examples where video games have been used to manage pain, the games providing a task which distracts the sufferer. DeMaria [8] (p. 34) cites a survey of casual game players, twenty-seven percent of whom claimed that distraction from pain was a benefit.

Finally, Loftus and Loftus [22] remark on some indirect benefits of playing computer games, of which the most important may be an introduction to the world of computers, and an incentive for children to learn computer programming.

IV. HEALTH RISKS OF VIDEO GAMES

A. Photosensitive epilepsy

Numerous studies have shown that playing video games carries a risk of seizure due to photosensitive epilepsy (PSE), (e.g. [33], [24]). Video game manufacturers are careful to point out this risk. For
example, Nintendo includes a Health and Safety Precautions booklet with every game [27].

The incidence of people with PSE is approximately 1 in 4000 [17] (p. 161). This figure is also quoted by Nintendo [27]. However, the incidence of epilepsy is much higher among people who have suffered a traumatic brain injury (TBI). About 35% of TBI patients experience a seizure, with an ongoing risk of seizure in 5% of open or penetrating head injury patients. Closed head injury patients have a 1% chance of seizures [31] (pp. 66-67).

Precautions can be taken to minimise the probability of a seizure due to PSE. The incidence of epileptic seizure correlates with the number of retinal cells stimulated, and the intensity of stimulation [29] (p. 158). Advice to reduce the likelihood of seizure is given in [29] and [17].

Nintendo recommend that a person known to have suffered a seizure in the past should seek medical advice before playing a video game [27].

B. Joint and muscle complaints

Nintendo [27] cautions that “playing video games can make your muscles, joints, skin or eyes hurt after a few hours”. Indeed, a condition called “Nintendo elbow” is identified by Bright and Bringhurst [4]. A variety of minor ailments of this type are reported by Griffiths.

Treatment for these conditions usually consists of taking a break from playing the game in question [14].

C. Other health issues associated with video games

Gwinup and colleagues [15] measured the cardiovascular effects of playing a video game (Berzerk) in 23 healthy young men. The mean heart rate and systolic blood pressure of the participants during play was significantly higher than the rate before or after. Gwinup offers the explanation that the playing of video games causes the release of catecholamines. Novice players experience greater anxiety, and so a greater rise in blood pressure than for the more experienced players. Gwinup cautions that, in view of these results, it may be expected that video game players will experience other cardiovascular effects, such as arrhythmias. He predicts from the results that such effects would be more pronounced in novice players.

Is it dangerous to play video games? Overall, “the evidence of serious adverse effects on health is rare”, however, “frequent players are the most at risk from developing health problems” [14].

V. VIDEO GAME ADDICTION

For many years it has been noted that someone who plays video games excessively may appear to be “addicted” (e.g. [39]). Indeed, the “addictiveness” of a video game is seen as a desirable quality by game players and designers, exemplified by the interview of a leading game designer in [34] (pp. 26-27).

Loftus and Loftus [22] examine video game addiction in terms of the psychology of reinforcement and of regret. Experiments with rats in Skinner Boxes have demonstrated that unpredictable reinforcement, such as that provided by video games, provide the longest extinction period (i.e., is addictive for longer). Furthermore, the regret a player feels when he or she “dies” prompts the player to try again, to “undo” the mistake.

Griffiths and Davies [14] have studied whether or not video game addiction exists. Griffiths’ opinion is that this is a real condition, because six major criteria for addictive behaviour can be seen in some people who play video games excessively.

Behavioural signs of addiction in adolescents which have been reported include stealing money to play arcade games or to buy game cartridges; truancy from school to play games and not doing homework [14].

Game-related crime is also reported in [22] (pp. 109-110). A thirteen year old boy in Des Moines, Iowa resorted to constant burglary to fund his Pac-Man habit. In Japan, a twelve year-old held up a bank with a shotgun, demanding only coins, for arcade games. And “cases of children becoming prostitutes specifically to earn money for video games have cropped up in several countries”.

VI. VIOLENCE AND VIDEO GAMES

Since the arrival of video games, concerns have been voiced over their violent nature and the possible effects on the player, echoing similar debates over violence on TV and in movies. Smith [36] notes that most video games – around 80% - feature violence, with this figure rising to over 90% for games targeted at mature audiences.

Within the class of “violent” games, Loftus and Loftus [22] distinguish between violence to aliens and violence to other people. With regard to games such as Defender, Galaxian and Space Invaders, (where “aliens” are “killed”), “Despite E.T., the idea of defending ourselves against aliens may well be so deeply ingrained in our collective psyche that it’s futile even to worry about it”. Much more worrying to them are “kill people games”, although no evidence at
the time was available to demonstrate that playing violent video games promoted actual violence.

Provenzo [32] (p. 65) feels that the criticism of games which emphasise violence is justified, but does not distinguish between games in which fairly abstract aliens are “killed” (e.g. Galaga), and games in which humans fight each other. Provenzo takes the view that in any case, violent video games “…do – at least on a short-term basis – increase the aggressive behaviour of the individuals who play them”.

Anderson and Bushman [2] undertook a meta-analysis of 35 studies of video game violence, and found that violent video games do increase aggressive behaviour in children and young adults. They concluded that “exposure to violent video games poses a public-health threat to children and youths, including college-age individuals”. This is seen as a strong view by other media researchers who believe that more studies are warranted [26] (p.232).

Other evidence suggests that video games are not devastating society to such an extent. DeMaria [8] (p.19) shows that while consumption of video games has risen linearly since the 1970s, the youth violent crime rate in the US remained steady, until it began to decline in the mid 1990s.

VII. VIDEO GAMES, GENDER AND SEXUALITY

Consalvo [6] finds that games from Donkey Kong (Nintendo, 1981) to Final Fantasy IX (Squaresoft, 2000) have presented not only an unquestioningly heterosexual theme, but also a stereotyped view of females who invariably need rescuing by a male.

Provenzo [32] analysed the cover art of 47 popular video games, finding that they routinely portrayed women as victims, having no initiative, and dependent on men. Smith [36] reports similar findings. Of the 54% of games featuring female characters, only two featured females on the cover, both portrayed “provocatively”. Within the games, Smith found that female bodies in games are sometimes hypersexualised, with unrealistically large breasts and small waists. Smith concludes that girls have fewer role models in games. The role models that they do have tend to be hypersexualized and disproportionately thin. These depictions may also affect boys’ social learning about women.

VIII. VIDEO GAMES AS CORRUPTING INFLUENCE

On Nov 9, 1982, then US Surgeon General, Dr C. Everett Koop, delivered a speech in Pittsburgh in which he declared video games evil, that produced “aberrations in childhood behaviour”, and which should not be played (quoted in [8], [22]).

Although this statement was not supported by any evidence, and was later retracted, the sentiment is presumed to have been commonly felt among parents at the time. Some communities have banned arcades on the basis of being an unwholesome environment where aggressive behaviour is encouraged. In the Philippines in 1981, then President Ferdinand Marcos banned arcade games for being a corrupting influence on children [32].

Loftus and Loftus [22] cite studies showing that heavy viewers of television (more than 4 hours a day) were found to have different conceptions of the real world than light viewers (less than two hours a day) -- agreeing with, for instance, portrayals of women as weak and passive. The concern is that heavy consumers of video games will suffer a similarly distorted world view.

Some games are unquestionably offensive by design. Provenzo [32] goes further and states that “video games have a history of being sexist and racist”, citing the particularly egregious game Custer’s Revenge (Mystique, 1982) as an example. This game attracted protests over its depictions of women and Native Americans.

The genre of ‘shocking’ games is now more popular than ever, with Grand Theft Auto IV breaking sales records on its release [3]. The GTA series of games is deliberately offensive and been duly criticised. For example, Hillary Clinton (quoted in [8]) complained that “The disturbing material in Grand Theft Auto and other games like it is stealing the innocence of our children and it’s making the difficult job of being a parent even harder”.

Game Ratings Organisations

The video games industry has responded to concerns over unsuitable content by creating regulatory bodies. In the US, the Entertainment Software Review Board (ESRB) has been rating games in the US since 1994. Under the ESRB scheme, there are five age-based categories and 30 content descriptors.

In most of the EU, the Pan European Game Information (PEGI) standard created in 2003 is used to rate games. This scheme similarly specifies age-based ratings and 6 content descriptors.

On its release, GTA IV was rated “M” (mature, for ages 17+) by the ESRB, who urged parents to observe their ratings [9].
The issue of unsuitable advertising and box cover artwork has been tackled by the Advertising Review Council of the ESRB, who issued guidelines for the marketing of video games [1]. Publishers must be “sensitive” in portraying violence, sex, alcohol and other drugs, offensive expression, and beliefs.

Smith [36] reports that these guidelines have been complied with on the whole, with the exception of the depiction of excessive violence.

IX. DISCUSSION

In any research, the participants must be informed of the known risks, to allow them to decide if the risks are acceptable.

The risks in this case are: health concerns, particularly PSE; the violent nature of some games; the attitudes expressed within the games; and the possibility that a player could become “addicted” to a game.

The incidence of PSE is approximately 1 in 4000 in the general population. However, for some groups of people the risk is much higher. TBI patients in particular have a high incidence of epilepsy – as high as 1 in 20 for some types of injury. [29] and [17] suggest ways of reducing the risk of seizure, which would be followed in a therapeutic setting.

Given that video game addiction exists, it would seem that enabling someone to play video games must carry the risk that the player will become addicted. This risk may be higher for people who, due to severe disability, are unable to balance game playing with a variety of other activities.

There is some evidence that playing violent video games encourages violent behaviour in children. Console games are rated by game industry bodies, such as the ESRB and PEGI, who have a strong interest in minimising these violent effects. Parents, and others who supply games to children, are encouraged to follow their guidance, but this practice is not believed to be stringently followed by all, shown by the ESRB’s plea to parents.

Informed consent

In the case of severely disabled participants, it may be impossible for the participant to communicate either his or her understanding of the risks or their consent. In order for informed consent to be granted, family members or carers of patients need to have risks and benefits explained to them. For example, Gnanayutham provided demonstrations for this group of people [11].

Applying this spirit to video game therapy, the suitability of a game could be assessed collaboratively by the researcher, the participant and his or her family. All parties could consider the game rating, the cover art, and the instruction manual. The game could be demonstrated by the researcher, and could be played and discussed with the participant and family members.

Debriefing

Debriefing a participant after playing a game is a necessity. If a participant does not do well at a particular game, he or she may well feel that they have “failed” a “test”. In fact the researcher will have chosen the wrong game, perhaps in an entirely inappropriate genre. Additionally, the researcher may not be providing a user interface which is suitable for the participant.

X. CONCLUSIONS

Video games have been shown to have therapeutic benefits when used appropriately. Successful therapeutic uses have included physical therapy for stroke patients, spinal cord injury patients, and traumatic brain injury patients. Additionally, games have been shown to help children with learning difficulties and behavioural problems, and are used for pain management. Playing games improves the player’s reaction time, coordination, and self-esteem. This final benefit may be of particular importance for people coming to terms with an acquired disability. Additionally, the authors are interested in investigating the further benefits which may result from playing online games with other players, fostering a sense of teamwork, belonging and fellowship.

Unfortunately, there are potential negative consequences to playing video games. Firstly, there are some health risks, notably that of epileptic seizure. There is some evidence that playing violent video games encourages violent behaviour in children. Other criticisms of video games are that they are sexist, racist, and perpetuate stereotyped views. Positive female role models in games are scarce, affecting both boys and girls. Heterosexuality is invariably the norm in games. Heavy users of video games may suffer a distorted view of reality. Finally, video games may be “addictive” to some people, leading to anti-social behaviour, and increasing the likelihood of other negative effects.

On the issue of violent and shocking games, it should be noted that games of this type are uncommon, but attract the most media attention. Some of the harshest critics of these games are careful to avoid tarring the entire output of the game industry with the same brush. For example, Walsh [40] notes that “With so
many good games available for children and youth, it is unfortunate that so much attention has to be paid to games which are inappropriate for all youth and harmful to some”. And even Anderson and Bushman [2] “...wonder whether exciting video games can be created to teach and reinforce nonviolent solutions to social conflicts.”

New therapeutic game software can of course be developed with the negative aspects avoided as far as possible. However, it would seem to be preferable to use commercial, off-the-shelf games for therapy where possible. One reason is to avoid the time and expense of developing a new game; and another is to offer more choice to the participant. But perhaps the most important is that therapeutic software is only therapeutic when it is used; for some, serious games will not be as engaging as commercial games, as they are relatively limited in content.

We have seen that video games have the potential to offer therapeutic benefits to many groups of people. Most of the negative effects of playing video games are dependent on the game content, and so selection of an appropriate game genre and title, in collaboration with the participant and his or her family, is essential. Engagement of the researcher with the participant during game play, and debriefing afterwards, are also necessary.

Finally, we return to the three main sources of ethical guidance: personal, professional and regulatory. All researchers seek to follow ethical standards set in law, by Institutional Review Boards, and by professional bodies. As to one’s personal value system, we have seen that video games are beneficial to certain groups of people. Knowing this, are we not ethically obligated to offer these potential benefits to all?

Acknowledgement: Thanks to SAE London for financial support.

REFERENCES

http://www.esrb.org/ratings/principles_guidelines.jsp
Retrieved 2 July 2010


http://news.bbc.co.uk/1/hi/technology/7379624.stm
Retrieved 2 July 2010


Retrieved 2 July 2010


Retrieved 2 July 2010


Chapter 2

Assistive Technology
Enhancing the Academic Achievement of Disabled Students in Greek Universities through Accessible Technology Services

Anastasia N. Gkouvatzi,
Democritus University of Thrace, Department of Physical Education and Sport Science, Greece
e-mail: anasgouv@otenet.gr; anastasia.n.gkouvatzi@gmail.com

Abstract - The increasing number of disabled students attending higher educational institutions in Greece has produced new data regarding the support of these students, which is matched by an increased interest in making university programs accessible to all students. The purpose of the present study was to provide an overview of potential access difficulties facing students with disabilities in higher education, in order to assist academic policy makers create reasonable provisions to facilitate the successful completion of their studies. The research documented the opinions of seventy disabled (24 hearing, 23 visual and 23 motor physical impaired) students aged 20-32, who attend nine Greek Universities, by means of semi-structured interviews. Our results showed that AT (Accessible Technology) adaptations could support: (I) environmental access, through: special circulation devices (navigators) inside and outside of campus; adapted lifts with haptic and vocal feedback; Braille signs; tactile maps; and e-information for transportation, (II) learning access through technical tools for adapted course materials and alternative examination provisions generally. More specifically, these include: Real Time Graphic Display, Greek Sign Language Translation System, Radio frequency modulated and FM Units Technical resources for the hearing impaired, Braille transformer and printers, screen readers and magnifiers, academic manuscripts in e-format for the visually impaired, and electronic pointing devices and technological solutions to support students who experience writing difficulties because of motor physical impairments. Thirdly (III) the effective use of ICTs, and specifically the implementation of digital libraries, accessible web pages, software accessibility validation programs, online courses for distance learning programs, electronic resources for communication, research and problem solving.

I. INTRODUCTION

The ratification of national conventions and European Union rules on Equal Opportunities for Persons with Disabilities has obliged each Member State to adapt its national legislation in order to be compatible with the principles and content of these provisions. The major reasons that have brought about an increase in the registration of Persons with Disabilities in higher education during the last thirty years are these legislation adjustments, alongside calls by disabled charities and organizations for the improvement of the provided quality of education, with the provision of special up to date, ad hoc measures.

Of an estimated 428,280 students with disabilities that were enrolled in 2-year and 4-year higher education institutions in USA, 5,571% were registered with hearing, 4,36% with visual and 14% with mobility impairments (PEQIS,1999) [1]. According to HESA (1999/2000) [2], in a total number of 525140 first year undergraduates in England and Scotland, 22,290 have a disability and of these 6% were deaf/hard of hearing, 4% blind/partially sighted and 4% wheelchair users/mobility impaired. The mass admission of disabled people into tertiary education was granted in Greece by law in 1975, allowing the enrolment of disabled people in tertiary education without national admission examinations, at a percentage up to 3% (now 5%, Law N3794/09) over the total number of admitted students. There are no official National Statistical data on the number of disabled students who enroll in tertiary education. Our research showed that out of a total number of 597,000 students attending Greek universities (303,000 have exceeded the planned semesters of study and have not yet obtained an academic degree) the number of disabled students is limited. In total we documented 144 deaf, 90 blind and partially sighted and 278 students with motor physical impairments attending Greek universities and Technological Institutions.

The Greek Constitution and national legislation consolidates Disabled Persons’ Accessibility Rights regarding infrastructure, services and goods accessibility. Furthermore, these specify that e-accessibility is linked to the provision of information and logistics to people with disabilities through using assistive technology1 to facilitate traffic accessibility. For this reason, Greek Universities and Technological Institutions must use AT as a tool to support disabled students in their studies in order to fulfil national and international laws regarding equal opportunities in education and everyday life.

II. LITERATURE

As information and communication technologies become more ubiquitous and an integral part of our everyday lives in the 21st century, the issue of inequality of access and usage of computers and the Internet becomes increasingly problematic. Findings

1 The term AT (Assistive Technology) covers a wide range of services, devices and software application developed to facilitate the communication and movement of persons with disabilities and grant them full participation in everyday life.
from the U.S. Department of Commerce (2000) [3] suggest that persons with disabilities have lower incomes, are older, and are less likely to be employed than those persons without disabilities. They also indicate that 21% of individuals with disabilities use the Internet, compared to 51% of people without disabilities. Almost 60% of Americans with disabilities had never used a computer and were only half as likely to live in homes with Internet access than those without disabilities [4]. While 40% of individuals with specific learning difficulties had Internet access from home or elsewhere, only 20% of individuals with visual impairments did.

The use of Information and Communication Technologies (ICTs) is ubiquitous due to the inclusion of assistive and instructional technologies designed to respectively improve both students’ and instructors’ functional capacity. The inclusion of students with disabilities in general education has generated a plethora of software and hardware programs designed to produce the necessary provisions and modifications that these students might need. Studies by Jackson [5] (2003) and Kaye [6] (2000) have showed that access and usage continues to be limited for many students with disabilities due to a lack of access, funding and trained teachers.

A. AT for deaf and hard of hearing students

The Canadian Council on Learning (2009) has developed innovative software that improves Deaf and Hard of Hearing writing quality, scanning and optical character recognition. [7]. Lang and Steely [8] (2003) indicated that interactive multimedia and web-based curriculum materials yielded significantly greater knowledge gains for deaf students as compared to traditional teaching practices. They presented the content of each lesson through a short text screen, a corresponding animation explaining the text and an American Sign Language (ASL) movie version of the text. The authors believe that efficacious instructional science programs for hearing students that have been well designed and properly tested can be successfully adapted for use by deaf students by interspersing text and ASL explanations with content animation and by providing additional practice on vocabulary and content graphic organizers.

A deaf or hard of hearing person may also have difficulty using the web. If a web site relies heavily on sound to convey concepts, a person with hearing disabilities may require adaptations, such as headsets and captions for visual descriptions of the audio content [9]. Also, AT web resources used to locate videos for use in content-area instruction can help students to integrate information from videos with information from other sources (e.g., textbooks, tradebooks and websites). Video reproduction is an excellent opportunity to teach note taking skills and to promote discussion in order to clarify and deepen students’ understanding of the content [10]. In addition to teaching content, video also offers teachers the opportunity to impart principles of media literacy [11]. ICTs, for example Real Time Captioning (RTC), promote essential student learning benefits to high and low reading deaf groups [12], television captioning enhances their literacy, even when they are at equivalent reading levels with hearing students [13] and video captioning strategies from qualitative data analysis were found to have similar benefits for deaf persons and students with language and learning difficulties alike. [14]. An automated gesture recognition system – sign2talk – was proposed [15] for both the deaf and hearing for further expansion of communication possibilities. Similarly, relating the individual way of signing with the signer’s level of deafness was proposed by means of a novel hybrid adaptive weighting (HAW) process applied to surface electromyogram (sEMG) and 3D accelerometer (3D-Acc) data.

B. AT for blind and partially sighted students

To serve the academic needs of blind students (VI) i.e., class participation, study and evaluation, a series of improvements and adjustments are required to facilitate such tasks as reading a computer screen, taking notes, and communicating via paper and mail. This would mean that VI students should be able to gain access to information from library catalogues, databases, websites and participation in any research process they wish. [16]. Srivilailuck, Beale, Murray and Kidd [17] (2005) studied the available literature and showed that blind and partially sighted students face problems in both the architectural design of institutions and the learning environment as a whole. VI students face obstacles in accessing learning materials delivered in web-based and other electronic formats [18]. Many librarians use software accessibility validation programs to determine whether a web page suffers from accessibility and compatibility issues [19]. Over the past few years, the navigational capabilities of this technology have become quite remarkable, but its use in practice has presented unexpected difficulties. So, Craven and Brophy [20] (2003) have suggested training over time to allow for familiarization with each new updated version of programs. Gkouvatzi, Mantis and Costa [21] (2010) studied the perceptions of 25 Greek blind and partially sighted students and showed that AT accommodations are first priority services. ICT services can create an effective academic learning environment providing AT such as: adapted lifts with haptic and vocal feedback, Braille signs, transformers and printers, screen readers and enlargers, adapted teaching
methods and materials, alternative e-examination adjustments, academic materials produced in Braille or in e-format, optical character recognition (OCR), software that improves writing quality, etc.

C. AT for Mobility impairment students

Mobility impaired students, for example those with carpal tunnel syndrome, arthritis, congenital deformities, amputations, cerebral palsy, spinal cord injury, ambulation difficulties, wheelchair dependence, and/or cerebral vascular accidents face tremendous problems in education accessibility. Issues such as physical access to the campus and the location of computers (e.g. room, desk, proximity to the computer, etc.) both require consideration.

Adapted keyboards and specialized software products can support learning, providing writing support in all educational settings. For example a keyboard with a reduced number of keys, or a keyboard that enables on-screen scanning decreases the distance and frequency of keystrokes needed to complete an activity [9]. Positive results are discussed by Mirenda, Turoldo and McAvoy, [22][19] (2006) in accordance with educational technology support for students with limited use of hands or arms, using a word prediction software program which improves writing quality, dictation and correct word sequences. Urbano, Fonseca, Nunes and Lopes [23] (2009) proposed new architectural solutions for an assistive powered wheelchair, tuning it to each individual in order to provide an effective and safe control system for people without enough strength to operate a joystick. Another interesting assistive device that can improve their confidence and daily independence was presented by Coelho and Braga [24] (2009), consisting of an electric powered wheelchair with a speech-based interface and monitoring instrumentation to help people to have a productive and fulfilled life.

D. ICTs and Distance Learning Courses

The widespread availability and flexibility of the Internet has led to an explosion in online offerings worldwide [25]. ICTs, including e-learning, can promote the inclusion of all students with various disabilities by means of distance learning (online) courses which provide enhanced opportunities for people who, because of climate, health, transportation or physical accessibility, face obstacles in attending classroom-based courses [26]. Similarly, in traditional classes, students who have print impairments can access course notes and handouts available on the course website without assistance. Burgstahler, Corrigan and McCarter [27] (2004) worked together as leaders in the Distance Learning Accessibility Project to identify and implement changes to policies and procedures that would lead to more accessible distance learning course offerings. They chose to institutionalize policies and practices assuring the accessibility of future courses to all potential students and instructors with as their ultimate goal the maximization of program access and minimization of special provisional needs.

The World Wide Web Consortium for international use developed Content Accessibility Guidelines in 1999 [28] (WCAG) (W3C, 1999) which examine not only the interfaces required to access a computer, but also the virtual environment in which a person will access information on the Internet. Wright and Dawson [29] (2003) and Schmetzke [30] (2001a), using an online tool called Bobby, analyzed and reported accessibility issues on Web pages. Completing several Web accessibility studies and analysing 1051 community colleges web pages showed that only 29% of the web sites were free of major accessibility barriers. Designing a web page with simple tab features between all options would be one example of increasing the ease of web access use for people with physical disabilities [9].

III. AIM

Many disabled students spend large amounts of their time in learning, because of insufficiencies in areas such as: the environment (architectural, assistive technology, infrastructural material), services (academic, administrative, social, cultural) and goods (e-library, tactile maps, digital information material/brochures). At the same time, these insufficiencies have also led to a high percentage of students not completing their studies (a figure fluctuating between 40 – 75% has been recorded) [1].

The purpose of the present study is to provide an overview of potential access difficulties facing students with disabilities in higher education, demonstrating the beneficial role of Assistive Technologies in the successful completion of their academic education.

The accessibility barriers encountered by disabled students will record and rank the most commonly presented needs in order of importance and will help Electronic Computer Engineers and academic decision makers formulate proposals which will be compatible with the “Design for all” principle to make universities a field for the evolution of social equality.

IV. METHOD

A. Sample

Our sample consisted of 70 students, aged 20-32, with hearing, visual and motor physical impairments attending 6 Greek universities and 3 Technological Institutions. Out of them, 24 (13 males, 11 females) were deaf, 23 (13 males, 10 females) were blind or
B. Analysis

For data collection, coding and synthesis we used Grounded Theory (a method of statistical analysis widely used in the Humanities and the Social Sciences) and content analysis for the interpretation and cross-correlation of data. Semi-structured interviews were carried out, documenting the perceptions of seventy disabled students and covering accessibility topics in areas such as: environment, infrastructures, AT material, web access, academic and administrative e-services. In Grounded Theory there is no predefined, unchangeable theory but only a theoretical schema that is derived from the data and its analysis. On this occasion, Grounded Theory contributed to the development of categories of meanings, the identification of their combining elements and the formation of suggestions. It took information from various sources, recognised sub-categories within the main categories, highlighting, connecting and bringing together all the different categories and variables that resulted.

The qualitative analysis of the context offered the possibility to decipher the suggestions while taking into consideration the evolving needs, characteristics and behaviour of the sample. The frequency of their appearance helped us to place needs in order of importance and to translate these into ideas for provision and adjustments in order to better satisfy the needs of disabled students. Before starting the interviews, the subjects were given some brief information about the procedure and aims of the project.

V. RESULTS

The present research shows that AT could support full participation in academic life, but that Greek universities are not yet equipped to deal with disabled students and provide them with quality education.

- For deaf, blind and MphI students the top priority needs for academic achievement are technological infrastructures, material and academic assistive technology services such as: e-accessibility for information and communication with academic teachers and administrative staff, e-libraries, well informed and easily navigated web pages, adapted to each disability category evaluation method, electronic writers for e-note taking, recording lecture devices, adapted to each disability manuscripts, e-manuscripts, and Distance Learning Courses (Tables I, II).

- For deaf students, teaching and learning practices could be expanded and improved through adapted manuscripts enriched by images, and libraries with plenty of visual material such as DVDs and illustrated books, digital Sign Language dictionaries and PCs with on-line Sign Language interpretation (Tables I, II). The target of all of the above suggestions regarding the support offered to hearing impaired students is to fill existing gaps in education as well as in communication. These are insufficiencies that cannot be remedied even by the constant presence of transcribers / interpreters during the various stages of tertiary learning as well as in everyday life at the university.

- For blind students, besides making adjustments to architectural designs such as Blind Walking Lanes marked by colour contrast or pavement plates distinguished by special textures, there is also the need to supply specialised devices that will support their full integration in education and in university life.

- These will include AT devices such as: screen readers and enlargers, Braille printers and labels, scanners, circulation devices and circulation tactile maps, adapted lifts with haptic and vocal feedback. In education, reasonable provisions should include alternative teaching methods and tools for the learning of shapes, mathematical symbols and maps and updated text transformers (screen readers software) from PDF to Word (Tables I, II).

| Table I: Environmental assistive technology and technical-material infrastructures |
|---------------------------------|---------|-------|-------|
| Total number                   | HI 24  | VI 23 | MI 23 |
| Adapted e-accessibility        | 9      | 19    | 7     |
| Lecture recording devices      | 7      | 14    | 4     |
| Technological infrastructures  | 3      | 20    | 4     |
| (PC, internet)                 |         |       |       |
| e-library                      | 3      | 7     | 2     |
| Electronic writer (notes)      | 3      | 6     | 3     |
| Adapted manuscripts (images)   | 14     |       |       |
| Libraries with visual material (DVD) | 3      |       |       |
| Screen readers                 | 18     |       |       |
| Screen enlargers                | 12     |       |       |
| Braille printers                | 11     |       |       |
| Braille labels                  | 11     |       |       |
| Technological infrastructures in library | 10     |       |       |
| Scanners                       | 10     |       |       |
| e-books                        | 14     |       |       |
| Circulation devices            | 7      |       |       |
| Adapted lifts with haptic and vocal feedback | 12 | 19 |
| Accessible classrooms (entrance-exit detectors, lifts) | 10 | 14 |
| Circulation/tactile or digital colored maps | 7 | 3 |
Table II: Academic assistive technology services

<table>
<thead>
<tr>
<th></th>
<th>HI</th>
<th>VI</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number</td>
<td>24</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Link to contact early</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information for Disabled</td>
<td>12</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Student Services online student</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>handbook.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-communication with teachers</td>
<td>7</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>and administrative staff.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic e- manuscripts</td>
<td>2</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Accessibility training sessions</td>
<td>7</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>(for AT familiarization)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapted on-line teaching</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(Distance Learning)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inquiry alternative teaching</td>
<td>2</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapted evaluation for each</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>disability category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-line Sign Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interpretation</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adapted manuscripts</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Alternative inclusive teaching</td>
<td>3</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>and learning practices(adapted)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecture recording</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Updated texts transformers</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>from PDF to Word</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing technical support was</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>provided by the Accessible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology Laboratories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increasing accessibility to web</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Mphi** AT services could facilitate environmental access for mobility-impaired students by means of digital circulation of colored maps, adapted lifts and vehicles for transportation inside and outside of campus. For learning participation, students stated a need for accessible web pages and ongoing technical support by the access Technology Lab. (Tables I, II).

Many of the disabled students indicated a need for specialized software and/or hardware to use a computer effectively. They also required alternative evaluation procedures for each disability type and extra training for each new version of programs to acclimatize themselves to AT familiarization (Tables I, II).

Students with visual and motor impairments have expressed the need for a specialised service that would be responsible for looking after, repairing as well as replacing AT technologies and equipment (i.e. seats, desks). Furthermore, for this service to be responsible for ensuring environmental accessibility through the removal of any potential obstacles.

VI. DISCUSSION – FUTURE WORK

Our research has highlighted the design and provision of customized services and assistive technology tools as a top priority need for the successful completion of studies by persons with disabilities. The needs encountered by disabled students for environmental and academic provisions and supported by AT are in agreement with the results of Srivilailuck, Beale, Murray, Kidd [17] (2005), who studied the available literature, and showed that blind and partially sighted students face problems in architectural design and learning environment. They therefore suggested collaboration between architects and electronic computer engineers in order to produce environmental and assistive technology adaptations capable of creating an effective learning environment.

We also observed limited use of the Internet by disabled students, although the majority of universities did provide this facility. Nevertheless, it is a phenomenon registered also in the results of Shapiro and Rohde [4] (2000) who found that generally, almost 60% of Americans with disabilities had never used a computer. Moreover Jackson [5] (2003) and Kaye [6] (2000) showed limited access and usage for many students with disabilities, probably due to a lack of access, funding and trained teachers. At this point, it should be mentioned that although a few students did exploit the facility for electronic communication with their professors by e-mail, the answers received to their questions were minimal. Although many students agree that e-learning and e-communication fulfill a basic academic need, they nonetheless suggest that personal communication with professors promotes learning and critical thinking, due to its immediate feedback and exchange of information. This argument is expressed in such phrases as: “the help of internet is important but without moderation its use can reduce or even completely replace the existence of any interpersonal communication”, “spending on services and AT is preferable to a financial scheme based on benefits”, and “I need human contact and respect for my cultural identity, not to communicate via the internet sitting on my sofa at home” (Mphi, 16; VI,16; HI, 24).

Out of 70 students with visual, hearing or motor impairments, 25 requested permission as well as the equipment necessary to record lectures. At the end of the lecture, or when deemed appropriate and in collaboration with a transcriber, the material will be reproduced in written form with explanations and remarks. For a number of visually and motor impaired students that cannot keep notes, the above-mentioned method is an integral part of their educational process.

---

2 Mphi, 16; VI,16; HI, 24, phrases reported in speech marks derive from the transcribed text of the sixteenth interviewee with motor impairment, of the sixteenth with visual impairment, and of the 24th with hearing impairment, respectively.
In the present investigation, deaf students stated that they avoided frequent use of Internet due to the fact that they do not know English (the majority of hardware and software is in English) and also because they find understanding the Greek language extremely difficult. From the interviews, it resulted that 22 out of 24 deaf students stated that interpretation was a basic need for efficacious teaching. The decision makers’ response to this request was negative, due to lack of funding resources and the reduced participation of deaf students in daily attendance. Literature states that this need is covered by ICTs with visual material (television, DVD, movies, images), accommodated real time captioning systems (RTC) and on-line interpretation procedures [9, 22, 11, 12, 13, 14]. For this reason, the development of specialized software or methods that will limit the difficulties that arise from the absence of a transcriber during lectures, extra tutoring as well as during involvement in university social and sport events. Hearing impaired students receive most of their information through visual channels. This is supported by the present research as 14 out of 24 subjects asked for academic manuscripts as well as picture rich texts.

Within the blind and partially sighted group, only the experienced and those actually studying in the field of ICTs frequently use the web. The same students, in spite of their familiarity with the Internet and web pages, declare that they spend a lot of time navigating, due to the complicated structure of sites and pages. These are viewpoints that concur with the conclusions of many investigators [16, 17, 18, 19] and for this reason Craven and Brophy [20] (2003) suggested extra training for each new version of a program. They furthermore declare that AT products like screen readers and enlargers, Braille printers and labels scanners etc., can support the full participation in learning of VI students, an attitude which was also indicated by study results by Gkouvatzi, Mantis and Costa [21] (2010). Three students with visual impairments expressed the need for a maintenance service responsible for the specialised equipment mentioned above. More specifically they mentioned that «six months after the acquisition of printers, enlargers, screen readers and such like, the equipment was out of order and remained like this for the next two years» (VI,2; VI,11)\(^3\).

In education, reasonable provisions should include alternative teaching methods and tools for the learning of shapes, mathematical symbols and maps. Very likely, further to the difficulties created by the lack of vision, this is an additional reason why blind students do not opt to attend studies in schools of physical sciences.

The need for accessible classes and the ongoing technical support provided by the Access Technology Lab, which is described by the MI students in the present investigation, has also been established by other researchers [9, 23, 24] who recommend a series of smart devices, systems and adapted furniture (desks, tables, chairs) in order to surmount specific obstacles.

Out of the 70 interviews, no requests were made by students for the improvement of the quality of written speech, contrary to the results of Mirenda, Turoldo and McAvoy [22] (2006) who proposed specific devices and procedures for its improvement. Also, out of the 70 students, only two were familiar with Distance Learning (DL) courses (one with HI and one with VI) out of a total of seven students who referred to DL teaching. This fact revealed how limited the information available on alternative method topics in the field of higher education and training is.

Eight students questioned have lobbied for inclusion in the committee responsible for the selection, assessment and provision of tools and supportive technologies for different categories of disability. Four of them stated that «we are part of human society, and constitute a cultural community with specific characteristics and we deserve the respect of the rest of the academic community», and «we should be included in decision making committees with our own representatives. Others should not be making decisions on our behalf».

Investigating the bibliography and a series of studies and statements of official state institutions and organizations as well as the websites of many universities, a significantly high number of methods and AT devices was recorded, which could support the environmental and academic access of disabled students. Comparing this data with the results of the research relative to needs in AT Services, as declared by disabled students, however, this data is not regularly updated.

After taking into consideration the limitations of different impairments, universities should systematically start working towards designing and creating appropriate internet sites. Through such sites, students should be able to access information regarding the following categories: the equal educational rights of students with SpLDs, the application of national law in higher education as well as in the market, the general principles of the university’s policy agenda, the curriculum, the provision on offer for the different types of impairment, the existing AT that will help them with their academic needs and aid them towards the successful completion of their studies. A well organised site can be a means to a better exchange of information, ideas and other useful practices amongst students with SpLDs within the same

---

\(^3\) VI,2; VI,11, phrases from the second and eleventh interviewees with visual impairments.
university as well as with other Greek and European universities.

Maybe future research that aims to investigate the reasons for this limited updating will lead to the planning and organization of a more efficacious pattern for the dissemination of specialized information.

VII. CONCLUSION

All seventy students that participated in this research agreed that ICT services are a priority they can support disabled students’ learning in tertiary education. At the same time, they can be a useful tool in identifying the strengths and weaknesses of ICTs accessibility on and off campus.

Studying the perceptions of disabled students allowed us to confirm that AT services are capable of creating an effective and accessible learning environment, thus improving students’ academic performance and providing economic and cultural benefits to individuals, and to society as a whole. Overall, there is no strong support provided to match the increasing expectations of students with disabilities and to help them to fulfill their academic potential.

For the development, management and efficient function of a specialised model of services and structures based on AT, the co-operation of field of disabilities along with financial and managerial services is essential. These groups will design the model and secure and manage the financial support. In order for this to function, the involvement of representatives with SpLDs in decision making processes is fundamental. In order to guarantee high quality provision for students with SpLDs, constant monitoring and evaluation of the effects of the different implementations (for accessibility, usability, and availability) is necessary.

It is thus clear that providing an environment with AT support services and academic facilities that meet the needs of all students would also create conditions necessary for the substantial equalization of deaf, blind and mobility impaired students with their non-disabled peers. These conditions would allow for an improvement in their academic profiles, thus creating more equal opportunities in the labor market and ultimately helping those with disabilities to claim larger share of the labor market, improving their quality of life.

REFERENCES


[16] R.A. Schiff, “Information Literacy and Blind and Visually Impaired Student”, Urban Library


Farfalla: a step toward an inclusive web
Andrea Mangiatordi¹, Riccardo Dondi²

¹Università degli Studi di Milano Bicocca, Centro QUA_SI, Italy, e-mail: a.mangiatordi@campus.unimib.it
²Università degli Studi di Bergamo, Italy e-mail: riccardo.dondi@unibg.it

Abstract—The concept of Architectural Barrier can be extended to the Web, in the sense that people with disability can experience problems in accessing pages which are designed for the able-bodied. Despite of this, many users with different kind of disabilities regularly access the Web using some kind of Assistive Technology. This means they can use a variety of resources and web 2.0 tools, but they need to be on their own computers, with their own accessibility software and configuration, in order to use them at best.

