Proceedings of the 2nd European conference on disability, virtual reality and associated technologies (ECDVRAT 1998)

The 2nd European Conference on Disability, Virtual Reality and Associated Technologies

Proceedings

Edited by:

Paul Sharkey (Programme Chair)
David Rose (Conference Co-Chair)
Jan-Ingvar Lindström (Conference Co-Chair)

10, 11 of September, 1998

Mount Billingen, Skövde, Sweden

ECDVRAT ‘98

Proceedings:

The papers appearing in this book comprise the proceedings of the Second European Conference on Disability, Virtual Reality and Associated Technologies, held on the 10^{th} and 11^{th} of September, 1998 at the Hotel Billingehus Conference Centre, Mount Billingen, Skövde, Sweden. The papers presented reflect the authors’ opinions and are published as presented and without change (formatting excepted). Their inclusion in this publication does not necessarily constitute endorsement by the editors, by ECDVRAT, or by the University of Reading.

Please use the following format to cite material from these Proceedings:

Author(s), “Title of paper”, in Proceedings of the 2nd European Conference on Disability, Virtual Reality and Associated Technologies, Sharkey, Rose & Lindström (Eds.), page numbers, Skövde, Sweden, 10-11 Sept. 1998.

Proceedings reference number: ISBN 0 7049 1141 8

Published by

The University of Reading

For information, contact:
ECDVRAT, Department of Cybernetics, University of Reading,
Whiteknights, Reading, RG6 6AY, UK

Phone: +44 (0) 118 931 6704 Fax: +44 (0) 118 931 8220 Email: *ecdvrat@cyber.rdg.ac.uk*

WWW: *www.cyber.rdg.ac.uk/P.Sharkey/WWW/ecdvrat/*

Printed in the UK.

Contents

Abstracts

Session I Virtual Life Skills

Session Chair: David Rose

Session II Interfacing to Virtual Environments

Session Chair: Paul Sharkey

Session III Training, Assessment & Rehabilitation

Session Chair: Albert Rizzo

Session IV Virtual Environments & Visual Impairment

Session Chair: Peter Jones

Session V Sign Language & Gesture Recognition

Session Chair: Rachel McCrindle

Session VI Enhancing Mobility

Session Chair: Francisco García-Ugalde

Session VII Spatial Orientation & Navigation

Session Chair: Jan-Ingvar Lindström

Additional Papers

Conference Committee

Conference Co-Chairs

Dr. Jan-Ingvar Lindström, Telia, Hanige, Sweden
Prof. David Rose, Dept. of Psychology, University of East London, UK

Programme Chair

Dr. Paul Sharkey, Department of Cybernetics, University of Reading, UK

International Programme Committee

Dr. David Brown, VIRART, Dept. Manufacturing Eng., University of Nottingham, UK.

Dr. Sylvie Collins, Dept. Psychology, University of Leeds, UK.

Mr. Martyn Cooper, Open University, UK.

Prof. Paolo Dario, Scuola Superiore D. Anna, Pisa, Italy.

Dr. Alistair Edwards, Dept. Computer Science, University of York, UK.

Prof. Toshio Fukuda, Dept. Mechano-Informatics & Systems, Nagoya University, Japan.

Prof. Ephraim Glinert, Computer Science Dept., Rensselaer Polytechnic Institute, Troy, NY, USA.

Prof. John Gray, Dept. Electronic & Electrical Engineering, University of Salford, UK

Dr. William Harwin, Dept. Cybernetics, University of Reading, UK.

Prof. Jim Hewit, Dept. Appl. Physics and Electronic & Manuf. Eng., University of Dundee, UK.

Dr. Javier Ibañez-Guzmán, ESA IGELEC, Saint-Nazaire, France.

Mr. Magnus Magnusson, University Karlstad, Sweden

Dr. Harry Murphy, Centre Disabilities, California State University, Northridge, USA.

Dr. Stuart Neilson, Centre Health & Disability Studies, Brunel University, London, UK.

Prof. Alan Newell, Applied Computer Studies Division, University of Dundee, UK.

Dr. Helen Petrie, Director, Sensory Disabilities Research Unit, University of Hertfordshire, UK.

Dr. Penny Probert, Dept. Engineering Science, University of Oxford, UK

Mr. Chris Ramsden, Clinical Engineering & Design Systems, Royal Hospital for Neuro-Disability, London, UK.

Dr. David Roberts, Dept. Cybernetics, University of Reading, UK.

Mr. Simon Rushton, Dept. Psychology, University of Edinburgh, UK.

Dr. John Wann, Dept. Psychology, University of Reading, UK.

Prof. Kevin Warwick, Dept. Cybernetics, University of Reading, UK.

Dr. Paul Wilson, Virtual Reality Research Group, Dept. Psychology, University of Leicester, UK.

Dr. Ali Zalzala, Dept. Automatic Control and Systems Engineering, University of Sheffield, UK.

Organisation Committee

ECDVRAT is very grateful to the Organising Committee for Mechatronics '98 through which all local arrangements were organised.

Introduction

The purpose of the 2nd European Conference on Disability, Virtual Reality and Associated Technologies (ECDVRAT '98) is to provide a forum for international experts, researchers and user groups to present and review how advances in the general area of Virtual Reality can be used to assist people with Disability. The initial Call for Papers generated considerable interest, with high-quality contributions from researchers in several countries, many from beyond Europe.

The International Programme Committee have selected 35 papers for presentation at the conference, collected into 6 plenary sessions, *[to be completed when final papers are in]*. The conference will be held over two and will run in parallel with the Mechatronics '98 Conference.

ECDVRAT '98 follows on from the success and of the first conference, held in Maidenhead, UK in 1996. Abstracts from both ECDVRAT '96 and ECDVRAT '98 are available online from the conference web site *www.cyber.reading.ac.uk/P.Sharkey/WWW/ecdvrat/* as are most papers from 1996 and all papers from 1998.

The Opening Address to the Conference will be given by Conference Co-Chair Prof. David Rose, of the Department of Psychology, University of East London. The conference will close with an address by the second Conference Co-Chair, Dr. Techn. Jan-Ingvar Lindström, of Telia Telecom, Sweden.

International Journal of Virtual Reality

The Proceedings of ECDVRAT '98 will be published in full on CDROM issued with two forthcoming Special Issues of the International Journal of Virtual Reality. In addition the two Special Issues will publish as full papers eight of the key papers from the conference – four of which will present research which concentrates on mainly technological aspects, and four of which will concentrate on psychological issues.

Acknowledgements

We would like to thank the Programme Committee, for their input to the conference format and focus, and to their commitment to the review process, the authors of all the papers submitted to the conference, the Organisation Committee for Mechatronics '98, especially Jeanette Karls and the students who have helped out over the three days of the conference.

On behalf of ECDVRAT '98 we welcome all delegates to the Conference and sincerely hope that delegates find the conference to be of great interest.

It is hoped that the ECDVRAT conference series becomes an established bi-annual event. To that end, ECDVRAT welcomes any feedback on this year's conference and in particular the collaboration with Mechatronics '98. We would also welcome suggestions for the venue/host country for ECDVRAT 2000.

Jan-Ingvar Lindström

David Rose

Paul Sharkey

Conference Chairs

Jan-Ingvar Lindström is the Corporate Area Manager for Telematics & Disability at Telia AB (Sweden) concerning activities on behalf of people with disabilities and elderly people. He was awarded the degrees of M.E. (1964) and Dr. Techn. (1967) from the Chalmers University of Technology (CTH), in Gothenburg, Sweden. Dr. Lindström has acted on the Programme Committee for a number of international conferences in the general area of assistive technology, and as a Board Member for a number of Institutes, among others the R&D Council of the Swedish Handicap Institute and the Swedish Dyslexia Foundation. He is a National Representative of ICTA within Rehabilitation International, an Expert on Technical Aids for the World Federation of the Deaf and has acted as an Advisor for the European funding programme TIDE (Technology for the Integration of Disabled and Elderly people in Europe).

David Rose is a Professor of Psychology and Head of the Department of Psychology at the University of East London (England). He completed his B.Sc. and Ph.D. at the University of London and worked for many years in the Department of Psychology at Goldsmiths College, London. Initially, Professor Rose's research was concerned with mechanisms of recovery of function following brain damage and, in particular, on environmental influences on the recovery process. He has published extensively on this subject. In recent years he has broadened his research interests to encompass Virtual Reality applications to brain damage assessment and rehabilitation.

Paul Sharkey is a Reader of Interactive Systems and Chair of the Interactive Systems Research Group of the Department of Cybernetics at the University of Reading (England). He was awarded the degrees of B.Sc.(Eng.) and M.A. from the University of Dublin, Trinity College (Ireland), and Ph.D. in 1988 from the University of Strathclyde (Scotland). After five years at the Robotics Research Group at the University of Oxford he joined the University of Reading in 1993 as a lecturer. Dr. Sharkey's main research interests lies in robotics, particularly the design and control of robotic mechanisms, teleoperation & telepresence, and more recently in the area of distributed virtual environments. Dr. Sharkey acted as Conference Chair and Programme Chair for ECDVRAT '96.

Artwork

Artwork and Conference Layout by sckelp, adapted from original ideas of Eric Phipps and David Angus (ECDVRAT Logo) and from original clipart from the CorelDraw Clipart Library (ECDVRAT '98 Graphic). Eric Phipps and David Angus (both of Project DISCOVER) and sckelp may be contacted through ECDVRAT.

Conference Sponsors

The principal sponsors of ECDVRAT '96 are the Institution of Electrical Engineers, the International Journal of Virtual Reality, and the Department of Cybernetics, of the University of Reading.

Additional help in publicising the conference has been gratefully received from Mechatronics '98 International Conference, the Mechatronics Forum of the Institution of Mechanical Engineers and the Institution of Electrical Engineers, uk-vrsig@mailbase.ac.uk, vrpsych-l@usc.edu, and Ability Magazine, amongst many others.

The general organisation for ECDVRAT '98 has been provided by the Mechatronics '98 Conference Organising Committee from the University of Skövde.

Abstracts

Virtual reality as a communication aid for persons with aphasia, E Ahlsén & V Geroimenko, Göteborg University, SWEDEN

The paper presents a prototype using virtual reality as the basis for a picture communication aid for persons with aphasia (acquired language disorder after a focal brain injury). Many persons with aphasia have severe word finding problems and can produce no speech or very little speech. This problem is often connected to problems of abstraction, which makes it problematic to use picture communication based on classifying pictures by semantically superordinate categories and searching for them via these superordinate categories. The use of virtual reality makes it possible to use purely visuo-spatial orientation as a search strategy in a picture database. In the Virtual Communicator for Aphasics, we are exploring these possibilities. By "moving around" in a virtual environment, based on video-filmed panoramas, the user can search for pictures essentially in the same way as he/she would search for objects in the real world and communicate by pointing to the pictures. Pointing for communication can be used directly in the panoramas, in overview pictures accessed by hot-spots, and in arrays of pictures of objects also accessed by pointing to hotspots in panoramas or overview pictures. Speech output is possible. Some of the potential advantages and limitations of using virtual reality based on photorealistic panoramas and pictures in this type of application are discussed.

Virtual reality in vestibular diagnosis and rehabilitation, D Alpini, L Pugnetti, L Mendozzi, E Barbieri, B Monti & A Cesarani, Scientific Institute S. Maria Nascente, Fondazione don Gnocchi, Milan/University of Sassari, ITALY

While vestibulo-oculomotor and vestibulo-spinal functions are usually investigated by means of electronystagmography and stabilometry, environmental exploration and navigation cannot be easily studied in the laboratory. We propose that virtual reality (VR) can provide a solution, especially for those interested in the assessment of vestibular influence over spatial cognitive activities. Subjects exposed to immersive VR show rotatory behaviors during exploration that are the result of both a lateralized vestibular dominance and of the interplay with ongoing cognitive activity. The effect of vestibular dominance over exploratory behavior disappears in non-immersive VR conditions, but certain patterns of exploratory movements still seem to be associated to cognitive performance. On these grounds, we propose the use of VR to improve current techniques of vestibular rehabilitation based on visual feedback. We describe a new equipment that combines the functions of a digital stepping analyzer with those of a PC VR workstation. The patient controls the navigation of virtual environments by means of appropriate displacements of the center of gravity. The combination of a closed-loop feedback to control displacements of the center of gravity and active exploration of the environments makes an otherwise static exercise contingent on a viridical representation of spatial navigation.

Anthropometric models of children, M T Beitler & R A Foulds, University of Pennsylvania/duPont Hospital for Children, Wilmington USA

The objective of the work presented in this paper is to create a complete database of Anthropometric Models of Children, which can not only describe the anthropometric attributes of the individuals, but the functional abilities and growth behaviors as well. This work also includes a prototype system, which is being used during an assessment session to automatically gather the anthropometric and functional information of the subject and incorporated it into the population data of the model database.

Virtual reality for blind users, R Berka & P Slavík, Czech Technical University Prague, CZECH REPUBLIC

The paper describes a solution to the problem how blind users can work with the three dimensional information available in the computer environment. This serious problem gains more and more importance as the use of three dimensional information is permanently growing. An experimental system that allows such communication has been designed and implemented. This system can be used as a sort of kernel for applications of various kinds that deal with the problem of communication between a blind user and a three dimensional model.

The development of the Virtual City: A user centred approach, D J Brown, S J Kerr & V Bayon, Nottingham Trent University/University of Nottingham, UK

This paper will develop the theme of the importance of community based involvement in the development of virtual learning environments (VLEs) for people with a learning disability. It is being presented alongside two other papers, one by the User Group, the other by the Testing Group, describing the design, testing and distribution of the Virtual City. This set of VLEs comprise a computer aided learning (CAL) tool to teach independent living skills to people with a learning disability. Our presentation will demonstrate the involvement of users in each of the stages of development of the Virtual City, and the benefits of this partnership, as opposed to a more tokenistic involvement.

Along side the development of this methodology, the presentation will concentrate on the demonstration of the Virtual City to show how users can learn skills such as the use of public transport, road safety skills, safety within a home environment, the use of public facilities within a cafe and the development of shopping skills within a large supermarket. Video will also be shown demonstrating users involving themselves in the various stages of production of this CAL tool.

Evaluation of virtual learning environments, S V G Cobb, H R Neale & H Reynolds, University of Nottingham/Metropolitan Housing Trust Nottingham, UK

The Virtual Life Skills project describes a user-centred design approach to building virtual environments intended to provide a practice arena for skill learning in children and adults with learning disabilities. In the first year of the project four modules of a Virtual City have been developed: a house, a supermarket, a café and a transport system (see Brown et al, this issue for a description of the project). Evaluation of the project has been concerned as much with the design of the virtual learning environments (VLEs) and issues of usability and access as with monitoring skill learning and transfer to the real world. Two approaches were taken to the evaluation the four virtual learning environments. For three of the VLEs, Supermarket, Café and Transport, a test-retest experimental design method was used. This compared user performance in real world tasks with the same tasks presented in the VLE. Expert assessment was used to evaluate the Virtual House, looking at usability and appropriateness of the learning scenarios. It was found that VLEs can provide interesting, motivating learning environments, which are accessible to users with special needs. However, individuals differed in the amount of support required to use the input devices and achieve task objectives in the VLE. Expert and user review methods indicate that the VLEs are seen to be representative of real world tasks and that users are able to learn some basic skills. However, it would be unrealistic to expect transfer of skill over a short time period of learning as used in this project. Further testing is needed to establish the longitudinal learning effects and to develop more reliable techniques to allow users to express their own opinions by themselves.

The use of haptic virtual environments by blind people, C Colwell, H Petrie, D Kornbrot, A Hardwick & S Furner, University of Hertfordshire/British Telecommunications plc Ipswich, UK

This paper describes two studies concerning the use of haptic virtual environments for blind people. The studies investigated the perception of virtual textures and the perception of the size and angles of virtual objects. In both studies, differences in perception by blind and sighted people were also explored. The results have implications for the future design of VEs in that it cannot be assumed that virtual textures and objects will feel to users, whether blind or sighted, as the designer intends.

Ambisonic sound in virtual environments and applications for blind people, M Cooper & M E Taylor, Open University Milton Keynes, UK

To date there has been much more effort directed to generating credible presentations of virtual worlds in a visual medium than in reproducing corresponding or even self contained worlds with synthesised or recorded sound. There is thus today a disparity in the relative fidelity of these 2 modes in most VR systems. While much work has been done in hi-fidelity and 3 dimensional sound reproduction in psycho-acoustic research and applications for audiophiles this has rarely been taken onboard by the VR community.

This paper describes work ongoing to apply Ambisonic techniques to the generation of audio virtual worlds and environments. Firstly Ambisonics is briefly outlined then principles behind its implementation in this context described. The design of the implementations to date is described and the results of trials discussed. The strengths and limitations of the approach are discussed in the light of this practical experience.

There is a range of applications envisaged for this technique that would be particularly of benefit to disabled people. The 2 principal areas under consideration by the authors is the extended use of the audio mode in HCI for people with disabilities and VR applications for blind and partially sighted people.

Both of these areas are described and specific examples of applications given in each. The limitations given available low cost technology are highlighted and the technology evolution required to make such applications widespread commented on.

The results of the development work and subsequent user trials undertaken to date are given and discussed. Lines of further exploration of this technique and its application are outlined.

Design of a non-contact head-control mouse emulator for use by a quadriplegic operator, E D Coyle, M Farrell, R White & B Stewart, Dublin Institute of Technology, IRELAND

A non-invasive scanning mechanism has been designed which is microprocessor controlled to locate and follow head movement in a defined zone about a persons head. The resulting head movements may then be mapped onto a computer screen and can further be used to control the cursor, thus replacing the necessity of using a standard PC mouse. To demonstrate the concept, a Graphic User Interface (GUI) of a push button telephone has been designed, and a procedure is outlined below by which one may select and highlight a digit or group of digits, and in turn dial the chosen number.

A practical example using virtual reality in the assessment of brain injury, R C Davies, G Johansson, K Boschian, A Lindén, U Minör & B Sonesson, University of Lund/Lund University Hospital, Höör, SWEDEN

Virtual Reality (VR) as a complementary tool for medical practitioners in the assessment and rehabilitation of people who have suffered a brain injury is discussed. A pilot-study has been undertaken on a prototype VR assessment tool. The design involved nine occupational therapists with expertise in the care of traumatic brain-injured patients and one (computer experienced) patient. The aim was to begin a dialogue and to ascertain the potential of a VR system. A common method for occupational therapists to assess function and ability is to ask a patient to brew coffee. From the performance of such a task, an individual's "functional signature" can be determined. The prototype was built using Superscape®, a personal computer based VR system, to be close to the real coffee making task, including effects of making mistakes, realistic graphics and sound effects. The world was designed to be as easy to use and intuitive as possible, though problems of mental abstraction level, transfer of training and realistic interaction have yet to be resolved. The comments from the test participants have highlighted problem areas, given positive insight and pointed out other scenarios where VR may be of use in the rehabilitation of people with a traumatic brain injury.

Development and evaluation of a virtual reality based training system for disabled children, M Desbonnet, S L Cox & A Rahman, University of Limerick, IRELAND

Children need mobility for normal development. An electric wheelchair can provide mobility for severely disabled children, but this requires training, which may be difficult. This virtual reality based training system is aimed at solving these difficulties in a practical, cost-effective way. The project involved the construction of two virtual environments for training these children. This was achieved by developing a software solution using WorldToolKit and AutoCAD.

Neural-network virtual reality mobility aid for the severely visually impaired, M R Everingham, B T Thomas, T Troscianko & D Easty, University of Bristol, UK

This paper describes a new approach to image enhancement for people with severe visual impairments to enable mobility in an urban environment. A neural-network classifier is used to identify objects in a scene so that image content specifically important for mobility may be made more visible. Enhanced images are displayed to the user using a high saturation colour scheme where each type of object has a different colour, resulting in images which are highly visible and easy to interpret. The object classifier achieves a level of accuracy over 90%. Results from a pilot study conducted using people with a range of visual impairments are presented in which performance on a difficult mobility-related task was improved by over 100% using the system.

Generating virtual environments to allow increased access to the built environment, G T Foster, D E N Wenn & F O'Hart, University of Reading, UK/University of Dublin, Trinity College, IRELAND

This paper describes the generation of virtual models of the built environment based on control network infrastructures currently utilised in intelligent building applications for such things as lighting, heating and access control. The use of control network architectures facilitates the creation of distributed models that closely mirror both the physical and control properties of the environment. The model of the environment is kept local to the installation which allows the virtual representation of a large building to be decomposed into an interconnecting series of smaller models. This paper describes two methods of interacting with the virtual model, firstly a two dimensional representation that can be used as the basis of a portable navigational device. Secondly an augmented reality called DAMOCLES is described that overlays additional information over a users normal field of view. The provision of virtual environments offers new possibilities in the man-machine interface allows intuitive access to network based services and control functions to a user.

Segmentation and classification of hand gestures for man-machine communication, F García-Ugalde, D Gatica-Pérez & V García-Garduño, FI-UNAM, MÉXICO/University of Washington, Seattle, USA

Man-machine communication in a natural way, it means without cumbersome gloves, is still an open problem. Keeping in mind the need to develop some friendly tools for helping people with disabilities to use the computer as a support tool for training into reinforced methods for learning to read or any other application. In this work we have addressed the problem of communication with a computer using some recognition of very basic hand gestures. From an engineering point of view our system is based on a video camera which captures image sequences and in a first time a segmentation of hand gestures is developed in order to provide information for its posterior classification and recognition. For classifying the segmented fields named e of gestures, for instance *hand # 1* and *hand # 2,* see figures 5a and 6a, we have proceed first to obtain a binary version of these segmented fields comparing them with a threshold, so rendering the classification faster, then based on the Radon transform (Lim, 1990), a computation of the projected sum of the binary intensity of gestures has been done at directions 0^0 and 90^0, see figures 1 and 2. For reducing the number of data to be processed a wavelet decomposition of the projected sum of the binary intensity for each orientation (0^0 and 90^0) has been done using Daubechies filters: d4 (Daubechies, 1988). This projected and wavelet decomposed information has been used for classifying the gestures: training our system with our dictionary and computing the correlation coefficient between the wavelet coefficients corresponding to *trained sequences* and others captured and computed in continuous operation, the computer is able to recognize the very simple gestures. The region segmentation has been done using a dense motion vector field as the main information then each region is matched to a four-parameter motion model (Gatica-Pérez et al, 1997). Based on Markov Random Fields the segmentation model detects moving parts of the human body with different apparent displacement such as the hands (García-Ugalde et al, 1997). The motion vector field has been estimated by a Baaziz pel-recursive method (Baaziz, 1991) and considered together with others sources of information such as intensity contours, intensity values and non-compensated pixels as inputs of the Markov Random Field model. The maximum a posteriori criterion (MAP) is used for the optimization of the solution, and performed with a deterministic method: *iterated conditional modes* (ICM). The complete segmentation algorithm includes initializing, region numbering and labeling, parameter estimation of the motion model in each region, and optimization of the segmentation field. So our probabilistic approach takes into account the fact that an exact displacement field does not exist (errors usually occur at or around motion boundaries), and that better results can be attained if an indicator of the quality of the vector field is known, this indicator is obtained from the non-compensated pixels as well as the intensity contours (García-Ugalde et al, 1997).

Can a haptic force feedback display provide visually impaired people with useful information about texture roughness and 3D form of virtual objects?, G Jansson, Uppsala University, SWEDEN

The aim was to investigate the usefulness of a haptic force feedback device (the PHANToM) for information without visual guidance. Blind-folded sighted observers judged the roughness of real and virtual sandpapers to be closely the same. The 3D forms of virtual objects could be judged accurately and with short exploration times down to a size of 5 mm. It is concluded that the haptic device can present useful information without vision under the conditions of the experiments. The result can be expected to be similar when observers are severely visually impaired, but this will be controlled in a separate experiment.

Virtual keyboard with scanning and augmented by prediction, P E Jones, University of Western Australia, AUSTRALIA

All our teenage users are confined to electric wheelchairs and are unable to speak or make any voluntary movements much beyond either moving their head against one of three switches mounted in the chair's headrest or to hit a large "banger" switch. Real-world devices are beyond reach, only devices in a virtual world are attainable. This virtual keyboard project was designed to meet their needs for interacting with commercial off-the-shelf software such as word processors, spreadsheets, electronic mail and Internet tools. The virtual keyboard uses scanning augmented by character prediction and word completion.

Implementation and capabilities of a virtual interaction system, A W Joyce, III & A C Phalangas, Alfred I. duPont Hospital for Children/University of Delaware, USA

Virtual Interaction refers to a technique for interacting with computer generated graphics. Graphical objects are overlaid on live video of the user. A chromakey separates the user (foreground) from the background resulting in a silhouette of user. The computer causes the graphical objects to move in relation to the silhouette so that they appear to interact with the user. This paper presents the implementations of the system, some techniques for interaction and discusses using the system as a tool for physical therapy.

S-TEL: An avatar based sign language telecommunication system, T Kuroda, K Sato & K Chihara, Nara Institute of Science and Technology, JAPAN

Although modern telecommunication have changed our daily lives so drastically, the deaf cannot benefit from them based on phonetic media. This paper introduces a new telecommunication system for sign language utilizing VR technology, which enables natural sign conversation on analogue telephone line.

On this method, a person converses with his/her party's avatar instead of party's live video. As speaker's actions are transmitted as kinematic data, the transmitted data is ideally compressed without losing language and non-language information of spoken signs.

A prototype system, S-TEL, implementing this method on UDP/IP, proved the effectiveness of avatar-based communication for sign conversation via a real lossy channel.

Sign language formal description and synthesis, O Losson & J-M Vannobel, Université des Sciences et Technologies de Lille, FRANCE

Special needs of deaf people appeal henceforward to sign language synthesis. The system presented here is based on a hierarchical description of sign, trying to take the different grammatical processes into account. Stress is laid on hand configurations specification thanks to finger shapes primitives and hand global properties, and on location and orientation computation issues. We then expose the results achieved from the corresponding written form of signs, leading to their computer virtual animation.

3D aural interactive hyperstories for blind children, M Lumbreras & J Sánchez, University of Chile, Santiago, CHILE

Interactive stories are commonly used for learning and entertaining purposes enhancing the development of several perceptual and cognitive skills. These experiences are not very common among blind children because most computer games and electronics toys do not have appropriate interfaces to be accessible by them.

This study introduces the idea of interactive Hyperstories performed in a 3D acoustic virtual world. The hyperstory model enables us to build an application to help blind children to enrich their early world experiences through exploration of interactive virtual worlds by using 3D aural representations of the space. We have produced AudioDoom, interactive model-based software for blind children. The prototype was qualitative and quantitatively field-tested with several blind children in a Chilean school setting.

Our preliminary results indicate that when acoustic-based entertainment applications are careful applied with an appropriate methodology can stimulate diminished cognitive skills. We also found that spatial sound experiences can create spatial navigable structures in the mind of blind children. Methodology and usability evaluation procedures and results appeared to be critical to the efectiveness of interactive Hyperstories performed in a 3D acoustic virtual world.