The Farfalla project is a web application which aims to allow users to set their accessibility preferences on a centralized website and then to apply them to every other website they will visit. This approach can work from any web browser, on (almost) any machine, even without administrative privileges: the only requirement is a javascript-enabled browser with internet connection.

In order to accomplish this, the Farfalla script only needs to be included in a target web page in the form of a small piece of HTML code. The attempt is to move the accessibility solutions from the user’s computer to the web itself; in this paper is explained why, if widespread, this practice could be the base for a truly inclusive web.

I. INTRODUCTION

The Farfalla project started in late 2006 as a Master Degree project. The story of its development is not linear mainly because of the influence that the author’s PhD thesis work had on the mindset on which Farfalla is based. Basically, it consists of a web framework which allows the use and the integration of some software assistive technology solutions in web pages, without the need of installing any additional piece of desktop software.

The aim of this project, besides the various forms it took during its development, has always been the same: to integrate typical Assistive Technology functionalities inside a web application, in order to provide disabled people with easier access to the information on the web. As long as the solutions implemented in this framework were tested on the field during this research process, this paper gives an overview of the project features and of the evolutions which have succeeded in its architecture. Actually, the software history can be split into two parts: an original one, when it was designed and a first prototype was developed, and a new one, where ideas emerged from fieldwork experiences were applied to a complete re-design of the software architecture.

This paper is organized as follows: section II explores the concept of accessibility, introducing our idea of “inclusive web” and presenting the state of the art for what concerns inclusive software solutions; section III exposes the foundations of the Farfalla project, describing the first prototype; in section IV and V the actual state of the software development is outlined, together with the technical specification about it. Section VI focuses on the evaluation of the project while section VII draws some conclusions about it.

II. A WORD ABOUT ACCESSIBILITY

The discourse on accessibility is often linked to the concepts of integration and inclusion. As explained in [11], [10] and [8] the two terms seem synonyms but have indeed a different meaning. The term inclusion is here preferred to integration because inclusion can be considered as an environmental trait, while integration is the result of an action from the people who live in it.

Web sites accessibility is traditionally measured in terms of adherence to standards, which are both the ones defined by the W3C¹ and the results of tests made with the most popular web browsers and assistive software solutions. This is indeed good under an ecological point of view: standards help websites’ developers by providing them with guidelines which grant quality to their work. But from the disabled person’s point of view, this also means adapting to a standard by purchasing the right software and by adapting to de facto standards in order to be able to access contents.

Our idea is that the ‘accessible web’ should gradually transform into an ‘inclusive web’: accessibility tools should be offered together with the contents, creating a truly accessible landscape in which users with different needs could move, being

¹ In 1999 the World Wide Web Consortium (W3C) published the Web Content Accessibility Guidelines WCAG 1.0. On 11 December 2008, the WAI released the WCAG 2.0, a recommendation which aims to be up to date and more technology neutral. The text of WCAG 2.0 is available at http://www.w3.org/TR/WCAG20/.
independent from the tools they use. In order to achieve that, websites’ developers and maintainers must be provided with an easier method to build accessible solutions and to implement them in their works.

One possible way to reach such a scenario is proposed in the rest of this paper: the Farfalla project is trying to simplify both the development and the implementation of web based accessibility tools, allowing websites to enhance their accessibility only by means of a ‘declaration of intents’. Websites that do not take it into consideration can still be made accessible on the user side, with few limitations.

A. State of the art

Some software with similar architecture and/or with functionalities which resemble the ones discussed in the following paragraphs already exists. This list of solution is not exhaustive: many other tools exist allowing accessibility enhancements and compatibility tests, the following described resources have been inspiring for this work. Among them there are ROKTalk\(^2\), the JISC TechDis Toolbar\(^3\), the ChromeVis plugin for Google Chrome\(^4\) and the AxsJAX library\(^5\) (see also [12], [3]).

ROKTalk is an automated text-to-speech tool which can selectively read the contents of a web page without any need of installing extra software.

The JISC TechDis Toolbar is a virtual toolbar which can load over any webpage and which allows some text effects (changes in colors and in font size), direct text-to-speech translation of selected areas, dictionary searches, spellchecking, automatic referencing. It is mainly intended as a support for studying and is also.

ChromeVis is a plugin for the Google Chrome web browser, providing selective magnification of text. It also provides keyboard shortcuts to change the selections moving from character to character, from word to word, from sentence to sentence and from paragraph to paragraph.

AxsJAX is a sort of meta-solution, which allows automatic correction of common errors in web pages and enhances compliance with W3C ARIA\(^6\). Those errors could cause problems to assistive software users, hence AxsJAX acts as a facilitator for them.

All the mentioned software has very interesting features, but all of them seem to be missing pluggability\(^7\), which is here considered fundamental. Farfalla intends to provide the developers with the possibility of writing new plugins to address the needs of people with different kinds of disabilities.

The available instruments seem to concentrate on blinds and users with low vision. Among them, the JISC TechDis Toolbar is particularly similar to Farfalla for its architecture and for the technologies on which it is based. The main difference between them is that the former allows the activation or deactivation of a variety of small tools but does not allow the creation of a user profile, which on the contrary is one of the base features of the latter.

III. THE ORIGINS OF THE PROJECT

As described in [6], the original idea of the project was to develop an interface allowing the users to write text documents and to save them using only a restricted number of keys. Later, this system turned out to be effective also for browsing web pages. This could improve the ability to retrieve information and to communicate, allowing users to access the multitude of services and applications that the web makes available.

After developing this assisted writing method, the project adopted an architecture based on a virtual web browser. It allowed moving from element to element in a web page and filling forms using only five keys. The aim of the virtual browser was to make direct interventions on the contents by actually copying them locally and displaying only a sort of mirrored version of them.

---

\(^{2}\) ROKTalk is a commercial solution by ROK Talk Ltd. The company website, http://www.roktalk.com/, offers a live demo of the high quality synthetic voice available.

\(^{3}\) The Joint Information Systems Committee (JISC) TechDis develops and maintain the project described at http://access.ecs.soton.ac.uk/ToolBar/

\(^{4}\) This extension of the Chrome browser allows some text magnification effects. It can be installed directly from http://code.google.com/p/google-axes-chrome/

\(^{5}\) The AxsJAX library was developed at Google Inc. and released as opensource code. The project is documented at http://code.google.com/p/google-axsjax/

\(^{6}\) The W3C WAI-ARIA, the Accessible Rich Internet Applications Suite, is a document providing guidelines for accessibility in rich web applications. It is available at: http://www.w3.org/WAI/intro/aria

\(^{7}\) In the case of JISC TechDis Toolbar, pluggability actually is an existing feature, but no guidelines for writing third-party plugins are available.
D. Text prediction

The text prediction module was used in combination with the scanning keyboard. In order to speed up the writing, Farfalla proposed word completions to the user. The possibility that one particular word would be suggested was determined by a score, based on both frequency and recency indexes. Taking account of the recency allowed a sort of context-sensible text prediction.

E. Text magnification

This module allowed the magnification of the text selected using the above mentioned scanning method. The enlarged version of fragments of text appeared in the upper part of the screen, using a white-on-black color combination. Links were highlighted using a different color. One problem with this approach was the fact that if the selection was very large, the magnified version covered the whole screen.

F. Text-to-speech (TTS)

The same text magnified with the previous module could also be spoken by a voice synthesizer. This module relied on the Festival TTS engine⁸, which was run on the same server that hosted Farfalla. The text was processed by Festival and then converted into an mp3 file, allowing it to be played from the browser by a small flash application.

IV. ACTUAL STATE OF THE PROJECT

The model explained above had one major flaw: it was designed as an application which was meant to replace another application. This has its obvious advantages but is also very expensive in terms of resources, both on the development side and on the user side.

The reflection about “ubiquitous accessibility” exposed in section II, lead to completely re-think and re-design the structure of this software.

The web is complex and ever-changing, with emerging technologies which have the potential to change the way websites are built in a relatively short amount of time. What remains the same is the base: HTML is the language of the web since its beginning and CSS has become a standard for layouts. Javascript is also a technology which has been part of the web development for years and is currently something very used for building web pages that are rich in functionalities but do not

---

⁸ The Festival Speech Synthesis System is a general framework for building speech-capable systems. It is developed and maintained by the Centre for Speech Technology Research of the University of Edinburgh. All the information about the project is available at: http://www.cstr.ed.ac.uk/projects/festival/
require plugins or other software in order to be viewable (as in the case of Adobe Flash).

The idea is hence to start from the HTML, using various techniques to inject a small code fragment into a web page. This fragment is a `<script>` tag pointing to a Javascript file and can be placed in the header of the page. The script relies on a remote database (communicating with it through AJAX and JSON techniques) to recall the user preferences and some plugins. Those plugins implement single functionalities, as the already mentioned virtual keyboard or text magnification, and can be combined in order to grant the user with the best experience, according to his/her needs.

In this case there is no need of replicating the browser functionalities: the ‘real’ browser does the browsing, while Farfalla only includes small stand-alone applications to the user’s request. Those applications work on the client side, hence there is no need of a huge amount of resources on the server side: the actual load is distributed on the computers of the final users.

V. TECHNICAL INFORMATION

The software architecture is composed by a backend and a frontend interface. The backend is a database where the profiles of the users are stored. Profiles are simply made of the list of plugins that each user needs, together with plugin-specific preferences. The frontend is visually composed by a toolbar, generated by the Javascript file included into web pages, which allows the user’s profile selection and activates the plugins.

![Diagram](image1.png)

Figure 2 - The scheme of the new architecture adopted

The backend is based on a MySQL database and has a web based graphical interface written using the CakePHP framework. It communicates with the frontend by producing a JSON output, including the list of plugins and options required by the logged in user. At the time of writing the backend is not yet publicly available: the demo available on the project’s website is based on a simpler script which simulates its responses. The frontend is mainly based on jQuery and jQuery-UI\(^9\), two Javascript libraries which grant developers with the possibility of abstracting from the web browser, by actually creating a compatibility layer between different rendering engines. Farfalla plugins are also based on the same technologies, even if by the means of AJAX techniques it is virtually possible to include almost any web-based application in the Farfalla plugin list.

The current plugin list includes:

1. a text-magnification system, which renders the elements under the mouse in a dedicated part of the screen, enlarging the text and setting its color to white on a black background, in order to make it easily readable (figure 3). Links are highlighted in different color and are assigned a number, which allows the user to visit them by pressing the corresponding key on the keyboard;

2. a virtual keyboard, which appears when the user clicks on ‘input’ or ‘textarea’ HTML elements (figure 4).

![Magnification Plugin](image2.png)

Figure 3 - The magnification plugin in action

![Virtual Keyboard](image3.png)

Figure 4 - An example of virtual keyboard

These plugins are based on a total rewriting of the modules which were present in the first prototype. The use of jQuery and jQuery-UI simplified a lot the

---

\(^9\) CakePHP is a framework for rapid development of web applications. The project website, [http://cakephp.org](http://cakephp.org), hosts the source code and the documentation.

\(^10\) jQuery is a javascript library for cross browser compatibility, while jQuery-UI is an extension of it, providing tools for easy development of web based graphical user interfaces. More information, including source code, live examples and documentation are available respectively at [http://jquery.com](http://jquery.com) and [http://jqueryui.com](http://jqueryui.com).
writing of the code, allowing more scalable and cross browser results.

Pluggability is a key feature in Farfalla: every functionality is offered by a piece of code which is specialized in offering only one particular service. Future development phases will include the publication of standard rules for building third party plugins, together with APIs for external services.

G. HTML injection methods

There are several techniques which allow to insert a piece of HTML code in a web page. Among them, three are already available:

1. manual insertion: the easiest and cheapest way of adding a piece of code to a web page depends on the webmaster work. Hence the Farfalla project primarily asks webmasters to adopt its solutions by adding a link to a Javascript library in the header of their web pages. The advantage for using this technique is that users are provided with Farfalla’s functionalities from every web browser they happen to have access to, without the need of installing anything.

2. bookmarklet: a bookmarklet is a piece of javascript code saved as a bookmark in a web browser. It allows a client-side code injection and installing it is as easy as adding a bookmark to a browser. The downside of this technique is that it is only useful for the page which the user is viewing when he or she activates the bookmarklet. After visiting another page, the bookmarklet effect vanishes and needs to be reactivated. The stable version of the bookmarklet is available at this URL: http://farfalla-project.org/bookmarklet/

3. Userscript: userscripts11 are Javascript files which can be semi-automatically installed in a web browser. Among the major web browsers, userscripts can be interpreted natively by Google Chrome, while Mozilla Firefox, Apple Safari and Microsoft Internet Explorer need a plugin application (respectively GreaseMonkey, GreaseKit and Greasemonkey for IE12) in order to apply them to web pages. The major advantage of this approach is that, once installed, the userscript is automatically run on any web page (or on any web page whose URL matches a regular expression). The user still has the possibility to exclude some pages or entire domains, which is useful in case of conflicts with other scripts. The major disadvantage of this technique is that the user depends on one particular browser. The bookmarklet technique is still available in case of necessity, as a practical fallback for this technique. The stable version of the userscript is available at this URL: http://farfalla-project.org/farfalla.user.js

There are some other ways to inject a fragment of code which have not been implemented yet. Among those, the following two seem to be the most interesting:

1. CMS plugins: many websites are built using Content Management Systems, such as Joomla13, Wordpress14 and Drupal15. Those softwares, as many others, allow expanding their functionalities by the installation of plugins. Those add-ons can be very complex and radically change the purpose of the original CMS system, but they can also be very simple. One of the mayor advantages in using a CMS for building a website is the possibility to completely manage it without knowing HTML. The idea, in this case, is to create a plugin for each of them which will allow webmasters to include the Farfalla script in their pages also without knowing HTML. This is a variant of the ‘manual insertion’ method described above, which is quite easy to develop and which an reach a large number of users.

2. Proxy server: proxy servers allow filtering the web traffic from a client to the rest of the internet. This can be done for security or network optimization purposes. An http proxy, such as Squid16, which is completely opensource, can also be extended in order to add some interesting functionalities. By the means of the Internet Content Adaptation Protocol (ICAP), which can be used with Squid or implemented in a brand new proxy, the Farfalla script string could be added, for example, to all the computers of a LAN. This would allow institutions like universities, schools and actually any network using a common gateway to provide all the local users with the possibility of using the resources offered by Farfalla.

These two solutions will be developed after the framework core software will be released as stable. The source code of the whole Farfalla project is open and accessible from the address http://gitorious.org/farfalla/, while the

---

11 The largest collection of userscripts is provided by the community site http://userscripts.org/. The website allows the direct installation of each script, together with documentation and with spaces for discussion.


13 http://www.joomla.org/

14 http://www.wordpress.org/

15 http://www.drupal.org/

16 http://www.squid-cache.org/
VI. PROJECT EVALUATION

This project opens up some interesting research possibilities, both in the technical and socio-pedagogical area. The effects of a regular use of this solution could be measured in both the autonomy of the users in browsing the Web and in the effects of this process on their social interactions with other people. The objective of such research should be measuring the efficiency of this particular tool and collecting data about needs in order to foster the development of new functionalities.

Interesting indicators of autonomy could be analyzed through the usage of longitudinal tests in quasi-experimental settings. Reduced experimental groups are here preferred to large-scale ones: what really matters is not the creation of a perfect instrument for a unique purpose, but of a set of tools capable of adapting to the widest possible range of needs. The research should hence try to involve people with different needs and the project should adapt to them, in a grounded perspective.

Project evaluation will adopt both quantitative and qualitative techniques. First, a method for univoque and objective evaluation of the ‘inclusivity score’ of a web site will be defined. This score will be based on a battery of different accessibility tests, involving Expert Review, User Testing, Subjective Evaluation and Barrier Walkthrough (see [2], [12], [4] and [6]). This part of the evaluation will provide a means to measure the compatibility of a sample of web sites with the mayor types of impairment. The same web sites will be tested back with the same methodology after applying the solutions made available by the Farfalla framework.

As a second step in the evaluation process, it will also be interesting to report the personal achievements of some people with disability, by interviewing them (following a semi-structured track) before and after a test period of web browsing with this aid. This user feedback could be used as a basis for subsequent work, together with Sociometric techniques. Those will be adopted to measure the quantity and intensity of the relations that those people will be able to develop and maintain in their immediate social context thanks to the technological aid.

VII. CONCLUSIONS

The present version of the Farfalla software design puts flexibility and accessibility together. In an inclusive perspective, accessibility solutions must be part of the environment itself. In the case of the Web this ecological perspective imposes that accessibility tools are offered by web pages, instead of being needed by them.

The injection techniques described in this article are in fact enabling techniques. Among them, the first, human based one is to be preferred: the main idea behind this architecture is that everybody can participate in transforming the Web into an accessible environment by adopting some simple measures.

If accessibility shifts from being a solution to being an environmental trait, the concept of disability itself can be partially or completely reconsidered. Moreover, if a change in accessibility practices involving inclusive solutions starts to spread, many people and categories who are not considered as ‘disabled’ or ‘in need of assistive technology’ would benefit of the increased accessibility of web pages, in the same way every pedestrian can benefit of a curb cut.

REFERENCES

[6] T. Lang, Comparing website accessibility evaluation methods and learnings from usability evaluation methods, Peak Usability, December 2003
in *Proceedings of the Canada International Conference on Education*, Toronto, Canada, May 2010


INTERACTION between FAMILIES and FAMILY SCHOOL RELATIONSHIP: the IMPACT of INFORMATION and COMMUNICATION TECHNOLOGIES on SPECIAL EDUCATION NEEDS

Carla Silveira Maia & Ana Margarida Pisco Almeida

Universidade de Aveiro, Departamento de Didáctica e Tecnologia
Educativa e Departamento de Comunicação e Arte, Aveiro, Portugal
Phone: +351 234 370 200, Fax: +351 234370985, e-mail: csmaia1@gmail.com, marga@ua.pt

Abstract: This study aimed to analyze the impact of Information and Communication Technologies (ICT), namely of a Web 2.0 tool, the blog, on the interaction between families of Students with Special Education Needs (SEN) and on their relationship with the educational context. Under this aim, a qualitative case study was driven over a group constituted by 7 relatives of students with SEN derived from limitations associated with a diagnosis of dyslexia. In the analysis of interactions occurred in the blog, interventions of participants external to the study were also covered, making a total of 44 participants. The results - based on the analysis of data from two questionnaire surveys, applied at the blog’s pre and post-interaction moments, and on the interactions content analysis – allowed to conclude that the ICT use promotes the interaction between these families and that this, in turn, seems to support a closer relationships between families and school.

Keywords: ICT; Web 2.0; Interaction between families; Family-school relationship; SEN.

I. INTRODUCTION

Continuously challenged by the diverse difficulties underlying the special condition of their children, families of students with Special Education Needs (SEN) still lack support and, therefore, still represent a major concern for the educational system.

These families social isolation and the benefits of parental involvement in the school context have been target of research in the education field. Although not innovative issues, these aspects still assume relevance in the educational debate and a novelty nature in our study approach for the adoption of Information and Communication Technologies (ICT) as supportive tools in these issues – an area that is struggling with a noticeable lack of studies. The need to spread and deepen studies in this field is particularly accentuated on the understanding of the potential of online support groups for parents of pupils with SEN. Therefore, the analysis of the impact of the available internet tools on the interaction between these families becomes relevant as well as its effects on the family-school relationship.

The recognized influence of the family on emotional and cognitive development of individuals [1] - particularly of the families of students with SEN - is naturally and undeniably reflected on the key role it plays on the academic performance of their children [2] [3]. With regard to children with SEN, literature has been growingly documenting that the efficacy of psycho-educational intervention requires family involvement [4] In addition, given the amplified stress and difficulties experienced by these families [5] – and its resulting feelings of anxiety, depression, disorientation, loneliness and hopelessness [6] – working with the family and encouraging their involvement is crucial. From this standpoint, in order to enable the extension and improvement of these families participation, the search for alternatives demands urgency.

Along with these evidences, ICT - particularly the social software within the Web 2.0 – emerges as an excellent support for school and families’ responses to the growing education needs, strengthening the relationship between families and between family and school. Given their potential in promoting the gathering and the simultaneous and immediate contact between people who share interests and concerns, regardless of time, geographic and even social distances, and also for providing easy and quick creation and editing interaction spaces without costs, these tools can create favorable conditions for improving these families situation. By opening up a range of opportunities for communication and interconnection, they may unprecedentedly allow the overcome of these families’ lack of support and assistance.

Considering the social interaction potential of blogs, this tool may promote the development of a network/ community that can become an important complement to formal treatment. Since parents are in a unique position to help each other [7], it is essential to enhance reciprocal emotional, social and practical support that has been proven as critical for the families of children with SEN [8]. More accessible and available, this sort of support [9] provides a feeling of acceptance, belonging and understanding, beyond the natural access to information and experiences sharing.
Within the main benefits of these experiences, literature stresses the reduction of anxiety and depression and an increased feeling of well-being [6] as a result of a growing sense of control and empowerment [10].

As the family welfare and the emotional and practical support are linked to an enhanced action power, this interaction between families enabled by the blog may, in turn, promote greater involvement in the school context.

Based on the strong communicative and collective nature of Web 2.0, its tools should be explored in this context. Therefore it is fundamental to understand how and to what extent they can encourage such practices.

From this standpoint, this study aimed to explore the potential of the existing technological tools – namely, the Web 2.0 - on the interaction between families of students with SEN and of these experiences over the relationship of these families with the educational context.

Accordingly, our research questions were:
1. To what extent does ICT trigger changes in the interaction between families of students with SEN?
2. Does this interaction between families promote a closer relationship between family and school?

II. METHOD

Aiming to analyze the impact of ICT on the interaction between families and on the family-school relationship, we developed a qualitative case study [11] supported on the creation of a blog for parents of students with dyslexia. This blog intended to be a virtual space for experience and information sharing between families who live similar problems.

In order to analyse the blog’s impact on the referred issues, we considered all the written interventions registered in the blog along a 2 months period and the data collected from 2 questionnaire surveys – answered by the recruited participants group before and after the interaction in the blog - on the habits of interaction with families of students with SEN and on their relationship with school.

A. Participants

The analysis unit of this study was mainly focused on a participants group composed by 7 relatives of students with dyslexia. Nevertheless, as a result of the blog’s open nature, new participants were brought into the network in the period considered for analysis. Their expressive adhesion and interventions relevance led to their integration in the blog’s content analysis, despite it was initially planned to be restricted to the 7 participants group. Thus, in the analysis of interactions occurred in the blog, a total of 44 participants were considered - integrating both the 7 study participants group and the 37 participants external to the study.

Similarly, this expanded group was also mostly constituted by parents of students with dyslexia.

As the enlarged group wasn’t intentionally recruited for the study, their demographic data was not accessible. Thus, the following description is confined to the 7 participants group recruited for the study considering as inclusion criteria to be in charge of the education of a student with dyslexia attending regular school and as exclusion criteria an insufficient Internet access.

The group of 7 participants was recruited through schools and institutions and included relatives of students of both genders with a varied severity of dyslexia, attending regular school between the 1st grade and the secondary education and aged between eight and nineteen years old. These relatives were all female, with various educational backgrounds and occupations and an average age of 43.83 years. Regarding the geographical data, the group included participants from different districts of the country - Oporto, Lisbon and Algarve.

B. Data collection and analysis

The data was collected from the questionnaire surveys answered by the 7 participants’ original group in the pre and post-blog moments and from the content of the 44 participants’ written contributions in the blog. The main purpose was to capture perceptions and opinions on the topics under investigation and the two months interaction period with other parents of students with dyslexia enabled by the blog impact on them.

Regarding the questionnaire surveys, these were mainly composed by multiple choice questions and aimed to gather information on patterns of interaction between families of students with dyslexia and on the nature of their relationship with the school. These questionnaire surveys were developed in three phases: information gathering on relevant topics; pre-validation test consisting on its application to an expert panel; application to the participants. Regarding the interaction developed, participants’ written contributions registered in the blog were analyzed. Publications from all participants in the blog were considered.

As analysis unit we considered the sentence, being its results organized into the following analysis parameters: (1) Identity, geographic origin and interventions’ schedules; (2) Theme diversity; (3) Interpersonal communication.

Thus, after handling with general aspects such as publications quantity - posts and messages -, the analysis was centered on the participants’ geographical data, their interventions schedule and their option for revealing or hiding identity. Such information intended to provide a better understanding of the relevance of the flexibility provided by this technological tool on the interaction promotion. For the thematic content analysis, themes or topics developed in the
publications were identified and quantified. Concerning interpersonal communication, based on indicators adapted from Philips (2000) [12] and Rourke et al. (1999) [13] scales and also on the literature regarding self-help groups, we analyzed aspects which characterize the degree of the participants’ social presence in the community, such as interventions/sentences expressing affection, cohesion and/ or interaction.

III. RESULTS

The main results, following presented, derived from the analysis of responses to the questionnaire surveys answered by the 7 initial participants and from the content analysis of interactions between the 44 participants registered in the blog, being respectively organized into two sections: (1) interaction between families and family-school relationship in the pre and post-blog moments (2) interaction between families occurred in the blog.

A. Interaction between Families and Family-school Relationship in the pre and post-blog moments

In order to understand whether there were relevant changes on the topics considered, from the pre to the post interaction periods, we proceeded to a comparative analysis of the responses to the two questionnaire surveys. In short, we intended to study how and to what extent the interaction enabled by the blog influenced the habits of interaction between families and their relationship with school.

![Fig. 1. Interaction frequency between parents of students with dyslexia (left); Number of families of students with dyslexia with whom the participants interact (middle); Addressing frequency of school topics in the contacts (right).](image)

The data in Figure 1 illustrates a considerable change in the interaction habits between parents of students with dyslexia after the interaction period mediated by the blog, evidencing a noteworthy increase in their frequency and size as well as in the addressing of school topics.

What Internet tool(s) do you use to contact other families?

![What Internet tool(s) do you use to contact other families?](image)

What would be/ was the main advantage of communicating by Internet?

![What would be/ was the main advantage of communicating by Internet?](image)

**Fig. 2. Situations in which interaction with other families of students with dyslexia occur (left); Most valued reasons for using the Internet to communicate (middle); Internet tools used in the contacts with other families (right).**

Figure 2 demonstrates that the pre-blog predominance of casual encounters in person was, on the post-blog period, replaced by intentional meetings through the "Internet". With reference to the advantages perceived on the use of Internet to communicate, participants reported the possibility of accessing anytime and anywhere in both of the periods. Only with regard to the post-blog moment, the online tool used for these contacts was, besides the blog, the "E-mail".

On a 0-5 scale, indicate your satisfaction level towards the following aspects (0 corresponding to “completely unsatisfied” and 5 to “completely satisfied”):

![School support](image)

![Access to information on education and support available](image)
Regarding the satisfaction with social and personal aspects related to the dyslexia problem, Figure 3 shows an average decrease of 0.28 in the satisfaction with the "school support", an average increase of 1.29 in the satisfaction with the "access to information on education and support available" and an average increase of 0.41 in the satisfaction with the "ability to cope with the child's problem".

Concerning the exchange of methods and strategies with the school:

The data illustrated in Figure 4 shows a decrease in the perception of "support provided by the school" to the student with dyslexia and an increased perception of "difficulties in keeping informed about their children's school daily life".

Regarding the perception on the "importance of acknowledging school daily life to better support their children", there was an average increase of 0.14 from the pre to the post-blog interaction moments (Figure 5).

Figure 6 demonstrates a slight increase in the parent’s attendance frequency to "school events (e.g. festivals, sporting events, etc.)" and to "school parents meetings".

Figure 7 refers to school and family information exchange concerning "methods and strategies that can help to improve their children academic performance or to cope with practical problems". The results show a clear increase in this sharing between the two contexts.

B. Interaction between Families Occurred in the Blog

The data collected from the blog created for the study purposes comprised a total of 177 publications from 44 participants. Although the publication average was 3.88 per participant, its distribution was not uniform, ranging from 1 to 34 publications per participant. The blog’s participants group included 7 participants (15.9%) belonging to the initial study group and 37 (84.1%) composed by participants who came into the network in the period considered for analysis – March and April, 2009.
Considering the short period of time and the little divulgation process – comments posted on online forums for parents informing about the blog – this number of participants was considered a success and, therefore, the previous intention of focusing the analyses on the 7 participants initially recruited was extended to the entire number of blog participants.

With regard to the participants’ geographical data, the distribution of the 44 participants reveals their spreading. The data collected registered 2 participants from the district of Braga, 9 from Oporto, 4 from Lisbon, 1 from Setubal, 2 from the Algarve, 1 from the Azores, 21 from unknown districts and 1 resident abroad.

The publications were also spread along a large time spectrum, ranging from 0:10 a.m. to 11:59 p.m. The greater participation frequency was registered between 9 p.m. and 11 p.m.

Concerning the 7 participants group, and as in this case it was possible to discern the use of true identities from nicknames, the data shows that 4 of them revealed their "true identity" and 3 opted for the use of a "nickname". In relation to those participants not included in the initial group, 24 identified themselves (65%) - although in these cases it is not possible to distinguish between the use of true identities and nicknames - and 13 participated anonymously (35%).

The expressions comprised in the "affection" category registered 335 occurrences along the 177 publications. Within this parameter, the category "Greetings" registered 152 occurrences, the expressions demonstrating "emotions/ state of mind" 88 and the "encouragement/ support" expression 47.

Regarding the interventions evidencing the interaction developed between the community members, a total of 286 occurrences along the 177 publications were registered. Within this category, the "discussion of issues already started by others" was the most frequent, with 113 occurrences (41%).

Related to cohesion, there were 476 expressions which refer to the community feeling. Among them, there were 152 occurrences of "vocatives" (31%) and 113 occurrences of “inclusive verbs" (24%) - Figure 9.

About the themes addressed along the comments (Figure 8), there was a clear predominance of the theme “Strategies and Information”. The offer or request of information or strategies regarding the issue of dyslexia registered 321 occurrences (33.9%) in a total of 946 thematic references.

Other themes considerably addressed were the "external support (provided by professionals external to school)” with 145 references (15.3%), the "special education services and policies" and the "parents attitudes and reactions" towards the problem with 123 (13%) and 120 references (12.7 %), respectively.

IV. DISCUSSION

The analysis of the data collected from the questionnaire surveys applied and from the publications registered in the blog suggest that in this participants group the blog has played a relevant role on the development of the interaction between families and of the family-school relationship.

With regard to the blog’s impact on the interaction habits between families of students with dyslexia, the considerable increase in the frequency and size of these contacts corroborates the documented Internet power in reaching further people, and multiplying interaction opportunities between them [14].
In agreement with literature statements, time and space flexibility were identified as the main advantages of online interaction. Accordingly, both the questionnaire surveys and the geographic and time dispersion verified on the participants' interventions in the blog seem to explain this interaction enhancement [15]. Apart from revealing the advantages of Internet on supporting individuals whose schedules or geographic circumstances do not allow the attendance of group meetings in presence, these data may also indicate a development of comfort feelings resulting from perception of permanent support availability - unlikely the formal support situation [9].

At the same time, it is also important to recall that the prevalence of casual contacts on the pre-blog period has been replaced by intentional contacts on the post-blog period - by the Internet. This may suggest that the pre-existing desire to interact was previously conditioned by a lack of opportunities, and that, consequently, this tool has enabled its accomplishment as it is demonstrated on the following participant's intervention: "I've asked myself hundreds of times, where are the dyslexic children parents".

In addition, the possibility of providing support anonymously seemed to be a determining factor for the development of these interactions, since anonymity/nicknames use was substantial. In this regard, it is also pertinent to stress that after contacting through the blog some participants began to meet in person and others to establish a more private contact by e-mail. This evidence contradicts previous data which pointed the lack of physical proximity as a negative aspect of online support [16]. Rather, and similarly to Baum (2004) [17] study's conclusions, it may evidence that the Internet use in this context is not a disadvantage, corresponding either to some participants need to remain anonymous, either catalyzing or facilitating meetings in physical context.

Besides, and also according to the literature [6], the interaction expansion from the blog to the E-mail may suggest the importance of the costless nature of the media used as, for example, the phone was not mentioned in this sense. Furthermore, as participants pointed beyond the blog, the e-mail as the online media of choice, we can assume the importance of tool's usability in promoting interaction, since the use of e-mail - instead, for example, of snn and forums - can be related to the possible familiarity of the participants' age group with this tool.

Regarding family-school relationship, the data that reports an increase of "addressing school issues in the contacts with other parents", of the attendance frequency to "school events" and "school parent meetings" and of the "methods and strategies sharing between school and families" suggest a more active parental involvement. On the other hand, data that evidences a decrease in the perception of the "school support" after the interaction in the blog and the greater perception of difficulties in "keeping informed about the child's school daily life" seem to suggest the opposite. However, if we analyze these data altogether with the increase in the "access to information regarding educational aspects" and in the "ability to cope with the problem", this dissatisfaction towards the "school support" can be explained by a more demanding positioning of parents - which means a greater ability to identify school performance deficiencies - resulting from the better access to information provided by the network. From this point of view, the satisfaction decrease may not be a negative aspect but a sign of greater parental involvement. Still, the increase of "perceived difficulties in keeping informed about the child's school daily life", if considered contiguously to the increase of "awareness about the importance of keeping informed about the child's school daily life to better support him/her", can suggest it has arisen from an attention increase in this regard.

With reference to the results reported in the publications' thematic analysis, beyond evidencing a huge diversity of topics discussed, these show a clear predominance of the "strategies and information sharing" thematic category. The higher addressing of this theme corresponds to the initial study participants' expectations and to theories which stress the importance of mutual peer support as a source of relevant and practical learning and of access to useful information [6]. This pragmatism and utility is obvious to the participants, as the following participant intervention shows: "I would like to tell D. that I've already read and printed your information and I'm already preparing to put it into practice".

Regarding the social presence level of the participants in the community, a high frequency of expressions referring to "interaction", "cohesion" and "affection" was found. This confirms that the blog effectively represented a favorable media to the interaction, to the information sharing and to the establishment of ties between the members of the community formed, as it is evident in the following participants' interventions: "Let's get together' and '...as you reported your experience with your child, it seemed I was watching the movie of my days ... I can't express how much I understand and share your feelings".

In short, although the experience reduced period of time, the results suggest a clear influence of the blog on the interaction between families and also evidence its possible influence on their relationship with school.

V. CONCLUSIONS

In this case, the blog has shown to be an important aid in providing and promoting interaction between families which experience similar circumstances and problems. It played an important role on the social isolation reduction and on the promotion of dialogue
on general issues related to key factors of academic performance and development of children with SEN.

Accordingly, and based on authors such as Baum (2004) [17] and Huws, Jones and Ingledew (2001) [18], we can reiterate that these support groups mediated by the Internet can be useful in addition or even replacement of the traditional groups. Comparatively to these groups which meet in presence, these online support groups evidence advantages mainly because of their potential in reaching a vast amount of people, of the time and space flexibility they offer and of the possibility of providing support anonymously.

The emotional, social and practical support claimed in the literature as key benefits of this type of online support networks, was in this study embodied by the satisfaction and well-being evidenced by the participants and the empowerment they revealed in coping with daily adversities.

We also believe that this beneficial contact between families of students with SEN may have promoted greater awareness of the parental role in the educational process and as such, led to an optimization of the participation of the families in the school context.

Thus, and considering as main study limitations the reduced number of participants and the short period of time considered for analysis, we consider it would be important to develop and deepen studies which could comprise a larger number of participants for a wider time span. We believe it would be a great contribute to a more comprehensive and complete analysis of peer support benefits and would allow confirming or refuting the data which suggest the closer family-school relationship induced by the interaction with other families in the blog. The impact of this kind of experience on the educational support that family is able to provide to the student and on the well-being and academic performance of student with SEN himself, would also be an important object of study.

VI. REFERENCES


Elderly Speech Collection for Speech Recognition Based on Crowd Sourcing

Ana Júdice¹, João Freitas¹, Daniela Braga¹,², António Calado¹, Miguel Dias¹,², António Teixeira³, Catarina Oliveira⁴

¹Microsoft Language Development Center, Portugal
²ADETTI – ISCTE, IUL, Lisbon, Portugal
³Dep. Electrónica, Telec & Informática (DETI)/IEETA - Universidade de Aveiro,
⁴Catarina Oliveira, Escola Superior de Saúde - Universidade de Aveiro,

{a-anju, i-joaof, i-dbraga, i-antonc, Miguel.Dias}@microsoft.com, ajst@ua.pt, coliveira@ua.pt

Abstract: In this paper, a preliminary experiment on elderly speech data collection based on a crowd-sourcing approach is described. From this experiment, a usability study of the web platform used for this collection was conducted and a few changes were made in order to meet senior users’ needs. The main goal of this work is to collect as much speech as possible from senior citizens in order to improve the current speech recognition accuracy rates, which does not perform well with speech uttered by the elderly. This work is under the scope of a larger project focusing on the development of technologies and services to support healthy, productive and active citizens, using multimodal interfaces and particularly embedded speech technology.

I. INTRODUCTION

Technology applications are often difficult to use for elderly people, either because they are not adapted to elderly’s specific needs, or because they have unfriendly interfaces in most cases, which contribute to their social isolation and exclusion. In terms of usability, it has been demonstrated that speech is the easiest and most natural way for human-machine interaction [1] and that this type of interface is often preferred, especially when mobility is reduced or the screen is too small for a touch interface.

Although speech interfaces are becoming more popular and covering more languages, the ASR (Automatic Speech Recognition) systems are still not adapted to deal with the voice of all population segments, specifically the ageing population, because: a) ASR acoustic models are often not trained with enough elderly speech, and b) senior citizens have not been seen as a relevant market sector for ICT related applications.

However, the growing senior population in Europe is worrying the European Commission, who is putting a great effort in funding Assistive Technology and Ambient Assisted Living solutions for ageing well and independent living. In a scenario of fast population ageing like the one we are facing, European policies are encouraging seniors to go online and to use technology in order to improve their quality of life and reduce their dependency on caregivers and social institutions. In this scenario, speech and multimodal interfaces must be available and usable by all, especially by the senior population.

Our voice characteristics change as we get older and that has a significant impact in speech technology performance. Current Speech Recognition systems are trained with hundreds of hours of adults’ speech, but they often don’t include elderly speech, which means the acoustic models are not prepared to recognize elderly speech voice. Recognizing elderly speech is even more challenging in certain languages (like Portuguese), due to the lack of speech training data. In section II, a discussion of what is elderly speech and a characterization of the senior population in Portugal is presented. In section II, it is also debated the lack of research and development on speech technology for elderly users in contrast with the growing need for this type of technology, especially embedded in Assistive Technology solutions (also described in section II) when we consider the fast growth of senior population.

It is well known that ASR systems based on statistical methods require vast amounts of transcribed speech data in order to achieve acceptable accuracy rates. Acquiring a lot of speech data is particularly difficult when addressing less-resourced languages or even any other language that is not amongst the big five (English, Spanish, French, German and Italian) in terms of market economic relevance. The main reason for this is that these corpora are expensive and recruiting speakers has proven to be quite costly and hard to manage. Besides, some speech databases lack quality because of the bad recording conditions, sample rates inconsistency, erroneous, inconsistent or inexistent transcription, etc.[2]. This paper describes
an ongoing work to tackle this issue by using a crowdsourcing approach (explained in Section III).

Besides speech technology, design of web interfaces for elderly people was another motivation for this work. When preparing a platform to collect elderly speech to improve speech recognition rates, we found that the web interface we had used in previous projects had to be adapted to this population segment. In sections IV and V, that experience is reported and discussed. In section VI, we describe the changes we made to the initial web platform in order to adapt it to the seniors’ users.

II. STATE OF THE ART AND BACKGROUND INFORMATION

A. What is Elderly Speech?

Until now, many disciplines, such as Psychology, Social Psychology, Sociology and Gerontology, have dealt with the subject of elderly people in various ways. The topic of elderly speech has established itself as an interdisciplinary research area with diverse approaches of investigation, which can be proved by the disparity of studies and methodologies, which are, in most of the cases, extremely difficult to compare.

Literature draws a divergent picture about how to characterize elderly speech. While some propose criteria to distinguish between elderly versus teenagers’ or adults’ speech, there are others denying that there are clear-cut differences [3]. The absence of a single deterministic phonetic cue, existent, for example, in gender determination, makes elderly speech classification inexact. Since aging increases the difference between biological age and chronological age and considering that biological aging can be influenced by factors such as, abuse or overuse of the vocal folds, smoking, alcohol consumption, psychological stress/tension, or frequent loud/shouted speech production without vocal training [4][5] it is not possible to determine an exact age limit for speech to be considered as elderly.

Conducted studies consider ages between 60 and 70 as the minimum age for the elderly age group [6]. Fiehler [7] follows the option that what is hastily called to be ‘typical for the speech of elderly people’ results from different situational circumstances, from which different registers are drawn, being this with respect to lexical and grammatical aspects. Though in contrast to teenagers’ speech, which is often used as a social identifying characteristic, seniors do not look for acceptation of a group in the same way, because of their experience of life, their (acquired) social status, and at the same time, because they are not necessarily dependent of a linguistic assignment from a peer-group.

Observations considering the voice of elderly people have proved that it is possible to state differences between elderly speech and teenagers or adults speech on an acoustic phonetic level [8]. With increasing age there is a deprivation of chest voice, general changes in frequencies, in the voice quality and the timbres. Changes in the heights of vowel formant frequencies particularly appear in older men, not only for biological reasons, but also because of social changes.

According with [8], [9], the speech rate is also slower. Simultaneously more breaks, more speech errors and a lower volume of speech were detectable. American studies [10] also conclude that elderly subjects in general produce fewer morphemes per utterance as well as fewer utterances per minute. Additionally, they assume that the subjects, while aging, eliminate more often compulsory grammatical morphemes as well as articles and possessive pronouns. Many studies agree that utterances get overall shorter with increased age, that seniors produce less correct verb tenses and also other correct morphological forms. It was also reported in [10] an age-dependent reduction of complex syntactic structures concerning written language.

B. Automatic Speech Recognition of Elderly Speech

Age is a key physiological characteristic of a speaker that must be considered in human-computer interfaces (HCIs) based on speech [12]. Although being a stable characteristic when compared with the awareness and emotional state of a speaker, age influences the performance of a SR engine, as several parameters of the speech wave form are modified, such as fundamental frequency, jitter, shimmer and harmonic noise ratio [13]. Additionally, with age, the cognitive and perceptual abilities decrease [14][15]. Studies show that SR accuracy of subjects aged below 15 or above 70 decreases dramatically, when using acoustic models specifically trained with young to middle aged adults. Experiments demonstrate a word error rate (WER) increase of 50% when comparing senior users with a middle age user group [16]. On the other hand, in [17], it was found that SR engines trained with elderly specific acoustic data showed a significant decrease (around 12% better) in WER when compared to engines trained with regular data. Similar results were obtained in [18] where an improvement of 2.9% in the WER was achieved when compared to engines trained with regular data.

In brief, generic trained ASR systems perform significantly worse when used by the elderly
population, due to various factors. The typical strategy to improve ASR performance under these cases is to collect speech data from elderly users in the specific domain of the target application and train elderly-only acoustic models.

C. Speech Interfaces for Assistive Technologies and applications

The existent companies and organizations that develop applications for seniors usually use elderly targeted places such as, nursing homes, government subsidized housing, retirement communities, senior’s centers and public libraries to develop, test and disseminate their software. The main objective of these applications is to provide access to web contents, to enable communication with friends and family, promote literacy, and overcome elderly reluctance towards media and electronic devices. According to [20], the number of seniors connected to the internet is rising hastily being the fastest-growing demographic group online. Internet is also increasingly becoming an important resource for information about health and health care options, communication and news. By being connected to the outside world, senior citizens become more socially integrated and have fewer depressive symptoms. The applications for seniors are characterized by friendly interfaces with buttons and font sizes above normal in order to tackle with low vision issues that are characteristic of the population in this age group. As stated in 0 speech can also be applied as an HCI in software that targets the elderly age group. Also, when used by elderly, speech interfaces are considered to be natural and practical to use [22].

When looking at the available speech-enabled solutions such as, Windows 7 accessibility features, QualiWorld platform [23], NIHSeniorHealth Website [24], Verbose Text to Speech [25], I2net Orion [21], etc., common features characteristics can be found. These applications focus on delivering to the user a simpler and easier access to daily tasks such as, using the internet. Amongst the common features are: Screen readers by using a TTS system; speech commanding and dictation to interact not only with the computer but also with household appliances; scaling the user interface to make text and graphics more readable; changing shapes, contrast and sizes of pictures; etc.

D. Characterization of the elderly population in Portugal

According to recent census and studies [26], the demographic evolution in Portugal is showing a continuous ageing, mainly because people live longer and the birth rate is declining. Since the beginning of the decade until 2008, the average life expectancy increased 2.26 years on average for both sexes – 2.46 years for men and 2.05 years for women. The estimated life expectancy in the period 2006-2008 is 75.49 years for men and 81.74 years for women. The overall birth rate used to be 1.56 children per woman in 2000, but that number declined to 1.33 children per woman in 2007, the lowest figure recorded thus far in Portugal. On December 2008, the elderly population in Portugal (65 and above) was 17.6% already. In 2004, the projections made in [27] showed that the increase of the population ageing is a phenomenon observed equally in all regions of the country. In general, an increase between 63.2% and 76.5% is to be expected. The same document predicts that the ageing rate will continue to increase until 2050, whichever scenario might be chosen. In the worst case scenario, the ageing index could reach to a point of 398 elderly (65 years and above) for every 100 youths in 2050, thus quadrupling the value of the ageing index in the course of 50 years. Even in the base scenario, the prediction is 243 elderly for every 100 youths, whereas the most optimistic scenario limits the number of elderly people per 100 youths to 190.

From the perspective of the WHO (World Health Organization), it is paramount that policies and programs to keep the older population active and productive are defined, so as to guarantee the sustainability of the current society. Maintaining an active aging is one of the biggest challenges our society will have to face in the coming decades in order to promote equal opportunities for all. In 2008 the working population rate continued to rise [28], which was mainly caused by the growing number of women joining the labor market, by the postponing of the retirement age and by the dynamics of migration flows. This increase of the working population rate also corresponded to a higher level of education and higher level of workforce qualification. Between 1998 and 2008, an increase in the total working population of around 529000 individuals corresponded to an increase of about 685000 individuals having at least completed secondary education, being noted that this relationship has intensified between 2004 and 2008 [29]. However, the proportion of assets with the corresponding level of education to higher education remained relatively low, standing at 14.8% in 2008.

As for as the usage of TIC is concerned, keeping the gender and the age group of the individuals in mind, we can conclude [30] that the computer and internet users are predominantly male, although there is a slight difference compared with use by female users. With regard to the age of users, note that the vast majority it is understood between 16 and 24 years, with one big difference from the older age (65-74), the latter being the least used computer and internet (8.1% and 6.6% respectively). We also can conclude that the
use of these technologies is directly linked to education level, and most users are located in higher education, followed by individuals with secondary education and, finally, individuals with lower education level (up to the 3rd cycle). The employment status is also a motivating factor to using these technologies, and most students who use computer (99.3%) and internet (96.7%), followed by the employee, unemployed, and finally, the inactive, with only 15% of computer usage and 12.1% of Internet usage.

III. THE CROWD-SOURCING CONCEPT

Crowd-sourcing is a term used to describe the leveraging of vast amounts of people to achieve a specific goal in a collaborative manner over the Internet. Many crowd-sourcing initiatives have been made possible due to the availability of Web 2.0 technologies, which enable massive collaboration projects to take place. Crowd-sourcing can be considered as a distributed process for the resolution of problems. Typically the process is as follows: an entity has a problem and needs to solve it in a cost effective way. The entity publishes the problem in the web and usually provides the tools to solve it. Users (the crowd) respond to the call and propose solutions to the problem. The publishing entity chooses the winning solution and rewards the user/users accordingly. Rewards vary from money incentives to just public recognition. The publishing entity will own the final winning solution. Multiple solutions can be found across the web in order to digitize old books[31], transcribe speech[32], classify tunes[33], classify galaxies from the Sloan sky survey[34], find ideas for proposed problems[35], image[36] and video [37] tagging and even build a summary of the entire Human knowledge [38].

For the elderly speech data collection described in this paper, we used the Doaravoz, a different version of the Yourspeech platform [39], adapted to senior users. Doaravoz is a tool based on crowd sourcing approaches [40], designed to collect speech data at negligible costs for any language. The concept behind this system is to provide the user with an entertainment reward in exchange for his/her speech. In a previous collection done using the Yourspeech platform, we invited Portuguese speakers to aid in the development of new ASR technology for European Portuguese. 25 hours of pure speech (with no silences or pauses) were collected showing that YourSpeech was a viable way for obtaining speech data at marginal costs given the fact that appropriate marketing and advertisement actions are taken. The preliminary collections reported in this paper show that a few aspects in the prior Yourspeech platform had to be changed, in order to adapt it to senior users, as described in section IV, V and VI.

IV. ELDERLY SPEECH DATA COLLECTION: PRELIMINARY TESTS

A. Sample group

The sample group selection was based on the following requirements: subjects should be over 65 years old, because that is the age group that is usually not present in the common speech data collections for Speech Recognition (such as SpeechDat [41]) and should be able to read, to hear, and to speak without any obvious voice pathology or disability. The sessions took place in Aveiro and in Lisbon, in April 2010. The first population sample consisted of 2 active female subjects, aged 75 and 76, from the Lisbon area, with a high level of education (secondary education and PhD). The second sample group was from Aveiro and was composed by 8 people over 65 years old (4 men and 4 women); 7 participants were at a public nursery house and one was a nursery house assistant. The 7 participants were over 80 years old, lived most of their lives in rural areas and had primary education level.

B. Description of the trial data collection session

In this preliminary data collection, all recording sessions were monitored by an assistant who guided the user and registered future improvements based on user experience and feedback. Then a preliminary test was conducted (to assess reading skills and hearing capabilities) in order to determine the user’s eligibility for the complete trial. A brief questionnaire to collect speaker information was also conducted. Then the user was invited to record 50 prompts using the Yourspeech data collection platform (hereafter Doaravoz 1.0) installed on a local PC and using a headset. The sessions took place in a silent room. Each recording session lasted roughly 30 minutes, except for one subject, who took longer.

A screenshot of the Doaravoz 1.0 data collection platform in action can be seen in Figure 1:
The data collection platform consists of an HTTP server and client application that performs the actual audio recording using an ActiveX control via Internet explorer. The architecture of the platform operates on the internet (in http://www.doaravoz.com), but in the specific case of this trial, it ran entirely in stand-alone mode in a laptop. Audio files were recorded at 16 KHz, 16 bits, no compression and in wave format.

The used prompts were extracted from public European Portuguese corpora such as Cetem Público and Portuguese classic literature.

V. RESULTS AND DISCUSSION

A. Tool usability

From the overall 10 sessions, only 5 participants (3 from Aveiro and 2 from Lisbon) have fully completed the recording sessions. The other subjects, even though they passed the preliminary testing, showed major difficulties in carrying out or completing the data collection task due to various problems associated to advanced age (such as poor sight capability), low literacy and reduced scholar level, fatigue, difficulty in keeping focused for a relatively long period of time. It should be noted that the 2 subjects from Lisboa were women with the highest level of qualification (the remaining subjects only had Primary education), and cultural level above the other subjects (with only primary school education level). Two other subjects required the Windows Magnifier tool support to be able to complete this task. It should be noted though that using the Magnifier tool slowed down the process, resulting in extreme tiredness of the participant after completing the task, added to the fact that the subject was 92 years old.

In brief, Doaravoz 1.0 speech data collection tool showed several usability issues in this age group, namely:

a) Font size was too small. The majority of the subjects had problems in reading the proposed prompts due to the used small font size. Windows Magnifier was used in order to enlarge certain parts of the screen. This added extra time cost to the session.

b) Control buttons were not positioned correctly, which again turned to be more time consuming because the user had to keep switching between prompts.

c) There were too many instructions. The data collection tool shows several on-screen instructions which didn’t prove to be more effective. Subjects often started to read the instructions instead of the actual text to read.

d) Some prompts’ content was inappropriate for this age group because: 1) some prompts were too long and used complex words, 2) some sentences contained sensitive words like “death” or family related terms, leading to an immediate emotional change in the user.

e) The end of the speech detection algorithm was incorrectly tuned, due to the naturally slower speech rate of the elderly. This caused some extra repetitions and increased sessions’ time.

f) The session proved to be too long (around 1h). The previous issues caused each trial session to get longer, which caused fatigue and boredom in the subjects.

B. Speech recognition of the elderly speech collected data

In order to make a first diagnosis of the existent SR systems’ performance in European Portuguese when dealing with elderly speech, a comparison test between two commercial SR systems was conducted. The experiment was done using the speech corpus that was collected from both senior groups.

The procedure was as follows: a) all valid sessions from the database and corresponding utterances were extracted; b) all unique words from all utterances were extracted and analyzed and a word list was built; c) for each utterance it was created a simple, one rule loop CFG grammar containing: all words contained in the prompt (N), N random words extracted from the complete word list; d) the SR engine was fed with recorded utterance and specific grammar for that utterance; e) the recognition result was extracted and compared with the original prompt (note that the utterances haven’t been transcribed; we assumed that the original prompts were correctly read by the speaker); f) the substitutions, deletions and insertions were calculated (as depicted in Table I); g) a recognition result that yielded a null value was considered a full prompt deletion; h) based on the total number of substitutions, deletions and insertions, the global WER was calculated using the formula: WER =
From the results shown in Table I, we can see that both SR systems still present a very high WER (36% and 28%), considering that the reference WER for English in Real Time Speech Recognition is less than 16% [42].

Table I: SR Engines comparison test results.

<table>
<thead>
<tr>
<th>Engine</th>
<th>Substitutions</th>
<th>Deletions</th>
<th>Insertions</th>
<th>Total words</th>
<th>WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Engine 1</td>
<td>269</td>
<td>492</td>
<td>66</td>
<td>2310</td>
<td>36%</td>
</tr>
<tr>
<td>SR Engine 2</td>
<td>103</td>
<td>498</td>
<td>50</td>
<td>2310</td>
<td>28%</td>
</tr>
</tbody>
</table>

Further tests need to be performed with the corpus correctly transcribed and with more sessions and utterances. Nevertheless, this test proved the need to collect elderly speech data in order to train the acoustic models with it and improve the SR performance.

VI. IMPROVED DATA COLLECTION TOOL FOR THE ELDERLY

Based on the results presented in previous section, an updated version of the data collection tool designed for seniors was produced and is already available online (http://www.doaravoz.com) (Fig. 3), by the time of submission of this article.

The following actions were taken in the latest version:

1) **Font size was enlarged.** The majority of the subjects showed several difficulties in reading the initial prompts due to the small font size.

2) **The position of some control and navigation buttons was changed** according with users’ feedback to a more intuitive and logic position.

3) **Instructions were simplified and made more clear using images.** The tool contained several text instructions that only increased confusion in the subjects. Often, the subjects tried to read the instruction text itself instead of the actual text to read. The graphical user interface was simplified and more clear instructions were provided, using images.

4) **Utterances were simplified.** The prompts were too complex and difficult from a syntactic and semantic point of view. **Content was filtered.** The prompts contained sensitive topics such as references to “death” or “dead relatives”. These words or references were filtered.

5) **Speech detection algorithm was adapted.** The speech detection algorithm was often prematurely triggered due to the paused and hesitant nature of elderly speech. The algorithm was adapted to deal with longer pauses that are normal in the seniors’ interaction with the system.

6) **The session was designed to be conducted in 30-40 minutes.** The number of prompts per session was reduced and the instructions simplified.

VII. CONCLUSIONS

The following conclusions can be derived from this pilot experiment and can be extended to web and software-based interfaces in general:

a) **Education level plays a significant role in the way the user interacts with computers.** In the Lisbon sample group, all subjects showed fewer difficulties in completing the recordings. Subjects from the Aveiro group took more time to understand the purpose of the test, as they weren’t familiar with the used technology. Therefore, a data collection task of this kind should be performed by an elderly population group with a higher literacy level.

b) **Elderly users need more time to deal with tools and technology.** This must be considered when programming the time for the users interaction with the systems.

c) **The data collection tool required several changes in order to deal with senior users.** Main changes included design, font and content changes and other technical aspects.

d) **The existent commercial SR systems have a very high WER when dealing with elderly speech,** because they are not adapted to seniors’ voices. Large amounts of senior speech data are required to improve SR accuracy and to allow Assistive Technology to make use of speech interfaces.

The speech data collection campaign has already started and is on progress. So far, 13 seniors from 60-80 years old have donated their voice. A preliminary usability study with the first 10 was conducted with very positive feedback. Nevertheless, a continuous improvement to the platform is envisioned.
REFERENCES


The Companion – Home Shopping From a General Purpose Assistive Technology Platform

Mary Breeze\textsuperscript{1}, Justin A. T. Halls\textsuperscript{2*}, Edward Varney\textsuperscript{2}, Heinz S. Wolff\textsuperscript{2}

\textsuperscript{1}Adult Community Care, Bristol City Council, Beam Street, Bristol, BS5 9QX
\textsuperscript{2}Brunel Institute for Bioengineering, Brunel University, Uxbridge, Middx UB8 3PH, UK.

\textsuperscript{*}Tel. +44 (0)1895 266909, Fax: +44 (0)1895 274608 e-mail: justin.halls@brunel.ac.uk

Abstract - The Companion provides an integrated approach to the supply of a range of assistive technology facilities in a single, easy to use package. It can provide support for care-worker monitoring, medication alerts and compliance monitoring, basic e-mail and display and storage of digital pictures. Since many of these facilities are not perceived to be of importance by the user the system is centred around the provision of home shopping where there is a clearly recognised direct benefit. The system is configured not to look like a computer and all user input is done by means of a barcode scanner in conjunction with a printed catalogue. No computer skills, keyboard or mouse manipulation are required.

I. INTRODUCTION

The demographics of the ageing population in the UK and elsewhere, is well established\textsuperscript{[4]}. It is widely recognised that assistive technology will be required to reduce the burden on carers and social workers in the UK and elsewhere\textsuperscript{[7]}\textsuperscript{[8]} and the UK government is committed to providing a mainstream telecare service\textsuperscript{[1]}\textsuperscript{[2]}.

A significant factor in the quality of life for the elderly is having a degree of independence, and more importantly being able to take care of themselves\textsuperscript{[5]}. Assistance with shopping is a service which has historically been provided by social services departments, but which is very costly in time and manpower. Internet shopping is becoming more widely available, but it has been found that older people do not find internet shopping easy\textsuperscript{[3]}. At the physical level, manipulation of the mouse and keyboard can be a problem, especially if the user has arthritis or suffers from Parkinson’s disease. However, even if a supporter was available to manipulate the computer, older people still found the process of internet shopping difficult. Barlow and Breeze\textsuperscript{[3]} compared several methods of providing home shopping services to people with limited mobility. On the basis of their trials they suggested that a simplified ICT-supported system was needed, based on intermediaries who handle the transaction, but without the need for the service users to have access to the internet.

The Companion has been designed to address both of these issues. It provides a platform for home shopping which does not require any computer expertise, and does not require the use of mouse or keyboard. The system is based around the use of a printed and illustrated catalogue. The concept of catalogue based shopping is a familiar one and presents no great challenges to the user. Items from the catalogue are selected by using a hand held barcode scanner. This provides a convenient and versatile input device which is easy to manipulate, and gives both audible and vibration feedback when a barcode has been successfully recognised. Additional barcodes in the catalogue allow the user to review their shopping list, change quantities and send the list to the local supermarket.

During selection the shopping list is displayed in large clear text on the LCD screen, and item descriptions are spoken aloud using a high quality synthetic voice. This, combined with the option of different display colours, enhances access to the system by people with poor eyesight. Although the system currently only uses British English, it is quite possible to extend it to use different languages or accents.

At this level, the Companion is just one of many assistive technology devices which are becoming available. Systems such as Eldy (www.eldy.eu), PointerWare, previously known as Softshell (mysoftshell.com) and BigScreenLive.

Development for The Companion has been, and continues to be, funded by Tools for Living Ltd. We wish to thank Bristol City Council Adult Community Care for seconding Mary Breeze to the project; the Dolphin Society for their support and funding for the trials of The Companion in Bristol; and Somerfield plc for their support in providing home delivery for the users of The Companion. We would also like to acknowledge the work of Mark Chapman who developed the original prototype version of The Companion under the direction of Prof Heinz Wolff.
(bigscreenlive.com) provide easy access to computers and the internet, either using dedicated, simplified applications or a simple gateway to standard applications. Numerous medication reminder systems are available and being developed[6], including the PillPhone (www.pillphone.com) and Medready (medreadyinc.com), as well as a growing number of caring home systems of varying degrees of complexity with or without additional telehealth monitoring, from providers like Tunstall (www.tunstall.co.uk), Tynetec (www.tynetec.co.uk), JustChecking (www.justchecking.co.uk) or Docobo (www.docobo.co.uk). These systems, while very worthwhile and valuable are not designed to provide a general purpose platform capable of supporting a range of different services to people with different needs. Where services overlap, multiple installations may be required increasing cost and complexity. The Companion is extensible to provide similar services to all of these products from a single integrated device.

Although it is distinctive in its ‘elder friendly’ interface, and the use of voice to accompany the screen display, its greatest strength lies in its versatility. As the range of assistive technology devices increases, and more and more of the needs of the elderly are addressed, there is a tendency to install numerous separate devices, each dedicated to a single specific task. Even where these are loosely tied together and operated as a ‘system’ there are many discrete devices, each requiring space and a power socket. This can become a significant problem in a small flat, such as is often found in sheltered accommodation, with limited places to locate equipment and minimal numbers of power outlets to supply them.

There are many services that technology can provide which are of great use to carers and supporters, but may not be of great perceived benefit to the user and The Companion is capable of providing a wide range of these. Many can be activated for minimal additional cost, simply requiring a small amount of extra software.

Care worker visits can be monitored by providing workers with ID cards incorporating a barcode which would be scanned on arrival and at departure. A list of action barcodes could be scanned to show which procedures had been completed during the visit. The results would be e-mailed directly to the care manager from The Companion.

The system can issue reminders to the user when it is time to take their medications and, with the cooperation of the pharmacy, they can have barcodes attached to the medication package which they would scan to indicate that they are taking the correct medications at the correct dose.

E-mails can be received by the system, displayed and read out. Accompanying pictures can be stored in a photo-album with barcode indexes for easy access. A number of standardised e-mail replies can be made available to acknowledge receipt of incoming e-mails.

With the addition of suitable hardware The Companion can also support more complex functions, from a simple video-doorphone to a comprehensive caring home monitoring system such as the Millennium Home[4].

The provision of a device which gives the user a large perceived benefit, such as a home shopping system, can make the provision of these additional services far more acceptable than would otherwise be the case.

II. MATERIALS AND METHODS

A. Hardware

The Companion platform is based on an IBM T23 Thinkpad laptop computer, although any similar machine would be suitable. Minimum requirements are for a 1GHz Pentium class processor and 256Mb RAM, with a DVD drive and Windows compatible sound card. 2 USB ports are desirable, but a single port is adequate for home shopping. E-mails are sent through the internal 56K modem connected to the user’s telephone system, or through a broadband modem if this service is available.

The keyboard and mouse-button, or mousepad, are blanked off by the addition of a plastic oversleeve. This prevents the user from inadvertently entering

Fig. 1. The Companion with the barcode scanner
invalid data, and also makes The Companion less like a ‘computer’. Figure 1 shows The Companion in use, with the barcode scanner, and displaying a typical shopping list.

The screen shows the list of selected items, the size of the items and the expected price. The number of items chosen is on the left and a running total of the cost so far is shown at the bottom of the screen. The prices given are guideline prices since the actual prices at the supermarket can vary on a day to day basis, but it is unusual for the actual prices to differ significantly from those shown on the screen.

The operating system is stored on a DVD, and the DVD drive is modified by the addition of a mechanical button which will open the DVD drawer when the machine is switched off. This allows the user to easily replace the DVD when an upgrade disk is issued.

Synthetic speech is an important component of the system, and is provided by AT&T Natural Voices. The preferred voice for British use is ‘Audrey 16K’, which provides very high quality speech. A user ‘dictionary’ is defined to improve pronunciation of some less common words found in the product list, and to provide more user friendly pronunciation of some of the abbreviations used by the supermarket.

For home shopping purposes a simple USB linear barcode scanner is adequate, although advanced applications may usefully benefit from advanced 2D and imaging barcode scanners. In use the barcode scanner is seen as a keyboard device, and all input appears to come from a keyboard. Thus no special drivers or interface software are needed.

The user is provided with a printed catalogue (Figure 2), which contains instructions for the use of the system, a few pages of ‘control’ barcodes, and an illustrated product list containing around 700 common items ranging from bananas, through frozen meals, to cleaning products and alcoholic beverages. The control codes shown allow navigation through the shopping list and additional barcodes allow the quantities to be changed, the list to be sent and to close down the system. For the trials the supermarket used was Somerfields and the supermarket supplied regular updates of their product list, with the associated barcodes. The Companion therefore recognises the barcodes for the full list of products stocked by the supermarket branch, not just those in the catalogue, and the user can scan the barcodes directly from products that they have around the house.

**B. Operation**

When The Companion is first configured for a user, the support worker uses special software to generate a set of barcodes defining user specific information. This
includes name, address, and internet service provider details which may vary depending on the telephone network used. The supporter then scans these barcodes to personalise The Companion for a specific user. In some cases a single Companion may have several users, in which case each user is provided with an identification (ID) card with a barcode which will allow them to ‘logon’ to the Companion and identify themselves to it.

When The Companion is switched on it boots in Windows XP embedded, directly from the DVD. This provides a more stable operating system than normal Windows XP and ensures that the operating system cannot become corrupted or modified. It also allows the operating system, as well as the Companion application, to be upgraded simply by issuing a new DVD to the user. The Companion starts directly into the Companion application, and either automatically logs in the default user, or waits for a user to scan an ID card.

When a user logs in, a new blank shopping list is created, but a barcode can be used to restore the previous shopping list if desired. Products are added to the list by scanning the appropriate barcodes from the catalogue or from existing items. As each product is scanned the item is displayed on the screen and the description and price are read out by the computer’s synthetic voice.

Barcodes are available to re-read the list or to browse through it, and the user can add, delete or change the quantities of items using the scanner. When the shopping list is complete a ‘send’ barcode would be scanned to dispatch the list by e-mail to the supermarket.

The supermarket delivers to users on a particular day each week, and the users know which day they will receive their goods. Payment is made automatically where the user has a bank account with direct debit facilities, or if they do not have a bank account then payment is made through the Companion support services.

Finally the user can log out, or switch off the Companion by scanning appropriate barcodes. In the event of a power cut The Companion will use its back-up battery to save the current shopping list and close down safely before turning itself off.

III. RESULTS

Initially ten prototype systems were deployed in the Bristol area starting in January 2005. These systems used a small single board computer and a small monochrome CRT monitor. These were replaced in Jan 2006 with the Mk II version described here. Between 2006 and 2010 a total of 118 users from Bristol, Stockport and Lincoln have sent 3504 shopping lists. In 2008, 1080 shopping lists were sent from 42 Companions, to a total value of around £27000. Users generally found the system very easy to use, but at least initially were grateful for the immediate telephone support that was available. In several cases users formed small social groups who would do their shopping together from a single Companion and would share collections of barcodes removed from their favourite products.

The recurring complaint about the system was over the lack of variety, or wide enough selection of items in the printed catalogue. Nevertheless the users found that the system was useful to them. A particular advantage, commented upon by several users, was that it enabled them to have the heavy, basic items delivered allowing them to shop separately for specialist items, or allowing their home care worker (who was restricted in the number and weight of items that they could get) to purchase additional items when required.

Reaction to the synthetic voice was varied. Most users found that it was clear and helpful, but a few users found that it was unnecessary, or were indifferent to it.

Although the majority of the users did have bank accounts, most of them were reluctant to allow direct debit schemes for payment of their groceries. During the trials, payment could be handled through the support services where necessary, but a reliable payment scheme is important to the wider deployment of the system.

IV. DISCUSSION

The system described here has been very successful, and very well received by users, during the trials taking place in Bristol and elsewhere. It is often suggested that the use of assistive technology can reduce the user's social connectivity. However, at least in this trial, it was observed that it encouraged users to get together socially. There was also anecdotal evidence to suggest that supporters who would otherwise have spent their contact time doing the shopping were now able to spend quality time with the user instead.

The most significant perceived problems with the system relate to the limited range of products available in the printed catalogue, and to the lack of access to special offers provided by the supermarket. These problems have been addressed through the development of an on-screen catalogue which provides a greater range of products, and which can be dynamically updated with special offers from the supermarket. This extension to the system has been well received by users who find it much more convenient than having to manipulate the paper based catalogue.
The key elements of this system are the use of a barcode reader for user input, and the combination of voice with the large screen display. There are several possible alternatives to the use of the barcode system as the input device, including i-buttons, rfid tags and touch screen. The barcode scanner was initially selected because it would also allow the existing barcodes on grocery products to be scanned in addition to the items in the printed catalogue. It is also very convenient to be able to print codes with a standard inkjet or laser printer, which makes it easy to produce catalogues, ID cards, and additional sheets of codes for controlling the system. Unlike touch screens, it is possible to have personal identification codes to allow users or supporters to log on, and identify themselves to the system. As the cost of touch screens is reducing, ongoing studies are in progress on the use of touch technology in combination with the barcode scanner to provide a more flexible user interface.

It is important that assistive living technology, which is designed to be used in existing homes needs to be unobtrusive and should not label the user as being dependent on the technology[7]. Although the current implementation only supports home shopping, it is clear that its real strength lies in the ability to support a much wider range of services at very little additional capital, or support costs. These additional services have additional discretion because the basic function is seen to be that of home shopping.

Medication compliance is a serious issue, especially amongst the elderly who may have complex regimes to follow. The Companion can provide reminders, confirm that the correct medication has been located, and remind the user of the correct dosage. It can also allow the user to report side effects or problems so that a qualified professional can contact them and try to resolve the problem.

Care Worker monitoring, and recording is very simple to implement by providing the care workers with an ID barcode and a list of barcodes for common tasks. The information is then sent directly to the managers by e-mail.

Services such as these may not always be perceived as beneficial by the user, although they may have great benefits for the supporting service providers. The balance provided by home shopping, interactive video doorphone and simplified e-mail support provide benefits which are not only beneficial to the user but are also directly useful to them, making acceptance of the system much greater.

The overall cost of the system has been kept as low as possible, for instance by using ex-corporate laptop computers as the basic component. However it has to be recognised that the capital equipment cost is not the most significant factor. It is not reasonable to expect to be able to provide a system such as this to a user and then to walk away and leave them to get on with it. There does need to be a support base to provide help with initial training and distribution, telephone support, assistance with upgrades and co-ordination with supermarkets and social services.

Facilities such as medication compliance or care worker monitoring involve the active co-operation of a number of different groups including social services, care providers, doctors and pharmacists. These links all need to be established when the system is initially installed in an area, and need to be maintained while the system is in use.

We estimate that a ratio of one supporter to every hundred users is reasonable. This immediately becomes the dominant cost factor, but is still far less than would be required for conventional home shopping support provided by a social services department.

REFERENCES


Chapter 3

Multimodal Human-Computer Interfaces
Abstract – The design of accessible Web user interfaces (UIs) is a complex challenge for software architects and developers since the requirements are often respected too late in the development process. UI modeling allows research on particular aspects of accessibility requirements during the development process – especially during analysis and design. It is based on the description of user’s workflow. Essential UI models are described. With Model-driven Design (MDD) and the Unified Modeling Language (UML) common standards of software engineering are used. In summary, the integration of accessibility in UI modeling helps to understand the needs during earlier development phases of software applications better.

I. INTRODUCTION

Despite accessibility guidelines for Web and software development exist since more than one decade; they do not meet practical needs of developers and designers in early project phases. According to e.g. DIN 15226 (DIN – German Institute for Standardization), the development process has four main phases: analysis, design, realization and test (see Figure 1). Typically, users are being mainly considered only during analysis. Afterwards, the user is not in the focus of attention. User-centered design (UCD) [1] aims to overcome this developer-user gap of classical software development and to ameliorate the usability of software and Web applications. Even, this gap is seen by many experts as a major cause for the lack of accessibility in UI design. Additionally, existing accessibility guidelines address typically concrete UI implementation and testing (e.g. [2]) but not analysis and design. User-centered design is recommended but not detailed in appliance.

The UCD approach engages to-be users in design and conception of applications. UI requirements are integrated early in the development process respecting the growing influence of usability in product design. Even, UCD has advantages for UI accessibility if users with disabilities are included in conception and design of applications. Accessibility has impact on the whole application architecture including database design. Therefore, early integration in application architecture avoids late changes and additional costs for development.

This paper is focused on the integration of accessibility requirements in UCD and model-driven UI design. Methods of user- and usage-oriented analysis are combined with model-driven design to examine the potential for engineering of accessible UIs. The UI modeling allows the description of single aspects in design as workflow, navigation or UI behavior. Research issues are:

• Influence of accessibility requirements on early UI design and conception
• Conception of the UI design process itself to facilitate the integration of accessibility requirements in software development
• Integration of expert knowledge about accessibility in design and modeling

The modeling of a portal for travel information serves as a demonstration example. The integrative application combines interests of people with and without disabilities to gain more information about travel destinations and connections.

The next section gives a short overview for the state-of-art in UI modeling. Then, the approach is demonstrated with the basic models for UI conception and an example. Finally, conclusions and an outlook are given.

II. RELATED RESEARCH

Publications for UI modeling can be separated into technical focused approaches as User Interface

![Figure 1: Accessibility and development process](image-url)
The ARIA [3] extends the WCAG concept with the requirements for Accessible Rich Internet Applications (WAI-ARIA). The content (captions, descriptions, subtitles, attributes, metadata etc.) to enable and facilitate the usage of the UI. Thus, new technologies as Silverlight, Flex or JavaFX enable to separate designer’s UI layout from developer’s core functionalities. Even, Web- and Desktop-based applications are merged on one technical platform.

Only few publications exist for integration of accessibility in model-driven design [11, [12], [13]]. The Dante-approach [11] allows semantic annotation of Web pages in a Web engineering process to provide navigation and to support screen readers to facilitate the audio presentation of the content. It is approximated that 85% of the annotations can be provided by models.

III. ACCESSIBILITY REQUIREMENTS IN DESIGN PROCESS

The Web Content Accessibility Guidelines (WCAG) [2] of the W3-consortium describe the requirements for accessible Web applications. The four basic principles require that information is perceivable, operable, understandable and robust. They are not focused on the development process itself nor do they describe the requirements in early development phases as analysis or design. One basic idea of the guidelines is annotation of text-based information for all non-text content (captions, descriptions, sub-titles, attributes, meta-information etc.). The Web Accessibility Initiative-Accessible Rich Internet Applications (WAI-ARIA) [3] extends the WCAG concept with the requirements for accessible interface components in Rich Internet Applications (RIAs). WAI-ARIA is using additional attributes for roles, states and properties to describe UI element functionality and behavior.

Besides the access to the main content, the application must provide access to navigation and functionalities for orientation and overview as breadcrumbs, site-maps etc. to enable and facilitate the usage of the UI. Thus, some general accessibility requirements with influence on the overall design are:

- Additional text-based annotations for interface components, interface structure etc. (WCAG). This may affect e.g. the database design.
- Interface components communicate role, state and state changing (WAI-ARIA). Especially, this has importance for RIAs but only relevant information should be communicated to avoid information overkill.
- Serial order of interface components corresponds with user’s workflow to support Assistive Technology (AT) as screen readers etc. Modern layout technologies as Cascading Style Sheets (CSS) allow a free composition of UI components but are not used by AT. Therefore, the order of UI components must correspond with the workflow.
- Clearly structured navigation through the UI – every HTML page has one main modus (one focused activity) in the main content area. Especially for users with AT it is more difficult to gain and keep overview and orientation. Clearly declared activities facilitate the usage. Complex nesting has to be avoided.