Virtual professional networks between speech pathologists, M Magnusson, University College of Karlstad, SWEDEN

All speech pathologists in one Swedish county will develop methods for supervision, therapy and development of professional methods, using ISDN-based videotelephony and Internet. The project is the first one of its kind and has initiated follow-up projects. It is primarily based upon research developed at the University of Karlstad, the Department of Disability and Language.

Multimedia INterface for the Disabled (MIND) project, R J McCrindle & R M Adams, The University of Reading, UK

The Multimedia Interface for the Disabled (MIND) project is concerned with developing a set of guidelines and authoring tools, for use by multimedia developers, to enable them to augment their products to encompass the specific needs of sensory impaired users. This paper presents the ethos behind the project and describes the MIND software prototype developed. The MIND prototype maximises the effectiveness of multimedia information delivery, through the provision of an adaptable and easily navigated user interface, which incorporates access to the augmented multimedia information created with the authoring tools.

User group involvement in the development of a virtual city, L Meakin, L Wilkins, C Gent, S Brown, D Moreledge, C Gretton, M Carlisle, C McClean, J Scott, J Constance & A Mallett, The Shepherd School, Nottingham, UK

In April l997 the Shepherd School, Virtual Reality Applications Research Team (VIRART) and the Metropolitan Housing Trust joined together to develop a Virtual City, which would help people with learning difficulties learn independent living skills. The Virtual City would be made up of many parts and each would be a Virtual Learning Environment.

At the end of the first year the Virtual City includes a house, a cafe, a supermarket and a transport system. People with learning difficulties can use a computer in a game playing way to learn lifeskills in these four learning environments.

Often when computer programmes or learning schemes are devised for people with learning difficulties, no one asks them what they would like to be included. However, in the Virtual City from the very beginning, the views of people who might use the programmes guided the whole project. This paper will examine in detail the User Group involvement in the development of the Virtual City.

Robotic travel aid for the blind: HARUNOBU-6, H Mori & S Kotani, Yamanashi University, JAPAN

We have been developing Robotic Travel Aid(RoTA) "HARUNOBU" to guide the visually impaired in the sidewalk or campus. RoTA is a motor wheel chair equipped with vision system, sonar, differential GPS system, dead reckoning system and a portable GIS. We estimate the performance of RoTA in two viewpoints, the viewpoint of guidance and the viewpoint of safety. RoTA is superior to the guide dog in the navigation function, and is inferior to the guide dog in the mobility. It can show the route from the current location to the destination but cannot walk up and down stairs. RoTA is superior to the portable navigation system in the orientation, obstacle avoidance and physical support to keep balance of walking, but is inferior in portability.

Making 3D models: A challenge in man-machine communication, A Osorio, LIMSI-CNRS, FRANCE

The purpose of this publication is to present a graphic 3D modeler with interactive interaction capabilities operating in real time. It may be used on both in Unix workstations and PCs. The modeler can be used to make and edit 3D reconstruction of articulated objects, a facility that is particularly useful in processing medical images generated by a CT scanner or magnetic resonance units. The modeler takes account of the physical characteristics of the objects manipulated and, in developing it, we have assumed that the prospective user will have the necessary expertise to be able to interact with and interpret the model.

Improving the mobility of severely disabled, P Peussa, A Virtanen & T Johansson, VTT Automation, Tampere/Adaptation Training Centre for Disabled, Launeenkatu, FINLAND

A modular mobility aid system is presented, which can be attached to most commercial electric wheelchairs to increase the independence and safety of the wheelchair user. The system consists of falling avoidance, obstacle avoidance, and beaconless navigation functions. Experiences from an evaluation period are presented. Also an idea is presented to combine map information to environmental perceptions in order to provide the wheelchair user a better awareness of his/her whereabouts. The navigation system can also be applied to different mobile robot applications.

Developments of a collaborative research on VR applications for mental health, L Pugnetti, L Mendozzi, E Barbieri, A Motta, D Alpini, E A Attree, B M Brooks & F D Rose, Scientific Institute S. Maria Nascente, Don Gnocchi Foundation, Milan, ITALY/Casa di Cura “Policlinico di Monza”, Monza, ITALY/University of East London, UK

The collaboration between or two scientific institutions is giving significant contributions to VR research into several fields of clinical application. Concerning the important issue of side-effects, future studies will clarify whether the encouraging results obtained in the recent past on patients with neurological diseases can be confirmed, and whether specific recommendations for the use of immersive VR in selected clinical populations can be made. Recent collaborative studies on the application of non-immersive VR to improve clinical testing of spatial memory provided evidence of good replicability of results in both healthy and neurologically affected groups. The development of retraining applications for spatial memory impairments and future studies aimed at assessing the impact of ambulatory disability on spatial cognitive abilities will be based on these findings. Finally, a newly approved transnational project will lead our groups into the field of the assistive technology to improve working skills and opportunities for employment of subjects with mental disabilities who seek a job.

Preliminary Findings on a Virtual Environment Targeting Human Mental Rotation/Spatial Abilities, A Rizzo, J G Buckwalter, P Larson, A van Rooyen, K Kratz, U. Neumann, C Kesselman & M Thiebaux, University of Southern California, Los Angeles, CA/Fuller Graduate School of Psychology, Pasadena CA, USA

Virtual Reality technology offers the potential to create sophisticated new tools which could be applied in the areas of neuropsychological assessment and cognitive rehabilitation. If empirical studies demonstrate effectiveness, virtual environments (VE’s) could be of considerable benefit to persons with cognitive and functional impairments due to acquired brain injury, neurological disorders, and learning disabilities. Testing and training scenarios that would be difficult, if not impossible, to deliver using conventional neuropsychological methods are being developed which take advantage of the attributes of virtual environments. VE technology allows for the precise presentation and control of dynamic 3D stimulus environments, in which all behavioral responding can be recorded. A cognitive domain where the specific advantages found in a virtual environment are particularly well-suited, is with human visuospatial ability. Our paper outlines the application of a virtual environment for the study, assessment, and possible rehabilitation of a visuospatial ability referred to as mental rotation. The rationale for the Virtual Reality Spatial Rotation (VRSR) system is discussed, and the experimental design that is being used to collect data from a normal, aged 18 to 40 population is presented. Our research questions are then outlined and we discuss some preliminary observations on the data that has been collected thus far with the system.

Transfer of training from virtual to real environments, F D Rose, E A Attree, B M Brooks, D M Parslow, P R Penn & N Ambihaipahan, University of East London, UK

Training is one of the most rapidly expanding areas of application of the technology of Virtual Reality (VR) with virtual training being developed in industry, commerce, the military, medical and other areas of education and in a variety of types of rehabilitation. In all cases such training rests upon the assumption that what is learned in the virtual environment transfers to the equivalent real world task. Whilst there is much anecdotal evidence there have been few systematic empirical studies and those that have been carried out do not lead to clear conclusions. This paper reports preliminary findings from a study, using a simple sensorimotor task, which seeks to establish not only the extent of transfer, but also the reliability and robustness of whatever transfers. The findings demonstrate a clear positive transfer effect from virtual to real training and suggest that the cognitive strategy elements and cognitive loads of the two types of training are broadly equivalent. However, caution is advised in the interpretation of these findings. The results are discussed in the wider context of models of transfer of training.

Japanese sign-language recognition based on gesture primitives using acceleration sensors and datagloves, H Sawada, T Notsu & S Hashimoto, Waseda University, Tokyo, JAPAN

This paper proposes a Japanese sign-language recognition system using acceleration sensors, position sensors and datagloves, to understand human dynamic motions and finger geometry. The sensor integration method realized a robust gesture recognition comparing with a single sensor method. The sign-language recognition is done by referring to a Japanese sign-language database in which words are written as sequences of the gesture primitives. Two recognition algorithms which are the automata algorithm and the HMM are introduced and tested in the practical experiments.

The invisible keyboard in the air: An overview of the educational, therapeutic and creative applications of the EMS Soundbeam, T Swingler, Norwich, UK

Soundbeam is a ‘virtual’ musical instrument, an invisible, elastic keyboard in space which allows sound and music to be created without the need for physical contact with any equipment. The system utilises ultrasonic ranging technology coupled to a processor which converts distance and movement information into MIDI. Originally developed for dancers - giving them a redefined relationship with music - Soundbeam has proved to have dramatic significance in the field of disability and special education, because even those with profound levels of impairment are, even with the most minimal movements, able to compose and to instigate and shape interesting, exciting and beautiful sounds. Individuals who may be especially difficult to stimulate can benefit from what may for them be a first experience of initiation and control. A continuum of applications - ranging from the fundamentals of 'sound therapy' (posture, balance, cause-and-effect) through to more creative and experimental explorations in which disabled children and adults become the composers and performers of enthralling musical collaborations, and beyond to interactive installations and multimedia performance - can be described.

Soundbeam first appeared in prototype form in 1984 and was taken up in a serious way in special schools in the UK and subsequently in Scandinavia, the Netherlands and elsewhere following its launch in 1990. A totally redesigned version of the machine will be ready in the Autumn of 1998.

User group involvement in the development of a virtual city

Leigh Meakin, Lianna Wilkins, Charlotte Gent, Susan Brown, David Moreledge, Colin Gretton, Michael Carlisle, Charlie McClean, Jane Scott, Jane Constance and Angela Mallett (Facilitator)

The User Group for the Virtual City, c/o The Shepherd School
Harvey Road, Off Beechdale Road, Nottingham, NG8 3BB, ENGLAND

shepherdschool@campus.bt.com

ABSTRACT

In April l997 the Shepherd School, Virtual Reality Applications Research Team (VIRART) and the Metropolitan Housing Trust joined together to develop a Virtual City, which would help people with learning difficulties learn independent living skills. The Virtual City would be made up of many parts and each would be a Virtual Learning Environment.

At the end of the first year the Virtual City includes a house, a cafe, a supermarket and a transport system. People with learning difficulties can use a computer in a game playing way to learn lifeskills in these four learning environments.

Often when computer programmes or learning schemes are devised for people with learning difficulties, no one asks them what they would like to be included. However, in the Virtual City from the very beginning, the views of people who might use the programmes guided the whole project. This paper will examine in detail the User Group involvement in the development of the Virtual City.

1. ORGANISATION OF THE VIRTUAL CITY PROJECT

Overseeing the whole project was the **Executive Group.** This consisted of one representative from Metropolitan Housing, one from the Shepherd School, one from VIRART and one Business Consultant.

The Executive Group were also members of a **Steering Group**, together with other individuals representing organisations interested in helping to develop the Virtual City. The User Group and people who were building the Virtual City also sent representatives to each meeting of the Steering Group. The main role of the Steering Group was to advise the Shepherd School, VIRART and Metropolitan Housing as to the development of the Virtual City and to monitor the work that was carried out.

At the very centre of the project was the User Group. People in the User Group had learning difficulties and they were representative of people who might use the Virtual City to gain lifeskills. Their main role was to decide :

a) what they wanted in the Virtual City;

b) what they wanted to learn from the Virtual City;

c) how they wanted the Virtual City to be designed.

They were also involved in writing articles and publicity.

A group involved in building the virtual city attended meetings of the User Group and they were then responsible for ensuring that Virtual Learning Environments were built according to the User Groups wishes. Once the initial Virtual Learning Environments were built, in order to find out whether people with learning difficulties could learn lifeskills from them, a **Testing Group** of five people with learning difficulties was set up.

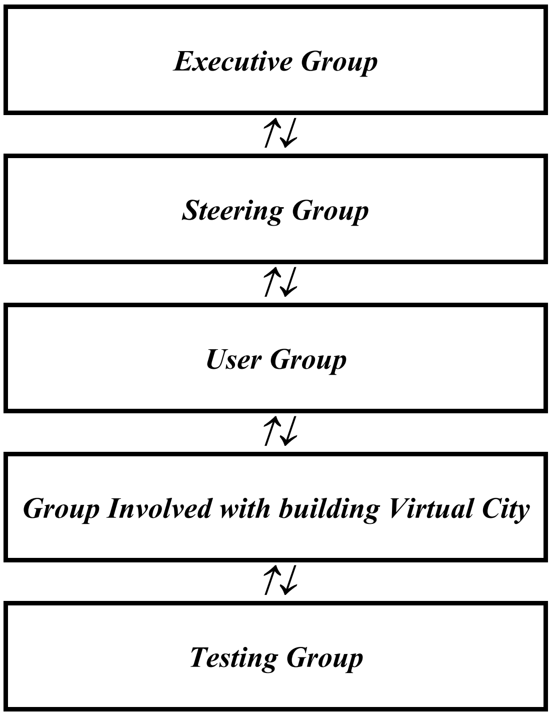


Figure 1. *Organisation of the Virtual City Project*

2. THE USER GROUP

Initially twelve volunteers with learning difficulties formed the User Group :

- One man lived with his grandmother and he did not attend college or any other placement.
- Two people lived at home and attended a course for people with learning difficulties at a College of Further Education.
- Three people lived at home and attended a unit in the Shepherd School for people aged sixteen to nineteen years with learning difficulties.
- Six people were tenants, who received support from Metropolitan Housing Trust projects, of these :
  One woman lived with her husband;
    a) Two men had part time jobs working in a restaurant;
    b) One man had an allotment that he tended daily;
    c) Two women attended college.

The group were aged between seventeen years and fifty five years and they felt that between them they had understanding and knowledge of a wide range of areas.

"I use a wheelchair"
"I don`t hear very well and I wear an hearing aid".
"I can`t see very well and I wear glasses"
"Sometimes people don`t understand what I`m saying"
"Reading and number are too hard for me"
"I`m from a mixed race family - my dad's from Jamaica".
"I live on my own with just a bit of support".

The User Group did not receive payment for their input to the project, but all transport costs were paid and substantial refreshments were provided.