Model-driven development is well-suited to integrate the first two needs into the design process since semantic annotations can often be taken from models and their behavior as well. Navigation modeling has to be focused on modes of usage and clearly structured navigation. UI models should describe workflow, UI structure, navigation, roles, behavior and states of components. User-centered design helps to identify workflows. Task modeling supports the serializing of workflow within the UI and clarifies the information the user need to control the application. Even, modes can be identified to describe the navigation.

IV. THE MODELING CONCEPT

A. The Basic Concept

User interface modeling as part of the design phase serves to design the Human-computer interaction (HCI) of a software application with the user. Particular models describe different views on the HCI as navigation, abstract presentation etc. Three essential models are common for most existing approaches [18] (see Figure 2):

Figure 2: Essential models in HCI

- Task Model
- Dialog Model
- Presentation Model

UI Metamodel

Human
Human-Computer Interaction
Computer

MARKUP LANGUAGE (UIML) [4], USER INTERFACE DESCRIPTION LANGUAGE (UIDL) [5], EXTENSIBLE APPLICATION MARKUP LANGUAGE (XML) and analytical approaches as USER INTERFACE EXTENSIBLE MARKUP LANGUAGE (UsiXML), USEWARE MARKUP LANGUAGE (useML) [8], UMLi, TASK MODELING LANGUAGE (TaskML) and DIALOG MODELING LANGUAGE (DiaMODL). Technical approaches are focused on a platform-independent, abstract notation of UI components and structure which is independent from other components of the application. UsiXML as an analytical approach is more abstract and includes a metamodel for all UI aspects as navigation, domain or resources modeling. This may be more powerful in future but until now only few applications exist and a formalized modeling process is missing. Publications on model-based design of UIs as useML, UMLi, TaskML and DiaMODL include the process of modeling as well.

New technologies as Silverlight, Flex or JavaFX enable to separate designer’s UI layout from developer’s core functionalities. Even, Web- and Desktop-based applications are merged on one technical platform.

Only few publications exist for integration of accessibility in model-driven design [11, [12], [13]]. The Dante-approach [11] allows semantic annotation of Web pages in a Web engineering process to provide navigation and to support screen readers to facilitate the audio presentation of the content. It is approximated that 85% of the annotations can be provided by models.
• Task Model: Describes user activities as actions and modes of usage
• Dialog Model: Describes the interaction of the user and the UI
• Presentation Model: Describes the UI structure and components and their behavior

The UI metamodel combines these different views on the HCI. It may contain further models to describe user profiles or application resources. The UI modeling is part of a development process. Here, the process is limited to software design and development – in particular Web applications.

Table 1 shows the integration of UI modeling in a UCD process. The UI modeling needs particular preparations as the workflow analysis in order to meet users’ expectation with the task model. Analysis includes user’s requirements and mental models of workflow. Task modeling includes typical use cases and activity diagrams. For graphical model notation the Unified Modeling Language 2 (UML) [14] is used. Since graphical notation is not accessible for visually impaired users, text-based notation with Human Usable Textual Notation (HUTN) [15] completes the UML diagrams. UML 2 and HUTN are based on the same metamodel Meta Object Facility (MOF) [16] and have the same descriptive fullness.

Even, an active participation of users with disabilities helps to integrate their needs in conception and design since the existing recommendations and guidelines do not covering all aspects of accessibility. After the conception, the models are used for the implementation with a Web application framework including accessible UI components.

Software development starts with requirements gathering. Interviews with experts and users give a first overview for the functionalities and design objectives. Sometimes it may be difficult for users to describe their needs or problems with existing Web applications. Methodologies as Appreciative Inquiry enable detailed analysis with user integration [17]. Even, accessibility requirements may differ and be antagonistic between particular user groups as visually or hearing impaired people. Thus, requirements’ gathering has to specify the user group at first.

B. The Travel Information System Example

The travel information system (TIS) provides users with additional information about travel destinations, connections and tourist services and is used as a case study. Similar to a wiki, users can read, create and edit articles for particular travel destinations and generate the content of the platform by their self. Thus, typical use cases are search a destination as well as read, create and edit an article for a destination (see Figure 3).

![Use case diagram for TIS](image)

**Figure 3: Use case diagram for TIS**

<table>
<thead>
<tr>
<th>Development phases</th>
<th>User- and model-centered design process for UIs</th>
<th>Accessibility requirements</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>Requirements gathering</td>
<td>Particular interests of all users</td>
<td>Appreciative Inquiry</td>
</tr>
<tr>
<td></td>
<td>Workflow analysis</td>
<td>Workflow analysis and mode identification</td>
<td>Structuring maps, interviews, participant observation</td>
</tr>
<tr>
<td>Design</td>
<td>Task modeling</td>
<td>Relevant information</td>
<td>Interaction &amp; usability pattern</td>
</tr>
<tr>
<td></td>
<td>Dialog modeling</td>
<td>Clear navigation, serial order, multimodality</td>
<td>Participant observation</td>
</tr>
<tr>
<td></td>
<td>Presentation modeling</td>
<td>Text-based annotations, enriched UI components</td>
<td>Prototyping, UI prototype evaluation</td>
</tr>
<tr>
<td>Realization</td>
<td>Implementation</td>
<td>Event-driven programming, international &amp; national accessibility guidelines &amp; recommendations</td>
<td>Prototyping, UI prototype evaluation, Web framework-based programming</td>
</tr>
<tr>
<td>Test</td>
<td>Test</td>
<td>Accessibility Guidelines</td>
<td>Automatic validation, heuristic evaluation, user tests</td>
</tr>
</tbody>
</table>
Different users were interviewed to obtain the needs of people with and without disabilities – particular visually impaired people. For people with disabilities particular demands were found:

- Information about hazard areas (e.g. clifey areas in nature)
- Information about parking places, additional assistive offers, restrooms etc.

The next step in preparation is analysis of user’s workflow for the documented use cases to support later on the task modeling. Structuring technique, participant observation and structured interviews are some techniques which support the analysis of user’s mental model of workflow.

C. Task Modeling

The UI design starts with the description of user’s workflow – the task modeling. The use cases (see Figure 3) are detailed with UML activity diagrams. UML has different categories for work packages as use cases, activities and actions [14]. Use cases define a global objective, intention and purpose of work and are described in use case diagrams. They consist of activities and actions which are described in activity diagrams. Activities include sequences of actions. Figure 4 shows the hierarchy of categories on the left side.

- Use Case defines the global objective, the purpose of work
- Activity serves for realization of a working task, delivers the result
- Action smallest (atomic) unit of activities has motive and intention
- Operation physical, dependent from technology, no intention, no autonomous use
- Inform query and view information from the application (text, images etc.)
- Select select one or more values from a set with possible triggering
- Activate direct triggering of an application function
- Input entering new data (text, images etc.)
- Edit modifying existing data with the UI changing values

The physical operations of user’s workflow as press a key or view an image are on the lowest level and not part of task modeling to stay independent from input or output modality. Here, user’s actions are described with five predefined basic elements to facilitate the task modeling and to clarify the level of description. The five elementary use objects are inform, select, input, edit and activate (see Figure 4 right side) [8]. Basic actions represent mini-dialogs between user and UI as well and can be detailed with sequence diagrams in UML.

Workflow of different people may differ and the UI should support various styles of workflow. Best practices of usability are also an appropriate source for workflow modeling to meet user’s expectations. Accessibility requirements in task modeling are the device-independent description of user actions. Later on, the task modeling serves to determine all relevant information in the UI. Relevant is information when needed by the user to follow his/her intended workflow.

Figure 5 shows the activity diagram for the case study. The basic actions (inform, input, edit, select and activate) are described with UML stereotypes.

Actions which are enabled at the same time are grouped in four activity regions (home, list, article, editor). The activity regions present the main modes [19] of the TIS-UI. A mode is a particular context of an activity as edit a text or a search. The UI may have various states within the same mode. Later on, modes are mapped on single HTML pages in Web applications. They may have submodes e.g. for auxiliary functions or navigation. Only one main mode is mapped to one HTML page to facilitate user interaction with assistive technology e.g. screen readers. Deep hierarchy of modes is avoided by the same reason.

Next is the modeling of the (macro)navigation between particular modes of the application usage by the user. Main and submodes are identified from the task model and described with a particular activity diagram – the navigation model. Figure 6 shows the navigation diagram for the TIS. Home, list, view destination and editor article are main modes whereas search and preview are submodes and included in other modes. On the right side three (sub)modes represent additional UI functionalities for accessibility as a sitemap and an accessible main and utility navigation. They are not only related to another particular mode since they are always included as sub-modes and the sitemap can be activated from every other mode.
D. Dialog Modeling

The dialog model describes the HCI and mediates between task model and presentation model [18]. Task and navigation model describe user’s point of view on HCI. In general, state chart are well-suited to design the UI behavior. Three methods are possible to derive the “other side” – the dialog behavior of the UI and to complete dialog modeling:

- Activity chain extraction [18]: An Activity Chain specifies a path of different dialogs to reach a certain goal. A dialog represents a set of enabled actions and an activity chain can be described as a State Transition Network (STN). Notation is supported by UML state charts.

- Mirroring of user actions in the task model: An UML activity diagram is used for the notation of the task model. Further on, swim lanes in the diagram can be used to mirror the user actions into the UI and to describe the UI actions. Later on, UI actions can be separated from system core actions with an additional swim lane. State charts are used to detail the UI behavior.

- State charts describing the UI behavior in a particular mode from the navigation model: The UI may have different states within the same mode. Figure 7 shows the state chart for the editor mode from navigation model (see Figure 6). At first, the editor interface is shown. A preview for a document is included if existing. Then the editor waits for text input. The input is shown until the user activates the preview or saves the document. An existing document is updated. Otherwise a new document will be created.
Presentation modeling includes the Abstract Presentation Model (APM). Modeling of accessible HCI dialogs can follow different strategies such as universal or multimodal design. Here, universal design is used meaning that perception and control of all relevant information in the UI is independent from input or output device. Information is relevant if needed by the user to follow his/her intended workflow. Additionally, interaction and usability patterns can be integrated as best practices of interaction design.

Modes from navigation model are mapped to single views. A view is restricted to contain only one mode and additionally some submodes to model site navigation, auxiliary functionalities such as search etc. This restriction supports accessibility since usage and orientation is easier with screen readers.

The hierarchy of UI components is described with a class diagram. Figure 8 shows a simple class diagram of the abstract presentation model of the TIS editor including navigation, search functionality and the editor part with a preview if necessary. It shows the static hierarchy of site components.

Views and UI components have attributes which are mapped to HTML elements and attributes later on to describe metadata and device-independent properties. Device-independent event-handler and WAI-ARIA-like elements are used to describe the behavior of the UI during runtime. Abstract UI components are anchor, link, text, buttons, image, text input, choice and container. Attributes are used to declare necessary metadata for UI elements such as title, name, type etc. The root component is the view-element.

F. Implementation

The Model-View-Control (MVC) architecture pattern separates database design (model) from control and UI (view). The separation of UI and application design is well-suited to implement the user-driven models for the UI. Meanwhile, frameworks for Web application development support the usage of accessible UI components. Here, Java Web frameworks with MVC architecture are evaluated – e.g. the Apache MyFaces (AMF) including Trinidad and Tobago subprojects and the Google Web Toolkit (GWT). AMF is an open source implementation of Java Server Faces (JSF).

Concrete presentation modeling and application layout are not part of the modeling process. For the implementation of the declared models an existing Web framework is used. There are different Web frameworks which combine the MVC-architecture pattern with libraries of UI components. AMF Trinidad and GWT have built-in support for accessible UI elements. GWT implements more client-side functionality to obtain application behavior similar to desktop applications. WAI-ARIA is also supported in GWT. The navigation model (see Figure 6) is used to generate the navigation configuration for the JSF application. 
notation completed with HUTN for accessible modeling. The approach is focused on Web applications. Nevertheless the basic ideas are valid for other platforms as well e.g. Desktop-application or UIs for automation systems.

Generally, task-oriented analysis is limited to recurring workflow when user can imagine the task in detail. Highly creative tasks such as painting or composing music have often no typical recurring workflow. In these cases, it is hard for users to describe the mental model of the task.

Until now, model data are translated manually into the runtime markup. Further work is necessary to design and implement a model compiler to automate the process. Further research questions are:

- Integration of accessible user-generated content
- Integration of visual and auditory media with XHTML+HTML, SVG or VoiceXML

Model-driven design supports the design of complex applications since particular aspects of structure and functionality are described in separate models. The integration of accessibility requirements in the process of development may help in future to integrate in and approach the knowledge of accessibility exerts and software designers for the easier design of accessible UIs.

REFERENCES

Abstract – The elderly undergo a problem of social and technological exclusion and there has been an increase of the elder population due to the enlargement in the average life span, creating more elders. Therefore, this problem is in need of higher attention.

A modular architecture is presented, designed to help resolve the identified problem, consisting in the modification of human-computer interaction by the means of voice command recognition, focused on senior citizens.

A system has been developed based on that architecture and tested with a group of subjects from the targeted population. The tests demonstrated that the tested subjects could easily use it the developed system, all of whom where seventy years of age and had no previous computer knowledge. Therefore, the system and architecture successfully permit elder users to interact with the computer.

The system design was focused on communication, allowing synchronous and asynchronous communications with the aid of e-mail and VOIP technologies; however, it has been designed with the modularity of the architecture, permitting the addition of new functionalities in real time execution.

The system can be considered an optimal solution for the presented problem, if applied with the correct hardware equipment.

I. INTRODUCTION

Many countries have an aged population due to the increase of life span expectancy. Consequently, the elder represent a large share in their current population [1] [2] [3]. The elder grew up in a time where illiteracy was common and technology was scarce, therefore creating large difficulties in following the rapid evolution of technological development. As the result, the majority of senior citizens have limited or null computer knowledge [4] [5].

As years go by, the individual suffers deterioration at a biological and social level [4] [5]. It is within this problem that Gerontechnology comes in, which aims at the study and development of new systems and technologies that can satisfy the necessities of the seniors and increase their quality of life [6].

There is a tendency for abandoning the elderly, leaving them alone or in a nursery home, isolating them from the exterior world [7]. Many depend on social workers, whose function is to give assistance those who need it, but they are limited to their work schedule. It is also common for the elder to live in remote locations, far away from the metropolitan areas where you can find the government sponsored services, such as social security, medical services or even entertainment services. In these situations it ends up being very expensive for the elder individual to travel to the site of one of these services, doing it only when strictly necessary [8]. From another point of view, if the services require contacting with an elder for the purpose of, for example, making a routine check-up, they encounter difficulties doing so [9].

Many systems and applications already exist that aid the computer interaction for people with disabilities, but the elder do not fall in that category. The information and communication technologies that can be manipulated to aid the senior citizen exist, such as speech recognition, electronic mail (e-mail) and Voice Over Internet Protocol (VOIP), that seem complicated to use, but just need to be manipulated in a way that allows simple interaction from the user. Most of the existing systems are very complicated and complete, and are not designed for accessibility.

This paper fits in the area of Gerontechnology [6], focusing on a solution for the referred problem, and consists of a scalable architecture that modulates the interaction with the computer through the recognition of voice commands, to execute simple system commands and fulfil tasks. It also focuses on the development of a system based on that architecture, which reflects on achieving synchronous and asynchronous communication with the resource of e-mail and VOIP technologies, and on granting the possibility of executing any system command with a voice command.

II. RELATED WORK

In order to understand the state of the art, some applications were analysed regarding voice commands and senior accessibility. A brief description of the analysis is followed.

A. Voice commands

Voice command recognition is accomplished with the use of a speech recognition engine. Windows Vista and Windows 7 come with one of those engines and with utility software that allows the user to execute many of the operating system commands and tasks through voice command recognition [10]. It tries to be a complete alternative to normal interaction methods.
(keyboard and pointing device), performing the tasks the same way; therefore, it is not considered a solution to the described problem.

There are also applications like TalkingDesktop [11] and Utter Command [12], but they also have the shortcomings mentioned above. These applications have been developed to function as an alternative to the normal input devices, to perform the same tasks, not to create simple tasks.

B. Senior citizen accessibility

Few systems have been developed with accessibility characteristics designed for the elderly. Out of those systems, worthy of note are ELDY [13] and SimPC [14], which are systems that allow informal communication (e-mail, small personal messages, social messages or VOIP calls, among others) and performing other useful tasks such as navigating the internet or sharing photographs and video, in a simple and effective manner, intended for the elder. However, the interaction is done with a pointing device and keyboard, not through the recognition of voice commands. The solution should be able to eliminate the keyboard and most of the pointing device from interaction.

III. VOICE COMMANDER

An architecture and a system called “Voice Commander” were created and they fall within the accessibility category. Their main goal is to modify the users interaction with the computer, allowing the execution of any preconfigured system command with optional parameters, through the recognition of voice commands. Specifically, it is oriented for enabling synchronous and asynchronous communication.

To achieve these functionalities, the system needed to be able to associate a voice command to a system command and recognize spoken voice commands. It was also necessary to create an address book system, to create an e-mail client and to use an existing VOIP client, taking advantage of the already existing e-mail and VOIP services.

The concept of e-mail used was of “Voice-E-Mail”, which is awaiting patent rights. It consists of sending a voice message from the sender to the receiver, by using the already existent e-mail technology. The user speaks out a voice message to send, which is recorded to an audio file, attached to an ordinary e-mail, and then sent to the destination e-mail account. The receiver then obtains the sent message, strips the audio file attached to the e-mail and plays the audio file.

Therefore, the synchronous communication was achieved via the use of VOIP services and the asynchronous communication was achieved by the use of “Voice-E-Mail”.

To achieve the recognition of voice commands, a speech recognition engine was used with a limited recognition dictionary, composed only of the valid voice commands in the current system context.

An address book was created in order to manage the contacts the system will communicate with, defining the contact information stored.

A. Architecture

A scalable architecture has been developed that allows the copulation of functionalities in real time, and it is represented in Figure 1.

![Voice Commander architecture](image)

This figure shows the main architecture, focused on functionalities. It consists of four Application Programming Interfaces (API’s) and a central Graphical User Interface (GUI). The architecture allows the addition of system commands dynamically, associated to a voice command, and has the functionality of speech recognition embedded. This specific scenario shows e-mail and VOIP functionalities, and has a link to the operating system to execute system commands. It also shows the address book, which defines unanimous contacts throughout the system.

The API’s and GUI are described in the following headings.

B. APICommunicationCenter

This API agglomerates the functionalities of speech recognition and sound recording, and defines the concepts of system command, voice command and the association between the latter.

A voice command consists of a word or set of words representing the phonetics of the voice command, i.e., what the user says. A system command consists of three things: the executable file name including the path, optional parameters and an image name related to the system command. An association is a one-to-one relation between a voice command and a system command.

Figure 2 shows the functional architecture for this API.
This figure presents the main functionalities available with this API, consisting of recording audio, recognizing voice commands through speech recognition, and availability of associations, allowing the addition, removal, modification and search of those associations.

The audio recording is achieved with the use of Microsoft DirectX technology, and encoded into a MPEG Audio Layer-3 (MP3) format, recording all audio from the default recording device to a small file.

The speech recognition is achieved with the aid of Microsoft Speech Recognition technologies, using a recognition dictionary composed only of the available voice commands in a current context. The recognition process always starts with a timeout, and if that timeout ends without a correct recognition, the process halts and the API informs the sender of the method through the invocation of a method described by an interface.

The API also has an adjustable confidence level, granting the ability to adjust the sensitivity of the speech recognition engine.

C. APIAddressBook

This API aggregates the functionalities related to the contacts used throughout the system and defines the concept and details of a contact, allowing the addition, removal, modification and search of contacts. It also allows the configuration of the information for the user of the system.

A contact consists of a name, an e-mail address, a voice command, a VOIP phone number and an image. This API grants the functionality of obtaining and modifying the users contact details, and adding, eliminating, modifying and searching contacts in the contact list.

D. APIMailManager

This API concentrates all the e-mail functionalities, acting as an e-mail client. It defines the concept of a mailbox, concedes the functionality of sending and receiving e-mail messages and stores all sent and received messages.

The mailbox is defined by two separate sets of e-mail messages, one for the sent messages and the other for received messages. It also contains a flag indicating the existence of unread received messages.

Figure 3 shows the functional architecture for this API.

This figure displays the main functionalities available with this API, consisting of sending, receiving and reading e-mail messages. The sending of e-mail is accomplished using Simple Message Transfer Protocol (SMTP). The receiving of e-mail is performed using Post-Office Protocol (POP). All sent e-mail messages and attachments are stored in the Outbox folder, and received e-mail and attachments are stored in the Inbox folder.

This API uses a thread that takes care of checking for and receiving new e-mail, communicating with the POP server, and supplies the means for starting and interrupting that thread. The API also concedes the ability to define the timeout between requests. When the e-mail updater thread receives a new e-mail, it signals the object that started the thread by invoking a method described by an interface.

The API also stores the information related to the user and e-mail servers, namely the SMTP hostname and port number, the POP hostname and port number and the user’s username and password. All this information is stored in binary.
E. APIVoipManager

This API incorporates all the VOIP functionalities, behaving as an integration module between the VOIP client and the system. The functional architecture for this API is presented in Figure 4.

![APIVoipManager functional architecture](image)

This figure shows the three main functionalities of this API, specifically making and receiving a VOIP call and transmitting video on an active call. As shown, it mainly acts as an integration layer between the VOIP client and the system.

Upon the change of status of a VOIP call, the API informs the object that instantiated it by invoking a method described by an API, passing the new status and the call object by parameter.

F. Graphical User Interface

The graphical user interface (GUI) must be aimed at the elderly, taking into consideration all the impairments that can be associated to them, such as poor sight and hearing, reduced eye-hand coordination, technological handicap and a low level of literacy. Therefore, the design of the GUI was envisioned by accord of the following features:

- Interaction oriented to be used with simple touch screen technology;
- Large and non-invasive buttons;
- High contrasting colours;
- Meaningful images representing all the actions;
- Use of text, only when necessary, with a large font without serifs;
- Pictures identifying the contacts wherever possible;
- All video in full screen;
- Feedback with notification sounds.

IV. Evaluation

The “Voice Commander” system was tested on a group of selected individuals, checking the amount of errors and time span of the tested operations determining if the elaborated system and architecture was usable by the individuals, and if it can actually aid in the solution of the exposed problem.

After the test, the users were interviewed with a simple questionnaire in order to acknowledge the system’s utility in real life, the impression of the senior citizens and improvements to make.

A. Evaluation scenario

The evaluation scenario consisted of eight senior citizens, divided into two rooms, executing the system operations. They were paired, and each pair had a desktop computer, with a large 17” screen, speakers, a microphone and a webcam, along with a mouse and keyboard.

The tests consisted of a brief explanation and demonstration of the system by the examiners, and then of the execution of the systems tasks in turns by the tested users, while the examiners took notes and counted the elapsed time and errors.

B. Tested users description

The tested users were all over seventy years of age, with the average age being eighty-one. Out of the eight users, six were female, three were illiterate and only one had ever had direct contact with computers.

C. Tasks and evaluation method

This application allows the user to execute commands vocally, and the tested operations were sending e-mail, receiving e-mail, making a VOIP call and receiving a VOIP call. To initiate communication (by sending an e-mail or making a VOIP call), the user must press a button to initiate the voice command recognition process, speak out the operation’s voice command and speak out the destination’s contact voice command. At this point, when sending an e-mail, the system records the voice message to send, and to send the message the user must press a button. When making a VOIP call, at the same point the user must press a button to start the call, and then press another button to terminate. To read an e-mail, the user must press the button to initiate the voice command recognition process and press a button to hear the received message. To receive a VOIP call the user must press a button to answer the call.

The examiners kept track of the necessary time to fulfill each task and the number of occurred errors. Only the time from when the desire of performing a task to the successful completion of that task was taken into consideration, e.g., the time elapsed trying to press a button or to recognize a voice command correctly.
Each time a user had to repeat a voice command due to improper recognition (false negatives) or could not understand which button he needs to press, an error was accounted.

D. Evaluation Analysis

The data present in the questionnaires and examiners notes was gathered and analysed, and the operations and results are described next, displaying the time average and error count average for each tested operation.

The objective of the analysis was to understand if the users could perform the operations, and if the fulfilment time was reasonable (under thirty seconds) with a low error rate (under five errors).

- “Send e-mail” operation – This operation consisted in pressing a button to initiate the voice command speech recognition, speaking the “send e-mail” voice command, speaking the destination contact’s voice command, dictating the message and pressing a button to send the e-mail message, and the results are presented in Figure 5.

This figure illustrates the time and error average for completion of the “send e-mail” operation, by gender. As presented, the operations were executed in an average time of eighteen seconds with an error rate of two. The female users had slightly poorer results than the male users.

- “Make video call” operation – This operation consisted in pressing a button to initiate the voice command speech recognition, the speaking of the make video call voice command, speaking the destination contact’s voice command, pressing on a button to make the call and then on another button to terminate the call, and the results are presented in Figure 7.

This figure shows the time and error average for completion of the “make video call” operation, by gender. In this operation, both genders completed the task within twenty seconds, with the females almost a second faster than the males, although the males had one fewer error, on average. Nevertheless, the operation was completed efficiently, within the pre-established values.

- “Receive video call” operation – This operation consists in pressing a button to answer an incoming video call and pressing another button to terminate it. The results are presented in Figure 8.
This figure illustrates the time and error average for completion of the “receive video call” operation, by gender. Worthy of note is that the error average is zero for both genders, and the average time of completion was six seconds, due to the simplicity of the operation. The female users were slightly slower than the male users.

E. User observations

The users started the experience very reluctantly, concerned about the system and technology, but after realising how simple and useful the tested system is, they became receptive and excited. Out of all the eight tested users, only one said he would not use the system in the future, while half the users affirmed the system could increase their quality of life.

Their final opinions on the most important features in a system of this type were the simplicity of interaction and task performance and the possibility of seeing the video of other person on the line during a VOIP call.

V. CONCLUSIONS AND FUTURE WORK

The elderly are somewhat excluded socially and technologically. In a world with an aging population, this problem becomes of a greater importance. A modular architecture and a system based on that architecture were presented, designed to solve that problem. The architecture consisted in the modification of human-computer interaction by the means of voice command recognition, focused on senior citizens.

The testing of that system demonstrated that it can be easily used by the test subjects. The subjects were over seventy years of age and without previous computer knowledge. Therefore, it can be concluded that the system successfully permits the elderly to interact with the computer.

In particular, the system was designed for communication, allowing synchronous and asynchronous communications with the aid of e-mail and VOIP technologies; however, it has been designed with the modularity of the architecture, permitting the addition of new functionalities in real time execution.

Although an ideal solution is utopic, this system can easily pass as an optimal solution to the presented problem, if applied with the correct hardware equipment. It is also rewarding to know the elderly tested were at ease with the system, commenting on its high utility and on how it can increase their quality of life.

A. Future work

In the future, the addition of new functionalities to the existing system is intended, having all the described characteristics in consideration.

VI. REFERENCES

“One-Button” Brain-Computer Interfaces
Jason Colman¹, Paul Gnanayutham²
¹SAE London, 297 Kingsland Road, London E8 4DD, United Kingdom
²School of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth,
PO1 3HE, United Kingdom
{j.colman@qantm.co.uk, paul.gnanayutham@port.ac.uk}

Abstract — The number of people with brain injuries is increasing, as more people who suffer injuries survive. Some of these patients are “locked in” their own bodies, aware of their surroundings but almost entirely unable to move or communicate. Brain-Computer Interfaces (BCIs) enable this group of people to use computers to communicate. BCIs tend to be hard to navigate in a controlled manner, and so the use of “one button” user interfaces is explored. This kind of interface is the simplest, and is the most universally accessible. It may be a useful “stepping stone” for a disabled person before he or she attempts to use a more sophisticated interface.

I. INTRODUCTION

People who have suffered a brain injury may have difficulties communicating. In the most extreme case, the patient may be non-verbal and quadriplegic. Some patients are cognitively intact but unable to communicate at all, which condition is termed “locked in syndrome”. The authors are particularly interested in improving accessibility for this neglected group of people, in areas such as communication, recreation, and controlling the environment.

This paper describes work, currently in its initial stages, which aims to provide access to off-the-shelf software, using a “one button” interface.

“One button games” are games in which the only control is a single button, which may be pressed or not pressed. At first, this seems a very limiting user interface. However, Berbank-Green [1] discusses one-button games and lists many ways in which games can be played using only one button.

A one-button interface, as the name suggests, has only one control: a button which can be pressed or not pressed. This is the most minimal control a user can exercise, and so is the most “universal”, in the sense of being accessible to the maximum number of users [15].

Such an interface clearly has its limits, and will not be suitable for all types of software. In this paper we discuss contexts in which a one-button interface will bring benefits to severely disabled people, by providing an immediately usable interface.

II. BRAIN INJURIES

A. Injuries to the brain

A traumatic brain injury (TBI) is an acquired brain injury caused by trauma such as a blow to the head, an impact with a blunt object, or penetration by a sharp object [23]. Common causes of TBI are motor vehicle accidents; bicycle accidents; assaults; falls; and sports injuries [23], [16] (p. 216).

The primary mechanism in many cases of TBI is diffuse axonal injury, i.e. widespread damage to axons (brain cells) caused by shearing or rotational forces [23]. At the microscopic level, the direction of the shear may be visible [16] (p. 218).

Other causes of acquired brain injury which are not classified as TBI include: cerebrovascular accidents (CVA, i.e. stroke), tumours, degenerative diseases (e.g. Parkinsons), demyelinating conditions (e.g. multiple sclerosis) and infectious disorders, (e.g. encephalitis) [19].

B. Numbers of people with brain injuries

Powell [24] reports that approximately one million people in Britain attend hospital every year as result of head injury. The incidence of disabled survivors is 100-150 per 100 000 – or more than 120 000 people in the UK suffering from long-term effects of severe head injury.

Improvements in road safety have reduced the number of people who suffer a head injury. For example, Cook and Sheikh [6] report a 12% reduction in bicyclist head injuries in England between 1991 and 1995, ascribed to the increased use of bicycle helmets over the period. Reductions in drink-driving and increased use of seat belts, crash helmets and air bags have reduced the incidence of head injury in many countries [16] (p.216). However, as medical care has improved, the number of people who survive a brain injury has increased [23]. Powell [24] reports that the number of brain injured people has increased since the 1970s, because the mortality rate has dropped since that time.
C. Assessment of brain injury

When a person suffers a moderate or severe brain injury, they will enter a comatose state. During this period, it is possible to assess the severity of the injury by gauging the responsiveness of the patient. The Glasgow Coma Scale, developed by Jennett and Teasdale, is commonly used [23]. Upon regaining consciousness, the patient will experience a period of post-traumatic amnesia (PTA). The period of PTA is judged to have ended when the patient is able to form new memories [23].

The periods of the coma and of the PTA give a reliable indication of the severity of the brain injury. A coma period of more than six hours, or PTA of more than 24 hours is classified as a severe injury, which accounts for 5% of all head injuries [24]. Other methods of evaluation are more suitable for assessing the patient’s longer-term prospects of recovery. These include the Rancho Levels of Cognitive Functioning [13].

Some patients remain in the comatose state, or transition to a persistent vegetative state (PVS). PVS patients are unable to move or communicate, and are not aware. Some other patients are cognitively intact and aware of their surroundings, but are unable to move or communicate. This condition is known as locked-in syndrome.

Recent cases have been reported of patients who were misdiagnosed as being in PVS, when they were in fact locked in. Monti and team [18] describe patients who are outwardly non-aware and non-communicative, but who can answer questions using MRI scanning. As patients diagnosed as PVS are more routinely scanned for cognitive activity, so the number of diagnosed locked-in patients may increase, and the number of PVS patients decrease correspondingly [18].

D. Consequences of brain injury

Powell [24] lists the effects of brain injury most often noted by relatives of the injured person. These effects include: personality changes; slowness; poor memory; irritability; bad temper; tiredness; depression; rapid mood changes; tension and anxiety; and threats of violence. Murdoch and Theodoros [19] note that disabilities due to TBI typically include physical impairments, cognitive, communicative and swallowing disabilities; and that motor speech and language disorders hamper interaction with family and friends, leading to isolation. There is a high frequency of depression and anxiety after an ABI [17].

E. Rehabilitation after a brain injury

As medical technology advances, more people survive brain injury. However, survival is not the same as quality of life. Rehabilitation is the process of regaining lost skills, or developing coping mechanisms to replace them, allowing cognitively impaired people to function safely, productively and independently [17]. Rehabilitation has two stages: the acute stage, where medical professionals stabilise the patient. The second stage is where family and carers take over [24]. Mateer [17] highlights some general principles for successful cognitive rehabilitation. These include: the need to take an individualised approach; to involve clients and caregivers in all aspects; and clear and realistic goal setting.

III. BRAIN-COMPUTER INTERFACES

A Brain-Computer Interface (BCI) is a system for controlling a computer that does not depend on the brain’s normal output pathways such as speech or gestures. Instead, a BCI will use any of the bio-potentials which are under the conscious control of the user [10]. For people with extremely limited motor ability, a brain-computer interface is the only way in which they can use a computer.

A. Bio-potentials

Bio-potentials are electrical signals originating in the brain and nervous system. The existence of electrical currents in the brain was first discovered in 1875 by Richard Caton [27]. These can be detected and used to control hardware and software.

Bio-potentials may be detected in two ways: invasive and non-invasive. Invasive methods involve surgery to place electrodes within the body or brain; non-invasive methods take measurements from the surface of the body. Invasive techniques provide higher amplitude signals with improved signal to noise ratio, but carry the risks of surgical procedures. In this study, we consider the use of only non-invasively measured bio-potentials: electroencephalography (EEG), electromyography (EMG), and electrooculography (EOG).

Electroencephalography (EEG) is the measurement of electrical waves produced by the brain. The existence of these regular waves was first published by Hans Berger in 1929 [2].

These waves have amplitudes ranging from approximately 1μV to 100μV at the surface of the scalp. The frequencies measured range from approximately 1Hz – 30Hz, the dominant frequency depending on the person’s mental state [5], [27].
Electromyography (EMG) is the measurement of electrical signals originating from muscle movement. These signals have the same frequency range as EEG and an amplitude range of 0.2 to 2000\(\mu\)V [12].

Electrooculography (EOG) is the measurement of electrical activity caused by eyeball movements. The range of frequencies is relatively low, from 1.1 to 6.25 Hz. The amplitude is higher than EEG, around 1 - 4mV [12].

Other non-invasively measured bio-potentials may be used for BCIs, but are not used in this study. These include evoked potentials, (e.g. P300 and N400); steady-state visual evoked potentials; and slow cortical potentials [12].

E. Commercially available Brain-computer interfaces

BCI hardware ranges from devices intended for playing computer games through to medical-grade EEG machines. The following table shows currently available consumer-level BCI hardware. These only measure non-invasive bio-potentials.

Table II: Commercially available BCI hardware. Prices are approximate.

<table>
<thead>
<tr>
<th>Name</th>
<th>Manufacturer</th>
<th>Approx Cost in £</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyberlink™</td>
<td>Brain Actuated Technologies Inc [3]</td>
<td>£1400</td>
</tr>
<tr>
<td>Neural Impulse Actuator™</td>
<td>OCZ Technology [22]</td>
<td>£85</td>
</tr>
<tr>
<td>Enobio®</td>
<td>Starlab [26]</td>
<td>£3150</td>
</tr>
<tr>
<td>EPOC</td>
<td>Emotiv [7]</td>
<td>£200</td>
</tr>
<tr>
<td>Mindset</td>
<td>Neurosky [20]</td>
<td>£130</td>
</tr>
</tbody>
</table>

In this study, the Cyberlink™ hardware with Brainfingers software has been used. Cyberlink was used as it has been successfully shown in studies [8] to enable locked-in patients to communicate.

Cyberlink/Brainfingers lets the user control the mouse cursor and mouse button clicks using bio-potentials. The software is configurable, so that different users can control the mouse using different EEG frequency bands, and also EOG and EMG, if appropriate.

IV. USABILITY OF BRAIN-COMPUTER INTERFACES

Participants invariably have a lot of difficulty in controlling the mouse cursor with Cyberlink. To move the mouse cursor at will, the user must be able to consciously control four separate 'channels' of bio-potential: one channel to move the cursor up, one to move it down, one for left, and one for right movement. Adding the ability to generate mouse button events further complicates the task facing the user. This difficulty means that in practice BCIs are difficult to use. Typically when using Cyberlink, the mouse cursor moves quickly to a corner of the screen and then stays there. This frustrates users, making it even harder to bring the cursor back under conscious control.

These difficulties have been addressed by developing the novel User Interface paradigms, Discrete Acceleration and Personalised Tiling [9]. Another approach, discussed here, is to make the interface easier to use by reducing the number of channels which the user must control. The simplest possible configuration is a one-button interface, requiring only one channel of information. To use this kind of interface, the user only needs to be able to consciously control one bit of information over time. The advantage of such an interface is its simplicity. Being the simplest kind of interface, it is as “universally accessible” as possible.

V. EVALUATING A ONE-BUTTON INTERFACE

To investigate the difficulty of using Cyberlink, a focus group was convened (six programming students, all male, age range early twenties to early thirties). The focus group participants were able-bodied.

A. Methodology

Standard methodologies for HCI design, e.g. Usability Engineering [21] or Contextual Design [14], stress the importance of “knowing the user” [21] and so evaluation with the intended users of the system is the norm.

Designing software for people who are severely disabled by brain injury is challenging, for reasons including the person’s communication difficulties and medical needs [11]. Because of this, in the case of designing for severely disabled people, a different methodology is called for. Gnanayutham and George [11] provide case studies where initial investigations are carried out with able-bodied participants, before evaluation with disabled participants begins.

In this study, a similar methodology is followed.

• The development process is iterative, as the most useful artefacts must be evolved and refined from earlier prototypes.
• Prototypes are initially tested using able-bodied participants.
• Summative evaluation is used to measure the usefulness of the prototypes.
Formative evaluation takes account of users’ perceptions throughout the development cycle.

The process could be thought of as a spiral, because we seek to iteratively improve a design based on feedback; and the circle of participants expands over time (fig. 1).

Figure 1. Methodology

B. Design

The focus group participants were asked to solve a “Fifteen Puzzle” [4] using Cyberlink. The fifteen puzzle was chosen for its familiarity and simplicity. The puzzle consists of 15 numbered tiles and a space arranged in a 4x4 grid. A tile horizontally or vertically adjacent to the space can be moved into it. The puzzle was “shuffled” by making 100 moves at random, the goal of the puzzle being to restore it to its initial state. Figure 2 shows the puzzle in its initial state (a) and shuffled (b).

The action required of the user was to generate a mouse-click event at the appropriate time, using Cyberlink. The tiles were “scanned”, i.e. highlighted one at a time, each for a period of one second, in numerical order. When the mouse button was “clicked”, the highlighted tile would move to the space. Figure 3(a) shows one tile highlighted. When the user generates a mouse-click event using Cyberlink, the highlighted tile moves to the space, as shown in fig. 3(b). The “scanning” technique has been used for numerous augmentative and alternative communication schemes [25].

Only moveable tiles were highlighted. It was recognized that the “artificial intelligence” (AI) of a user interface must not be intrusive. However, in the case of the 15 puzzle, most tiles cannot be moved. Scanning all 15 tiles would result in a lot of wasted time and frustration for the user; and so it was decided to only scan the tiles which could be moved into the space.

Variations in the UI elements were tested, to see if some visual cues would improve the one-button interface. One visual cue was to show, underneath the puzzle, the tiles which would be highlighted, in the order in which they would be highlighted. The other was to show a “progress bar” on the highlighted tile, showing how much longer the tile would be in the highlighted state. The variations were numbered as shown in table II.
The mean time per click and mean number of errors per click were measured, for the four different UI types. An error was recorded if a tile was moved twice in succession. The design of this evaluation was within-subjects, with the order of the UI variations randomized, to counterbalance learning and fatigue effects.

Pilot testing revealed that solving a thoroughly randomized 15 puzzle was difficult for some participants, so the puzzle goal was simplified. The new goal was to make the top row of tiles all the same fruit type. It was felt that this change would not affect the measurement of mean time per move and mean number of errors per move, and would put participants under less pressure to “perform”.

C. Results

Mean time per tile movement, for the four UI variations, is shown in table III.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Mean time to move a tile for UI version</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>2.22</td>
<td>2.10</td>
<td>2.06</td>
<td>2.46</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2.64</td>
<td>2.38</td>
<td>2.00</td>
<td>2.13</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2.37</td>
<td>3.03</td>
<td>2.13</td>
<td>2.37</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2.36</td>
<td>2.05</td>
<td>1.93</td>
<td>2.01</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1.88</td>
<td>2.78</td>
<td>2.38</td>
<td>2.14</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>1.74</td>
<td>2.18</td>
<td>2.08</td>
<td>2.13</td>
</tr>
</tbody>
</table>

Mean number of errors per tile movement, for the four UI variations, is shown in table IV.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Mean no. errors per tile movement per UI version</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0</td>
<td>0.08</td>
<td>0</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.09</td>
<td>0.05</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0</td>
<td>0.08</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Feedback from the focus group members was that the one-button interface was immediately usable, compared with 2-axis mouse cursor control.

Mean time to move a tile was close to 2 seconds, regardless of UI variation. For each puzzle run, 15% or fewer moves were mistakes, counted as a tile being moved and then immediately moved back to its former position. Half of the puzzle runs had no mistakes.

Participants commented that highlighting the tiles in a consistent order, e.g. always clockwise, would be an improvement over numerical order. The participants also suggested other ways to improve the interface by reducing the amount of time spent waiting for the
chosen tile to be highlighted. These were to add a “double click” or “hold” action to speed up scanning.

D. Interpretation of results

Adding a progress bar to the highlighted tile received favourable comments from the participants. However, this did not result in any significant improvement in mean time to move a tile, or number of errors. Neither did displaying all tiles which would be highlighted. Indeed, this may have been a distraction.

Scanning time was one second per tile. On average there are three tiles which may be moved into the empty space. The mean time of just over 2 seconds per tile movement suggests that users were able to move a tile the first or second time it was highlighted. It may be that this time would be hard to improve upon, whatever UI improvements were made.

The low error rate and low time per tile movement shows that the one-button interface was easy to use compared with a 2D cursor-control interface. However, the participants’ comments and ideas for speeding up scanning suggest that it can become frustrating waiting for the chosen tile to be highlighted.

VI. DISCUSSION

The focus group findings show that one-button interfaces are quickly usable and offer a low error rate. This suggests that it may be fruitful to design a one-button brain-computer interface that would work with off-the-shelf software. One might call this an “accessibility layer”.

It is the authors’ view that it is better to attempt to make existing, “off-the-shelf” software accessible, rather than to write new software with accessibility features. The reasons are that writing new software is expensive and time consuming; and a small number of researchers cannot hope to provide every type of software required.

The “accessibility layer” would be used in two phases: a configuration phase, and a run-time phase. During the configuration phase, rectangles representing clickable areas would be drawn on the screen, over the software to be used. This phase would probably be carried out by an able-bodied person. During the run-time phase, the rectangles would be scanned, i.e. highlighted in turn, and the mouse cursor moved to that location. This would enable a disabled user to click on a button or other UI element in the application by generating a mouse-click event, using BCI hardware such as Cyberlink.

This kind of interface would only be usable with certain types of software, i.e. those based on clicking buttons in dialog boxes. Many applications also require typing. The interface could be extended to also emulate key presses by scanning a software keyboard when required. The software keyboard could use a scanning algorithm designed to reduce the waiting time for the user as much as possible, e.g. one of the algorithms described in [25].

VII. CONCLUSIONS AND FUTURE WORK

The number of people with brain injuries is increasing, as medical care has improved. Some of these patients are cognitively intact, but cannot communicate, except by using a brain-computer interface (BCI). The number of people diagnosed with this condition may increase if diagnostic tests such as those described in [18] become widespread.

BCIs can be difficult to use, and can require a lengthy training period. A “one-button” interface is simpler, and so easier to use, with less training. This type of interface is limiting due to its simplicity, but could find use as a first “stepping stone”. When a user outgrows the one-button interface, he or she is ready to move on to an interface that is more sophisticated. It is the authors’ belief that the confidence gained by successfully using the one-button interface would help the user, as learning to use a more sophisticated interface may be difficult and frustrating. A one-button interface would not replace a 2D cursor interface, but rather would complement it.

We have outlined a design for an “accessibility layer” allowing a one-button interface to be applied to off-the-shelf software. The types of software to which this could be applied are currently limited. Future work would concentrate on designing accessibility layers for more varied types of software, and on making common applications and operating systems more accessible.

REFERENCES


http://www.xs4all.nl/~hc11/15puzzle/15puzzen.htm
Retrieved 6 July 2010


[7] Emotiv/EPOC headset
http://www.emotiv.com
Retrieved 6 July 2010


http://rancho.org/patient_education/cognitive_levels.pdf
Retrieved 9 July 2010


[20] Neurosky/Mindset headset
www.neurosky.com
Retrieved 6 July 2010


[22] OCZ Technology/NIA headset
www.ocztechnology.com
Retrieved 6 July 2010


[26] Starlab/Enobio headset
www.starlab.es/products/enobio
Retrieved 6 July 2010

Running Graphical Desktop Applications on Tactile Graphics Displays Made Easy

Christiane Taras, Michael Raschke, Thomas Schlegel, and Thomas Ertl
Institute of Visualization and Interactive Systems, University of Stuttgart,
Universitaetsstr. 38, 70569 Stuttgart, Germany
Email: taras, raschkml, schlegel, ertl@vis.uni-stuttgart.de

Abstract—Tactile graphics displays have improved over the years and their importance for blind people and their education has been shown. Unfortunately, most of these devices can be used only with applications specifically developed for them. In this paper, we present a new framework for tactile rendering which makes it easy to run graphical desktop applications on a tactile display. Our framework utilizes a filter program to analyse the application’s user interface without having access to the application’s source code. The output is converted to a tactile user interface with the help of configurable rendering methods. As the rendering is based on widgets that can adapt their layout to the target output size, the tactile rendering can easily be adapted to devices of different sizes. We present the concept of our framework and demonstrate how it can be applied to existing applications using Windows Solitaire as an example. The framework was developed within the HyperBraille project, but is encapsulated into independent components that can be easily reused in other projects. We believe that our framework will ease promoting the development and spreading of tactile graphics displays.

I. INTRODUCTION

Graphical information plays an important role in our daily life. Whether it is a diagram in information materials, a graph of a mathematical function, a part of a graphical user interface or something like a tile in a web-based scrabble game - graphics are everywhere. For visually enabled people, graphics support understanding and give joy to the life. The same is for blind and visually disabled as for example Aldrich and Sheppard [12], [24] showed. For that reason, a lot of research is done in the field of preparing tactile graphics for school books, maps and many more (see for example [4], [5], [6], [7], [8] and [18]).

Due to the complexity of most graphics, interactive exploration of tactile graphics is much more useful than working with static ones alone. Therefore, for several years different groups of researchers have been working on the development and improvement of tactile graphics displays. Vidal-Verdú and Hafez [25] gave a good overview on previous work. Watanabe, Kobayashi et al. showed that tactile graphics displays can heavily support the communication between teachers and blind children [27] and between blind persons when talking about graphical information [15]. Furthermore, tactile graphics displays reveal the possibility of rapidly changing the presented graphic which helps in keeping the graphics up-to-date. So it’s worth performing research in the field of tactile graphics displays.

Unfortunately, this field has two major problems. The first is, that the displays themselves are quite expensive or not at the market at all. The second is, that there is not enough useful software that drives these displays. Until now, only some special applications like [14], [21] and [22] were developed for tactile graphics displays. The framework that we present in this paper is meant to contribute in solving the second problem. It does not provide guidelines how to design an optimal tactile user interface. But it provides a technical means to easily convert ordinary graphical user interfaces into tactile representation. Thereby, different possibilities to optimize the rendering according to users preferences or characteristics of special applications were implemented. Only if a large number of ordinary desktop applications can be run on tactile graphics displays easily those devices will be accepted as valuable working equipment for visually impaired people. This would raise market demand and thereby through higher production output the prices of such displays probably would decrease. So, besides improving the displays and their production processes themselves, providing frameworks for valuable software is a very important aspect to promote tactile graphics displays.

The framework was developed as part of the HyperBraille project [1]. The goal of this project is to develop a touch-sensitive tactile graphics display combined with a graphical screen reader to better support blind and visually impaired people in controlling graphical user interfaces. The current prototype of the display is the BrailleDis 9000 [26] which was enhanced during the project. The BrailleDis 9000 provides 120×60 pins that can either be lifted or lowered. It’s surface is touch-sensitive. So, performing input via gestures on the display is possible.

II. PRESENTING GUIs TO BLIND USERS

Currently, blind people mostly use screen readers to access graphical user interfaces. The screen reader analyses the user interface and gives information about it via speech or Braille output. Soon it was recognized, that not only real text is important for the user but also information about other interface elements has to be presented. So, for example, also the type and state of an interface element is spoken or presented in Braille. For the Braille output it is also possible to present simple layout information. Thereby, the user can, for example, find out if an element is aligned to the left of a dialog or if a number of elements are aligned in one line. As Braille displays
can only present a small number of signs the type information is often shortened. For example, the buttons of the message box "Do you want to save the changes?" are presented as "btn Yes btn No btn Cancel". The active button is often underlined or marked with a blinking cursor. But, all in all, the information stays textual. The user has to distinguish if the presented information originates from real text or describes a user interface element.

For tactile graphics displays one can go a step further. There, one is not bound to textual information but can also use borders to surround text or group elements or use small graphics that can even exceed the bounds of ordinary Braille signs. So a kind of tactile widgets can be created that combine graphical elements with Braille text and thereby ease accessing graphical user interfaces with tactile displays. Graphical elements can represent types, states and grouping of interface components in a much more compact and flexible way than text alone can do. For this reason, in the context of the HyperBraille project, Kristin Albert designed a set of tactile widgets [11] to represent ordinary graphical user interfaces. Therefore, she analyzed visual widgets for their interaction possibilities. From this she created a modular widget set with a small number of tactile modules which represent the interaction that is possible for a widget (see 1). Those modules are combined with different border patterns to compose the tactile widgets (see 2).

Albert’s widgets were utilized in different user studies during the HyperBraille project (e.g. [17]). First user feedback has shown, that tactile widgets are a useful means for presenting graphical user interfaces on tactile graphics displays. Of course, the widgets are further investigated and improved. For example Denise Prescher evaluated a set of scrollbars for a tactile windowing system [20]. Furthermore, Albert’s widgets were partly derived from TAWIS, the graphical screen reader of Friedrich Lühti. A blind man who has long time experiences with working on a tactile graphics display. Unfortunately, his work is unpublished.

To further promote the tactile widgets and so establish a common "Look&Feel" for tactile user interfaces, we also use them in our tactile rendering framework for the default rendering. For the composition of a whole tactile user interface out of those widgets, we try to keep the original layout. By this, on the one hand, we keep groupings that are expressed only through positioning of widgets and not logically in the user interfaces widget hierarchy. On the other hand, keeping the original layout supports communication between users of the graphical interface and users of the tactile interface. User studies during the HyperBraille project have shown that this is a useful approach (see for example [23]).

III. OUR TACTILE RENDERING FRAMEWORK

As figure 3 shows, our tactile rendering framework supports the whole process necessary to run a graphical desktop application on a tactile graphics display. This includes configurable tactile rendering, as well as interaction handling. Thereby, it is neither bound to a certain application nor to a special display.

A. The Rendering Process

To apply our framework to a graphical desktop application only its executable is needed. This is plugged to the framework using an application filter. The filter analyses the application’s user interface at runtime and provides the rendering framework with an object-based representation of the interface elements that have to be rendered. This includes information about the hierarchical composition, as well as attributes like type and name of the user interface elements. Using an object-base representation instead of just the graphical output of the application facilitates redesigning the user interface for the tactile representation. Different filters can be connected to the renderer to support various applications. In HyperBraille we have developed a couple of them for different applications like Microsoft PowerPoint, Microsoft Internet Explorer, Mozilla Firefox and ordinary Windows applications that support UIAutomation [9]. The filter also has to track changes in the application’s user interface and inform the rendering framework about this. Figures 4 and 5 illustrate the output of a UIAutomation filter for an ordinary dialog.

The filter’s output is converted into a tactile user interface via exchangeable rendering methods. The tactile user interface itself is built up with user interface components as known from graphical user interfaces for sighted people. In short, building the tactile output with our framework is like composing an ordinary GUI. By this rendering rules which were defined by other developers can easily be influenced. If the rendering methods only provided bitmaps, it would for example be hard to change the cell spacing of a table. Using widgets this is just a matter of changing the appropriate attribute. Furthermore, the tactile output can easily be adapted to differently sized displays by providing appropriate layout algorithms.

When the composition of the tactile user interface is finished, it is provided to the tactile graphics displays as bitmap. Different displays can be connected to our framework and even be driven synchronously at the same time. The output size can be configured according to the connected displays.
Display handlers converting the bitmap into appropriate pin state definitions can be registered with the renderer. Of course, for the BrailleDis 9000 this is provided with the framework. In our implementation we have used a special bitmap format encoding pixels originating from Braille output and blinking pixels in addition to the pin up or down information. We also provide a graphics library to handle this special format and convert it to ordinary black and white bitmaps. The Braille encoding mainly was realized to support sighted developers with a readable view of the tactile output. But it can also be used to provide speech output for the tactile user interface. In the following chapter, we have used the developer’s view to present some of our results. This view is synchronized with the tactile output supporting the communication between sighted and visually impaired users.

For the implementation of the interface components, we have utilized Windows.Forms. This provides everything we needed. By inheriting from Windows.Forms.Control [3] we can easily build up a widget hierarchy and also use predefined layout controls of the .NET Framework. Using the DrawToBitmap method, the tactile user interface can be rendered to a bitmap without presenting it on the screen. Furthermore, developers can test their widgets’ rendering behaviour with an ordinary Windows.Forms project.

For the default rendering of common user interface elements, like buttons, menu items and labels, our framework provides implementations of the widgets mentioned in section II. Thereby, the widget modules used to mark the element type are realized by our interpretation of tactile icons, so-called tactons [13], for tactile graphics displays. Our tactons are fixed combinations of lifted and lowered pins that always have to be shown in full size and may not be scaled or cut. So the tactile impression always stays the same and serves to identify the widgets. For the different borders to present enabled and disabled state or special groupings an implementation of dash patterns is provided. Those patterns can be defined separately for each of the four borders (top, left, right and bottom). Furthermore, layout algorithms were implemented, to align widgets according to the original layout but at the same time...
generate a compact presentation that requires as less scrolling as possible. Of course, our framework is not bound to the default widgets. New widgets can be integrated easily.

B. Influencing the Rendering Result

The rendering result currently can be influenced in different ways. One way is to change the tactons used as markers for the different widget types. This can easily be done via replacing the tacton library used for rendering, which is a simple XML file. There the tactons are defined by patterns of '0' for lowered pins and '1' for lifted pins. A name is given to identify each tacton. Changing the tacton library influences the rendering of each user interface. It is meant to easily adapt the "Look&Feel" of the tactile user interface representations to a single users preferences or to new findings from user studies. It was kept as simple as possible so that it can also be changed by blind people themselves.

Of course, replacing tactons can influence the rendering only literally. The layout of the interface remains the same. Furthermore, it is also not possible to overcome filtering problems. For example, a clickable label will be reported as label by a UIAutomation filter and not as button. So for this label the press marker will not be used, although it should be. Changing the layout and implementing special rendering for single interface components is possible by creating own rendering methods. Such rendering methods can be registered with the render via the event handling mechanism of the .Net Framework. In those methods own widgets can be created and used. They only have to be derived from Windows.Forms.Control and implement the Paint method to provide an appropriate bitmap output for the tactile rendering. The event handling mechanism was used for the registration as it is the most flexible way and applies almost no restrictions. Furthermore, it is compatible with the Add-In pipeline that is used in the HyperBraille software and based on the Add-In-architecture of Microsoft .NET [2]. Self-defined rendering methods can also retrieve the widget representation of a certain object as it would be if this method was not registered. So, small changes to the existing rendering can be applied easily. That means, if for example someone just wants to change the representation of a label within a dialog box, it is not necessary, to implement the rendering for the whole dialog, but only for the label to change. The other dialog elements as well as the layout of the elements will be provided by the default rendering of our framework. The interface defined for the rendering methods is not restricted to a special input type. The rendering can be adapted to whatever content the application filters generate. Be it a UIAutomation tree or a string including the serialization of a graphical user interface.

Furthermore, the Braille output can be influenced easily, even without programming. The Braille font used for rendering is also defined in a simple XML file that assigns unicode characters to pin state patterns. So one can easily switch between different languages or define Braille signs for special characters in mathematical expressions. As Braille text is quite large related to the size of tactile displays, we also implemented the possibility to shorten text easily. Therefore, another XML file was defined containing pairs of text patterns to search for and appropriate replacements. So one can, for example, shorten "Replace with" by "Repl. with" or "User manual" by "Help". This saves much space and so helps to reduce the scrolling effort.

C. Input Handling

Our framework not only handles different displays as output but also as input devices. Therefore, display handlers can be connected that provide the renderer with a scrolling and a click event. The scrolling is used to move the output of the tactile user interface on the display. Of course, not every user interface will fit on the display at once. The click event is used to control the rendered graphical desktop application via the tactile graphics display. Besides text input, click handling is sufficient for controlling the majority of graphical desktop applications. Every user interface element that was converted into a part of the tactile user interface can be clicked and the click will be forwarded to the original application. Therefore the tactile display handler has to provide the renderer with the coordinates of the clicked pin. The renderer then retrieves the widget of the tactile user interface that was rendered to this pin and requests the screen coordinates to click on for this widget. In our implementation these coordinates are set for the widget during its creation in the rendering method. So, the original user interface element does not have to be analyzed once more.

IV. Results

In the following, we use the Solitaire game (see fig. 6) to illustrate how the framework eases rendering a graphical desktop application to a tactile display. To analyze the elements of the user interface, we have used a generic UIAutomation [9] filter that is part of the HyperBraille software and converts UIAutomation elements into components of the HyperBraille off-screen-model [16].
As can be seen in figure 7, showing the tactile output for the options dialog of the Solitaire game, the default rendering of our framework produces quite good results for ordinary dialogs with standard controls. For rendering check boxes and radio buttons, we have used instances of the TactonWidget that is provided with our framework. Unfortunately, the borders of the groups are not visible in the tactile rendering as the groups are not defined via hierarchical composition of the group and the items they visually contain. So the information about the grouping is not available to the UIAutomation filter. This can also be seen in figure 5. Although, the groups are also represented in the UIAutomation tree (e.g. in the line "group" "Draw"), the items that are visually positioned within the groups are not children of the group elements. Also for the layout of the tactile user interface only the position of the elements can be used.

The default rendering of the Solitaire game itself is shown in 8. It shows that depending on the tactile display’s size, much scrolling and significant cognitive effort is needed to grasp the whole playing field. The UIAutomation filter reports all the cards in the game as buttons and panes containing the description of the cards in their names. As the default rendering uses the buttons’ names as captions for button widgets, it does not really result in an interface that supports users of the display. Of course, it is possible to interact with the resulting interface and so with the original application also with this rendering. But because of the long button texts it is quite hard to keep the presented cards in mind. Furthermore, it may also be confusing, that the cards are sometimes represented as buttons and sometimes as panes, although they can be clicked in both states. So for the Solitaire game itself, special rendering methods have to be provided to turn it into a useful tactile representation.

With just one day of work and 400 lines of code for special rendering methods, we have managed to realize a really useful representation for our 120×60 pins sized tactile graphics display. Thereby most of the time was spent in investigating the UIAutomation output of the game to figure out, how the different buttons and panes are named in different situations. This was necessary to create an optimized layout. Therefore, we have not only changed the rendering of the cards themselves, but also adjusted spacing between different areas of the playing field. Furthermore, we have rearranged the title, the menu and the status bar to fit into one single line. This illustrates how useful it can be to be independent of the original user interface’s layout. As you can see in figure 9 utilizing the new rendering methods the whole user interface fits onto the display at once. Scrolling is only necessary if the card stacks get really large.

To demonstrate how easy it is to use the framework, we finish this section by providing some excerpts of the code that is used for the rendering of the Solitaire game. The following lines realize the nice rendering of the cards. They are part of a small C# class, that in it’s constructor registers the rendering methods with the renderer that is provided by the framework as singleton. This automatically calls the rendering methods, puts the returned widgets into the widget hierarchy and creates the bitmap that is sent to the tactile display.

```csharp
// Define shortenings for the card types.
```
Dictionary<String, String> cardTypes = new Dictionary<String, String>() {
    {"Hearts", "h"}, {"Spades", "s"},
    {"Clubs", "c"}, {"Diamonds", "d"}};

// Shortenings for the values of the cards.
Dictionary<String, String> cardValues = new Dictionary<String, String>() {
    {"two of ", "2"}, {"three of ", "3"},
    {"king of ", "k"}, {"ace of ", "a"}};

// Rendering methods are defined as event
// handlers. They can access the renderer
// via the sender object. The content to
// render is given in the GetWidgetEventArgs.
// The widget for the tactile user interface
// is defined via assigning it to e.Widget.
// If e.Widget is left undefined, other event
// handlers are called with the same
// rendering content.
void WidgetForCard(object sender,
    GetWidgetEventArgs e) {
    // this method handles only buttons
    if (e.ToRender is Button)
    {
        String text = String.Empty;
        // Get the name of the button.
        String name = (e.ToRender as Button).Name;
        // Retrieve the shortenings to represent
        // the card's type and value.
        foreach (var kvp in cardTypes)
        {
            if (name.IndexOf(kvp.Key) > -1) {
                text = kvp.Value; break;
            }
        }
        foreach (var kvp in cardValues)
        {
            if (name.IndexOf(kvp.Key) > -1) {
                text = text + kvp.Value; break;
            }
        }
        // If the button really represented a card,
        // create a TextWidget with the shortenings
        // as text, a surrounding border and the
        // screen coordinates of the card as click
        // point as widget of the tactile user
        // interface.
        if (text.Length > 0)
            e.Widget = new TextWidget(text) {
                TopBorder = new SolidBorder(),
                LeftBorder = new SolidBorder(),
                Size = new Size(10, 6),
                Tag = GetClickPoint(e.ToRender)
            };
    }
}

Here we have used the class TextWidget that is provided by the framework for rendering text into Braille surrounded by a tactile border line. You can also see that the programmer does not have to care about the overall layout. Although, it can be influenced. But if the default layout is satisfying, it also does not have to be reproduced when changing some small parts of the rendering.

V. Evaluation

The rendering part of our framework is used to create the graphical tactile rendering of the HyperBraille software - the HyperReader. Thereby, it has been tested by at least ten developers everyday since one and a half year. A number of Add-Ins have been implemented by the project partners that register rendering methods to apply special rendering for applications like Visio (special rendering for UML shapes) and Excel (special rendering for formulas).

For the development of the Add-In methods the framework was often used independently of the HyperReader. So the framework proved to be useful for a kind of rapid prototyping for tactile representations of desktop applications.

Naturally, all adaptations to the default rendering were presented to our blind project partners and also other blind subjects interactively on the BrailleDis 9000 and were improved on basis of their comments. User evaluations showed that representations should be as compact as possible. Representations should always try to avoid scrolling (especially horizontal scrolling) as far as possible just like shown in the Solitaire example. Also the configurable text shortenings have been accepted very well as they support compact layout. For known GUls users often only need to read a couple of the first characters of an element’s label to know which element is shown. Blind users performed configurations for text shortenings and type markers (tactons) by themselves.

VI. Conclusion and Future Work

We have presented ongoing work on a framework for running graphical desktop applications on tactile graphics displays. The example of Solitaire illustrated how easy it is to use the framework and adapt the default rendering for special components. Only a few lines of code are necessary to achieve really good results. We have shown that also for tactile user interfaces the use of widgets and layout routines is valuable as thereby applying small changes to a default rendering is easy.

As the rendering interface is open to arbitrary objects, it is possible to handle any application. The developer only has to provide an appropriate filter to analyse the user interface and rendering methods to handle the filter’s output. Of course, analysing the graphical user interface is easier the more accessible it is.

We believe that our framework pushes the work on tactile graphics displays one step forward from the software engineering view. Tactile graphics displays get more useful the more applications can be controlled with them. As the framework implements a set of default widgets and layout algorithms a number of ordinary desktop applications can be translated to tactile user interfaces easily. Furthermore, a common Look&Feel is promoted for tactile user interfaces, which makes it easier for users to cope with new applications. Of course, our framework cannot create an optimal tactile rendering for each and every application automatically. But as it will always produce a representation for each interface
element, it reveals the possibility that the users themselves can adjust the rendering by creating own rendering methods. So blind people can play a very active role in the design process of tactile user interfaces, just as they can currently improve the output of ordinary screen readers.

In the future, we want to provide more predefined widgets that can be configured and mapped to GUI elements without any line of code but just by XML configuration files in an even more flexible way than it is possible now. Of course, our framework is integrated in the graphical screen reader that is developed in the HyperBraille project and so will be available with this in the future.

ACKNOWLEDGMENTS

We thank our colleagues of the HyperBraille project for their support in the design of the rendering framework and the design of the tactile output itself. The HyperBraille project is funded by the German Ministry of Economy and Technology (BMWi) under the grant number 01MT07004. Only the authors of this paper are responsible for its content. The Consortium consists of Metec AG, F.H. Papenmeier, Human-Computer Interaction Research Group at the Technical University of Dresden, Visualization and Interactive Systems Institute, IZFM institute and IMS-Chips at the University of Stuttgart, the Research Group for Information Processing and Communication at the University of Potsdam, as well as Blista Marburg and T.O.P. GmbH & Co. KG Freiburg.

REFERENCES

Enhanced Autonomous Control in Interactive Multi-Sensory Environments

Kenneth Brown, Phil Ellis
Faculty of ADM, University of Sunderland, Priestman Building, Green Terrace, Sunderland, SR1 3PZ, UK
Phone: +44 191 515 3406, Fax: +44 191 515 2415, e-mail: {kenneth.brown, phil.ellis}@sunderland.ac.uk

Abstract - iMUSE (Interactive Multi-Sensory Environments) promote expression and wellbeing with disabled and older people through the expressive use of sound. This is enabled through the use of non-invasive technology so that participants have the opportunity to exert a degree of control. They can use hand or arm movements to generate sound with accompanying images and vibration. However control of secondary parameters such as the audio timbre and visualisation colour scheme currently have to be adjusted by the facilitator.

This paper describes methods of enabling autonomous control for selected aspects of iMUSE in order to enhance the health and wellbeing of the participants. The project focuses on research into potential user interface types and the development of corresponding hardware and software enabling control using simple participant actions. The ‘vibration level’ and ‘visualisation colour’ were the parameters chosen as being most relevant to the users.

The system is currently being trialled and some preliminary participant observations are given. It is already clear that the MIDI-to-control mapping methodology devised for this project has potential to extend both to other aspects of iMUSE, and to other projects requiring a similar methodology.

I. BACKGROUND

A. Current iMUSE Control

iMUSE1 is an environment that promotes expression and wellbeing with disabled and elderly people through the expressive use of sound. This is enabled through the use of non-invasive technology so that participants have the opportunity to exert a degree of control. Control is important for developing a sense of cause and effect which is a central aspect of the activity. They can use hand and arm movements to generate sound and accompanying images [1]. Sensing is enabled using a system called the Soundbeam2, a device which uses an ultrasonic beam. Movement in the beam generates MIDI signals which are then used to drive other MIDI controlled applications which generate the audio, visual and vibratory stimuli. A history of iMUSE and fuller description of its technologies and benefits are given in [2]. Control of secondary iMUSE parameters such as audio timbre and visualisation colour are currently adjusted by a facilitator: the person who operates the iMUSE PC using a keyboard and mouse according to the directly communicated (verbal) or indirectly communicated (facial expression/body language) preferences of participants.

B. Reason for Enhancing Autonomous Control

Rodin [3] states “interventions that enhance options for control by nursing-home patients promote health”. It is hoped that providing autonomous control for selected aspects of iMUSE will enhance the health and well being of participants by promoting further engagement and exploration within iMUSE and thus increase the effectiveness and benefits it has been shown to provide [4],[5].

C. Project Aim

The aim of the project is to research potential user-interface types and the corresponding hardware and software that will allow a relatively simple participant action such as pressing a button or turning a knob to perform potentially complex actions within iMUSE, resulting in changes to the participant’s perceived feedback from the relevant output. A number of physically different ‘styles’ of human interface have been prototyped for evaluation by two groups of participants, each group having differing cognitive and physical abilities. Because iMUSE is being increasingly adopted in schools for people with disabilities and care centres for the disabled or older people, the control needs to be implemented using an appropriately cost-effective technology, as these institutions often have limited funding available. Two particular parameters were selected for the provision of this enhanced autonomous control.

D. Choice of Parameters

Parameters of the iMUSE system called the “vibration level” [VL] and the “visualisation colour,” [VC] were chosen because they are two of the feedback stimuli most strongly perceived by participants, yet there is a high degree of variation in

---

1www.sunderland.ac.uk/imuse
2www.soundbeam.co.uk

The authors would like to thank The Linbury Trust without whose generous support this project could not have taken place.
participants’ subjective preferences relating to these elements. For this reason these controlling parameters are those that facilitators most often have to adjust on behalf of participants during a session. A decision was made to not provide too many nor too complex a set of options, as this would tend to confuse rather than help the participants.

II. POSSIBILITIES FOR CONTROL

A. Potential Devices and Interfacing Technologies

Accessible human interfaces can translate non-challenging participant actions into the parameter changes that cause the appropriate sensory feedback. Various technologies both simple and complex could be employed, yet whatever is chosen must be optimised for accessibility and ease of use of the participant interface, whilst having the ability to generate the required control within development time and cost budgets. Human interface devices of various types that could potentially be employed include:

- Commercial wired and wireless switched and continuous controllers from accessible interface suppliers,
- Commercial MIDI control surfaces,
- Additional Soundbeam ‘beams’,
- Presentation pointers,
- Joysticks,
- Bluetooth HID devices, e.g. ‘WiiMotes’,
- Accessible computer keyboards or keypads,
- Webcams with appropriate gesture recognition software.

At an early stage we hoped that wireless controls would be feasible. However, for control of audio levels such as the vibration level a wireless method would be overly complex and beyond the resources of this project to implement. Evaluation of potential devices and their various interfacing methods determined that none of the items listed above were able to perform the required control without additional hardware or software development. There were no ready-made commercial solutions that would grant the desired control, because this project requires provision of remote control for non-standard aspects of the computer system. Pragmatically we rejected high-cost or solutions that would take an overly long time to implement.

Due to the fundamental differences in the elements we had chosen to control, we had to use a different solution for each one. The elements and solutions are outlined in the following sections.

III. IMUSE TECHNOLOGIES - VIBRATION

During iMUSE sessions, audio signals from one of the iMUSE PC’s two soundcards (stereo and multichannel respectively) pass to separate analogue channels on an external mixing desk. The mixing desk uses its main faders to set the level of sound to the main speakers, and it uses an ‘aux-send-A submix’ to drive the soundchair / soundbox output. The audio signals generated by the applications in the PC that comprise the different operating modes of iMUSE, are split between the main speakers and the soundchair (vibration transducer) at the mixing desk stage. Any control that would change the audio level at an earlier stage would thus also affect the main speaker volume.

For this reason we determined that control of the vibration level was best achieved by inserting a participant controlled volume control into the Aux-send-A signal chain.

IV. IMUSE TECHNOLOGIES - VISUALISATION

COLOUR

In the majority of the iMUSE modes of interaction, colourful abstract visuals are derived from audio using a software application called G-Force written by the company SoundSpectrum. G-Force is particularly useful for iMUSE because it can respond to both the amplitude and the pitch of the incoming audio, and also is able to utilise ‘live’ rather than pre-recorded sounds, an essential feature of the iMUSE environment. G-Force is also ‘facilitator friendly’, having shortcuts for often-used actions. For instance there are thirty-six ‘key-macros’ (the control key plus an alphanumeric key) for selecting preset visualisations. These factors mean that while G-Force is not the only possible visualisation application possible for iMUSE, it is at present the most suitable. The visualisation colour has to be controlled by controlling G-Force. G-Force does not support programmatic accessibility frameworks such as Microsoft ‘UI-Automation’, however an alternative way to control inaccessible applications such as these is to use software that emulates typing on the facilitator keyboard, i.e. generating ‘pseudo-keypresses’. Using this methodology means that not just G-Force but potentially other applications could be controlled in this way. A block diagram of the elements necessary for control is shown in Fig. 1.

3 www.soundspectrum.com
4 msdn.microsoft.com/en-us/library/ms726294/
Fig. 1. Generic VC Block Diagram

The interface layer in the diagram above digitises the physical device adjustments from the switched or continuous controllers, and then the Application Control Software translates this data into the appropriate simulated user-input sequences needed to alter the relevant parameter in the running application.

A. Hardware Interface

With respect to potential interfaces (e.g. wifi, bluetooth, USB, etc.,) we reasoned that to utilise the ‘MIDI-switch’ interface (supplied as an add-on to ‘Soundbeam 2’ or built-in as standard to the ‘Desktop Soundbeam’ interface) would the most cost-effective solution in iMUSE, since most installations would already have one.

Fig. 2. Soundbeam2 Switch Box (left) and Desktop Soundbeam (right).

The Soundbeam interfaces have eight inputs which can accept either a switch or a voltage. They can be configured to map each switch signal to a different MIDI-controller parameter. These interfaces are suitable for four reasons:
- firstly that standard accessible switches or relatively easy to produce potential divider devices can be used as inputs;
- secondly that a five volt voltage source is provided on each socket - so there is no need for the HIDs to have their own power source;
- thirdly that MIDI signals are relatively straightforward to decode in the software applications that are appropriate to provide the MIDI-to-pseudo-keypress translation required;
- fourthly that the devices are hot-pluggable - which is convenient for evaluation of different styles of HID with participants.

B. Visualisation Colours

Control of the visualisation colour allows the participant to choose from a number of what G-Force call “colormaps”. There are over two hundred of these supplied in the default G-Force installation, but we decided that a smaller number of choices would be more suitable for iMUSE. After field trials twelve were chosen as the optimum number.

Table 1. Visualisation Colours

<table>
<thead>
<tr>
<th>Colour</th>
<th>Key</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqua Spectrum</td>
<td>Aqua</td>
<td>6</td>
</tr>
<tr>
<td>Brass and Aqua</td>
<td>Brass</td>
<td>7</td>
</tr>
<tr>
<td>Candy</td>
<td>Candy</td>
<td>8</td>
</tr>
<tr>
<td>Day Dream</td>
<td>Day</td>
<td>9</td>
</tr>
<tr>
<td>Dragonfly</td>
<td>Dragonfly</td>
<td>g</td>
</tr>
<tr>
<td>Eggplant</td>
<td>Eggplant</td>
<td>h</td>
</tr>
<tr>
<td>Orange to Blue</td>
<td>Orange to Blue</td>
<td>j</td>
</tr>
<tr>
<td>Sea</td>
<td>Sea</td>
<td>k</td>
</tr>
<tr>
<td>Strawberry</td>
<td>Strawberry</td>
<td>l</td>
</tr>
<tr>
<td>Sword</td>
<td>Sword</td>
<td>b</td>
</tr>
<tr>
<td>The Haight</td>
<td>The Haight</td>
<td>n</td>
</tr>
<tr>
<td>Unripe Plum</td>
<td>Unripe Plum</td>
<td>M</td>
</tr>
</tbody>
</table>

C. HID Requirements

The only commercially available accessible human interface devices compliant with this interface are binary switches, which with appropriate software can be used to increment and decrement through a list of colours. However with this method there is no way of selecting a particular item other than by continually pressing the switch until the choice is correct. We decided to construct a custom HID that would allow the possibility of making a specific selection.

D. A Low-Cost Custom Multiple-Selection HID

We decided to use a simple rotary switch as the basis for the HID (see Fig 3.) Whilst various digital position encoders are commercially available, they are expensive, need an external power supply, and do not easily interface to a standard PC. 12-way make-before-break rotary switches usually have an end-stop (i.e. do not rotate through a full 360 degrees.) Trials with this type of switch indicated that it is much more convenient for the participant to be allowed to use a continuous rotation action. Participants and facilitators often loosened the pointer knob by accidentally attempting to move it past its end. However we discovered that one
particular low cost switch (type Lorlin CK1034) can be adapted to have continuous rotation by shaving off the internal plastic end-stop bush and then re-assembling the switch, or it can be specially ordered as such. Eleven resistors (each of 4K7 ohms) are soldered between the switch terminals and it is connected so that the interface will receive a voltage from 0 to 5V proportional to the switch position.