3. SUPPORT FOR THE USER GROUP

Each member of the User Group made their own decision as to whether they needed individual support. At the beginning of the project three members each had a Support Worker, who accompanied them to meetings, helped with social and communication problems and gave emotional support. After attending three meetings one other man found it difficult to continue due to personal problems. However, after investigation, a Support Worker was provided and he has remained with the project until the present day.

4. FACILITATOR PROVISION

Before the User Group was formed a facilitator was employed in order to:

- Initially give information about the role of the User Group to possible members.
- Ensure meetings were accessible to the User group by ordering taxis, providing lifts and arranging suitable times and venues.
- Keep the User group informed about times, venues and agendas of meetings.
- Organise materials and refreshments for meetings.
- Enable the User Group to put forward their ideas and views and to make their own decisions.
- In agreement with the User Group take minutes of the meetings and make written reports.
- Accompany representatives of the User Group to Steering Group meetings and feedback information from User Group Meetings.

The User Group valued the help provided by the facilitator.

"I`m glad you do things for us and help us with what we want to say".

"You make things nice for us".

"You help us to get here and buy food for us".

"You're very chatty and help us to talk"

5. USER GROUP MEETING

The User Group met approximately once per month with additional meetings held whenever necessary. The first meeting was felt to be very important. The facilitator wanted everyone to feel comfortable with each other, to have an enjoyable time and to be motivated to attend further meetings. In addition, it was felt that the first meeting should help the User Group become aware of their importance in the project. A Cocktail Party was felt to be appropriate.

The Cocktail Party was very successful and everyone was keen to come to further meetings.

"I had a great time - it was really good"

"I liked meeting all the new friends"

"The food was alright - I liked it"

"I can't wait for the next meeting"

During the Cocktail Party a full explanation of the project was given and members of the User Group discussed the organisation of the meeting - the following decisions were made:

- Meetings would always be on a Monday between 5 and 7 p.m.

  "It's easier to remember if they're on the same day"

- Three different venues would be tried out for meetings and then the User Group would make an informed choice as which they preferred.
- The following timetable would take place for each meeting :

  5.00p.m. to 5.15p.m. - Coffee and Greetings.
  5.15p.m. to 6.15p.m. - Working on the Project
  6.15p.m. to 7.00p.m. - A Party Meal

  "It makes it better if you have a party".

  "I don`t mind working, but its boring if you talk for too long".

- Dates for meetings would be decided on an ongoing basis as people did not know what they would be doing too far in advance.

"I never know what I`m doing - I could be on my holidays or going out".

All future meetings did in fact follow these initial guidelines made at the Cocktail Party. At every meeting there was themed or party food varying from meat pie parties to birthday parties and these proved very successful. At the end of the first year only two people had stopped attending meetings. One of these now had a regular girlfriend and the other finished college late and felt too tired. All other members of the group were highly motivated to attend and their commitment to meetings was excellent.

"I should be going to a Youth Club - but I`d rather come here".

"I am tired as I go to College, but I don`t want to miss the meeting".

The Project Manager and people building the Virtual City also attended each meeting of the User Group.

6. USER GROUP INVOLVEMENT IN THE STEERING GROUP

Meetings were held monthly between l.00p.m. and 2.30p.m. Throughout the Project two people from the User Group were keen to be involved in the Steering Group, but one had difficulty speaking in front of a large group

"There are too many people and I get scared".

and the other man had a tendency to speak when other people were talking.

Three people, who never attended a steering group meeting said they did not like meetings, where people talked all the time. Three people attended two or three meetings, but had other things to do during the afternoon and were loathe to give up other activities. One man said he did not want to attend because the meetings were too boring and long. The facilitator spoke for the User Group at meetings of the Steering group with a little input from the representatives.

Towards the end of the development of the first phase of the Virtual City, it would seem that the User Group involvement in the Steering Group could be improved. This might be by altering the times and the style of the Steering Group and/or by improving the communication skills of the User Group representatives.

7. USER GROUP DECISION MAKING REGARDING CONTENT OF THE VIRTUAL CITY

The first decision made by the User Group was to choose all the different areas they wanted in a Virtual City. Their ideas were creative and thoughtful and forty six areas were chosen .

City	*Skiing*	*Bowling*	*Cinema*	*Pub*
Football	*Cricket*	*Post Office*	*Station*	*Snooker*
Police Station	*Fire Station*	*McDonalds*	*Hospital*	*Fairground*
Housing Trust(Metro)	*Disco*	*Driving*	*Jobs*	*Swimming*
Health & Beauty	*Cycle Track*	*Motorbike*	*Airport*	*College*
Pizza Hut	*River*	*Office*	*Leisure Centre*	*Roller Skating*
Workshops	*Bank*	*Shops*	*Supermarket*	*Dentist*
Doctor	*Hairdressers*	*Restaurant*	*Stadium*	*Day Centre*
Houses	*Church*	*Gym*	*Night Club*	*Theatre*

8. USER GROUP MEETING - WHAT THEY WANT IN THE VIRTUAL CITY

This was achieved by a process of brainstorming and then by eliminating any areas not agreed by the majority of the group

The User Group then decided upon ten priority areas:

- City layout (include transport)
- House
- Supermarket
- Post Office
- Leisure Centre
- Cafe/bar
- Health Centre
- College
- Emergency Services (police, fire)
- Facade of other buildings

"These are the ones I use the most".
"We wanted everything that was in Nottingham".
"I chose places I might find difficult to use".
"I chose places I go to".
"I do not go to all the places, but I`d like to learn how to"

At a meeting of the Steering Group, after taking into account building considerations, it was decided that the first Virtual Learning Environments would be:- Transport - The Cafe Bar - The Home - The Supermarket.

9. USER GROUP DECISION MAKING REGARDING LEARNING OBJECTIVES IN THE VIRTUAL CITY

The next task for the User Group was to decide upon specific learning objectives for each of the learning environments. It was agreed that a different environment should be discussed each time the group met. The Group were very confident and sure and they had no problems in deciding what lifeskills should be learnt in the different environments. Occasionally they were too general, but if the facilitator helped to expand their ideas, they quickly responded with more details,

Example	*"We want to learn about safety".*	
Facilitator -	*"What sort of things would you like to learn about?"*	
	-	*"Burglar Alarms"*
	-	*"Electricity"*
	-	*"Gas"*
	-	*"Fires"*

If there were any disagreements, the User Group members were noticeably ready to compromise or change their minds.

A typical example:

User 1	*"We would like to learn about burglar alarms*"
User 2	*"Nobody has burglar alarms"*
Facilitator	*"Some places do - like group homes or schools or hospitals"*
User 1	*"Yes, they do*"
Facilitator	*"So, do you want to learn about burglar alarms?"*
User 1	"*I don't mind"*
User 2	*"I don't mind"*
Facilitator	*"What does everybody else think?"*
User 1	*"Lets have a vote"*

The User Group chose to use voting when they had different views.

If ever the vote was equally divided, the User Group had their own simple strategy to settle issues. One man stood in the middle of the room with his eyes closed and he spun around with one arm pointing out. Everybody changed position and when the man stopped spinning, the suggestion of the person nearest his pointed arm was the one taken up. During the time the User Group met, nobody ever had cross words and all agreements were reached amicably. Consequently, there were no problems in regards to agreeing learning objectives for the environments in the Virtual City

Example of Learning Objectives for the Virtual City - Cafe Bar

1. Making choices and decisions - ordering drinks from a list for self and others.
2. Social skills when ordering
3. Communication with staff and public
4. Money handling - paying for drinks
5. Appropriate behaviour - table manners, etiquette.
6. Appropriate dress.
7. Toilet use in public situation
8. Dealing with alcohol
   - What drinks you can order, at what ages
   - What affects these drinks have on you

10. USER GROUP DECISION MAKING REGARDING DESIGN OF THE VIRTUAL CITY

The User Group led the design of the Virtual City. Story boards were used to achieve this. As each design or detail was verbally described by the User Group, these were pictorially illustrated on large sheets of paper to form visual images of each environment or part of environment. The group especially liked using the story boards.

"I like using pictures, because I can't read"

"It's easier because you can see what people mean"

"If it's not right, I can say so and then it can be changed"

"I've made a storyboard to say what sort of house I'd like to live in"

Steady progress was made in all areas, until food for the supermarket shelves was discussed. The User Group described the items they would like on the shelves, but did not mention any multi-cultural foods or products. The Facilitator asked if they had forgotten about them, but the User Group refused consistently to give them any consideration. The Facilitator explained that some people liked different foods from other countries. Also that the families of some of the User Group were from Jamaica and Italy and they might like different foods. The User Group then compromised by agreeing to have pizzas, but they did not want to extend this.

User 1	*"I never have Chinese or Indian food or anything like that.*
User 2	*"We don't want it - we just want ordinary food"*
Facilitator	*"Would anybody like any other foods?"*
Everybody	*"No - No - No"*

The Facilitator was concerned, but since it was a majority decision, it was taken back to the Steering Group. The Steering Group suggested that a wider sample of people with learning difficulties should be asked about their food shopping. The User Group agreed with this decision and one User Group member volunteered to do a research with the Facilitator.

A small study of the food shopping of 20 adults with learning disabilities from a nearby Day Centre took place. The results of this study were used to decide food items on the supermarket shelves.

11. CHANGE OF USER GROUP DESIGNS AND CONTENT

Although the views of the User Group were usually followed, there were occasions when the Steering Group felt obliged to change them. In the example of the Supermarket food, the Steering Group had committed themselves to a visually multi-cultural city and this had to be paramount in the design and content. Most of the design and content of the Virtual City was, however, guided by the views of the User Group.

12. PUBLICITY

At the end of the first year of development, the User Group were fully involved in publicising the Virtual City. They submitted an article to a magazine designed to provide new information about disability. More importantly, they were key speakers at an afternoon conference on the Virtual City. Three hundred people listened to their presentation and they all spoke extremely well, even though it was an unnerving experience.

"Its the hardest thing I have ever done"

"I was really scared"

"My hands went all sweaty"

"I wanted to go home"

After their presentation the User Group also demonstrated the use of the computer programmes and talked to people about the advantages of the Virtual City.

"When I use the Virtual City programmes, it helps me to learn before I go out and then I can do things properly"

"You can practise choosing a bus and if you get it wrong, its safe"

"The programmes help you with going out. It helps you to walk into a cafe and feel comfortable"

"Using the programmes is fun so its easier to learn"

13. THE IMPORTANCE OF INVOLVING THE USER GROUP IN THE VIRTUAL CITY

Very rarely do people with learning disabilities provide input to computer programmes or learning schemes. However, in this project they proved without doubt that they do have valuable opinions about their own needs and desires. Instead of a Virtual City based on the assumptions of other people, the User Group involvement has meant that the whole project has been built with informed input from real experts in the field of learning disabilities. The Virtual City has developed in response to the needs of people with learning difficulties, not the assumed needs. Consequently this is a valuable step forward in this field.

"Because its doing it for yourself - thinking of yourself - not others doing it for you"

"We know what we need"

"We have our own things to do - the programme can help us to do these things - not things other people do"

"When other people do things - it's for them, not us"

14. THE ADVANTAGES OF BEING A MEMBER OF THE USER GROUP

Although the User Group was of great advantage of the Virtual City Project, it was a two-way process and the project was also very beneficial to the User Group. Whilst they all put a great deal of effort into the work, it was seen by them first and foremost as an enjoyable social event.

"I love coming to the User Group - it's good fun"

"I'd rather come to the User Group than go to the Youth Club"

"I never go out anywhere else, so I like coming"

"Its somewhere to go and meet new friends"

The User Group all took pride in their appearance for the meetings and three of the group were extremely fashion conscious. Their clothing was usually the focus of conversation for the first few minutes of each meeting.

The User Group realised their importance to the project and this gave them a sense of self esteem and worth.

"I know I'm useful here"

"I can't read or write, but it doesn't matter here - I can still help"

It was noticeable as the year progressed that the User Group became more self assured. At the time of the Virtual City Conference, most of the group were confident enough to speak in front of a large audience.

Only two women said they preferred not to speak, but they wanted to attend the Conference and individually talk to people about the Virtual City.

The group dynamics were very amenable to the requirements of the group. Everyone was friendly and at ease and social conversations took place between all members of the User Group.

15. THE FUTURE OF THE USER GROUP

The User Group will meet again in July 1998 at a large party for everyone involved in the Virtual City Project. After this time, officially the group will not exist. However, six members of the group will attend a Conference in September 1998 and three members will join a new User Group for a Housing Options section of the Virtual City. It may also be that as the Virtual City continues its development, members of the group will again be involved.

All members of the User Group have exchanged addresses and phone numbers and it is arranged that in three months time, the group will meet again purely on a social friendship basis.

The User Group have all expressed regrets at the ending of their meetings, but they all are pleased with their involvement and some of them feel it might influence their lives in some way.

> *"Since the User Group, I am ready to go back to College - I don't just want to sit at home any more"*
> *"I liked helping in the User Group and I want to help people again now"*
> *"When I leave school I am going to phone everybody and visit them".*

16. CONCLUSION

The original decision of the executive group to have user participation in the design of the Virtual City proved to be a great success. The User Group worked co-operatively and creatively together to provide the guidance and specialist information necessary for the building of an effective learning environment for people with learning difficulties. The project organisation enabled the User Group to make their own decisions and support workers and a facilitator helped in this process.

Although the User Group worked on a voluntary basis, party food and social time formed part of the User Group's meetings and this ensured their motivation and willingness to give their time and expertise to the project. The User Groups motivation to attend Steering Group meetings was less effective and it may be that a different approach to meetings would have been preferable. The use of story boards proved an ideal method to assist understanding and recording for people with learning difficulties and the User Group themselves devised techniques for solving problems of decision making.

The success of the User Group was beneficial to themselves in terms of personal achievement and to the Virtual City project in ensuring that the content, learning objectives and design were all appropriate and necessary to the people who would be using the programmes.