Using this method it is feasible to have a colour coded legend and a knob with a pointer or other position indicator to provide visual feedback on the choice being made, although this also makes the module much larger in size. The version shown above could have a legend added by printing out and laminating segments of the colours in Table 1.

E. Mounting

Each custom switch is mounted in a small plastic module with a lead attached that plugs into the interface. The module can be attached to the chair arm, mounted on a boom-microphone stand, or can be hand-held by the participant, according to preference. Commercial accessible switches can also be used without them necessarily being housed in a module, and some of these e.g. the “AbleNet Specs Switch”\(^5\) are supplied with velcro and elastic straps to facilitate mounting.

F. Application Control Software

Various potential software solutions were evaluated. Whilst it would be possible to code from scratch in a high level language such as C++, it was felt that ready-made scripting applications designed for this type of task would provide a faster implementation, at the expense of flexibility. Two applications were selected that appeared suitable: the input-output mapping application ‘GlovePIE’\(^6\) and the scripting application ‘AutoIT’\(^7\). It was verified that each can generate the relevant ‘pseudo-keypresses’ capable of controlling G-Force. Each application has some advantages and disadvantages however.

GlovePIE has native MIDI-receive capabilities and keypress and mouse emulation, but has no ability to programmatically select a particular graphical user interface [GUI] element in order to send pseudo-keypresses to the correct window. Each application has some advantages and disadvantages however.

AutoIT on the other hand has excellent programmatic window focusing capabilities including the ability to send certain keystrokes direct to a non-focused window. AutoIT does not have a native MIDI-receive capability. However, a compromise is possible using a separate thread running a simple MIDI-to-TCP translator. AutoIT then receives the data via TCP messages instead. Systems using both GlovePIE and AutoIT have been constructed and evaluated. Whilst the GlovePIE script was relatively easy to develop and implement, trials highlighted problems. We found that usability for the facilitator was impaired due to the requirement that they manually re-enable G-Force after adjusting any other auxiliary parameters (echo or reverb levels for instance.) In the instances where this was neglected, the participant’s actions did not result in any colour change, which tended to discourage them from attempting further autonomous control.

This problem of allowing ‘dual-control’ between the facilitator and the participant, within an inherently single-user focused window paradigm, cannot be solved using GlovePIE. For a better solution we used AutoIT to provide an application control script which is described in more detail below.

The following sections give more details of the technical implementation firstly of the VL and then of the VC.

V. IMPLEMENTATION DETAILS - VL

A. Construction

The VL comprises of a monophonic logarithmic (type ‘A’) 10K\(\Omega\) potentiometer inserted between the ‘aux-send-A output’ and the ‘tape-input’ of the Alesis Multimix desk which is generally used in iMUSE systems. The participant-held end of the assembly is modular (swappable) to allow for ergonomic trials of differently constructed potentiometer modules, without having to construct multiple audio breakout leads.

![Fig. 4. Prototype ‘slider’ VL Control](image)

Two modules have been built to date - one slide type (shown in Fig. 4.) and one rotary type. The

5 www.inclusive.co.uk

6 http://glovepie.org/

7 http://www.autoitscript.com/autoit3/
electronic circuit is the same in both instances, and is shown in Fig. 5. The potentiometer module is inserted into the vibration-only signal path by means of a mono ¼” jack plug replacing the original plug occupying the mixing desk ‘aux-send-A’ output.

![Fig. 5. The VL Circuit](image)

The signal is routed via twin-screened audio cable to a stereo plug, which allows a potentiometer module to be connected. The return signal path is via the second core of the cable and terminates with a ¼” mono line socket, which accepts the original lead that was connected to the desk. The potentiometer modules are constructed using a plastic housing containing the potentiometer, the stereo socket and associated wiring.

![Fig. 6. Participant on Vibroacoustic Chair showing the Vibration Level Control (Participant’s left,) Visualisation Colour Control (Participant’s right) and Projected Visualisation (Inset.)](image)

B. Ergonomics of the VL Control

The VL module can be hand held or mounted on the chair arm (see Fig. 6.) Each module housing was designed so that it could potentially be hand-held. Constraints on the minimum overall size are governed by the commercial availability of suitable potentiometers. The linear slide module is approximately 10cm long by 1.5cm square and the rotary module is approximately 6cm x 4cm x 1.5cm. Participant evaluations were conducted over several weeks to determine the effectiveness of each type of module. After interviews with the participants, it was found that they preferred the slide version to the rotary version. Satisfaction was such that further ergonomic trials for this module were deemed unnecessary.

Whilst the VL is relatively simple to implement using soldering and lead construction skills, the VC design is more involved, in that both hardware and software is required.

VI. IMPLEMENTATION DETAILS - VC

A. HIDs

For evaluation purposes we have implemented four different HIDIs. The custom continuously-rotatable twelve-way switch shown in Fig. 3., and described previously, a single push-switch, a pair of push-switches as shown in Fig. 7., and a slide potentiometer (not shown). All these devices are compatible with the Soundbeam MIDI switch interface.

![Fig. 7. Two of the VC Modules](image)

B. Ergonomics

The control can be hand held, or mounted on the top or side of the chair arm, or can be mounted on a boom microphone stand. For the rotary control, a number of different control knobs were sourced and trialled. Participants preferred the larger easier to grip knobs, probably influenced by factors such as arthritis and increasing frailty associated with the ageing process.

C. MIDI Input allocation

Since different types of switch can be plugged into the eight MIDI inputs of the Soundbeam interface, we allocated inputs one to four for binary switches, and inputs 5-8 for multi-level switches or potentiometers. Configuration (using the Soundbeam user interface) involves selecting
between level and voltage, allocating a unique MIDI channel for each input, enabling the relevant MIDI controller (see the next section on protocol), and setting the offset and depth. For new presets we devised a template which can be used as a starting point. In the case of the desktop Soundbeam, the save-files are textual so cut-and-paste methods allow inserting the required functionality into pre-existing save-files.

D. Protocol

As well as defining what type of switch is used for grouped inputs, we have had to devise a MIDI protocol to encapsulate the data. The protocol used is essentially arbitrary, in that it is used purely internally to MUSE, but we have used the most appropriate MIDI “control change” messages for our purposes. We designated MIDI control numbers 16-19 inclusive (General Purpose Controllers 1-4) to encode the variable level data from potentiometers or multi-way switches, and MIDI controllers 80-83 inclusive (General Purpose Controllers 5-8) to encode on-off switch values.

E. Application software

The software which converts the switch actions to feedback control is written using the AutoIT scripting language. AutoIT can compile scripts to produce an executable file suitable for use on a computer that does not have AutoIT installed. A prototype facilitator user interface to the script is shown below in Fig. 8.

![Fig. 8. Prototype Facilitator User Interface](image)

AutoIT allows development of custom GUIs like the one shown above. Buttons have been provided to start G-Force and perform other common facilitator actions.

F. Application software - MIDI Input

The desktop Soundbeam connects to the PC using USB and has its own MIDI device driver, whereas the Soundbeam2 requires a separate MIDI interface (we use an M-Audio “Midisport 2x2”) which has its own driver. To allow for device changes, the facilitator can select a particular MIDI device by name.

The application software spawns a small TCP client thread ‘netmidic.exe’ with appropriate parameters to perform a conversion from MIDI to a network protocol that is receivable by AutoIT.

The incoming TCP data stream is parsed and decoded to determine what switch operation has occurred.

The pseudo events that are generated following each switch operation are hard-coded at development time. For instance a switch off-on-off transition on input 3 generates pseudo keypresses that result in the decrementing of the G-Force visualisation colour to the previous one in the set of twelve. Input 4 similarly increments the colour. A change in the multi-way switch position connected to input 5 results in a selection of the particular colour represented by that switch position, out of the twelve colours.

Other inputs are free for future developments, such as selecting the G-Force ‘waveshape’ parameter for instance. Using the same devices but changing the software could be made to cause a different outcome, so the apparent limit of eight inputs is not problematic, as it allows eight items to be changed simultaneously, which is more than enough for current needs.

G. Application software - Focusing and Sending Keys

A library of useful AutoIT functions has been developed for G-Force, for instance setting focus to a particular window or attempting to send an emulated keypress.

![Fig. 9. Example AutoIT script](image)

```bash
Func GEF_KeyMacroForceFocus( $key )
  if GEF_SetFocused() then
    Send( $key )
  return true
Endif
EndFunc
```

8 http://www.m-audio.com
9 http://public.sreal.com:8000/~div/MIDI-utilities/
The example section of script shown in Fig. 9 checks whether the G-Force window has focus. If not it attempts to set the current keyboard focus to G-Force window (with a time out). If this succeeds it can proceed to send the emulated keypress, if not the function returns without performing any action.

H. Application software - Warn Mode

As was described earlier, some form of arbitration is required to ensure the facilitator and participant do not attempt to use the single resource that is the window that currently has keyboard focus.

In earlier versions of the application control software this was done using a facilitator operated 'mode' button (top-centre in Fig. 8.). Warn mode would sound an audio prompt if the facilitator had taken focus away from G-Force. It would also sound a different 'error' tone if the participant moved a switch (the movement in this instance does not result in a pseudo-keypress because it would not arrive at the correct window).

I. Application software - Concurrent control

The latest version of the software uses a newly developed key-sending function that succeeds in posting the requisite keypresses reliably to an unfocused G-Force window. This version of the software is still being tested but if reliability is satisfactory it will be a large improvement over the older methodology, in that both the participant and the facilitator can adjust parameters at the same time with no unintentional side-effects. The lack of side effects means it is more reliable for both the participants and the facilitators. It is also much more user-friendly for facilitators: demanding fewer mouse clicks during the course of a normal iMUSE session.

VII. PRELIMINARY PARTICIPANT EVALUATIONS

The VL has been readily adopted by our initial test subjects. Comments on the use of the slider style VL module have been very positive from all users. Tests on the different styles of VC HID are ongoing and preliminary interviews have again recorded very positive responses from the users. Typical comments include “It's very easy to use”, “I like it better when I can do it - [the facilitator] doesn’t know which ones [colours] I prefer.”

VIII. CONCLUSIONS

At this stage of evaluation, the controls appear to be effective and cost effective additions to the existing iMUSE participant interfaces.

The VC was initially less successful because of the small amount of unreliability. By increasing the reliability of the system the participants have shown increased autonomy with control of colour.

One factor we have already noted in the HID trials is that not all participants prefer the same device. We may find that the modular approach is in fact necessary, allowing participants where feasible and where costs allow, to select the particular HID style, mounting position and orientation that they prefer from a choice of suitable modules.

A useful spin off from this project is that the VC methodology can be adapted by adopting a similar script that is tailored to some other software application. AutoIT provides tools to facilitate identifying the necessary window names that might be required for other applications, and with a small amount of development, both keyboard and mouse operations can be emulated. This means that for applications which do not support a built-in accessibility framework, the concept provides a general method of enhancing accessibility through the use of switches in performing arbitrary windows human interfacing tasks.

IX. FURTHER WORK

A problem was found with the linearity of the digitised voltage on the Soundbeam SB2 interface, (see Fig. 10.) The voltage-derived data values output by different Soundbeam interfaces using the same rotary switch are not the same, hence the discriminating function in the application control software has difficulty in determining which choice is represented by a given level.

![SB2 Non Linearity with 12 way 51K rotary Switch](image)

Fig. 10. Soundbeam 2 linearity

To alleviate this, some form of auto calibration needs to be done; for instance the facilitator might run a special calibration script involving rotating the 12 way switch through several rotations (in both directions), in order for the exact switching thresholds to be calculated. This would only need to be done once for installation, or if the interface was changed.
REFERENCES


Chapter 4

Computer Vision for Accessibility
Fiducials Marks detection to Assist Visually Impaired people Navigate

Paulo Costa*, Hugo Fernandes, João Barroso, Paulo Coelho, and Leontios J. Hadjileontiadis

P. Costa is with the Polytechnic Institute of Leiria, Department of Informatics Engineering, Leiria, Portugal; H. Fernandes is with the University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal; P. Coelho is with the Polytechnic Institute of Leiria, Department of Electrotechnical Engineering, Leiria, Portugal; J. Barroso is with the GECAD – Knowledge Engineering and Decision Support Research Center, Porto, Portugal; L. J. Hadjileontiadis is with the Department of Electrical and Computer Engineering, Division of Telecommunications, Aristotle University of Thessaloniki, Thessaloniki GR-54124, Greece

* Corresponding author: Paulo Costa, Phone: +351 244820300, Fax: +351 244820310, e-mail: paulo.costa@estg.ipleiria.pt

Abstract — Assistive technology enables people to achieve independence in the accomplishment of their daily tasks and enhance their quality of life. Visual information is the basis for most navigational tasks, so visually impaired individuals are at disadvantage due to the lack of information or given insufficient information about their surrounding environment. With the recent advances in inclusive technology it is possible to extend the support given to people with visual disabilities during their mobility. In this context we propose and describe the SmartVision project, whose global objective is to assist visually impaired people in their navigation through unknown indoor and outdoor environments. This paper is focused mainly on the Computer Vision module of the SmartVision prototype, were we propose a new algorithm to recognise fiducials marks suitably placed on sidewalks, revealing to be a promising solution.

I. INTRODUCTION

Assistive technology enables people with disabilities to accomplish daily living tasks and assists them in communication, education, work and recreation activities. Principally though, it can help them to achieve greater independence and enhance their quality of life. Of the various assistive technologies available, a special focus was put on those that help blind or visually impaired people with their mobility.

The World Health Organization estimates that there are 37 million blind people worldwide [1]. Blind or visually impaired people have a considerable disadvantage, as they need information for bypassing obstacles and have relatively little information about landmarks, heading, and self-velocity. The main issue on using assistive technologies is to provide additional information useful to blind people during their mobility process, i.e. walking.

Human mobility can be distinguished between Orientation and Navigation. Orientation can be thought of as knowledge of the basic spatial relationships between objects within the environment. Information about position, direction, desired location, route, route planning etc. are all bound up with the concept of orientation. Navigation, in contrast, suggests an ability to move within the local environment. This navigation implies the knowledge of immediate objects and obstacles, of the formation of the ground (holes, stairs, flooring etc.), and of dangers both moving and stationary.

The aim of the present work is to propose a new algorithm for the computer vision (CV) module that will be integrated with the SmartVision project, described later. The new image processing algorithm is intended to extract useful information from an outdoor scene in the University of Trás-os-Montes and Alto Douro (UTAD) campus, and put the blind user correctly positioned on the sidewalk along a predefined route. In order to reduce the image complexity features for extraction/detection fiducial marks were placed along the sidewalks. The algorithm we propose uses Ensemble Empirical mode decomposition (EEMD) for image processing and basic correlation for template matching.

The paper is organized as follows. Section II presents a classification of navigation systems to assist visually impaired people. Some projects that represent the state of the art are presented. Section III presents the proposed algorithm and the related techniques used. Section IV presents and discusses the results. Finally, section V concludes the paper.

II. NAVIGATION SYSTEM TO ASSIST VISUALLY IMPAIRED PEOPLE

An Electronic Travel Assistant (ETA) has to supply the visually impaired with the necessary routing information to overcome obstacles in the near environment with minimum errors. This displacement between the origin and the destination is varies according to the programmed route. A distinction must be made between primary support systems such
as guide cane and guide dogs, and the secondary ones that use the most recent technologies.

These secondary systems are the focus of current study and consist of a wearable or handheld computer with a Global Positioning System (GPS) responsible for the macro navigation. In order to prevent collision with obstacles (micro navigation) these secondary systems also make use of the services of primary navigation systems. In the mid-eighties, Collins and Loomis independently proposed the use of GPS to assist navigation for the visually impaired, in their navigation systems [2].

According to the proposed model by Loomis [2] a system to assist navigation for visually impaired people is organized in three basic components: 1) The position unit and orientation is responsible for supplying the navigation system with the user’s spatial location, in the form of local and global coordinates. Due to the strong dependence on the environment in which the system is used, this is the functional block that more specifically characterizes the navigation systems; 2) The geographic Information System (GIS) contains main geo-referenced database system data. This functional block is an essential component of the navigation systems. Its main function is to store additional information about user’s position, navigation maps, object positions and possible dangers; 3) The user interface is the most critical component in the navigation system for assisting the visually impaired because it acts as a substitute for vision sensing (or attempts to). The user interface must be user-friendly in such a way that the user does not encounter difficulties which would impede daily use. Typically interaction with the visually impaired is through audio interfaces, like Text-To-Speech (TTS) or virtual audio (sonification) and tactile displays like Braille keyboards or vibrotactile devices.

Figure 1 shows the Loomis’s proposed block diagrams for his navigation systems. The position unit and the orientation block can be seen, with several sensors for macro and micro navigation, the GIS block for route planning and the user interface block to provide feedback to the blind.

---

A. Navigation Systems and related work

Navigation systems to assist visually impaired people can be classified in three groups based on their usage. The indoor systems are to be used in structured environments with less complex scenes, typically inside buildings or in isolated controlled campuses. The outdoor navigation systems are intended to be used in exterior open space, typically on the street. The indoor/outdoor systems can be used in both indoor and outdoor spaces, switching functionalities based on environment operation.

The following presents some commercial research and development (R&D) projects that currently describe the state of the art in outdoor navigation systems for assisting visually impaired people.

1) Navigation systems without local obstacle information, the systems BrailleNote GPS [3], StreetTalk [4], Trekker [5], NOPPA [6], Navigator [7] and Drishti [8] are GPS based systems to assist the navigation of visually impaired people. Their primary components are a PDA or Laptop especially designed/adapted for people with visual disabilities, a Bluetooth GPS receiver and specially developed software for configuration, orientation and route mapping. The output user interaction can be from Braille display or a speech synthesizer.

2) Navigation systems with local obstacle information provide better knowledge of the local scenario, increasing the information quality provided to the blind user to overcome local obstacles.

Several techniques are used to detect and measure object distances, multiple ultrasonic sensors (sonar) [9], Laser Range Scanner (LRS) [10] and computer vision (CV) techniques [11], [12], [13], [14] and [15].

B. The SmartVision: active vision for the blind

A system to assist the navigation of blind or visually impaired people is currently being developed at the University of Trás-os-Montes and Alto Douro (UTAD). This project is named SmartVision and its main objective is to develop a system that helps visually impaired people to navigate, providing ways to get to a desired location and, while doing so, giving information about obstacles and various points-of-interest (POI) like zebra-crossings, building entrances, etc. The system is built in a modular structure, combining several technologies, as seen in Fig. 2. The SmartVision Module is responsible for managing and establishing communication between all the available modules. This module also receives inputs from the user and makes decisions on what information the user should get from the system.

The Location Module is responsible for providing regular updates on the user’s current geographic coordinates to the SmartVision Module. To provide this information both in indoor and outdoor environments, this module makes use of different technologies: Global Positioning System (GPS) for outdoor environments and Wi-Fi for indoor environments.
Radio-Frequency Identification (RFID) and Computer Vision are common to both indoor and outdoor environments and are based on the detection of landmarks placed in the ground. Each location technology has a specific accuracy and the Location Module always chooses the one with the best accuracy from the ones available in each moment. In terms of hardware, the RFID reader is placed in the white cane and the camera is chest-mounted. The GPS antenna is connected via Bluetooth and the Wi-Fi antenna is a built-in component of the mobile computer.

The Navigation Module is responsible for route planning and providing information about surrounding points-of-interest (POI). It connects to the SmartVision Module and requests two different data inputs: GIS data and location data. To get the GIS data, the SmartVision module queries the GIS server in order to get maps and POIs. The user location is fed from the Location Module. After analyzing the requested data, the Navigation Module feeds back the SmartVision Module with navigation instructions. The amount and accuracy of the GIS data stored in the GIS server is critical in order to feed the blind user with the most appropriate instructions.

The Computer Vision Module provides orientation instructions by detecting known landmarks in the ground and keeping the user within safe routes. Using a stereo vision system, disparity information is extracted from the captured image frames and can be used to create a depth map. This information is useful to know the distance between the user and detected landmarks. So, in addition to giving orientation instructions to the SmartVision Module, with this distance information, the Computer Vision Module has the possibility to feed the Location Module with location information.

Finally, the Interface Module provides user interface using two outputs and one input. The two outputs are text-to-speech software and vibration actuators. Since the hearing sense is very important to blind users, the vibration actuators are used for navigation guidance and the voice interface is used for menu interaction and to provide POI information. The user gives inputs to the system by using a small four-button device to scroll between the menus, apply the options and go back to the previous menus.

The user interacts directly with the SmartVision Module through the Interface Module and all other modules are independent. This way, the user can get information even when some modules are not available, or cannot provide information. For example, if GPS is not available or if the user is in an indoor environment, the Location Module can get information from the RFID tags, Wi-Fi or Computer Vision Module. Redundancy is, therefore a very important factor to increase the reliability of the system.

C. Computer Vision Model for SmartVision

Several image processing techniques are used to extract useful information from the scene, i.e., object identification, recognition and scene description. This information is very important for tracing a route between the scene objects.

In the context of assisting visually impaired people, the computer vision model must deal with the large amounts of image data acquired (high bandwidth process) and provide useful scene information to the user (Human Computer Interaction - HCI) which is typically a low bandwidth process.

Several Computer vision techniques have been used in navigation systems to assist people with visual disabilities. The Principal Component Analysis (PCA) was used in ASMONC [9], the Tyflos system [10] uses Fuzzy Like Reasoning segmentation technique, Expectation-Maximization (EM) algorithm was used by Zelek [12], stereo images for measuring distance from object were used by Meers [14] and Hadjileontiadis [15], a Neural Network technique was used in NAVI [16], and later the same project authors also tested Fuzzy Learning Vector Quantification (FLVQ) to classify objects in the scene.

The computer vision model of the SmartVision is one of the most critical because it deals with large and heterogeneous amount of data and in general requires a high computing power. All the computations are made on a laptop computer and for image acquisition we use the stereo vision Bumblebee 2 developed by Point Grey Research.

The Bumblebee is a packaged system that includes two pre-calibrated digital progressive scan Sony ICX084 CCD cameras with a baseline (the distance between cameras) of 12cm, and a C/C++ Software Development Kit [PGR 2003], and a 400 Mbps IEEE-1394 Firewire interface for high speed communication of the digital video signal. Gain and shutter control can be set to automatic or adjustable manually. The calibration information is preloaded into the camera allowing the computer software to retrieve it for XYZ coordinate calculations and image correction.

The FlyCapture SDK was used for image capture and camera control. The image size used in this work is 512 by 384 pixels. For the calculation of the
disparity and the correction of the images we used the Triclops SDK, and will be used for future improvements. Both this SDKs are provided together with the Bumblebee 2 stereo vision system.

III. EMPIRICAL MODE DECOMPOSITION

In the real world, data from natural phenomena like life science, social and economic systems, are mostly non-linear and non-stationary. Fourier and wavelet transform (built upon predefined basis functions) are traditional methods that sometimes have difficulty in revealing the nature of real life complex data. The adoption of adaptive basis functions introduced by Huang et al. [17] provided the means for creating intrinsic a posteriori base functions with meaningful instantaneous frequency in the form of Hilbert spectrum expansion [17]. This approach is embedded into a new decomposition algorithm, namely Empirical Mode Decomposition (EMD) [17], which provides a powerful tool for adaptive multi-scale analysis of non-linear and non-stationary signals. EMD is a method of breaking down the signal without leaving the time domain; it filters out functions which form a complete and nearly orthogonal basis for the signal being analysed. These functions, known as Intrinsic Mode Functions (IMFs), are sufficient to describe the signal, even though they are not necessarily orthogonal [17]. IMFs, computed via an iterative 'sifting process' (SP), are functions with zero local mean [17], having symmetric upper and lower envelopes. The SP depends both on an interpolation method and on a stopping criterion that ends the procedure. Some updates of the 1D-EMD have been proposed which address the mode mixing effect that sometimes occurs in the EMD domain. In this vein, 1D-Envelope EMD (1D-EEMD) has been proposed [18], where the objective is to obtain a mean ensemble of IMFs with mixed mode cancelation due to input signal noise addition.

A. 1D-Empirical Mode Decomposition (1D-EMD)

1D-EMD considers a signal \( x(t) \) at the scale of its local oscillations [17]. Locally, under the EMD concept the signal \( x(t) \) is assumed as the sum of fast oscillations superimposed to slow oscillations. On each decomposition step of the EMD, the upper and lower envelopes are initially unknown; thus, an interactive sifting process is applied for their approximation to obtain the IMFs and the residue. The 1D-EMD scheme is realized according to the following steps [17]:

1) Identify the successive extrema of \( x(t) \) based on the sign alterations across the derivative of \( x(t) \);

2) Extract the upper and lower envelopes by interpolation; that is, the local maxima (minima) are connected by a cubic spline interpolation to produce the upper (lower) envelope. These envelopes should cover all the data between them;

3) Compute the average of upper and lower envelopes, \( m_k(t) \);

4) Calculate the first component \( h_1(t) = x(t) - m_1(t) \);

5) Ideally, \( h_1(t) \) should be an IMF. In reality, however, overshoots and undershoots are common, which also generate new extrema or exaggerate the existing ones [17]. To correct this, the sifting process has to be repeated as many times as is required to reduce the extracted signal as an IMF. To this end, treat \( h_k(t) \) as a new set of data, and repeat steps 1-4 up to \( k \) times (e.g., \( k = 7 \)) until \( h_k(t) \) becomes a true IMF. Then set \( c_t(t) = h_k(t) \). Overall, \( c_t(t) \) should contain the finest scale or the shortest period component of the signal;

6) Obtain the residue \( r_1(t) = x(t) - c_1(t) \);

7) Treat \( r_1(t) \) as a new set of data and repeat steps 1-6 up to \( N \) times until the residue \( r_N(t) \) becomes a constant, a monotonic function, or a function with only one cycle from which no more IMFs can be extracted. Note that even for data with zero mean, \( r_N(t) \) still can differ from zero;

8) Finally,

\[
x(t) = \sum_{i=1}^{N} c_i(t) + r_N(t),
\]

where \( c_i(t) \) is the i-th IMF and \( r_N(t) \) the final residue.

B. 1D-Envelope Empirical Mode Decomposition (1D-EEMD)

One of the major drawbacks of the original 1D-EMD is the appearance of mode mixing, which is defined as a single IMF consisting of signals of widely disparate scales, or a signal of similar scale residing in different IMF components. The effect of adding white noise scales uniformly through the whole time-scale or time-frequency space, will provide a reference distribution to facilitate the decomposition method. The added white noise may also help to extract the true signals in the data, a truly Noise-Assisted Data Analysis [18]. The 1D-EEMD is implemented as follows:

1) Add white noise series \( w(t) \) to the data \( x(t) \), \( X(t) = x(t) + w(t) \);
2) Decompose the \( X(t) \) data with white noise into IMFs, \( X(t) = \sum_{j=1}^{N} c_j(t) + r_h(t); \)

3) Repeat step 1 and step 2 several times with different noise series \( w(t) \), \( X_i(t) = x(t) + w_i(t) \), and obtain corresponding IMFs, \( X_j(t) = \sum_{j=1}^{N} c_j(t) + r_h(t); \)

4) Finally, the ensemble means of corresponding IMFs of the decomposition are

\[
c_j(t) = \frac{1}{N} \sum_{i=1}^{N} c_{ij}(t),
\]

where \( N \) is the ensemble members.

C. 2D-Empirical Mode Decomposition (2D-EMD)

The sifting notion is essentially identical in 1D and 2D cases. Nevertheless, due to the nature of the 2D data of the images, some issues should be handled with care.

In particular, in a 1D space, the number of local extrema and zero crossings of an IMF must be the same or differ by one [17]. In a 2D space, the IMFs typically use the definition of symmetry of upper and lower envelopes related to local mean [19]. There, many ways of defining the extrema are in use; hence, different local extrema detection algorithms could be applied. Fast algorithms use the comparison of the candidate extrema with its nearest 8-connected neighbours [20], while more sophisticated methods, like morphological reconstruction, are based on geodesic operators [21]. Furthermore, the interpolation method should rely on proper 2D spline interpolation of the scattered extrema points. In [19] the thin-plate smoothing spline interpolation is used. In Bi-dimensional Empirical Mode Decomposition (BEMD) [21] Radial Basis Functions are used for surface interpolation. This combination of 2D extrema extraction and 2D surface interpolation represents very heavy computation power, suitable neither for real-time implementations, nor for use in portable devices.

D. The Proposed Computer Vision Approach

In order to provide useful information to blind people navigation the vision system must be able to detect relevant features in the scene and help the blind user to keep safe courses. The first approach was intended to reduce the image complexity for the processing algorithms and enhance the detection. Fiducial marks were made on sidewalks representing safe paths along the user route. From several geometric marks, circles were adopted because in this application they are scale and rotational invariants. Several captured images with different circle radius were tested and to minimize the size and maximize detection rate the 15 cm circle radius were chosen.

The proposed CV phases are described as follows:

1) Decompose the captured image with Ensemble Empirical Mode Decomposition

2) Image filtering to eliminate the higher frequencies containing noise and fine details. This process is achieved in the Ensemble Empirical Mode Decomposition reconstruction phase by eliminate the first 4 IMFs according to (1)

\[
x(t) = \sum_{i=5}^{N} c_i(t) + r_h(t), \quad \text{were } i \text{ start at } 5 \text{ IMF.}
\]

3) Define a region of interest (ROI) near the blind user, in our case we chose to analyse the first half of the image and a quarter image size for each side of the blind user. Data outside the ROI area is set to zero.

4) Preform a data binarization of the ROI image using a global threshold using Otsu’s method, followed by canny edge detection.

5) Finally the ROI image is passed to a circle detection procedure using a simple correlation template matching.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to test the proposed algorithm to detect fiducial marks in the sidewalk a set of different images were taken on UTAD campus. We present two images that represent different areas of the campus as can be seen in Figure 3 a) and Figure 3 e), the second one represents a more difficult task for any image processing algorithm. Figure 3 b) and Figure 3 f) are the EEMD filtered images as can be seen the higher frequencies were removed, with this procedure prior image binarization image artifacts were minimized. Figures 3 c) and Figure 3 g) are the ROI near the user that are processed to circle detection, consider that the user is centered at the image bottom. Finally Figure 3 d) and Figure 3 h) represents the circle detection of the respective images, all circle detected are marked with a rectangle. In order to improve the visualisation results these two images are at a bigger scale.

From the circles detected in the image, the blind user must go in the direction of the nearest circle. This ensures that he/she will not get out of course. Based on the relative position of the blind user and the detected nearest circle its possible to compute the trajectory correction and output it to the blind user. The interface to the user uses five microvibrators corresponding to the five directions, i.e. left, left-diagonally, straight, right diagonally and right.
V. CONCLUSION

In the present work a Computer Vision module for the SmartVision project was proposed. For an efficient assistance to blind user’s navigation the CV must detect accurately specific features in the environment. In outdoor navigation due to very different scenarios we adopted to mark sidewalks with fiducial marks to improve the CV feature detection efficiency. To the fiducial mark detection the EMD Template Matching method was implemented and the system has proven to be able to detect the defined landmarks and provide valid and simple instructions to the blind user.

The SmartVision prototype is also composed by other modules, as seen in section II, and, at the moment, they are all being integrated. A set of tests done to the assembled system by blind users will be performed in order to validate and improve the system.

Further work is needed to enhance the method accuracy, future improvements will continue to use EMD image analysis. Range image information (disparity map) will be integrated into CV model to add the obstacle detection feature.

ACKNOWLEDGMENTS

This research was supported by the Portuguese Foundation for Science and Technology (FCT), through the pluri-annual funding of the Institute for Systems and Robotics (ISR/IST) by the POS_Conhecimento Program which includes FEDER funds, and by the FCT project PTDC/EIA/73633/2006 - SmartVision: active vision for the blind.

REFERENCES


The SmartVision Navigation Prototype for the Blind

1 Vision Laboratory – Inst. for Systems and Robotics (ISR), University of the Algarve (FCT/ISE), Faro, Portugal
2 GECAD, Knowledge Eng. and Decision Support Research Group, Inst. of Engineering of Porto, Porto, Portugal
3 University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

Abstract — The goal of the project "SmartVision: active vision for the blind" is to develop a small and portable but intelligent and reliable system for assisting the blind and visually impaired while navigating autonomously, both outdoor and indoor. In this paper we present an overview of the prototype, design issues, and its different modules which integrate a GIS with GPS, Wi-Fi, RFID tags and computer vision. The prototype addresses global navigation by following known landmarks, local navigation with path tracking and obstacle avoidance, and object recognition. The system does not replace the white cane, but extends it beyond its reach. The user-friendly interface consists of a 4-button hand-held box, a vibration actuator in the handle of the cane, and speech synthesis. A future version may also employ active RFID tags for marking navigation landmarks, and speech recognition may complement speech synthesis.

I. INTRODUCTION

Similar to the Greek project SmartEyes [9], the goal of the Portuguese SmartVision project is to develop and integrate technology into a portable device for aiding blind persons and those with severe visual impairments, about 180 million of which 40-50 million are completely blind. This device must be relatively small and cheap, and easy to assemble using off-the-shelf components. It must be extremely easy to carry and use, yet providing help during navigation and locating objects. It should be stressed that the device cannot replace the white cane; it will be an extension of the cane, issuing warning signals when approaching a possible obstacle or when the footpath in front is curved and the heading direction should be adapted. The device will not employ headphones as these block surround sounds. We assume that the user has adapted to the cane and relies on hearing. In this sense the device will be "non-invasive" such that normal surround sounds and the device's signals, including verbal communications with queries and answers, are integrated in a natural way.

Acknowledgements: This research was supported by the Portuguese Foundation for Science and Technology (FCT), through the pluri-annual funding of the Institute for Systems and Robotics (ISR-Lisbon/IST), by the POS_Conhecimento Program which includes FEDER funds, and by the FCT project SmartVision: active vision for the blind (PTDC/EIA/73633/2006).
Below we discuss technological solutions in general (Section 2) and the ones being developed in the project: the prototype (S. 3), GIS and the navigation module (S. 4), the vision module (S. 5), the white cane with passive RFID tags (S. 6) and the interface module (S. 7). We conclude with a discussion in Section 8.

II. TECHNOLOGIES AND INFRASTRUCTURES

Technology can provide solutions to the three broadly-defined applications, even with off-the-shelf hard- and software, but computing power may be a limiting factor in portable and therefore small devices. Another problem is the necessary infrastructure to use the technology, like outdoor GPS and indoor Wi-Fi reception in combination with a GIS, the level of detail of the latter, plus extra infrastructure specifically devoted to blind persons. For example, many railway stations have bands of rippled pavement at the platforms. Such bands can be employed at many other locations to mark the centres of footpaths. If too expensive, are there cheaper solutions? Since most roads are marked by white stripes which reflect light, it is also possible to mark footpaths, for example by reflecting circles a few metres apart (see Fig. 4). During the day these can be easily detected by a miniature camera, but during the night a small light source is necessary, for example a powerful LED, carried by the user, which flashes every two or three seconds.

A large metro and railway station in Lisbon is equipped with loudspeakers which emit sounds of birds, with different sounds in different areas for global navigation, and most sighted passengers do not even notice the sounds. Spoken messages stored in electronic tags are provided by the “Talking Points” system [13] and the tags can be fixed to anything, from an entry of a building to its elevators, also to bus stops and busses. The latter system is a commercial one, the company only providing “turnkey” solutions.

There are many possibilities, but the main question is: how much is a civilised society willing to invest in infrastructure for relatively few users, knowing that the infrastructure will make a tremendous difference for those users. For example, the reflecting circles mentioned above could be substituted by a solar-power driven LED in a small armature with a white-glass cover in the pavement, such that blind persons do not need to carry a light source when it is dark. One armature including a small solar panel, day-night sensor and other electronics might cost a few euro if mass produced. Electronic parts are so cheap that LED modules (solar) power supply, which are more expensive but have a large action radius, and active tags with battery or external (solar) power supply, which are more expensive but have a large action radius. To be useful, passive tags must be drilled into the pavement at rather small distances if the tip of the cane is equipped with a small sensor and the swaying cane should encounter the tags with a reasonable probability (see Section 6). Active tags, on the other hand, are more suitable for marking important landmarks like entries of buildings and shops, with a distance of 5 to 10 metres. Since both passive and active RFID tags emit unique codes, the codes can be stored in a database. Passive tags can be used for local navigation at footpath crossings etc., whereas active ones are better for global navigation, both in- and outdoor: the name of a landmark like a shop in the GIS for position calibration if GPS or Wi-Fi reception is not available. See also Section 3.

Concerning object recognition in a supermarket or household pantry, if the blind person is wearing a camera at chest height, perhaps three shelves can be analysed with sufficient optical resolution. But in a supermarket these shelves normally carry the most expensive products, whereas the cheapest ones are on the bottom shelves. In addition, all shelves are cramped with products, because shelf space is money, hence products are not well separated and this hampers object segregation and identification (see Section 5). In a pantry, this problem can be avoided and both top and bottom shelves can be reserved for bigger things which can be easily identified by hand. In any case, it should be possible that a blind person takes an object in his hand, holds it at arm length in front of the camera attached at chest height, and uses a special command to ask the system to identify the object.

An alternative is to label objects with a QR barcode, a square one which can hold more information [8]. A mobile phone with a camera and QR barcode software sends the code to a URL server and receives an audio file, created by text-to-speech, with a description. Of course, this solution can be applied anywhere, software for printing and reading QR barcode labels is freely available, the price is almost nil apart from server and database maintenance, but the labels must always be visible. Instead of putting the labels on the products it is possible to put them on the shelves, provided they are checked on a regular basis and no one puts products at wrong locations. A complication when using a camera at chest height is the long distance between camera and barcode label, which also implies that there may be many labels in the camera’s field of view. If a portable computer is used instead of
a mobile phone, all software can be integrated and no wireless communications are required (text-to-speech may be required anyway; see Section 7).

For localising few but very important objects, like a bunch of keys at home or a suitcase on a conveyor belt at the luggage claim area of an airport, there already exist good solutions: an electronic label which can be attached to the object and a small hand-held gadget with a push button. The label beeps and it works with ultrasound or radio signal. It is rather easy to make a gadget with a few buttons for localising various objects. Here we can assume that blind persons do not put very important objects nor such a gadget at arbitrary locations, but: errare humanum est!

There are other technologies which can be employed, notably ultrasound detectors for spotting possible obstacles at foot- or head-height [19]. A recent development is Nokia’s indoor positioning technology [3] based on DoA (direction-of-arrival), with an accuracy of 2-5 m. Another technology employs ToA/AoA (time/angle of arrival) with UWB (ultra-wideband), aiming at 10 cm. Pending availability of such advanced technology, one must select what is currently available and integrate this into a working system which is modular, such that future developments can be easily and rapidly adopted. Below we describe the components of the SmartVision system.

III. PROTOTYPE OVERVIEW

The system integrates GPS (in the future the European Galileo equivalent) and Wi-Fi with a GIS. GPS is for outdoor navigation and Wi-Fi mainly for indoor, specifically in bigger buildings like a shopping centre or a university’s administration. However, Wi-Fi access points are still not ubiquitous and GPS may not be available in narrow streets and in bad weather. Therefore, a fail-safe solution is required such that the user can still be assisted: by active RFID tags with calibrated positions, and if these cannot be located by visual object recognition, and, if all else fails, by the user asking a passer-by to bring him to the nearest landmark which is included in the GIS. The basic idea is that the user’s system always has a global map with enough detail along the planned route. The actual position is constantly calibrated using GPS, Wi-Fi or RFID tags, whatever is locally available, and each time Wi-Fi reception is possible and a significant distance has been traversed, the local map is updated by the GIS server, for example in a radius of a few kilometres. This way the user can always consult the GIS for taking the best or shortest path to a destiny and for knowing the major landmarks along the route.

The system consists of a few modules; see Fig. 1. The central SmartVision module is responsible for managing all other modules and the communications with these. This module also receives input from the user (not shown) and takes decisions about what information the user should get from the system. Also not shown are the speech and other interfaces (see Section 7), nor is the electronic compass which will be integrated in the future.

The GIS and navigation modules provide regular updates of the user’s current geographic coordinates to the SmartVision module, either using GPS in outdoor or Wi-Fi in indoor environments. Both GPS and Wi-Fi are complemented by RFID and computer vision, and if GPS or Wi-Fi signals cannot be received the SmartVision module must rely on RFID and computer vision. The latter are common to both in- and outdoor environments and serve to detect specific landmarks. Each location technology has a specific accuracy and the navigation modules always choose the best accuracy from the ones which are available at any moment.

Concerning hardware, one small RFID reader is placed in the white cane for detecting passive tags in the pavement, see Fig. 6, and the camera is chest-mounted. The GPS unit is connected by Bluetooth and Wi-Fi is a built-in component of the portable computer. In the future, a second RFID reader will be connected to the computer in order to detect active tags attached to landmarks.

The navigation modules are responsible for route planning and for providing information about surrounding points-of-interest (POIs). They connect to the SmartVision module and request GIS data and location data. To get the GIS data, the SmartVision module queries the GIS server in order to get a local map and its POIs. The user's position is obtained from the location modules. After analysing the requested data, the navigation modules feed the SmartVision module with navigation instructions. The amount and accuracy of the GIS data stored on the GIS server is critical in order to provide the best instructions.

The vision module provides local orientation instructions by detecting sidewalk borders and possible obstacles in front on the path, for guiding the user safely. It also detects already known landmarks (GIS in the immediate surrounding) in order to confirm encountered POIs and re-calibrate the user’s position if
GPS or Wi-Fi cannot be received. In addition, this module can be used to detect and recognise objects on demand. The camera used is the Bumblebee2 from Point Grey Research Inc. Being a stereo camera, it is possible to extract disparity information. This information is calibrated in order to estimate the distance between the user and a detected landmark (in the future to be combined with heading information from an electronic compass attached to the camera unit).

Finally, the interface module is the link between the SmartVision module and the user. At the moment it serves two outputs and one input. The two outputs are text-to-speech audio and vibration actuators. The vibration actuators are used for local navigation, i.e., obstacle avoidance and heading direction. The audio interface is used for navigating the menus and providing POI information. The user provides input by using a small four-button device to navigate a simple menu system and to select options.

IV. GIS AND NAVIGATION MODULES

Given the fact that Wi-Fi is not everywhere available, all or most information required for global navigation must be stored on the SmartVision prototype. This way it is possible to have access to the information whatever the scenario, both in- and outdoor, without a need to regularly query the GIS server. The geographic information stored on the prototype is updated when Internet connection is available, through the use of webservices. The information is stored in digital map files, or shapefiles [7], and a MySQL database which holds landmark details. For the distribution of geographic information, the adopted architecture is three-tier client/server [22]. In this model, the client application, i.e., the SmartVision prototype, handles and provides geographic information to the user. For a detailed description see [7, 10]. Figure 2 illustrates the client/server architecture.

The navigation modules handle aspects related to the computation of the route that the user must follow, from the initial or actual position to the chosen destiny. In terms of functionality, all local POIs in the database are transmitted to the user by the interface module, in order to choose a desired one. Then, navigation instructions are issued using Dijkstra's Shortest Path First (SPF) routing algorithm. According to Ertl [4], this algorithm is able to calculate, in a graph, the shortest path from a starting vertex to a destination vertex, and this algorithm provides a balanced solution between calculus efficiency and implementation simplicity. Obviously, the road layer in the map is built in a manner similar to a graph, i.e., the points where a road begins, is intersected and ends correspond to the graph's vertices and these are linked by the graph's edges. See [7, 10] for further details.

Concerning the use of Wi-Fi, this technology is also used to compensate for the lack of GPS reception inside buildings. By using the calibrated locations of Wi-Fi access points (APs) in the GIS in combination with a triangulation method based on the APs' signal strengths, it is possible to approximate the user's position with an error margin of 7-8 m, which is worse than Nokia's margin of 2-5 m [3]. Implicitly we assume that there are at least three APs within reception range, i.e., at the time the localisation by Wi-Fi is requested. This solution provides an indoor localisation system very similar to outdoor GPS. However, if less than three APs are within reception range, an alternative to re-calibrate the user's actual position is required. Therefore, in the near future tests with active RFID tags will be conducted in combination with visual landmark (object) recognition by the Vision module. These extensions are required in any case, i.e., if no or few APs are present inside a building or GPS is temporarily not available outdoor.

Figure 2. Three-tier client/server architecture.

V. VISION MODULE

The Vision module provides information for local navigation: the heading direction to keep the user within safe boundaries in corridors and paths including footpaths (sidewalks), possible obstacles in front and beyond the reach of the white cane, and known objects which may serve as navigation landmarks or which can be a destiny (bus stop, taxi stand, cash machine, post office, etc.). The latter aspect must still be studied by using real test video sequences, although first results on the recognition of household items are available (Fig. 5). This module consists of complementary sub-modules: (A) sidewalk and path detection; (B) obstacle detection; (C) landmark detection; (D) detection of moving persons or other objects possibly on collision course; and (E) object recognition.

A. Sidewalk and path detection

There exist a few methods to detect the borders of sidewalks [18]. We detect them by using Canny's edge detector in combination with a tracking mask to obtain straight lines, from the bottom of each frame to the top, characterised by slope, length and proximity to the left and right boundaries of the frame. The de-
ected borders define the horizontal position and width of an obstacle detection window (ODW). As its name indicates, the latter is used for obstacle detection. For a detailed description see [2]. Although first results were good, a new approach is being explored: borders are detected using an adaptive Hough-transform space, which is faster and yields more robust results [12]; see Fig. 4 (bottom).

B. Obstacle detection

Two different methods are applied within the ODW. The first one counts variations of gray values, i.e., local maxima and minima, on each horizontal line inside the ODW. Then, outliers are reduced by averaging counted variations over groups of lines, and variations over these groups are combined into a single value which indicates whether the frame has a possible obstacle or not. Final confirmation is obtained by combining the results of a few subsequent frames. The second method is based on irregularities in vertical and horizontal histograms of the binary edge image. An obstacle can lead to two different signatures: if the pavement is smooth, an obstacle may appear as a local excess of edge points, but if it has a strong texture there will be a huge amount of edge points and an obstacle may appear as a local deficiency (lack or gap) of edge points. The second method is used to confirm the result of the first one, but it also serves to detect the size and position of an obstacle in order to guide the user away from it. Figure 3 shows successive frames 1 to 14 of a test sequence with their annotation, i.e., image type and obstacle alert.

C. Landmark detection

By using a stereo camera, disparity information can be extracted from the captured image frames. Calibrated disparity yields a distance map, for example for knowing the distance between the user and special navigation landmarks on the pavement or, when such landmarks are not present, for detecting the correct heading direction in front but also vertical obstacles like trees, light poles and traffic signs.

First experiments were conducted with high-contrast circles (Fig. 4, top-left) which can easily be segmented by thresholding, followed by edge detection and the Hough transform for circles. Combining the left and right frames, the distance can be computed and also the orientation for advising the blind user to maintain or correct his heading direction [6]. The more general solution, not based on high-contrast circles on the pavement, is already being pursued: general image disparity but, because of limited CPU power available on the portable device, restricted to matching left-right images only at the height of the obstacle detection window. This solution is closely related to optical flow as described below.

Figure 3. An annotated test sequence with the ODW and a detected obstacle.

Figure 4. Top-left: input frame with markers. Top-right and middle: a moving person is picking up one marker and has been detected, also the person’s shadow (white rectangles). Bottom: detected sidewalk borders with and without partial occlusion.

D. Detection of moving objects

Keypoints are 2D singularities like edge crossings and line ends [17]. Keypoint annotation (edge types, orientations, etc.) in combination with multi-scale processing facilitates various processes such as image matching (stereo and optical flow) and object segregation and tracking. Here we focus on optical flow in
order to detect moving objects in front of the user, both approaching objects and those which cross the user's path.

Instead of working with mere point (keypoint) clouds in combination with mathematical matching methods, there is a growing interest in using additional information [20]. There are four reasons for doing this: (1) By definition, keypoints at any scale are caused by junctions of lines and edges at the same scale, and this line/edge information is tractable for keypoint annotation. (2) At coarser scales the bigger filter kernels will lead to few detected lines, edges and keypoints, and the same effect will occur in image pairs - stereo for disparity; sequential for optical flow. (3) Going from a coarse scale to the finest one, few keypoints are located at the centre of an object, then more keypoints are found at object parts, and finally many keypoints at object (and contour) details. (4) Matching annotated keypoints is much easier than matching mere point clouds, especially when one starts at a coarse scale and the matching results there are used to steer the matching process at a finer scale, the procedure being repeated until the finest scale.

The solution of the correspondence problem based on annotated keypoints combines left-right (stereo) and successive (optical flow) frames, comparing keypoints in an area with a radius proportional to the filter kernel's size, and starting at coarse and going to fine scale. At each scale the corresponding vectors are computed or refined. In object segregation and object tracking the processing is the same, matching keypoints being linked over and within scales by using a multi-scale tree structure [20]. With this information we can track the movements of persons, animals or other obstacles (Fig. 4 top and middle row); for more details see [12]. In the future, the goal is to detect which object is moving - man, animal or other - and the type of movement - laterally crossing or approaching. It should be stressed that the goal is not general scene analysis; the processing is only applied between the detected path borders and at short range.

**E. Object recognition**

An important goal is to locate household items, for example on shelves in a pantry, like toothpaste or a bottle of ketchup. To this purpose we use the well-known algorithm coined SURF (Speeded Up Robust Features), which is a scale- and rotation-invariant interest-point detector and descriptor [5]. Bay et al. [5] claim that it outperforms previously proposed schemes like SIFT with respect to repeatability, distinctiveness and robustness, and that it is faster. All this is achieved by relying on integral images for image convolutions and on the strengths of the best existing detectors and descriptors - Hessian matrix-based detector and distribution-based descriptor - simplifying these methods to the essential.

Extensive tests are being conducted with many typical household objects, using the OpenSURF library [1], a C/C++ implementation. Depending on the complexity of each object, a single or a few different views are used for training, their features being used in realtime for locating and identifying the objects arranged on shelves. Figure 5 shows typical images with recognised objects. As expected, the recognition rate is very good, provided that the camera is at reasonably short distance, about 0.5 m. Of course, the user must be trained in holding and pointing the camera, also in communicating with the system in order to optimise performance. In the near future, similar tests will be conducted to recognise bigger but important objects for navigation, first indoor (elevator, toilet, welcome desk, furniture, plants) and then outdoor (cash machine, telephone booth, etc.).

**VI. WHITE CANE AND PASSIVE RFID TAGS**

Above, the use of white markers on the pavement for aiding local navigation was discussed. One of the main problems of such markers is the necessity of a light source when it is dark. An alternative solution is to use passive RFID tags in the pavement, which always respond provided that a signal-emitting antenna is at close range, less than 10 cm. Experiments were conducted with an RFID antenna in the very tip of a cane made of light PVC, similar to the ones used by blind persons. The handle contains a vibration actuator and close to the handle was mounted the tag reader with Bluetooth interface; see Fig. 6 (top-left) and [11].

The RFID tag reader used is from Phidgets, mainly because of previous experience and easy implementation. The passive tags have a working frequency of 135 kHz. The tags have the shape of a nail or golf tee, 2.5 cm long and 4 mm wide, and the electronics are fixed inside a hard and resistant plastic encapsulation;
see Fig. 4 (top-right). The tags can be easily inserted into drilled holes in the pavement.

In the experiments we tested different tag layouts, both in a laboratory and on UTAD Campus, in order to find the best layout. The main problem is the combination of the walking speed of the user, the swaying of the cane and the distance of the cane’s tip to the surface of the pavement, i.e., the probability of “hitting” a tag with some regularity. Initially the tags were distributed on the pathway along three parallel lines: two lines close to the path borders and one line in the centre to alert the user to special events like fixed obstacles (trash cans, trees) and change of direction. To this end the codes of the tags are stored in a small database together with event types, and the cane’s vibrator is activated with different temporal patterns. In a second configuration, a single line along the path’s centre was tested in combination with four-by-four tag arrays at bifurcations etc.; see Fig. 6 (bottom). The latter solution proved to be better, only requiring additional tests in order to optimise the tags’ spacing for increasing the “hit” rate. Results of the tests and further details are described in [11].

Figure 6. Top-left: PVC cane with RFID reader. Top-right: golf tee-shaped passive RFID tags. Bottom: proposed layout of the tags in the pavement.

VII. INTERFACE MODULE

When developing a navigation aid for visually impaired users, the user interface has to be intuitive and extremely easy to use [21]. The interface can only rely on touch and hearing, but it may not interfere with normal hearing of surround sounds nor the normal use of the white cane. The actual version of the SmartVision prototype employs speech synthesis, a vibration actuator in the handle of the cane, and a small handheld box with only 4 pushbuttons. Each interface has a specific use. The speech module is used to guide the user with brief messages, or longer messages but only when intentionally requested by the user. The vibration module is used when very simple instructions like “turn left” or “turn right” are issued. One study proposed that vibration actuators could be used in the

see Fig. 4 (top-right). The tags can be easily inserted into drilled holes in the pavement.

In the experiments we tested different tag layouts, both in a laboratory and on UTAD Campus, in order to find the best layout. The main problem is the combination of the walking speed of the user, the swaying of the cane and the distance of the cane’s tip to the surface of the pavement, i.e., the probability of “hitting” a tag with some regularity. Initially the tags were distributed on the pathway along three parallel lines: two lines close to the path borders and one line in the centre to alert the user to special events like fixed obstacles (trash cans, trees) and change of direction. To this end the codes of the tags are stored in a small database together with event types, and the cane’s vibrator is activated with different temporal patterns. In a second configuration, a single line along the path’s centre was tested in combination with four-by-four tag arrays at bifurcations etc.; see Fig. 6 (bottom). The latter solution proved to be better, only requiring additional tests in order to optimise the tags’ spacing for increasing the “hit” rate. Results of the tests and further details are described in [11].

Figure 6. Top-left: PVC cane with RFID reader. Top-right: golf tee-shaped passive RFID tags. Bottom: proposed layout of the tags in the pavement.

VII. INTERFACE MODULE

When developing a navigation aid for visually impaired users, the user interface has to be intuitive and extremely easy to use [21]. The interface can only rely on touch and hearing, but it may not interfere with normal hearing of surround sounds nor the normal use of the white cane. The actual version of the SmartVision prototype employs speech synthesis, a vibration actuator in the handle of the cane, and a small handheld box with only 4 pushbuttons. Each interface has a specific use. The speech module is used to guide the user with brief messages, or longer messages but only when intentionally requested by the user. The vibration module is used when very simple instructions like “turn left” or “turn right” are issued. One study proposed that vibration actuators could be used in the
already been approved. This extension suffices to optimise all modules and to include the newest and best localisation solutions like Nokia’s indoor with a better precision. In addition, extensive field tests are planned in collaboration with ACAPO, the Portuguese association for amblyopes and the blind. The goal is to convert the prototype into a real system which can be assembled and maintained by non-experts, with detailed instructions. This system is coined Blavigator, from blind navigator.

REFERENCES


A vision system for detecting paths and moving obstacles for the blind

J. José, M. Farrajota, J.M.F. Rodrigues, J.M.H. du Buf

Vision Laboratory, Institute for Systems and Robotics (ISR), University of the Algarve (FCT and ISE), Faro, Portugal

{jose, jrodrig, dubuf}@ualg.pt and elsio_farrajota@hotmail.com

Abstract — In this paper we present a monocular vision system for a navigation aid. The system assists blind persons in following paths and sidewalks, and it alerts the user to moving obstacles which may be on collision course. Path borders and the vanishing point are detected by edges and an adapted Hough transform. Optical flow is detected by using a hierarchical, multi-scale tree structure with annotated keypoints. The tree structure also allows to segregate moving objects, indicating where on the path the objects are. Moreover, the centre of the object relative to the vanishing point indicates whether an object is approaching or not.

I. INTRODUCTION

Navigation of blind people is very arduous because they must use the white cane for obstacle detection while following the front sides of houses and shops, meanwhile memorising all locations they are becoming familiar with. In a new, unfamiliar setting they completely depend on people passing by to ask for a certain shop or the closest post office. Crossing a street is a challenge, after which they may be again disoriented. In a society in which very sophisticated technology is available, from tracking GPS-RFID equipped containers in an area of hundreds of meters to GPS-GIS car navigation to Bluetooth emitting the sound of movie trailers to mobile phones in front of cinemas, one can question what it may cost to provide the blind with the most elementary technology to make life a little bit easier. This technology may not replace the cane, but should complement it: alert the user to obstacles with their position and size, such that the user can be alerted and informed about the best way to avoid them.

There exist some methods to detect the borders of paths and sidewalks, see e.g. [9]. In a previous paper [2] we presented a detection method for paths with fixed obstacles. The system first detects the path borders, using edge information in combination with a tracking mask, to obtain straight lines with their slopes and the vanishing point. One of its modules serves to help the blind navigate outdoors on paths and sidewalks. It must alert the user to possible obstacles, both fixed objects and moving ones like persons and animals which may be on collision course, and how to avoid them.

Acknowledgements: This research was supported by the Portuguese Foundation for Science and Technology (FCT), through the pluri-annual funding of the Institute for Systems and Robotics (ISR-Lisbon/IST) by the POS_Conhecimento Program which includes FEDER funds, and by the FCT project SmartVision: active vision for the blind (PTDC/EIA/73633/2006).
II. PATH DETECTION

In the SmartVision project, a stereo camera (Bumblebee 2 from Point Grey Research Inc.) is fixed to the chest of the blind, at a height of about 1.5 m from the ground. Results presented here were obtained by using only the right-side camera, and the system performs equally well using a normal, inexpensive webcam with about the same resolution. The resolution must be sufficient to resolve textures of the pavements related to possible obstacles like holes and loose stones [2] with a minimum size of about 10 cm at a distance of 3 to 5 m from the camera (the first metre is not covered because of the height of the camera; this area is covered by the cane swayed by the user).

Detection of path borders is based on: (A) defining a Path Detection Window (PDW) where we will search for the borders in each frame; (B) some pre-processing of the frame to detect the important edges and to build an Adapted Hough Space (AHS); and (C) the highest values in the AHS yield the borders.

A. Path Detection Window PDW

Let \(I(x, y)\) denote an input frame with fixed width \(W_I\) and height \(H_I\). Let \(HL\) denote the horizon line close to the middle of the frame. If the camera is exactly in the horizontal position, then \(HL = H_I/2\). If the camera points lower or higher, \(HL\) will be higher or lower, respectively; see Fig. 1. The borders of the path or sidewalk are normally the most continuous and straight lines in the lower half of the frame, delimited by \(HL\).

The value of \(HL\) is computed dynamically, by averaging the values of the previous five frames: for \(HL_i\), i.e., frame number \(i\) which still must be analysed, \(y_{VP,i} = (\sum_{j=1}^{i-5} y_{VP,j})/5\). This cannot be done in the case of the first five frames, for which we use \(y_{VP} = H_I/2\). This is not a problem because the first frames are mainly used for system initialisation and a frame rate of 5 fps implies only one second.

B. Adapted Hough Space AHS

The Canny edge detector and an adapted version of the Hough transform are applied to \(I_{PDW}\) for the detection of the borders and the vanishing point. In order to reduce CPU time, only gray-scale information is processed after resizing the window to a width of 300 pixels using bilinear interpolation, maintaining the aspect ratio. Then two iterations of a 3x3 smoothing filter are applied in order to suppress noise.

The Canny [3] edge detector is applied with \(\sigma = 1.0\), which defines the size of the Gaussian filter, in combination with \(T_s = 0.25\) and \(T_H = 0.5\) which are the low and high thresholds for hysteresis edge tracking. The result is a binary edge image \(I_p(x_p, y_p)\), of width \(W_p = 300\) and height \(H_p = H_I(W_p/W_I)\), with \(x_p = [1, W_p]\) and \(y_p = [1, H_p]\), and with the extrapolated horizontal line (vanishing point) at \(y_{VP} = y_{VP}(W_p/W_I)\). Figure 2 (left) shows one original frame together with the resized and low-pass-filtered PDW (top-right) and detected edges (bottom-right).

The borders of paths and sidewalks are usually found to the left and to the right, assuming that the path or sidewalk is in the camera’s field of view; see e.g. Fig. 4. We use the Hough transform [12], where \(\rho = x \cos \theta + y \sin \theta\), to search for lines in the left and right halves of the PDW for border candidates, also assuming that candidates intersect at a vanishing point.

As we want to check straight lines in the two halves of the window using polar coordinates, we use a different reference point. Let \(l_P'(x'_P, y'_P)\), see Fig. 3 (top), be the new origin at the bottom-centre of image \(I_p\), where \(y'_P \) is related to \(y_P\) by \(y'_P = y_P - (y_p - 1)\) and \(x'_P\) is related to \(x_P\) by \(x'_P = (x_P - 1) - W_p/2\), with \(y'_P = [1, H_p]\) and \(x'_P = [-W_p/2, W_p/2 - 1]\).
The Hough transform is applied to \( I'_{p} \), yielding the Adapted Hough Space \( I_{\text{AHS}}(\rho, \theta) \) with \( \theta = [0^\circ, 180^\circ] \) and \( \Delta \theta = 0.5^\circ \), and \( \rho = [0, W_{p}/2, \cos \theta + H_{p}, \sin \theta] \). For \( \theta = [0^\circ, 44^\circ] \) we apply \( \Delta \rho = \cos \theta \), and for \( \theta = [45^\circ, 90^\circ] \) we apply \( \Delta \rho = \sin \theta \), in order to adjust the interval to polar coordinates such that no lines are repeated or missed.

For \( \rho \) we apply \( \frac{\rho}{H_{p}} \), and for \( \theta \) we apply \( \frac{\theta}{H_{p}} \), in order to adjust the interval to polar coordinates such that no lines are repeated or missed.

\[ \begin{align*}
\text{Figure 3. Top: PDW with detected edges and coordinate systems. Middle: AHS with zoomed areas (at bottom). The left and right borders are marked in red and green, respectively.}
\end{align*} \]

Since we have “mirrored” lines for \( \theta \) and \( \pi - \theta \) for the same values of \( \rho \), we only calculate the lines for \( \theta = [0^\circ, 89^\circ] \) for the right border, which we denote by \( L_{p,\theta}(x'_{p,r}, y'_{p}) \). Similarly, the lines for \( \theta = [91^\circ, 180^\circ] \) are computed for the left border, \( L_{p,\pi-\theta}(x'_{p,l}, y'_{p}) \), because of the \( y'_{p} \) mirror axis.

For \( L_{p,\theta} \) we use \( x'_{p,r} = (\rho - y'_{p} \sin \theta)/\cos \theta \) and \( y'_{p} = (\rho - x'_{p,r} \cos \theta)/\sin \theta \). For \( L_{p,\pi-\theta} \) we use a vertical line with \( y'_{p} = [1, H_{p}] \) and \( x'_{p,r} = 0 \), and \( I_{\text{AHS}}(0,180^\circ) \) has the same \( y'_{p} \) but \( x'_{p,l} = -1 \). For obtaining the maximal number of pixels on the projected lines \( L_{p,\theta} \) and \( L_{p,\pi-\theta} \), we increment by 1 the \( y'_{p} \) and compute the corresponding \( x'_{p,r} \) and \( x'_{p,l} \) for \( \theta = [0^\circ, 44^\circ] \). For \( \theta = [45^\circ, 89^\circ] \) we increment by 1 the \( x'_{p,r} \), with \( x'_{p,l} = -x'_{p,r} - 1 \), and calculate the corresponding \( y'_{p} \).

The \( I_{\text{AHS}} \) space is filled by checking the pixels in \( I'_{p} \) from top to bottom: left-to-right for the right border \( (L_{p,\theta}) \) and right-to-left for the left border \( (L_{p,\pi-\theta}) \). As for the normal Hough space, \( I_{\text{AHS}} \) is a histogram which is used to count the co-occurrences of aligned pixels in the binary edge map \( I'_{p} \). However, there are two differences.

First, vertical and horizontal lines in the image cannot be borders of a path or sidewalk (see e.g. Fig. 1). Hence, we restrict the Hough space to \( \theta = [20^\circ, 69^\circ] \cup [111^\circ, 160^\circ] \) such that vertical and almost vertical lines in the intervals \( \theta = [0^\circ, 20^\circ] \) and \( [160^\circ, 180^\circ] \) are ignored; the same is done for horizontal and almost horizontal lines in the interval \( [70^\circ, 110^\circ] \). This yields a reduction of CPU time of about 30%.

Second, longer sequences of edge pixels count more than short sequences or not-connected edge-pixels. To this purpose we use a counter \( P \) which can be increased or reset. When we check each pixel in \( I'_{p} \) for a projected line \( L_{p,\theta} \) and find the 1st ON pixel, the corresponding \( I_{\text{AHS}}(\rho, \theta) \) bin will increment by 1. If the 2nd pixel following the first ON pixel is also ON, \( P \) is incremented by 2, and \( I_{\text{AHS}}(\rho, \theta) \) is incremented by \( P = 3 \), and so on. If a next pixel is OFF, the variable \( P \) is reset to 0 and \( I_{\text{AHS}}(\rho, \theta) \) is not changed. In other words, a run of \( n \) connected edge pixels has \( P \) values of \( 1, 3, 5, 7, \ldots \), or \( P = n - 1 \) + 2, with \( P = 1 \), and the run will contribute \( n^2 \) to the relevant \( I_{\text{AHS}} \) bin.

The final value of the \( I_{\text{AHS}}(\rho, \theta) \) bin is the sum of the \( P \) value(s) of all sequences of ON pixel(s): \( I_{\text{AHS}}(\rho, \theta) = \sum_{k=1}^{k} P_{n_{k}} \), with \( k \) the number of sequence(s) of ON pixel(s) and each sequence having at least one ON pixel.

An \( I_{\text{AHS}} \) is shown in Fig. 3 (middle) together with magnified regions (bottom). The left and right borders are marked in red and green, respectively, also in the edge map \( I'_{p} \) (top).

C. Path borders

Until here we explained the computation of \( I_{\text{AHS}} \), but only during the initialisation phase of the first 5 frames. After the initialisation phase, for optimisation and accuracy purposes, we will not check the entire \( I_{\text{AHS}} \) space. Each border \( (\rho, \theta) \), both left and right, is stored during the initialisation in the array \( M_{i}(\rho, \theta) \), with \( i \) the frame number.

After the fifth frame \( (i = 6) \), we already have five pairs of points in \( M \), which define two regions in \( I_{\text{AHS}} \). These regions indicate where in \( I_{\text{AHS}} \) the next border positions are expected. The regions are limited by the
In frames $i \geq 6$, we look for the highest value(s) in $I_{\text{HS}}(\rho, \theta)$ in the regions between $\rho_{\text{min,}i/r} - T_\rho$ and $\rho_{\text{max,}i/r}$, and between $\theta_{\text{min,}i/r} - T_\theta$ and $\theta_{\text{max,}i/r} + T_\theta$, on the left and on the right side, respectively, with $T_\rho = 10$ and $T_\theta = 5^\circ$. This procedure is applied for all $i \geq 6$, always considering the borders found in the previous five frames.

In the two regions as defined above we look for the highest values in $I_{\text{HS}}$. We start by checking the highest value, and then the 2nd highest value. If the 2nd highest value represents a border which is more similar to the border of the previous frame, we still check the 3rd highest value and so on. If a next highest value does not correspond to a border which is more similar to the border of the previous frame, the search is terminated and the best match is selected.

Borders are considered more similar if the intersection of the new candidates $(VP_i)$ and the intersection of the borders of the previous frame $(VP_{i-1})$ have a smaller distance $d$, with $d = (\sum_{j=1}^{5} \rho_j)^2 + (\sum_{j=1}^{5} \rho_j)^2)^{1/2}$.

In this search, all combinations of left and right border candidates are considered. If in the left or right regions where the $I_{\text{HS}}$ values are checked there is no maximum which corresponds to at least one sequence of at least 10 connected ON pixels, the border is considered not found for that side. In this case, the average of the last 5 borders found is used: $\bar{\rho}_i = \sum_{j=1}^{5} \rho_j / 5$ and $\bar{\theta}_i = \sum_{j=1}^{5} \theta_j / 5$ on the corresponding side. Figure 4 shows the results of path and border detection in the case of two image sequences.

### III. DETECTION OF MOVING OBJECTS

Apart from detecting path borders and possible stationary obstacles on the path, see also [2], it is necessary to detect and track moving obstacles like persons and animals. To this purpose we use multi-scale, annotated, and biologically-inspired keypoints. Keypoint detection is based on Gabor filters [6], and provides important image information because keypoints code local image complexity. Moreover, since keypoints are caused by line and edge junctions, detected keypoints can be classified by the underlying vertex structure, such as K, L, T, + etc. This is very useful for matching problems: object recognition, stereo disparity and optical flow.

The process for tracking moving objects consists of three steps: (A) multi-scale keypoints are detected and annotated; (B) multi-scale optical flow maps are computed and objects are segregated; and (C) the regions that enclose objects allow us to track the objects’ movements and their directions.

A. Keypoint Detection and Annotation

Gabor quadrature filters provide a model of cortical simple cells [6]. In the spatial domain $(x, y)$ they consist of a real cosine and an imaginary sine, both with a Gaussian envelope.

Responses of even and odd simple cells, which correspond to real and imaginary parts of a Gabor filter, are obtained by convolving the input image with the filter kernels. Responses are denoted by $R_k^E(x, y)$ and $R_k^O(x, y)$, $s$ being the scale given by the wavelength $\lambda$ ($\lambda = 1$ corresponds to 1 pixel), and $j$ the orientation $\theta_j = j\pi/N_\theta$ with $N_\theta$ the number of orientations. We use 8 orientations, $j = [0, N_\theta - 1]$, and 8 scales equally spaced on $\lambda = [6, 27]$ with $\Delta \lambda = 3$.

Responses of complex cells are modelled by the modulus $C_{x,y}$, which feed two types of end-stopped cells, single $C_{x,y}$ and double $D_{x,y}$; for details see [6]. Responses of end-stopped cells in combination with sophisticated inhibition schemes yield keypoint maps $K_{x,y}$.

In order to classify any detected keypoint, the responses of simple cells $R_k^E(x, y)$ and $R_k^O(x, y)$ are analysed, but now using $N_{\phi} = 2 \times N_\theta$ orientations, $\phi_k = k\pi/N_\theta$ and $k = [0, N_{\phi} - 1]$. This means that for each Gabor filter orientation on $[0, \pi]$ there are two opposite keypoint classification orientations on $[0, 2\pi]$, e.g. a Gabor filter at $\theta_1 = \pi/N_\theta$ results in $\phi_1 = \pi/N_\theta$ and $\phi_1 = \pi/N_\theta$.

Classifying keypoints is not trivial, because responses of simple and complex cells, which code the underlying lines and edges at the vertices, are unreli-
able due to response interference effects [8]. This implies that responses must be analysed in a neighbourhood around each keypoint, and the size of the neighbourhood must be proportional to the scale of the cells, i.e., the size of the filter kernels.

The validation of line and edge orientations which contribute to the vertex structure is based on an analysis of the responses, both $R^E_{x,y}$ and $R^O_{x,y}$, and consists of three steps: (1) only responses with small variations at three distances are considered, (2) local maxima of the responses over orientations are filtered and the remaining orientations are discarded, and (3) even and odd responses are matched in order to filter the orientations which are common to both. The same processing is applied at any scale $s(\lambda)$.

In step (1), for each orientation $\phi_k$ the responses of the simple cells on three circles around the keypoint positions, with radii $\lambda/2$, $\lambda$ and $2\lambda$, are compared, considering also orientation intervals $\phi_k \pm \pi/N_\phi$. The three maximum responses $R_{k,r}$ in the orientation interval around $k$ and at the three radii $r$ are detected, and their maximum $R_k = \max_{r} R_{k,r}$. Only responses with small variations at the three radii are considered, i.e., $R > 0.6 R_k$.

In step (2), the average $\overline{R} = (1/N_\phi) \sum_k R_k$ is computed and, for validation purposes, all $\overline{R}_k$ below $0.95\overline{R}$ are suppressed. If there exist maximum responses $\overline{R}_k$ for the two neighbouring orientations $\phi_{k-1}$ and $\phi_{k+1}$, they will be removed if $\overline{R}_{k-1} < 0.95\overline{R}_k$ or $\overline{R}_{k+1} < 0.95\overline{R}_k$. The above values were determined by analysing simple objects like triangles, squares and polygons.

The analysis in step (3) only concerns the matching of equal orientations, discarding any orientation which is not detected in the responses of both even ($R^E_{x,y}$) and odd ($R^O_{x,y}$) simple cells. Remaining orientations $\phi_k$ are attributed to the keypoint, plus the junction type L, T, + etc.

In the above procedure there is only one exception: keypoints at isolated points and blobs, especially at very coarse scales, are also detected but they are not caused by line/edge junctions. Such keypoints are labeled “blob” without attributed orientations.

Figure 5 illustrates responses of simple cells in the case of a black square in a noisy background. It shows two scales, $\lambda = 6$ (column 1 and 2) and $\lambda = 15$ (column 3 and 4), only three orientations of all 8, even cells in columns 1 and 3 and odd cells in columns 2 and 4. Dark levels are negative and bright ones are positive. Also shown is one detected keypoint at each scale with (in red) the three distances $\lambda/2$, $\lambda$ and $2\lambda$, at which the responses are tested.

Figure 6 shows keypoint detection and annotation results together with optical flow. At top-left it shows one frame from the sequence shown in Fig. 1, and at top-right a combination of two successive frames. The second row shows keypoints detected at two scales, $\lambda = 6$ (left) and 15 (right). The third row shows annotated keypoints at the two scales, and the fourth displacement vectors between the successive frames.

### B. Optical Flow

To compute the optical flow, we do not consider each scale independently for two reasons: (1) non-relevant areas of the image can be skipped because of the hierarchical scale structure, and (2) by applying a multi-scale strategy, the accuracy of keypoint matching can be increased, thus increasing the accuracy of the overall optical flow. Therefore we apply a multi-scale tree structure in which at the coarsest scale a root keypoint defines a single object, and finer scales add more keypoints which constitute the object’s parts and details. As stated before, at a very coarse level a keypoint may correspond to one big object. However, because of limited CPU time the coarsest scale applied will be $\lambda = 27$, which is a compromise between speed and quality of results. Hence, at the moment all keypoints at $\lambda = 27$ are considered to represent individual objects, although we know that several of those may belong to the same object.

Each keypoint at the coarsest scale can be linked to one or more keypoints at one finer scale, which can be slightly displaced. This link is created by down-projection using an area with the size of the filter ($\lambda$). This linking is repeated until the finest scale is reached. Hence, keypoints at a finer scale which are outside the “circle of influence” of keypoints at a coarser scale will not be relevant, thus avoiding unnecessary computations. Figure 7 illustrates the linking principle by cones.

At any scale, each annotated keypoint of frame $i$ can be compared with all annotated keypoints in frame $i-1$. However, this comparison is restricted to an area of radius $\lambda$ in order to save time, because (1) at fine scales many keypoints outside the area can be skipped since they are not likely to match over large distances.
and (2) at coarse scales there are less keypoints, the radius $\lambda$ is bigger and therefore larger distances (motions) are represented there. The tree structure is built top-down (Fig. 7), but the matching process is bottom-up: it starts at the finest scale because there the accuracy of the keypoint annotation is better.

Keypoints are matched using three similarity criteria with different weight factors: the distance $D$, the attributed orientations $O$, and the tree correspondence $C$. The distance $D$ serves to emphasise keypoints which are closer to the centre of the area. For having $D=1$ at the centre and $D=0$ at radius $\lambda$, we use $D = (\lambda - d)/\lambda$ with $d$ the Euclidean distance. The orientation error $O$ measures the differences of the attributed orientations, but with a relaxation of $\pm \pi/N_9$ of all orientations such that also small rotations are allowed. Similar to $D$, the summed differences are combined such that $O = 1$ indicates good correspondence and $O = 0$ a lack of correspondence. Obviously, keypoints marked “blob” do not have orientations and are treated separately. Parameter $C$ measures the number of matched keypoints at finer scales, i.e., at any scale coarser than the finest scale. The keypoint candidates to be matched in frame $i$ and in the area with radius $\lambda$ are linked in the tree to localised sets of keypoints at all finer scales. The number of linked keypoints which have been matched is divided by the total number of linked keypoints. Hence, parameter $C$ describes the consistency of the matching at a candidate's position at the finer scales, thereby influencing the matching of the candidate at the actual scale.