The leading role played by the User Group in developing the Virtual City gives respect and understanding to people with learning difficulties by empowering them to play the fullest possible role in producing materials for themselves.

> *"Because its doing it for yourself - thinking of yourself - not others doing it for you"*

Written by the facilitator on behalf of and with full agreement of the User Group. All quotes provided by the User Group.

CAFE/BAR - CONTENT AND FEATURES

ITEMS

- Well displayed signs (toilets)
- Clean tables

- Staff (friendly, helpful)
- Food and Drink
- Menu Board and Prices
- Menus on Tables
- Seating – Comfortable
- Tables and Chairs
- Space between Chairs
- Moveable Chairs
- Glasses, Crockery

 ISBN 0 7049 1141 8

The development of the Virtual City: A user centred approach

David J Brown[1], Steveno J Kerr[2] and Victor Bayon[2]

[1]Department of Computing, Nottingham Trent University, Nottingham, ENGLAND.
[2]VIRART, University of Nottingham, Nottingham, ENGLAND.

[1]*epzdjb@epn1.maneng.nott.ac.uk*

ABSTRACT

This paper will develop the theme of the importance of community based involvement in the development of virtual learning environments (VLEs) for people with a learning disability. It is being presented alongside two other papers, one by the User Group, the other by the Testing Group, describing the design, testing and distribution of the Virtual City. This set of VLEs comprise a computer aided learning (CAL) tool to teach independent living skills to people with a learning disability. Our presentation will demonstrate the involvement of users in each of the stages of development of the Virtual City, and the benefits of this partnership, as opposed to a more tokenistic involvement.

Along side the development of this methodology, the presentation will concentrate on the demonstration of the Virtual City to show how users can learn skills such as the use of public transport, road safety skills, safety within a home environment, the use of public facilities within a cafe and the development of shopping skills within a large supermarket. Video will also be shown demonstrating users involving themselves in the various stages of production of this CAL tool.

1. BACKGROUND

VIRART have enjoyed a long association with community groups and national organisations developing the role of VLEs for special education (Cromby et al, 1996, Brown et al, 1997, Cobb and Brown, 1997). Some of he outcomes of these collaborations have included:

- *Lifestyles* and *Makaton* distributed by Rompa (UK, with International Distributors).
- The *Virtual Factory* for the UK Health Education Authority.
- The *Virtual Tenancy*, produced for Metropolitan Housing Trust and distributed by Pavilion (UK).

This development has been accompanied by a continual evaluation programme to determine the usability of these VLEs and the degree of transfer of this experience into real life skills (Brown, et al, 1998, Standen, et al, 1998).

In March 1997 the National Lotteries Charities Board (NLCB) funded the first stage of the production of the Virtual City in which people with learning difficulties could practice independent living skills. On a National and European level it is recognised that the development of such skills is important not only for the self-esteem of users, but also to off-set the rising cost of social care.

The NLCB commissioned a project to build the Virtual City in which the skills required by people with a learning disability for some degree of independent living could be learnt and practised. The commissioning partners for the project were the Metropolitan Housing Trust (a housing association for people with learning difficulties), The Shepherd School (a school for some 170 students with severe learning difficulties) and VIRART (The Virtual Reality Applications Research Team). This was to be a user centred project, with an emphasis on members of the learning disabled community being involved in all stages of the project.

The importance of a user centred approach in project development for people with learning difficulties is well understood and documented in the special needs community. This is exemplified by the activities of such groups as People First and the Acting Up Team. In a recent public address Gordon Grant, Professor of

Cognitive Disability, University of Sheffield (Grant, 1998) expounded models of care and how these can be adapted so that university based research for people with disabilities is carried out in partnership with relevant community groups. Indeed he went further, and suggested that money from sponsoring organisations should be given to community groups and that they should decide how to spend these on research and development. Universities and other research groups could then be held totally accountable for their research and development in the disabled and rehabilitative fields. This approach is central to the development of the Virtual City to teach independent living skills, where research, development and distribution is directed by a consortium representing community based groups, schools and research institutions.

2. DEVELOPMENT METHODOLOGY

During the early stages of this project a development methodology was established to guide the production of these virtual learning environments (VLE's). This process placed the user at the centre of each of the decisions made regarding the project development.

The framework for development is in brief:

User Group to determine the components developed (transport system, recreation centre, house, supermarket, cafe, etc.), the learning objectives for each component (safety, communication, financial skills, etc.), the types of interaction that are possible and the dialogue that will occur.

Steering Group to incorporate the input from experts in the learning disabilities field in the components and learning objectives of the Virtual City.

Story-boards developed from the User and Steering Group input.

Building programme for the Virtual City based on these story-boards.

Expert Review of the Virtual City whilst under construction to ensure that they fit the original vision of the User and Steering Groups.

Evaluation phase in which users and experts combine to determine the nature of testing that will help to establish whether the use of the Virtual City delivers the original learning objectives.

Refinement phase in which the results from the evaluation phase are used to edit the Virtual City to ensure that the design best fits the original vision of the User and Steering Groups.

Dissemination phase the User and Steering Groups decide how to best market and publicise the Virtual City.

Each of these processes involved in this developmental methodology will now be discussed in further detail.

3. USER GROUP

The User Group consists of fifteen people with a learning disability, together with a facilitator to provide effective advocacy. The age range of the group varies from fifteen to sixty, representing Nottinghamshire special schools, housing associations and colleges. The group acts as the voice for users contributing their ideas to the project and every step is taken to ensure their continued and enjoyable participation. To this end all attendance costs are met, including travel, a communal meal and carer expenses should they wish to be accompanied by a mentor.

In initial meetings the User Group considered their ideal city, including the types of places they would like to visit but perhaps were restricted from doing so. They then produced a list of thirty-eight components to such a city, but were persuaded to choose their favourite twelve, in a prioritised building list. The facilitator then worked out with the group what was important to them in each of these components and what they would like to be able to do within them. These activities are called learning objectives and it is interesting to note that one of their overriding concerns was personal safety. Whilst this process was going on rough story boards were drawn up by the Project Manager so that the Users had a visual interpretation of their ideas as they made them, supported by the facilitator providing a language based interpretation.

4. STEERING GROUP

The Steering Group for the project is made up from around fifteen professionals working in the learning disabled community around Nottinghamshire. Their input provides further ideas on components for the Virtual City and learning objectives, as well as providing overall guidance for the project, on areas such as ethics and testing. This group meets at around the same regularity as the User Group; once per month, with more meetings scheduled at periods of high activity.

Users views are further represented on this group by one or two users selected from the User Group, who attend the Steering Group meetings supported by the facilitator. The Facilitator and user representatives plan this input, having received an agenda before the meeting.

In the second stage of development of the Virtual City, the Steering Group review the ‘rough’ story boards generated by the User Group. They may add extra learning objectives to some of the components, or even suggest alterations to the components. In this way the Steering Group suggested that the house designed by the User Group be altered to a single story dwelling to ensure wheelchair access, and added extra learning objectives such as the chance to practice dressing skills depending on the weather conditions outside the bedroom window.

5. STORY-BOARDS

The story boards are produced in much the same way as film story boards are produced. These detail the components (house, supermarket, café, etc.) and the learning objectives. They also detail the ways in which the Virtual City will ‘scaffold’ the user so that as each outcome (as a result of an interaction with the Virtual City) is introduced the user is encouraged to choose the correct one. The user can experience other outcomes and learn the risks, or dangers, associated with these lines of action. The ability to experience different outcomes from a range of actions is particularly pertinent and powerful in learning objectives such as road safety.

Dialogue within the Virtual City will also be shown on the story boards, usually associated with the learning objectives. This ensures that dialogue will be appropriate by giving language and speech therapists the opportunity to input their ideas on how language can best help to interpret what is happening in the Virtual City.

Because these story boards are visual and easy to understand they can serve as a means by which users can check the building progress of the Virtual City, to ensure that it is developing within their vision. Some of these story boards will be shown as PowerPoint overheads during this presentation.

6. BUILDING PROGRAMME

The story boards are then examined by the Project Manager and Builders and a building programme is devised. The platform used is Superscape for dual reasons. The first is that the associated hardware and Visualiser are cheap, especially important for a product that will be distributed to the resources starved special needs community. The second is that there are still too many unresolved health questions associated with head mounted displays to risk exposure on a group with such specialised needs (Wilson, 1995).

There are moves afoot concerning the possibility of delivering the Virtual City on the Internet, possibly via Viscape. Other programming implications concern the possibility of using distributed virtual environments, and the learning implications that could ensue from users interacting in the Virtual City from different locations.

7. EXPERT REVIEW

Selected members from the User and Steering Groups attend review meetings whilst the building of the components of the Virtual City is taking place. This allows users and their representatives to check the building progress. Users can compare the developing Virtual City with the original story boards, using the latter as their metric.

8. EVALUATION PHASE

The design, execution and results of the evaluation phase are being presented by our co-researchers and will not be detailed here. The purpose of this phase is to provide information from the users on the usability of the Virtual City and whether they can use it to learn independent living skills. Other benefits may accrue from its use, including increased confidence in the use of computers, and increased confidence generally, such as autonomous moves to use services such as a supermarket or café. One of the members of the User Group was encouraged to join a computing course at a local college in response to working on the Virtual City.

9. REFINEMENT PHASE

Again the refinement process is fully described in our colleagues' paper. However, we can say that testing results have been used to heavily improve the design of the transport system, café, house and supermarket. It is this process of users' ethical inclusion in the testing process that further underlines the importance of representing user views in all stages of the development of the Virtual City.

10. DISSEMINATION PHASE

The most ideal method of dissemination of news on the development, testing and distribution of the Virtual City would be one involving the users. We do, however, recognise that users would not want to be involved in the production of some of the academic papers describing the development and testing methodologies, although their contribution to these is certainly acknowledged. The dissemination strategy at the end of year one of this project is:

- *Newsletter*: The User Group have produced a newsletter describing the project in accessible language and Makaton Symbols. Topics include the User Group meetings, the testing programme, profiles of people involved in the User Group and a description of the Virtual City.
- *Web-site:* This is based on the newsletter, and linked to the Shepherd School site @ http://www.campus.bt.com/campusworld/orgs/org1573/index.htm.
- *Conference papers:* These include the presentations made at ECDVRAT'98 by the User Group, Testing Group and ourselves.
- *Disabilities Press:* The User Group have written an article entitled 'The Virtual City' for the Way Ahead Magazine, distributed through the Children's Disability Register.
- *Open day:* This was held at the Notts. County Ground on 18th July, where over 300 service providers and users attended, and over 200 more expressed an interest in the project. Attendees were asked if they wanted to contribute ideas to the project and these replies are currently being processed. Within the next stage of the project, a nation-wide testing programme is planned based around those who wish to join the project.
- *Journal papers:* Several Journal papers are planned, the first of which is currently in publication discussing the proposed testing methodology (Brown, et al, 1998).
- *Distribution:* We are negotiating with national (Pavilion) and International (Rompa) distribution companies to distribute the Virtual City. In the next stage of the project we are planning to develop a production company formed of users from the project to handle the burning of CDs, manual production and packaging. We also want to extend user involvement into distribution, with users demonstrating and marketing the Virtual City at special needs exhibitions and conferences.

11. CONCLUSION

The first stage of the Virtual City project has produced:

- A User and Steering Group to guide all stages of development of the Virtual City, and providing ideas for future development far out-stripping current funding levels.

- A Development Methodology to guide the collaboration of service providers and people with a learning disability working with research groups to produce effective virtual learning environments.
- A Virtual City, running on desktop PCs under the Superscape platform and at the moment consisting of a transport system, house, café, supermarket and housing advice centre in which people with a learning disability can learn independent living skills.
- A Testing Methodology to ensure that the design of the Virtual City is suited to the abilities of people with a learning disability, and can be used to develop independent living skills.
- A Dissemination Strategy which will result in the first stage of the Virtual City being distributed to the special needs market by Summer 1999.

12. REFERENCES

Brown, DJ, Neale, H, Cobb, SVG and Wilson, JR, 1998. Structured evaluation of VE for special needs education. In Press in *Presence: Tele-operators and Virtual Environments*, MIT Press. 1998.

Brown, DJ, Kerr, SJ, and Wilson, JR, 1997. VE in special needs education: The LIVE programme at the University of Nottingham. *Communications of the ACM*, August, vol. 40, no.8, p72-75. 1997.

Cobb, SV and Brown, DJ, 1997. Health and Safety in Virtual Factories. *In Proceedings of the First Annual International Conference on Virtual Reality, Education and Training*, Loughborough, 1997.

Cromby, JJ, Standen, PJ and Brown, DJ, 1996. The Potentials of Virtual Environments in the Education and Training of People with Learning Disabilities. *Journal of Intellectual Disabilities Research*, Vol. 40, part 6, p489-501, Dec. 1996.

Grant, G, 1998. 'Hearing the voices of people with learning disabilities'. The Anne Craft Memorial Lecture, QMC, University of Nottingham, June, 1998.

Standen, PJ, Cromby, JJ and Brown DJ, 1998. Can students with learning difficulties transfer skills acquired in a virtual environment to the real world? In Press in the *Journal of Mental Health Care*. 1998.

Wilson, JR, 1996. Effects of participating in virtual environments: A review of current knowledge. In *Safety Science*, 1996; Vol. 23, No 1, pp39-51.