The three parameters are combined using the similarity measure $S = \alpha(\beta P_0 + (1 - \beta) P_7) + (1 - \alpha) P_9$ with parameters $\alpha = 0.7$ and $\beta = 0.6$. These values were determined empirically. The candidate keypoint with the highest value $S$ in the area ($\lambda$) is selected and the vector between the keypoint in frame $i$ and the matched one in frame $i-1$ is computed. The remaining candidates in the area can be matched to other keypoints in frame $i-1$, provided they are in their local area. Keypoints which cannot be matched are discarded. Figure 6 shows, at the bottom, the vectors between matched keypoints at $\lambda = 6$ (left) and $\lambda = 15$ (right). Since optical flow in this example is mainly due to movement of the camera, it can be seen that there are some errors. In principle, outliers could be removed, but improvement of the matching process is still subject to ongoing research.

IV. TRACKING OF OBJECTS ON COLLISION COURSE

As mentioned above, at a very coarse scale each keypoint should correspond to an individual object. However, at the coarsest scale applied ($\lambda = 27$) this may not be the case and an object may create several keypoints. In order to determine which keypoints may
belong to the same object we combine saliency maps with the multi-scale tree structure.

Figure 8. A sequence of 4 frames (top) with saliency maps at left and optical flow at right, both at $\lambda=6$.

A saliency map can be based on keypoints as these code local image complexity [6]. Such a map is created by summing detected keypoints over all scales $s$, such that keypoints which are stable over scale intervals yield high peaks, but in order to connect the individual peaks and yield regions a relaxation area is applied. As applied above, the area is proportional to the scale and has radius $\lambda$. Here, in order to save CPU time, the process is simplified and saliency maps are created by summing responses of end-stopped cells [6]. Figure 8 (bottom-left) shows three examples scaled to the interval [0, 255], but only at scale $\lambda=6$.

The saliency map of a frame defines, after thresholding, separated regions-of-interest (RoI) and these can be intersected with the regions as defined by the tree structure (Fig. 7). Hence, neighbouring keypoints are grouped together in the RoIs and their displacement vectors after the matching process yield the optical flow of segregated image regions, i.e., where an individual object or a combination of connected (sub)objects is or are moving. In order to discard small optical flow due to camera motion, optical flow vectors are only computed if at least 4 scales the matched keypoints in successive frames have displacement vectors with a length which is bigger than 1 pixel.

Figure 8 (top) shows four successive frames of a sequence. The bottom part shows saliency maps of the last three frames (at left) together with optical flow vectors (at right), both at one fine scale. By using the intersections of the thresholded saliency maps and the areas defined by the tree structure, the optical flow vectors of the keypoints in segregated regions can be averaged, and the intersected RoIs can be approximated by curve fitting, here simplified by a rectangular bounding box. The centre of the bounding box is used for tracking moving objects over frames, also indicating where exactly the object is on the path. This is illustrated in Fig. 9. It shows parts of two sequences, and in each frame the detected path borders and the bounding boxes plus one combined image which shows the tracking of the centre of the bounding box. The flow vector of the centre allows us to detect the lateral movements left-to-right and right-to-left, and the distance between the centre and the vanishing point of the path borders indicates whether the object is moving towards or away from the camera. This is shown by the colour of the arrows in the top-right corners of the frames: yellow means movement at the same distance, green means going away, but red indicates an approaching object such that the user can be alerted to be even more cautious than in the yellow or green cases.

V. CONCLUSIONS

We presented a system for detecting path borders and the vanishing point, together with a biologically inspired algorithm for optical flow based on multi-scale keypoint annotation and matching. Moving obstacles can be detected and tracked, such that the blind user can be alerted and informed about the approximate position on the path and whether the object is approaching or not. Detection of moving obstacles complements detection of static obstacles in front on the path, just beyond the reach of the white cane [2].
Having a reasonably fast algorithm for optical flow, the same algorithm can be applied to stereo disparity in order to also estimate the distance to objects, both moving and static. The algorithms will be integrated in the SmartVision prototype, which also employs a GIS with GPS, WiFi and passive as well as active RFID tags [4].

The developed vision system is not unique. Recently, a similar system has been developed for intelligent cars, for tracking roads and lanes and for detecting possible obstacles like pedestrians [11]. The basic concepts like borders, vanishing point and optical flow are the same, but the implementation is completely different. This is also due to the different requirements: a car may have a speed of 100 km/h, but blind persons with the white cane do not exceed 1 m/s. However, all CPU power of a portable computer will be required because the ultimate goal is to substitute a big part of the functionality of a normal visual system.

REFERENCES


A Non-Visual Photo Collection Browser based on Automatically Generated Text Descriptions

Frode Eika Sandnes
Faculty of Engineering, Oslo University College
N-0130 Oslo, Norway
e-mail: frodes@hio.no

Abstract—This study presents a textual photo collection browser that automatically and quickly analyses large personal photo collections and produces textual reports that can be accessed by blind users using either text-to-speech or Braille output devices. The textual photo browser exploits recent advances in image collection analysis and the strategy does not rely on manual image tagging. The reports produced by the textual image browser gives the user a gist about where, when and what the photographer was doing in the form of a story. Although yet crude, the strategy can give blind users a valuable overview about the contents of large image collections and individual images which otherwise are totally inaccessible without vision.

Keywords—image tagging; photo browsing; visual impairment; textual image descriptions

I. INTRODUCTION

Low cost and high quality camera equipment combined with inexpensive and large memory have resulted in an explosion in the size and number of personal image collections. Most users enjoy taking photographs. The management and retrieval in such large collection is an emerging challenge [5] – even for users with perfect vision as the sheer volume of images is overwhelming. Manual tagging of individual images is tedious, labourous and seems to be a lost battle, although some researchers are working on creative and playful ways, or games, for tagging images [3, 4]. In fact, the image tagging problem is well known in the universal design and accessibility community as the problem of missing textual, or alt-text, descriptions of images [5]. However, both face recognition and landmark recognition seem to be a lost battle, although some researchers are working on creative and playful ways, or games, for tagging images [3, 4]. In fact, the image tagging problem is well known in the universal design and accessibility community as the problem of missing textual, or alt-text, descriptions of images [5].

Images are highly visual objects that obviously are not accessible to blind individuals. From a blind users’ perspective an image collection is simply a hierarchy of folders with numbered filenames. These filenames carry no useful information beyond indicating the sequence in which the photographs were taken. Date and time at which the images were taken are easily available, but carry no significant meaning to users without additional information or a context. Perhaps the key valuable clue in such collections may be the folder names as many photographers occasionally empty their cameras on their computers and create a new folder with some approximate short description of the context or setting.

Blind users may be interested in learning about the contents of image collections for a number of reasons. The prime reason is that images are social objects that we like to share, and blind users may not like to be completely excluded from the activity, yet they may want independence and not having to rely on others explaining the contents of images. Second, a blind user who is a part of a family with non-blind family members may want to know the contents of the images stored on the family computer.

Other multimedia formats such as audio and video are more accessible to blind users because of the sound. Moreover, the most common application, music collections are also often coded with text such that each audio file has a descriptive name and additional meta information coded such as the artist, album and song. This allows for non-visual browsing even without having to listen to the audio itself.

Recent advances in image analysis can also benefit non-visual interpretation of images. Google recently introduced face recognition in their Picasa photo browser. Clearly, information about which people are present on what photographs could be of interest to blind users. Knowing who is present in each picture can help establish a better understanding of the image collection structure. Moreover, if the blind user is familiar with a particular person the image will have a totally new meaning to the user. Moreover, landmark recognition in images is an ambitions, but promising, approach that may prove useful [14]. Blind users may find it interesting to know if an image contains a well known landmark – especially if the user is familiar with the landmark, has read descriptions about it, etc. Again, the image may become more meaningful to the user. Moreover, landmarks are associated with locations and locations are interesting when establishing the context and storyline for a sequence of photographs. Faces and landmarks are important as they make up a majority of the contents in images. However, both face recognition and landmark recognition are computationally expensive and not easily available. Moreover, landmark recognition requires a large database of landmarks to be useful.

In this study a similar approach is taken with very simple means based on the authors own recent results [8-11]. Image metadata are analyzed and the information are used to put an image collection into context as a comprehensible story suggesting where, when and what the photographer was doing.
II. METHOD

Most modern cameras store information about each photograph taken in special EXIF headers [7]. Such information includes the time and date the image was taken, optical characteristics such as shutter speed, aperture, iso film speed, the focal length of the lens and often special maker notes are provided with camera specific metadata. The EXIF format also support geo-coded images, that is, images taken together with a GPS measurement [1]. However, few cameras are equipped with GPS receivers.

It is the information stored in the EXIF headers that is the basis for the text based photo browser presented herein. There are two particular characteristics that are used for analysis, namely temporal features and exposure related features.

A. Temporal analysis

1) Events

Temporal features are well exploited by photo browsing applications. Applications such as Picasa primarily organize images chronologically into year and date. However, this is often done quite statically. In this study we propose to organize chronologically into events. An event is here defined as a physical event experienced by the photographers such as a two week holiday trip to Lisbon, Portugal, a three day conference trip to Oxford, England, etc. Photo collections often carry this signature, that is, long time intervals with no photographs, then a few days with bursts of photograph indicating an interesting event.

To identify the events we simply scanned all the images in chronological order. If the difference between the timestamps of two images exceeded three days they were marked as belonging to different events, while if separated by less than three days then they were considered as belonging to the same event.

This organization allows for abstract descriptions such as “During 2010 there were 3 events – one during March 6-7, one during July 7-10 and the last during October 23-25” to be automatically composed. Such descriptions can be further enhanced with additional information such as if it was a weekday or weekend. For example “During 2010 there were 3 events – one during the weekend of March 6-7, one during the week July 7-10 and the last during the weekend October 23-25”. Whether it is weekend or weekday can greatly affect our perception and interpretation as we usually have free time in the weekend for leisure, while activities during the week are often work related. Next, this linguistic information is helpful as few people are able to determine whether a date occurred during the week or during a weekday.

Further, linguistic additions could be added such as the season (Spring, Summer, Autumn, Winter, and special holidays such as Christmas, Easter, Chinese Lunar New Year, etc. For example, “During 2010 there were 3 events – one during the Spring weekend of March 6-7, one during the Summer week July 7-10 and the last during the Autumn weekend October 23-25”.

2) Local time

People travel more than ever before and we typically take many photographs when we are visiting a new place. Assuming that we approximately know the time-zone of the location for a particular event then we can deduce the local time of the photographs. How to determine the time-zone is outlined in section 2.2. Given the local time for an event we can provide the following textual summarizations “During the first day of the Spring weekend 123 pictures were taken before noon, and 14 in the afternoon. During the second day 23 pictures were taken in the afternoon and 45 in the evening”. This information tells us that the first day was probably spent outdoors in the morning and that there was some party the second in the evening. One could soften the statistics and give a more vague human-like description to ease comprehension and overview. For example “During the first day of the Spring weekend most pictures were taken before noon, and a few in the afternoon. During the second day a few pictures were taken in the afternoon and most in the evening”.

B. Exposure analysis

Exposure values stored in image files provides valuable information about the lighting conditions when the photograph was taken. Although two images may appear to be subject to the same lighting condition this may not be the case. Cameras typically adjust the shutter speed, aperture and the film speed to adjust to various lighting conditions. A measure known as exposure value has been proposed as a standard for indicating the lighting conditions of a scene and it is easily computed using the shutter speed, aperture and film speed [2, 6]. The exposure value falls in the range of 0 (total darkness) to 16 (bright sunshine) or above.

1) Outdoor/indoor

The exposure value is a very simple way of determining if an image is captured indoors or outdoors as generally an image with an exposure value of less than 10 is indoors or at night and an exposure value of 10 or more usually have to be an outside photograph taken during daytime. Outdoor classification can also be performed using just image contents, but the contents based approach is more computationally expensive and complex [12, 13].

If the exposure value is below 10 and it is during the morning of afternoon, it is a good chance that the photograph is taken indoors. Whether images taken at night are outdoors or indoors is harder to determine.

Given the capability to determine if images are taken outdoor or indoors one can create summarizing statistics as in the following example: “During first day of the Spring weekend 98 outdoor pictures and 25 indoor pictures were taken before noon.”. Alternatively, a vaguer summary can be provided, namely “During the first day of the spring weekend most pictures were taken outdoors before noon.”

Next, the exposure value can also be used to give a rough estimate of the weather. If the exposure value is high, that is above 12 then it is likely that it is sunny and if it is 12 or below it is likely that it is cloudy. However, these factors depends on the time of day, as it is brighter in the middle of
the day than in the morning or in the afternoon, and whether the scene is in the shadows or not.

2) Geographical location

It has been shown in [11] that it is possible to derive the approximate geographical coordinates for a set of images based on the exposure values for the images in that event. In simple terms this is done by fitting a sinusoidal curve with a 24 hour period on top of 24 hour windows of image data such that one is able to estimate the approximate midday and sunrise and sunset times (interested readers are referred to [11] for a detailed description of the approach). Given these estimates it is possible to compute the approximate latitude and longitude for the event. Experiments showed that a latitudinal accuracy of approximately 30 degrees and a longitudinal accuracy of approximately 15 can be achieved. This is at least enough to give a linguistic estimate of the continent where the images were taken, and in most cases it is also capable of determining the correct hemisphere. Research is underway to significantly increase this accuracy down to a couple of degrees [8]. If one is so fortunate to have geo-coded images then this information can be read directly from the EXIF information with high accuracy.

Given such estimates it is possible to provide descriptions as provided in the following example: “The spring weekend of 2010 was probably photographed in North America (34 degrees North, 116 degrees West)”. This information makes the report much more interesting to the user.

C. Focal length

Many digital cameras come equipped with zoom lenses allowing the focal length to be adjusted. This information also usually encoded into the EXIF headers. Typically, we use a small focal length, that is, a wider angle lens, to capture large objects in the view or if the objects are very close. Long focal lengths represent zoom magnification and are usually used to capture very small objects, or objects far away. Unfortunately, few cameras store the autofocus distance measurements as also these would be immensely useful.

Focal length information can also be summarized in a useful manner. For example, “During the first day of the Summer week 34 of the images had a long focal length (55mm or more) and 22 images had a short focal length (7.5 mm or smaller)” or in a more abstract form “During the first day of the summer week most of the images were taken with zoom.”

D. Manual information

People are unlikely to tag individual images in their collections. However, many users tend to organize their images in folders with meaningful names “USA conf 2010”. These manually added textual descriptions can be added to the textual reports. For example, “During 2010 there were 3 events – one during the weekend of March 6-7 labeled ‘USA conf 2010’, one during the week July 7-10 labeled ‘Australia conf’ and the last during the weekend October 23-25 labeled ‘Mountain trip’.”

III. EXPERIMENTAL EVALUATIONS

The textual image collection browser was run on a subset of the author’s personal image collection including 3,046 unique images taken over a period of four years at various locations around the world. The textual image collection browser was implemented in java and run on a Dell personal computer with an AMD Athlon Dual core processor and 4 Gb RAM running Windows Vista Personal edition. Drew Noakes’ freely available (EXIF) metadata-extractor library (available at http://drewnoakes.com/) was used for extracting EXIF information from the images. It took about 3 minutes to generate the image collection report. The Java code was not optimized and significant speedups could be achieved with simple tuning.

Figure 1 shows a short extract of the output produced by the textual image collection browser.

This collection comprises 23 events.

The 6-day event entitled Asia-Taiwan-Tainan-iceer contains 615 images. It occurred during the winter week, starting February 28, 2005 in AUSTRALIA AND SOUTHEAST PACIFIC (77 degrees South, 154 degrees East)

On the first day (Monday) 120 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the second day (Tuesday) 80 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the third day (Wednesday) 88 images were taken. Most of the photographs were taken at night

On the fourth day (Thursday) 114 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the fifth day (Friday) 155 images were taken. A majority of these photographs were taken indoors during a mostly sunny day

On the sixth day (Saturday) 53 images were taken. A majority of these photographs were taken indoors during a mostly sunny day

The 9-day event entitled Americas-USA-SouthernIllonois contains 768 images. It occurred during the summer week, starting June 23, 2005 in NORTH AND SOUTH AMERICA (61 degrees North, 101 degrees West)

On the first day (Thursday) 37 images were taken. A majority of these photographs were taken indoors during a mostly sunny day

On the second day (Friday) 67 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the third day (Saturday) 91 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the fourth day (Sunday) 96 images were taken. A majority of these photographs were taken indoors during a mostly sunny day

On the fifth day (Monday) 60 images were taken. A majority of these photographs were taken indoors during a mostly cloudy day

On the sixth day (Tuesday) 91 images were taken.

Figure 1. An extract of automatically generated textual image collection descriptions.
The extract shows that the data is presented in a semi hierarchical manner. First a statement regarding the entire collection is provided. Next, an overview of each event is provided followed by details on a day-based level of detail.

This extract represents a relatively large collection of images and the listing is relatively verbose and repetitive. A more compact form could be used, but then the data would be more cryptic. The linguistic representations are easy to interpret. Moreover, it should be possible to present this data in a semantically structured way such that users with screen readers could jump directly to the desired items. It is also possible that one could have made the day descriptions optional and instead present an overall trend for the event regarding the time of the day the photographs are taken and whether this is outside or inside and the overall weather.

The strategy can also produce a description for individual photograph. General information is then collected from the event as a whole and added to the individual information about the image. The examples in Figure 2 illustrate this.

The images in Figure 2 were chosen arbitrarily and the results show that the descriptions are more or less correct. The local time estimate in the first image is slightly incorrect as it is claimed that this image is taken in the early afternoon, when it was actually taken a few hours before lunch. Moreover, the weather predictor tags this as a “cloudy outside motive” while this day had clear skies. The reason is that this image is dark as the motive depicts a street shadowed by tall buildings. The second and third images are quite correctly explained. However, the simple linguistic image classifier could with simple means have resolved that this is a motive taken in Africa by eliminating Europe due to the image occurring on the southern hemisphere.

Note that the three images are taken in three totally different time-zones and the example illustrates the
important of adjusting the presentation of time into local time as humans more naturally relate to local time.

A. Implications

Besides allowing non-visual insight into the contents of large image collections the strategy could also be used to help improve the accessibility of web pages. First, the strategy could be run on an image collection and each individual image could be labeled with the information extracted for its respective event and day and the information written back to each of the image files into the maker note of the EXIF header. Here a unique label such as for example “auto-alt” could be used to identify the automatic image tag. Web publishing tools could then automatically extract the automatically generated description and insert this into the alt-attribute of the image tags on the web pages. This would not be as good as a manual alt-text, but it would be more informative than no alternative image text at all.

IV. LIMITATIONS AND FUTURE WORK

The current strategy assumes that the camera has a correctly set clock where its relative position to UTC is known. Moreover, it is assumed that the camera stores exposure attributes as described herein. A study of a large set of current compact and digital SLR cameras revealed that they all store this information – this includes modern mobile phones.

One ultimate danger of automatic analysis is that there is a false positive rate, that is, the algorithms occasionally produce erroneous results, emphasizes insignificant details and overlook details that would be perceived to be important to a non-blind user. However, also humans err and occasionally miss important details, or misinterpret when visually inspecting visual objects.

The current approach is limited in terms of analysis and the information extracted. This is because only image-meta information was considered. Future work should also consider the image analysis and incorporate some of the existing image analysis research results within face recognition, object recognition, landmark recognition, etc, to provide more rich descriptions of the images. More work should also be put into extracting high level summaries from the low level image analysis results.

V. CONCLUSIONS

This study has presented a textual image browser. The browser summarizes the contents of image collections in the form of stories comprising chronologically ordered events suggesting, when the event to place, where the event to place and what was done. Although still very limited, the approach extracts useful information about highly visual image object into a textual domain accessible by blind users. The current approach does not actually look at image contents and are therefore computationally highly effective. The current approach also considers individual images and their relationship to their respective events. This could for instance be used to automatically populate alt attributes of web-page images. However, further work is needed to derive information about the contents of individual photographs.

Another important future research question is what information blind users find useful in context of image collections.

REFERENCES


I. ABSTRACT

The question of making games accessible for the colorblind becomes important when one considers that approximately 8 to 9% of the market is colorblind. The type of problems that colorblind gamers encounter while playing a video game are many and they can be found in different kinds of games, if not all. These problems can easily be solved without having to create an entire new game for them and by understanding how their condition works. The use of cognitive psychology, symbols, color theory and patternicity are useful ways of creating design solutions that can accommodate the needs of the colorblind gamer with the rest of the market. At the same time lessons should be learned from the very few game studios that are doing something to help. The following paper will expand on these topics and provide with basic solutions that can be implemented when possible.

II. INTRODUCTION

The multiple definitions of color blindness can lead to confusion and false statements of how we, the color blind, really see the world. Perhaps the most accurate definition for color blindness in the context of this research is: The genetic inability to distinguish differences in hue.

Color blindness is almost a male exclusive condition as it seems to affect 1 in 20 men for 1 in 200 women. Nevertheless, women are the undeniable carriers since the condition is passed on through a defective x chromosome. Consequently, it is extremely curious that the first ever reported case of colorblindness to be found in literature comes from a woman.

“A case of Dr. Truberville’s (1684) has been quoted and referred to as one of colorblindness. In a letter to the Royal Society, London Aug. 4 1684, he says: A maid two and twenty years old came to me from Banbury, who could see very well, but no color beside black and white. She had such scintillations by night (with the appearances of bulls, bears etc) as terrified her very much. She could see to read sometimes in the greatest darkness for almost a quarter of an hour” (Jeffries, 1879)

Though vague and inconclusive, Dr. Truberville’s report is the first ever recorded case of colorblindness which was followed by 3 others in 1777, 1779 and most important of all in 1794. This last report was made by the English Chemist John Dalton.

“His report excited so much attention, that his name became attached to this chromatic defect, which, by general consent, is now among nearly all nations called colorblindness, not only from a desire not to connect so distinguished a man’s name as Dalton with a physical defect, but also being nearer the truth, since only the red-blind are Daltonians.” (Jeffries, 1879)

So in reality how many people are affected by color blindness and its variations?

Tritanopia\(^1\) 0.008% of the world’s population.

Monochromacy\(^2\) 0.00001% of the world’s population.

Daltonism\(^3\) 8 to 10% of the world’s population.

This means that approximately 8 to 9% of the gaming market will be red and green colorblind and will be affected by design and lighting that does not take their condition into consideration. In effect, these players will have issues recognizing important

---

1 This is very rare and commonly known as Blue-Yellow color blindness. It is estimated that 0.008% of the world’s population are affected by this type. (Gegenfurtner, Sharpe-2001)

2 colorblindness, or as how the name suggests and monotone type of color perception, is very rare. Affecting 1 in 33,000 to 100,000 or 0.00001% (Gegenfurtner, Sharpe-2001)

3 Roughly 8 to 10% of the world’s population is affected by this type. Red and green color blindness is by far the most common condition. Also men are 100 times more often affected than women, making it a "men's world". (Gegenfurtner, Sharpe-2001)
elements in a game which are crucial to beat or simply enjoy the experience. Therefore a comprehensive approach to game design that helps convey the information of the game properly can help improve the experience of a game.

“If everyday design were ruled by aesthetics, life might be more pleasing on the eye but less comfortable; if ruled by usability, it might be more comfortable but uglier.” (Norman, 1988)

It would be fair to claim that color blindness is not something that is constantly on game designers and artists’ mind when approaching a new project. A bolder statement could argue that video game designers and artists do not take into account usability and accessibility in their designs. Designers in all fields are constantly encouraged to create products that win aesthetics awards and that neglect the needs of the user and how the product will be interacted with. As designers, one becomes an expert in using the products one creates, knowing the state of the product, its boundaries, constraints etc. This, undoubtedly, can detach the designer from the targeted user’s perspective as the latter will not have all the aforementioned knowledge thus have problems interacting with the product if not designed properly.

“In 1937 architect Frank Lloyd Wright built a house for industrialist Hibbard Johnson. One rainy evening Johnson was entertaining distinguished guests for dinner when the roof began to leak. The water seeped through directly above Johnson himself, dripping steadily onto his bald head. Irate, he called Wright in Phoenix, Arizona. 'Frank,' he said, 'you built this beautiful house for me and we enjoy it very much. But I have told you the roof leaks, and right now I am with some friends and distinguished guests and it is leaking right on top of my head.' Wright’s reply was heard by all of the guests. ‘Well, Hib, why don’t you move your chair?’” (Ochshorn, 1999)

Just as in the Fallingwater example, video game players should not have to adapt to the products they buy in order to enjoy them.

III. AIM AND OBJECTIVES

The aim and objectives of this research are to point out existing efficient practices as well as provide additional solutions and guidelines that can help facilitate the accessibility and usability in games for color blinds whilst the non color blind would remain unaffected.

If one considers the first commercial game ever made (Computer Space, by Netting Associates 1971), one could argue the game industry is still at its dawn. It is also not a perfect science but rather composed of many sciences such as; mathematics, architecture, sociology, psychology and many more. Therefore solutions will be extracted from the fields of cognitive psychology, patternicity, game design, lighting principles and color theory.

Since this is an issue that affects the user, the gaming public’s opinion will be of extreme importance. Consequently, questionnaires about the topic will be created to be distributed in gamer networks (forums) to gather crucial information.

With these solutions the author seeks to encourage other designers and artists to think about the color blind user when creating their content in order for “all” players to be subjected to assimilating large amounts of information very quickly. In many cases any person with very good eye sight can easily misinterpret elements in a game if they are not properly brought to their attention.

Sales could be affected by the improvement and implementation of these solutions since 8% of the market will benefit from a better experience. This will benefit studios and encourage designers to continue using this approach, making “everyone” a target group.

Finally one of the most important aims of this paper is to provide with proof that there is no need to spend extra resources or money to make games accessible for the colorblind. Good design and constant play testing will provide with the essential elements for accessible games without affecting the non colorblind market.

IV. COLORBLIND GAMING

Call of Duty MW2 is an FPS (First Person Shooter) where the player takes the role of two soldiers in fight against a terrorist group. The player is given

---

4 The tendency to find meaningful patterns in meaningless noise. (Shermer, 2008)
clear objectives in every level or mission, which needs to be reached by using a compass in the HUD (Head up Display) and the environment. During every mission the player is surrounded by complex architecture, nature, civilians, fellow soldiers and enemy soldiers.

**Figure 4.1**

This game is a great example of how much information is being given to the player in hardcore games and how much can go wrong if this information is not properly conveyed. Players need to be aware of their surroundings as well as who is around them at all times, so they can avoid shooting civilians or fellow soldiers.

This level of concentration becomes even more demanding when the game is played in its multiplayer mode. In this mode the action becomes a lot faster and the player has to be able to identify friendly soldiers from enemies to help the team win the match. This recognition of enemies is done by the difference in uniform color, name tag color and the use of the radar. These is a good start since the player is getting the information they need in more than one way, but the way the information is conveyed to the player is not necessarily the most suitable one. Nametag color is the most important piece of information to recognize friends from enemies. Friendly nametags will be of a desaturated green and enemy nametags of a desaturated red, making it very easy to confuse them and shooting the wrong character.

But not only complex games pose a problem to colorblind gamers. There are many kinds of circumstances that make it very hard or even impossible to play a game. The following images are from games that do a poor job at conveying the information needed for colorblind gamers, the images on the left show the normal color palette of the game and the images on the right a simulation of how a colorblind person sees the game.

The colorblind simulation has been achieved by using Adobe Photoshop’s Proof Setup mode.

**Figure 4.2**

**PUZZLE BOBBLE**

Puzzle Bobble is a game where the player needs to clear the bobbles on the screen. The players do this by shooting bobbles and creating groups of three or more of the same color to make them disappear. For a colorblind person the green and yellow bobbles are identical and the subtle graphics inside the bobbles are not enough to tell them apart.

**Figure 4.3**

**MASS EFFECT 2**

Mini games are known for being very fast and used to unlock special areas in a larger game. This is the case for the Target Code Segment mini game in Mass Effect 2. The game shows a sample code on the top of the screen and ask the player to recognize the same segment of code in the array of codes moving from the bottom to the top of the screen. The player can move through the array of codes to select the right one with the directional pad while avoiding the red codes. The colors used to recognize the elements in the mini game often lead to confusion and make the code recognition a harder and longer process.

**Figure 4.4**

---

5 Any game in which successful completion or demonstration of skill requires a significant investment of both time and effort on the part of the player. (www.urbandictionary.com)
A. LEARNED HELPLESSNESS

“Learned helplessness is a phenomenon that refers to the situation in which people experience failure at a task, often numerous times. As a result, they decide that that task cannot be done, at least not by them: they are helpless and they stop trying” (Norman, 1988)

Colorblind gamers often experience this learned helplessness phenomenon when they find no explanation to why they keep on failing at the implest tasks in a game. This will make gamers not recommend the game, stop playing this type of games or even worse stop playing games all together.

V. WHAT IS BEING DONE?

Colorblind gamers are asking for changes and hoping their prayers are heard. But in reality only a handful of studios are doing something about an issue that deserves more attention and can be solved by thinking ahead and listening to the player. Studios like PopCap Games and Valve are one of the few who are doing something to help colorblind gamers and perhaps PopCap is a pioneer at it. With extremely successful games like Bejeweled, Plants vs Zombies and Peggle, PopCap Games has gained an excellent reputation in the accessible and fun games category.

Bejeweled, for example, is a time based puzzle game where the player’s objective is to swap one gem with an adjacent gem and form either horizontal or vertical chain reactions. With every new chain reaction the player is rewarded with points and more time. In other simple versions of the game there is no time limit and the game is simply over when there are no more possible moves.

In a game such as Bejeweled, the player is constantly busy with a small amount of information that is helping make the decisions and beat the game. The information comes in the form of color, shape, position of objects and time. Besides this, the player is provided with a Hint button which gives extra information by pointing at the gem which has a possible move available.

Combining the shape and color information, the designers of Bejeweled make sure that all players (especially color blinds) have more than one way of recognizing the objects.

On the other hand if the shape information is removed from the game, color blinds would most certainly have major issues playing the game. Next to this Non color blinds would be slower at playing the game since they would have to focus more to beat the game, having less information to their disposal.

Valve has implemented a colorblind mode to some of their games but unfortunately this is only available in the PC versions. This can be a problem since colorblind gamers need these options in all versions of the game. Nevertheless the option is there and it does significant improvements to many aspects of the game which makes it a lot more accessible.

The images above show the differences and contrast between playing Left 4 Dead 2 in normal and colorblind mode. The image on the left shows the game being played on normal mode and the image on the right being played with colorblind mode. Notice that the major improvement is the fact the enemies now have a white outline which makes them easy to recognize from friendly characters. Next to this, other significant improvements have been done to the UI. The first and most important being the crosshair which is now white, also the health and ammo bar have a different color which gives them higher contrast.

Another great example that should be followed and learned from is Naughty Dog’s Uncharted 2, starting by their rich and playful color palette, which almost mocks the average shooter with their grey and unwelcoming colors. Their multiplayer mode solves the character recognition issue with a very simple
but smart solution. Just as in Call of Duty Modern Warfare 2, the most important piece of information to recognize allies from foes is the name tag color, friends are green and enemies are red. But just as the general color palette of the complete game, these two colors are far apart from each other, one being a saturated and vibrant yellowish green and the other a strong saturated red.

Besides the color information the game helps you recognize characters by keeping friendly nametags (green) always visible and no name tags on enemies (unless they are being shot at). This lets the player know where his allies are at all time and an intuitive way of knowing that a character with no name tag on is not friendly and can be shot at.

VI. SOLUTIONS

The solutions suggested in the present document will not be applicable to all situations. There will be multiple cases in which the design of a game will simply not allow for some solutions to be acceptable.

A. SPATIAL MAPPING

Cognitive physiology gives us a set of rules that one can follow on how the human brain works and understands things. One can exploit these rules and ideas to his or her advantage to facilitate the experience and interaction with the game.

“Mapping: Technical term meaning the relationship between two things” (Norman, 1988)

Spatial mapping is one of these ideas that help the player understand how something works and how it needs to be interacted with. The spatial mapping example can easily be seen in the God of War games.

When in detective mode, important elements are clearly visible and have a strong and warm color which stands out from the environment immediately. Enemies can easily be recognized because they are the only red objects on the screen and a small user interface gives you extra information which is extremely useful when planning your next move.
palette is used. This transparent image suggesting which button to press can be easily missed or misunderstood.

God of War 3 by Santa Monica Studios, 2010 Figure 6.2

On the other hand, when God of War III was released, the quick time event mechanics were significantly improved. The first major improvement was changing the button information to a solid color with a glow around it so it cannot be blended with the background. And perhaps most importantly the use of spatial mapping is being implemented by displaying the button that needs to be pressed on the side of the screen that represents position of the button on the controller.

With these mechanics, spatial mapping is being exploited to our advantage helping the player focus on other more important things in the game and use the knowledge acquired from the game and knowledge they have in their heads to complete the quick time events.

Spatial mapping can also be used in other stages of the game to facilitate interaction. For example, menus and user interfaces that do not use spatial mapping can be frustrating and confusing, leading the player to make the wrong choices by accident.

Bioshock 2 by 2K Games, 2010 Figure 6.3

One really good example of this can be found in Bioschock’s 2 hacking mini game. The button B (Red) will abort the mini game, X (Blue) will buy it out if you have the possibility to do so and A (Green) will hack it. The positions of the buttons are based on the design of the mini game and not on the peripheral used to input the actions. This can lead to making the wrong choice really quickly because gamers will use the knowledge they have on their heads about where the buttons are placed on the controller. Therefore the player can easily make the mistake of hacking when they want to abort, or even worse, to abort when they want to hack.

Uncharted 2 by Naughty Dog, 2009 Figure 6.4

On the other hand, Uncharted’s 2 user interface design bases the positioning of their buttons on the position they have on controller, therefore making an action as simple as accepting or denying a multiplayer invitation very intuitive. This same treatment has been giving to many more aspects of the game and therefore making it more accessible and fun.

B. COZTUMIZATION

A really good way of making information to play the game accessible is by giving the player the freedom to map their own controls. Almost all of the EA sports franchises have learned this valuable lesson by listening to the players and giving them the responsibility of choosing where they want their actions mapped. Players can also choose from 2 standard mappings that have been derived out of pure standardization. With this freedom the player can create their unique knowledge on how to play a game that gives very little opportunity to use spatial mapping.

C. FEEDBACK

Another important form of freedom that can help make the game more accessible is the ability to modify the display settings. This way the player can easily change the contrast, saturation, brightness and gamma of the whole game. This can only be done properly when feedback is given while interacting with these settings, otherwise the player will not know the effect it’s having on the display.
D. PATTERNICITY

“Patternicity: The tendency to find meaningful patterns in meaningless noise” (Shermer, 2008)

According to Michael Shermer’s book *How We Believe* (2008), our brains are belief engines and pattern recognition machines that connect the dots and create meaning and patterns out of what we see in nature and all around us.

These patterns that our brains are so good in recognizing are sometimes harder to identify and need a little help.

If we outline the pattern on the image on the left it is easily recognized and therefore we can see it even without the help of the outline. This pattern recognition ability can be exploited to our advantage especially in the shooter genre by helping the player find meaningful patterns in a chaotic situation and therefore conveying only the right information when the player needs it.

Halo ODST is an FPS (First Person Shooter) where the player assumes the role of an Orbital Drop Shock Trooper or ODST. As part of the gameplay the player’s heads up display contains a VISR mode which outlines the enemies in red, allies in green and important objects in yellow. Though a poor choice of colors for the outlines, it does help with pattern recognition making objects and enemies pop out in the dark and monochromatic environment.

This solution is applicable for this specific design or kind of game. The futuristic settings allows for VISR mode to be accepted and believed by the player and therefore not breaking the suspense of disbelief.

Some other games will not be flexible enough to allow this to happen and other solutions should be applied to work around the issues.

E. COLOR CODING

The use of colors to identify elements in any product can work when used properly, the same can be said about video games and their color palette. If the important elements don’t stand out from the rest of the scene or product, any user (especially colorblind users) will have problems interacting with it.

When analyzing the simple calculator design one can easily notice that one of the most important and useful keys stands out from the rest (the equal sign). At the same time other keys (functions) in the calculator have a unique color coding which can be easily identified from the frame and numbers. This helps the user to interact with the product in a more intuitive manner.

A very useful way of coming up with an appropriate color palette is by using basic color theory principles. For example, the use of complementary colors is extremely useful to create both harmony and contrast between colors.
Good use of color harmony can be found in Rembrandt’s work. He creates a general atmosphere with a range of analogous colors which represent the burning of Jerusalem and uses a complementary color to make Jeremiah’s tunic stand out.

In the case of complementary colors, red and green can pose a problem but only if used incorrectly. Certain shades of red and green can be too close to each other and cause confusion, while rich and saturated versions of the same colors will stand out.

By using Photoshop’s Proof Setup tool anyone can test different shades of red and green and simulate how a red and green colorblind person would see them. This way it is possible to use red and green for color coding when the design of the game does not allow for any other possibilities.

The use of color to communicate information to the player should always be complemented with one or more ways of recognize the item. If the design of the game simply does not allow for implementation of symbols, spatial mapping, natural mapping, shapes and or patterns, then enough time should be invested in finding a color palette which can accommodate every user. Once again, this can be achieved by using simple color theories and testing out the color palette in Photoshop’s Proof Setup tool.

---

6 Analogous color schemes use colors that are next to each other on the color wheel. They usually match well and create serene and comfortable designs. (www.tigercolor.com)
VIII. REFERENCE

E
ESA- Essential facts about the computer and video game industry
   F - 2010

G
Gegenfurtner, Sharpe - Color Vision from Genes to Perception -
Cambridge University press – 2001

N
Norman, D-The Design of Every Day Things- Basic Books- 1988

O
Ochsorn, J- Des-integrating technology and design -

S
Shermer, M - Patternicity- Scientific American Magazine -

U
Urban Dictionary- Definition of hard core gamer and hard core game-