Evaluation of virtual learning environments

S V G Cobb[1], H R Neale[2] and H Reynolds[3]

[1,2]Virtual Reality Applications Research Team (VIRART),
University of Nottingham, UK

[3]Metropolitan Housing Trust, Hazlemont House, 11 Gregory Boulevard
Nottingham NG7 6LB, UK

[1]*epzsvc@epn1.maneng.nottingham.ac.uk*, [2]*epxhn@epn2.maneng.nottingham.ac.uk*

ABSTRACT

The Virtual Life Skills project describes a user-centred design approach to building virtual environments intended to provide a practice arena for skill learning in children and adults with learning disabilities. In the first year of the project four modules of a Virtual City have been developed: a house, a supermarket, a café and a transport system (see Brown et al, this issue for a description of the project). Evaluation of the project has been concerned as much with the design of the virtual learning environments (VLEs) and issues of usability and access as with monitoring skill learning and transfer to the real world. Two approaches were taken to the evaluation the four virtual learning environments. For three of the VLEs, Supermarket, Café and Transport, a test-retest experimental design method was used. This compared user performance in real world tasks with the same tasks presented in the VLE. Expert assessment was used to evaluate the Virtual House, looking at usability and appropriateness of the learning scenarios. It was found that VLEs can provide interesting, motivating learning environments, which are accessible to users with special needs. However, individuals differed in the amount of support required to use the input devices and achieve task objectives in the VLE. Expert and user review methods indicate that the VLEs are seen to be representative of real world tasks and that users are able to learn some basic skills. However, it would be unrealistic to expect transfer of skill over a short time period of learning as used in this project. Further testing is needed to establish the longitudinal learning effects and to develop more reliable techniques to allow users to express their own opinions by themselves.

1 INTRODUCTION

The aim of the Virtual Life Skills Project was to develop a virtual city containing a variety of real world settings to enable individuals with learning disabilities to learn about and practise important living skills with an ultimate goal of preparation for independent living. Four components of the Virtual City were completed: Virtual Supermarket; Virtual Café; Virtual House and Transport System. These are referred to as Virtual Learning Environments (VLEs).

The project as a whole implemented a user centred approach to design and evaluation of the VLEs. Design specifications for the VLEs were made according to the User Group requirements and learning objectives defined by the project Steering Group (full details are given in Brown et al, this issue). When completed, a programme of testing was carried out to assess the suitability of the VLEs for their intended users. The results produced recommendations for design changes to each of the VLEs.

The ultimate objective of the Life Skills project was to assist development and improvement of real world skills. However, it was recognised that there may be many factors influencing learning from use of VLEs. For example, usability and access issues would have a huge influence on learning from any computer-based programme.

This aspect was perhaps the most influential in determining support worker/advocates' initial impressions – if users couldn't control the computer then how could they learn anything from it? The evaluation study had to consider this and so background information concerning users' abilities and experience of computers was obtained and measures of computer skill were tracked throughout the testing programme.

It was also recognised that it may be too ambitious to expect to see changes in real world skill levels within the short time scale of this project. Other outcomes which may be necessary foundations for later skill learning should also be identified. One example would be enjoyment from interacting with the VLEs. If a user enjoys the VLEs then they would be more motivated to use them again and to explore new features within the programme. This self-motivation, together with the advantages of 'learning by doing' and exploration, is ideal for learning in any context. The evaluation study therefore had to be broad enough to identify any benefits from using the Virtual City irrespective of their influence on skill level.

In addition to testing the suitability of the VLE designs for users with learning disabilities, the evaluation study was set up to identify benefits of using the Virtual City. We have identified four desirable outcomes of a VLE (Brown et al,1998):

- Usability – that users can access the computer programmes appropriately
- Enjoyment – that they like using them and want to explore the VLEs
- Skill learning – that from exploration and practice in the VLE users are better prepared to carry out certain real life tasks
- Transfer of skills – that users can apply their new knowledge and skills into their everyday life

2 EXPERIMENTAL STUDY

Figure 1 illustrates the evaluation procedures used to assess each of the virtual learning environments. For three of the VLEs, Supermarket, Café and Transport, a test-retest experimental design method was used which compared user performance in real world tasks with the same tasks presented in the virtual environment. This was not possible for the Virtual House and so the assessment was based on expert and user comments on usability and design features of the Virtual House in support of the learning objectives.

2.1 Selection of Testers

In line with the user-centred approach to design and development taken in this project, the evaluation study was based on user trials. It is important to acknowledge that these 'users' were not being assessed in their use of a completed product but were contributing to its development. For this reason 'users', representing the target user population, are described as 'testers'.

It was important to the project that a range of testers with different abilities and backgrounds took part in the study. Individuals from a variety of community centres were invited to participate in the evaluation study. Background demographic information (age, gender, reading ability, numeracy, comprehension, physical disability and computer use) was obtained via questionnaire and 20 testers were selected for the experimental evaluation study.

Figure 2 shows the range and background of testers who took part in the Virtual House evaluation study. This covered representative user population and expert representatives from the learning disabilities community.

2.2 Method

An introductory meeting allowed testers to complete 'habits questionnaires'. These were relevant to the VLE that they would be testing and provided useful information concerning the testers' skill levels and also a basis upon which to assess the potential relevance and impact of VLE training for each individual. For example, if a tester goes shopping on a weekly basis but cannot go unassisted, then training in the Virtual Supermarket is relevant to them. If, at the end of the project or at some later date, they are able to go shopping independently, then the VLE training may also have had high impact on their life skill development.

Figure 1 shows the test-retest method used in this section of testing. At the first scheduled testing session the tester, together with their support worker, completed a number of tasks in a real environment. The experimenter recorded how much support the tester requested or was offered by the support worker for each activity with specific interest in who was making decisions and how much prompting the tester needed to complete tasks.

One week later the tester and support worker visited the University of Nottingham to start the VLE training sessions. They completed tasks, similar to those completed in the real environment, in the VLE. These sessions were video recorded to allow the experimenter to further analyse the activities. The experimenter also observed specific difficulties faced in using the computer program.

One week after completion of the training sessions the tester and support worker repeated the tasks in the real environment. The experimenter recorded the activity in exactly the same way as before. When all testing sessions had finished the support workers completed attitude and opinions questionnaires.

2.3 Enjoyment and Usability

User enjoyment was assessed taking data from observing tester use of the VLE and using questionnaire answers from testers and support workers.

User attitudes, opinions and comments indicated that:

- There was a very high overall level of enjoyment.
- The testers experienced low levels of anxiety and frustration. Highest levels were felt in the first real world and VLE sessions.
- Navigation, although having been found as one of the most difficult tasks to do, was often stated as the most enjoyable aspect of using the VLE.

The support workers further consolidated this information by rating tester enjoyment on a seven point Likert scale. There was a significant change in attitudes before and after use of the VLEs reflecting that support workers reported that testers did enjoy using the VLEs more than expected. As an example, one support worker wrote " Very much enjoyed using virtual environments and still talks about using them".

Usability assessment for the Virtual House was based on expert and user responses to questions regarding how easy or difficult they found it to complete the tasks in each room. The responses are summarised in Tables 1,2 & 3. It was found that to explore all of the activities in the Virtual House could take up to an hour. Not everyone could afford this much time and so some questions could not be answered.

Table 1. *Expert review of Usability for themselves*

	Easy/v. easy	Difficult/ v. difficult
Find your way around the house	79%	7%
Know where you were in the house	86%	0
Move around using the joystick	50%	21%
Position the cursor over objects	79%	7%
Activate objects using the mouse	71%	14%
Understand what you were expected to do	93%	0

Table 2. *Expert review of Usability on behalf of users*

	Easy/v. easy	Difficult/ v. difficult
Find their way around the house	31%	38%
Know where they were in the house	54%	15%
Move around using the joystick	0	54%
Position the cursor over objects	23%	46%
Activate objects using the mouse	23%	31%
Understand what they were expected to do	38%	23%

Table 3. *Users review of Usability for themselves*

	Easy	Difficult
Find your way around the house	100%	0
Know where you were in the house	100%	0
Move around using the joystick	66%	33%
Position the cursor over objects	50%	50%
Activate objects using the mouse	66%	33%
Understand what you were expected to do	100%	0

It can be seen that the experts experienced few difficulties in using the Virtual House but anticipated that users from the special needs groups that they represented would. However, the users themselves found that they could use the Virtual House – the only difficulties reported were related to use of the computer input devices.

In the experimental study it was found that all of the testers could use both the joystick for navigation and the mouse for interaction. However different levels of support were needed to use the input devices and complete tasks. One tester had a physical disability which meant that she could not use the mouse without physical assistance but understood the mouse 'concept' and would have been able to use a different device on her own. Ability to use the input devices was seen to improve during the course of the experiment. Observation of testers using the computer yielded a positive change in support worker attitudes concerning tester ability to use the computer and its input devices.

A usability content analysis was performed on the observation and questionnaire data (summarised in figure 3). The categories summarise the type of problem and display the number of this type of usability problem found. For each problem type, design properties relating to the VLEs were defined. These, if refined, could increase support given in this activity and increase its usability. For example, user interactions could be supported by providing standard coloured symbols with a simple text voice over, replacing a text box. Suggested refinements to improve VLE usability in all categories are summarised.

Many usability problems of the same type occur in these three VLE's. The most frequent types of usability problems were reading text (10 incidences) and the VLE not providing enough/ the same clues as the real world (7). One main recommendation made was for standard design features (e.g. green for yes, red for no, a move forward arrow) and use of Makaton symbols in place of text. This method was used to advise and give evidence for important design modifications.

2.4 Skill learning and transfer

To attribute any real world improvement to VLE use we need to also look at the testers performance in the VLE training sessions. This was done by looking at how performance changed over the real world and VLE sessions and recording tester and support worker feedback. Each learning objective was broken down to a set of skills and these were further divided into basic components. In each testing session the interactions between the tester and the support worker were observed and certain behaviours were monitored for each component. Example behaviours are; who makes the decisions, who takes control and how much help the tester requires to do each task. A 5 point scale was produced which could be used to record the level of support worker involvement in the task, the scale ranges from no support worker involvement to physical prompts given by support worker and support worker does task for tester. This allowed the change in support worker involvement over time to be monitored and any change in behaviour linked to specific components of tasks. The methodology allowed us to compare performance, behaviour and attitude. This meant any (potentially important) changes may be noticed e.g. increased involvement in and awareness of shopping in the real world or an increased confidence in performing certain tasks.

The results show definite examples in tester skill transfer from VLE to real world in only a handful of activities. There may have been many more, less obvious skill/knowledge development from using the VLEs but they have gone undetected. Skills learnt from these sessions may not be evident in the real world straight away but noticed by support workers at a later date.

One example from the Virtual Supermarket exemplifies transfer of skill. A tester learnt to do task 2 (collect shopping trolley) alone in the VLE. Less support worker prompts were recorded in the second real world session. Her support worker commented "The testers' life skills regarding collecting and returning the trolley <u>when out shopping</u> have noticeably improved since the beginning of the virtual supermarket programme."

One tester used the café VLE to learn which toilet they should use in a public situation. In the first real world session she tried to enter the female toilets, but the VLE is used to put across the concept that her wheelchair will not fit in here and she must use a toilet designed with wheelchair access. This knowledge was demonstrated in the second real world session.

Using the Virtual Transport facilitated the first formal 'travel training' for this set of testers. All of the testers appeared more familiar and confident in doing the tasks in the second real world session after having practised in the VLE. One tester learned to put the coins in the correct slot on the bus in the VLE and repeated this skill in the real world second session. Another tester needed no prompts to collect the bus ticket in the real world second session, a procedure learnt in the VLE. Their support worker commented that he felt more comfortable with taking students out to cross roads and use public transport after they had trained using the transport VLE.

3 CONCLUSIONS

This project found that VLEs can provide interesting, motivating learning environments, which are accessible to users with special needs. However, individual differences determined how much support testers required to use the VE input devices and achieve task objectives in the VLE.

The reported opinions of support workers changed over the testing period. Questionnaire responses suggested that they gained a more positive attitude towards the use of VR in teaching life skills. This was also demonstrated by the expert testers of the Virtual House who thought the technology would be suitable for teaching a multitude of topics. Twenty-eight different additional activities were suggested for just one component of the virtual city.

An important outcome of this project is that it has enabled development of a unique evaluation methodology using a user-centred design and evaluation approach. The testing programme uncovered a number of usability issues of VLEs used in special needs applications. Many usability problems had common causes and the Usability Content Analysis in figure 3 shows their categorisation. Design refinements were suggested for each usability category and made to the VLEs. Further research would aim to provide design guidelines for the building of VLEs for special needs applications to decrease time to final product and minimise difficulties with usability.

Expert and user review methods indicate that the VLEs are seen to be representative of real world tasks and that users are able to learn some basic skills. However, it would be unrealistic to expect transfer of skill over a short time period of learning as used in this project. Using the VLE over a longer period may have allowed greater skill learning and real world transfer. Further testing is needed to establish the longitudinal learning effects and to find out the optimal number of VLE training sessions for skill learning.

Most of the evidence collected from testers was from questionnaire answers, often interpreted by a support worker. This study found the need for needs further development to allow users to express their views by themselves. This may involve a multimedia/animation based questionnaire.

Acknowledgements: Testing was carried at the University and in the community. Thanks are expressed to all parties who supported the testing program including: Sainsbury's, Castle Marina; Café Lautrec, University Arts Centre; Nottingham Buses; Shepherd School; Clarendon College; Metropolitan Housing Trust; Nottingham Social Services.

Special thanks are expressed to all of the individuals, teachers, parents and support workers who contributed to the testing programme. This required many hours' commitment and a good deal of answering questions. We thank everyone involved for their patience and effort.

4 REFERENCES

Brown, D.J., Neale, H.R., Cobb, S.V.G. and Wilson, J.R. (1998) Structured evaluation of virtual environments for special needs education. *Presence* (to appear)

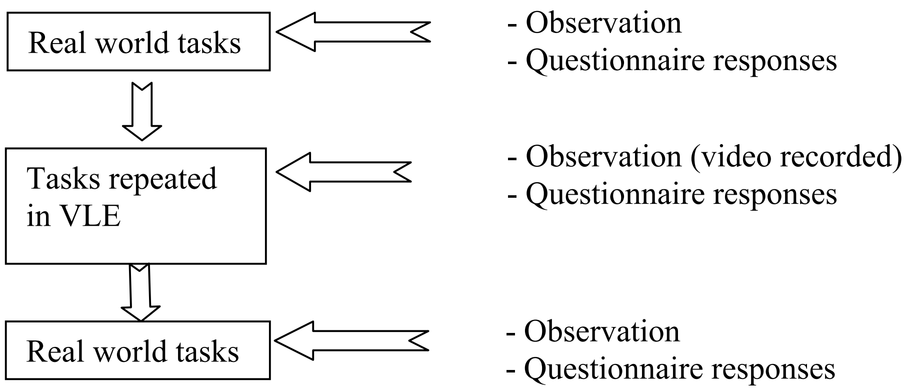
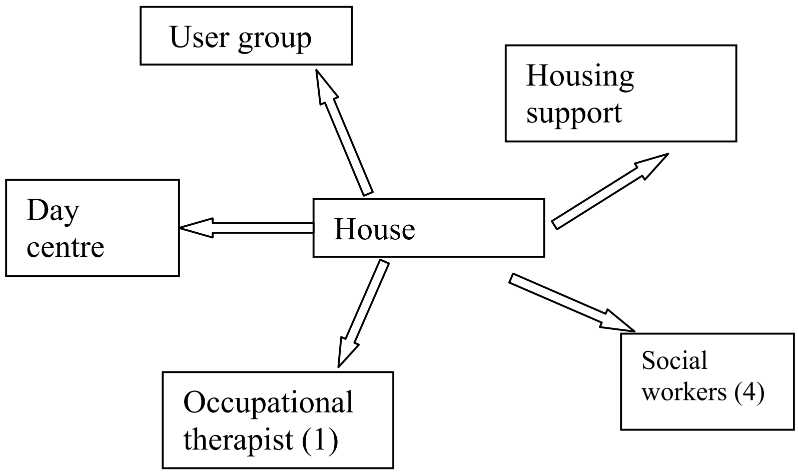


Figure 1. *Test re-test method*

Figure 2. (below) *Evaluators of the Virtual House*



Usability problem	Category	Refinement to VLE design
S2.1 Create a list option S2.2 Choose food categories C2.1 Enter personal details C2.3 Sit here? Screen text overlay C2.4 Menu – screen text overlay C2.6 Wash hands – screen text overlay T2.1 Show and try buttons T2.2 Use of text boxes T2.3 Select destination T7.3 Get off bus at correct stop by clicking on text box	**Reading text problems (10)**	**Increase use of Makaton symbols** **Standardise 'yes' 'no' 'move on' with colours, symbols and position.** **Speech therapist to simplify any text and suggest symbols.**
S2.3 Select item activates small number C2.5 Click food activates small tick in box S2.14, C2.8 Paying – click on coin activates small numeral (denotes how many chosen)	**Unsure of effect of action (4)**	**Highlight object –red outline when selected** **Transfer coin selected to representation of hand – real world clue provided.**
C2.2 Bump into table to sit down S2.11 Bump into cash desk to allow loading of goods	**Not naturalistic action/interaction metaphor (2)**	**Allow interaction using mouse** **Can position trolley in larger area next to cash desk**
S2.7 Collecting trolley – small area to click on C2.5 Choose the food – small area to click on T7.2 Confusing to click on coin box, ticket machine and driver – due to arrows C2.11 Use toilet in café overlapped/confusing instructions as to what to click on and in what order.	**Problems to interact with object (4)**	**Enlarge object/provide closer automatic viewpoint** **Highlight object by making it red and flashing** **Clarify verbal instructions given – speech therapist input**
S2.5 Enter doors of supermarket C2.10 Enter the toilets T4.1 Position at shelves	**Navigation problems (3)**	**Make doors wider, double doors open together automatic close is slowed/stopped** **Provide auto viewpoint at shelves**
S2.4 Current state of list is not known when creating it S2.6 Difficulty finding product areas and individual products S2.9 In VLE extra step needed to use shopping list S2.14, C2.8, T4.2 Payment – no opportunity to select different coins and then change your mind – use different ones. T4.2 Recognition of coins	**VLE does not provide enough/ same clues as real world (7)**	**Provide clues which reflect those given in the real world e.g. – can see representation of list when creating it, coins/notes more realistic** **Increase clues (more than you would have in real world) given to help usability of VLE – e.g. picture/symbol signs in supermarket**

Figure 3. *Usability Content Analysis*

Multimedia INterface for the Disabled (MIND) project

R J McCrindle and R M Adams

Department of Computer Science, The University of Reading
Whiteknights, Reading, RG6 6AY, UK

R.J.McCrindle@reading.ac.uk, RyanAdams@compuserve.com

ABSTRACT

The Multimedia Interface for the Disabled (MIND) project is concerned with developing a set of guidelines and authoring tools, for use by multimedia developers, to enable them to augment their products to encompass the specific needs of sensory impaired users. This paper presents the ethos behind the project and describes the MIND software prototype developed. The MIND prototype maximises the effectiveness of multimedia information delivery, through the provision of an adaptable and easily navigated user interface, which incorporates access to the augmented multimedia information created with the authoring tools.

1. INTRODUCTION

Multimedia has been the subject of intense development over recent years. This development coupled with the continually decreasing costs of the enabling technology has resulted in a significant expansion in the use of multimedia. Indeed, it is already noticeable that multimedia applications are impacting significantly upon our environment whether in the workplace, in education or in entertainment (Sloan, 1996).

Multimedia by its very nature of being a multi-modal mix of text, graphics, audio, video, and animation, stimulates our senses of sight and hearing, often in harmony to maximise information delivery. For the majority of the population, a multi-modal mix thus adds value to the information dissemination process. However, there is a small but significant group of the general and computing population who through some form of sensory impairment are unable to exploit fully the capabilities of multimedia.

Research into interface technology for users with various forms of sensory deprivation has revealed a number of specialised techniques, tools and applications that can be used to better facilitate computer interaction for the disabled user. These include Braille screen readers (Alva, 1998), screen magnification tools (Dolphin, 1998), acoustic environments (Lumbreras et al, 1996), speech synthesisers (Dolphin, 1998) and virtual keyboards (Istance, 1996). However, many such developments rely on specifically designed hardware peripherals, which must therefore be 'bolted onto' the software application with which they are to be used. This is due to the fact that the broad spectrum of software products currently in use, and particularly those of a multimedia nature, do not include, interfaces specifically designed to assist the user who has sensory impairment.

Evidence suggests that multimedia applications will become continually intrinsic to our work and everyday lives. Hence there is a very real need to research and define an approach to multimedia product development, such that mainstream multimedia products can be effortlessly used by the full spectrum of the population (Edwards et al, 1995).

2. AIMS OF THE MIND PROJECT

The MIND project was instigated to further investigate and invite discussion about the perceptions and needs of individuals throughout the entire spectrum of the sensory impaired population, particularly with regard to their use of multimedia applications. The key aims of the MIND project were threefold addressing the authoring, end-user interaction and technological aspects of multimedia development:

1. To develop a set of guidelines and authoring tools that can be used by multimedia developers to augment their products to encompass the specific needs of sensory impaired users. These guidelines and tools address navigational and interactive aspects of the interface, as well as provision of

compensatory methods for presentation of multimedia information from which sensory impaired users may presently be precluded.

2. To provide the user with an interface that can be adapted personally by themselves to enable navigation, interaction and presentation of previously non-perceivable information in alternative forms according to the specific requirements of their disability.
3. To design and develop a software prototype that addresses the first two aims, whilst operating on any 100% MPC level-2 (Sloane, 1996) compliant personal computer running the Microsoft Windows 95 operating system, to enable maximum uptake and evolution of the system.

Central to meeting these aims was the mapping of the general psychological issues associated with user interface design (see Table 1) to the identified limitations and requirements of the sensory impaired users (see Table 2) within the scope of affordable technology (Eysenck and Keane; Goldstein; McKnight et at, 1993; Preece, 1994).

Table 1. *Examples of psychological user interface design issues considered*

Psychological Issue	Interface Considerations
Cognitive Issues	Can established information, theories and methods be applied to the MIND interface? For example, can information processing theory be considered in an effort to conceptualise user behaviour and consequently make predictions about user performance?
Perception & Representation of Information	How information is perceived and represented is of paramount within a system such as MIND. For example, how can the interface enable information projection via particular media types to be accessible to all individuals across the range of disabilities considered?
Attention, Memory & Learning	How can the MIND system directionally focus attention and enable the user to quickly and easily learn and remember the operation of the interface, whatever the disability?
Input Devices	Which input devices are most suitable for a particular disability and how can the interface accommodate these?

Table 2. *Examples of possible physical and psychological limitations related to impairment*

Sensory Impairment	Possible Physical & Psychological Limitations
Blind	Unable to see the VDU. Complete screen contents may have to be stored In memory for fluent interface navigation. May feel unconfident using the mouse.
Partial Sight	May only be able to see large objects present on the VDU. May be able to only focus on one particular area of the screen at any particular time.
Deaf	Unable to hear sound emitted from PC soundcard.
Hearing Impaired	May be unable to hear certain frequency ranges. May only have monaural hearing with consequent loss of stereo sound perception.
Co-ordination Difficulties	May be unable to utilise the mouse and therefore unable to accurately locate and select controls within a two dimensional screen definition. May be unable to utilise more than a few control keys.

3. THE MIND SYSTEM APPROACH

In addition to the design and development of the system architecture, it was deemed important to consider the position of the developers and the end users within a proposed typical development and usage pathway and to account for this with the prototype design. For brevity, this is summarised in Figure 1.

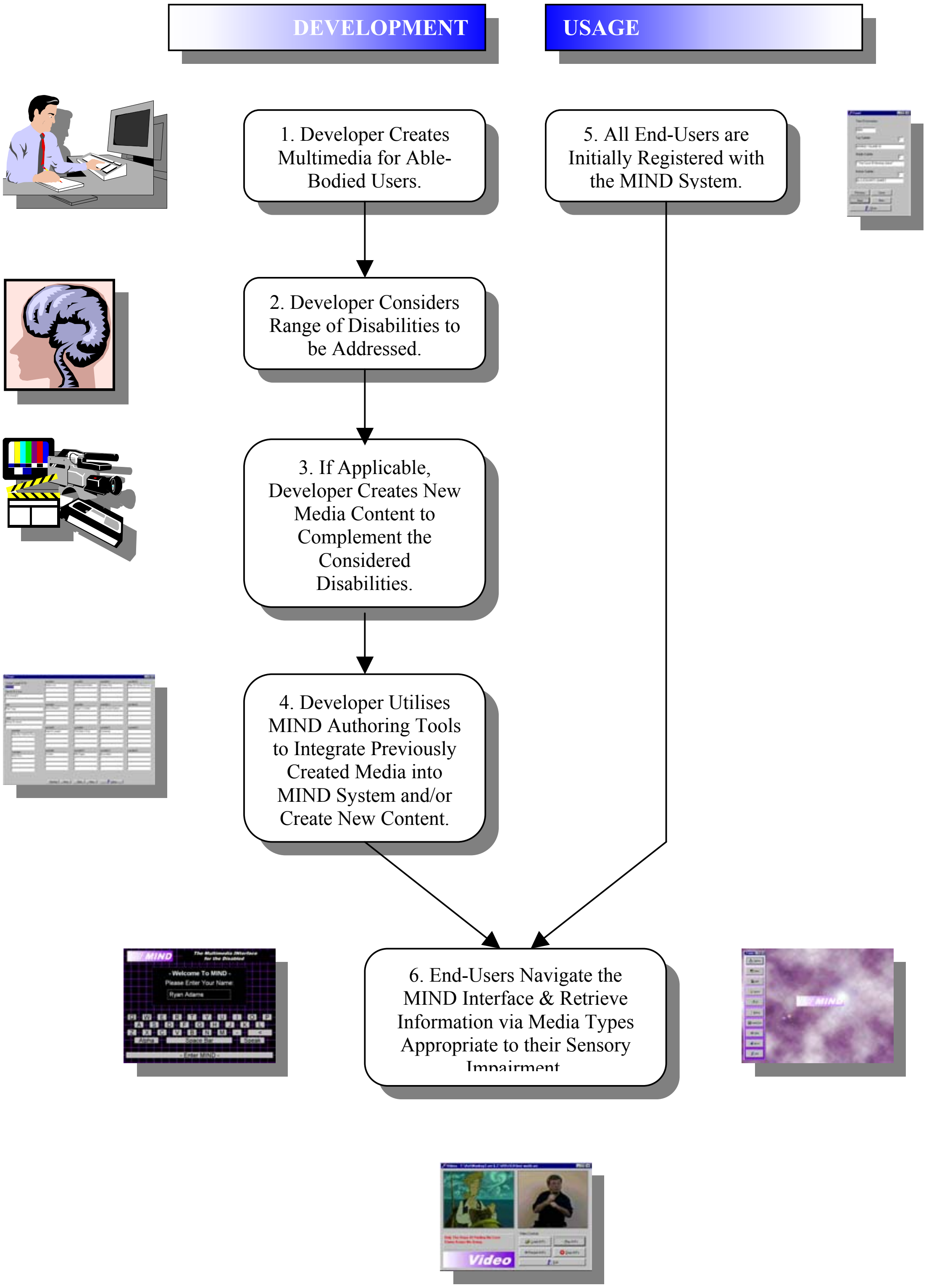


Figure 1. *Diagram showing a possible development and usage pathway.*

4. THE MIND SYSTEM PROTOTYPE

The MIND system is being designed and developed via a combination of rapid evolutionary prototyping and incremental development techniques (Sommerville, 1995; Connel and Shafer, 1995). This approach has enabled a flexible framework to be constructed into which the various interface elements and authoring tools can be progressively integrated. Such an approach has also enabled user feedback to be incorporated into the system features from an early stage. To facilitate rapid and flexible development, within a visual integrated development environment that offers customisable multimedia components, Borland's Delphi version 3.0 (Borland, 1997; Osier et al, 1997) was chosen as the development language. To enable maximum uptake the MIND system has been designed to operate on any 100% MPC Level 2 compliant personal computer running the Microsoft Windows 95 operating system. The key features currently incorporated into the prototype are summarised in the following subsections and Figures 2 to 11.

4.1 Logging onto the System

The 'Log-in' form provide the means of entering the MIND system. It consists of a virtual keyboard and a data entry box. A toggle-audicon enables the Querty arrangement (see Figure 2) of the keyboard to be changed to an alphabetical format (see Figure 3) and vice versa. Each time a letter of the alphabet, represented by an audicon, is selected, the chosen character is added to the end of the name string in the data entry box. A blind user can check the text currently entered in the data entry box by selecting the 'Speak' audicon. Navigation of the keyboard may also occur via tabbing from one audicon to the next.

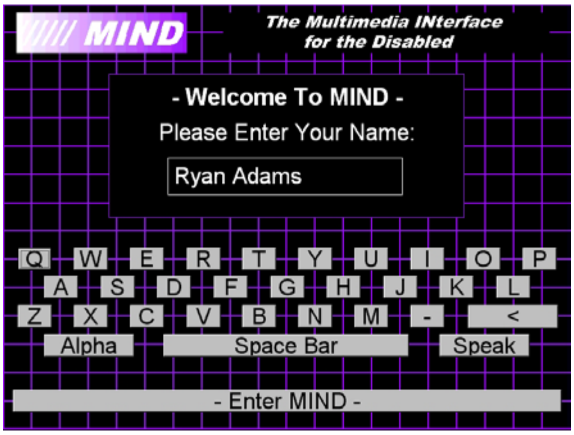


Figure 2. *The Log-in Form showing Qwerty arrangement of the virtual keyboard*

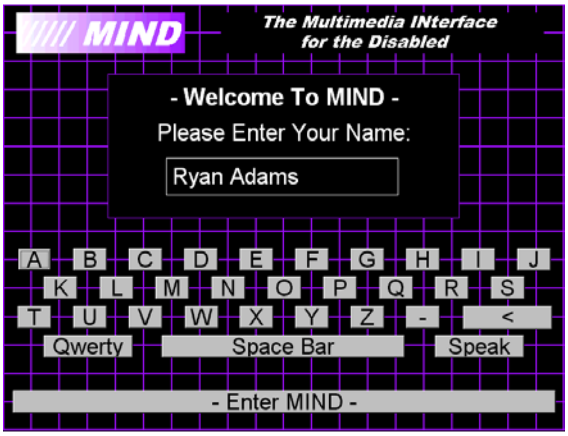


Figure *3.* *The Log-in Form showing Alphabetical arrangement of the virtual keyboard*

Alternatively, the MIND system incorporates Microsoft's Speech Recognition engine (1998) thereby enabling the user to enter their log-in name verbally by means of an external microphone. If recognised as a registered user, the system acknowledges the user and automatically loads all associated preferences saved under that user's login name. If a particular name is not recognised then the system asks the user to try again, advising them of the need to speak as clearly as possible.

The preferences of each user are stored in a 'User Registration File' via the 'User Registration Tool'. This tool allows the addition and editing of users currently registered with the system. It also compiles users' names into the special 'grammar file' needed by the speech recognition engine to determine which log-in names are valid.

4.2 The Main Menu and Interface

The multimedia interface consists of a resizable and adaptable 'Main Menu' window or form containing the buttons from which all facilities that provide access to the multimedia components can be selected (see Figure 4). This form is completely resizable such that the buttons also resize, thereby maintaining their spatial relationship along the vertical axis. The font describing the buttons' functionality also resizes in order to provide partially sighted users with the largest typeface possible within the vertical screen definition. Resizing the main menu to cover the full width of the screen also aids in the spatial location of buttons for blind users since they only have to consider the vertical axis in order to locate a particular button with the mouse cursor.

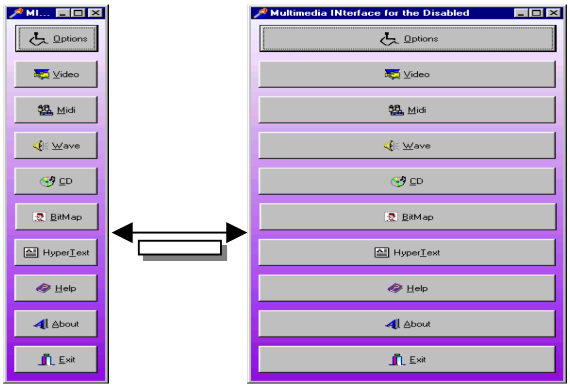


Figure 4. *The Resizable Main Menu Form*

Figure 5. *The Underlying Backdrop with Windows 95 Taskbar Removed*

To prevent accidental selection of applications outside of the MIND system by blind users, the Windows 95 Taskbar is automatically removed from view while the MIND system is running. The Main Menu Form is also placed upon a backdrop (see Figure 5) so as to eliminate the possibility of blind users selecting Windows 95 desktop items with the mouse cursor.

4.3 Selection of Features

All features may be selected in the normal manner by positioning the mouse cursor over the corresponding screen button and depressing the left mouse button. However, the buttons are also implemented as ‘audicons’, which are entities that represent a certain functionality by auditory as well as visual means. As the mouse cursor is moved over a button a digital sample of a voice describing that particular button’s operation is emitted. Alternatively, the options available within the main menu or tools may be located by means of ‘tabbed’ keyboard input coupled with explanatory auditory information.

This simple keyboard interaction has also been designed to assist users with co-ordination impairment since it eliminates the needs for physical dexterity and accuracy often associated with the use of a mouse. Additionally, blind users who do not feel completely confident with the mouse may also prefer this method of interaction. Future developments of the system may also consider ‘headset’ selection of features.

4.4 Navigation within MIND

Many of the navigational aspects of the MIND system have been determined by the needs of visually impaired and blind users. The MIND system follows a simple 2-tiered hierarchical design. This means that in order to navigate from any one form or multimedia representation to another, the user must return to the main menu. Additionally, when the user selects an option from the main menu the resultant form or window is displayed ‘modally’. This means that the current form must be closed before the user can reactivate the main menu. Although, this may appear to be somewhat tedious, it ensures that the user is prevented from becoming ‘lost’ within the system.

The selection of, or exit from, any of the forms available within the main menu is also accompanied by auditory information describing the users actions and the current status of the system. Additionally, to prevent accidental selection of applications outside of the MIND system, the Windows 95 Taskbar is automatically removed from view whilst the MIND system is running.

4.5 Consistency and Defaults

To provide a consistent interface, by default all the attributes pertaining to the buttons on the ‘Main Menu’ form are extended to every button throughout the MIND system. However, the ‘Options Menu’ that is currently being implemented, will provide methods whereby the attributes of each button within the interface can be adapted to suit the preferences of individual users.

4.6 The Video Form

One of the key multimedia features implemented to date is the ‘Video Form’. This form enables the synchronous display of two videos and a corresponding subtitle file (see Figure 6). The left video constitutes the ‘Main’ video, whilst that to the right may be an accompanying ‘Signing’ video, thereby translating the audio within the main video to sign language for comprehension by deaf users. Additionally, a subtitle file may be created using the Subtitle Authoring Tool and then displayed in parallel with the main video. The subtitles may also be displayed in different colours to indicate which individual within the main video is currently speaking or in order to depict background sound. The video form and its components are completely resizable to aid partially sighted persons, and if required the subtitling may be expanded to fill the entire lower half of the form (see Figure 7).

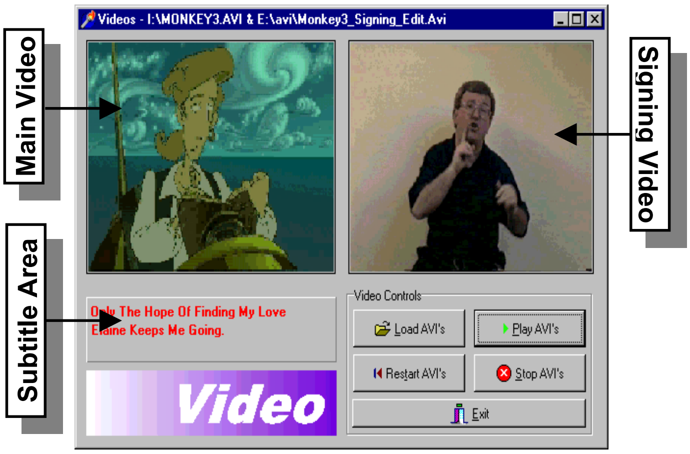


Figure 6. *The Video Form*

Figure 7. *Video Form with Enlarged Subtitle Area*

At present, the video form only supports media in AVI format but future developments will allow the presentation of MPEG encoded video. For demonstrative purposes, a suite of media was assembled. The main video was a short promotional video for a computer game. The second video was a British sign language (BSL) translation of the main video which was also complemented by an associated subtitle file.

The 'Subtitle Authoring Tool' (see Figure 8) provides an interface whereby subtitle text can be easily entered along with a time of activation and a particular colour. The use of colour within the subtitles helps identify the different characters speaking within a video and may be defined by assigning colours to an integer value which is then specified using the tool. For example the value '1' corresponds to the Colour 'Black' for the subtitles defined in Figure 8. The developer can browse through the list of subtitle records by selecting the 'next' and 'previous' buttons.

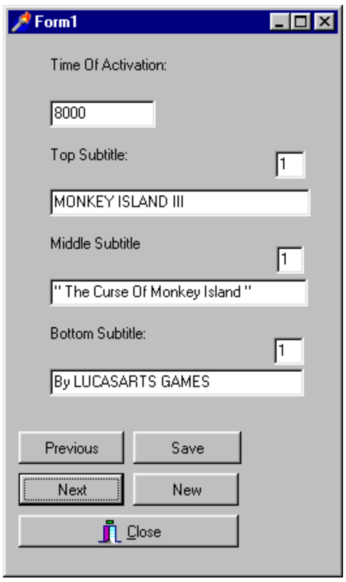


Figure 8. *The Subtitle Authoring Tool*

4.7 The CD Form

The Audio CD Database Tool has also been successfully implemented. This enables the provision of a CD form and associated interface from which audio CD tracks can be played in parallel with a corresponding video and subtitle file (see Figures 9 and 10). The CD form consists of: a main resizable window, a display area for subtitles, an area for AVI presentation and an area where current CD information can be displayed.

Figure 9. *The CD Form with an Associated Video and Subtitle File*

Figure 10. *The CD Form without an Associated Video and Subtitle File.*

All CDs that are registered with the MIND system by use of the CD Database Tool (see Figure 11) will automatically be recognised and relevant files loaded. The information that can be stored about each CD is currently:

- The title of the CD and a corresponding auditory description.
- The artist of the CD and a corresponding auditory description.
- The label of the publisher who released the CD and a corresponding auditory description.

Each individual CD audio track may also be associated with:

- Textual and auditory information describing the title of the track.
- A MIND subtitle file that contains textual information describing the contents of the track, or which is a 'dictation' of the voiced words within the CD track.
- A Video file, which may contain relevant information such as a signing translation or a music video.

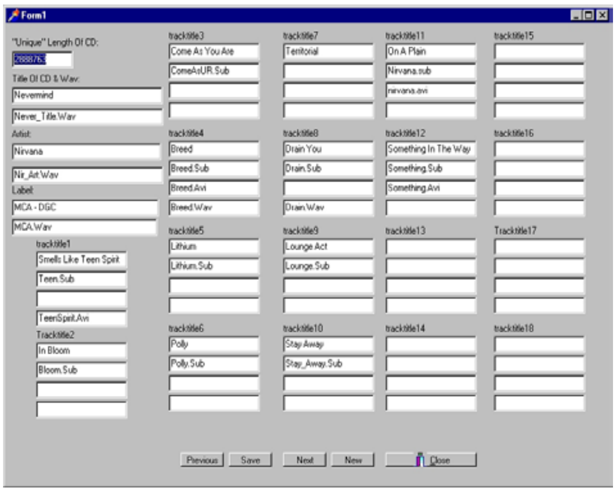


Figure 11. *The Audio CD Database Tool*

The 'Speak Info' audicon informs a blind user of the current track playing. If a WAV file has not been associated with the currently selected track then the system will read out the text by utilising the Text-to-Speech (TTS) engine. The MIND system utilises both Microsoft's Speech API (1998) and a set of shareware components collectively called DTalk (1998; Barbosa, 1998). The DTalk multiple component set consists of a Text-to-Speech component to enable talking applications and Grammar/Speech Recognition (SR) components to make applications listen and understand. The DTalk components interface to the Microsoft TTS and SR engines present in their Speech Development Kit (SDK) (1998).

5. WORK IN PROGRESS

Work within the MIND project is ongoing. The features currently being developed and incrementally incorporated into the MIND system are summarised in Table 3.

Table 3. *Features under development and incorporation within the MIND prototype*

Feature	Tools being Considered
The MIDI Form	A designated video area where information about the current MIDI tune playing can be displayed (e.g. display of a composer purveying the current tempo and auditory experience). A visual piano keyboard, whose keys change colour, indicating the current notes in play. A scrolling musical score, which highlights the notes currently in play. Textual information describing the style, tempo and key of the music and who it was composed by.
The WAVE form	A designated subtitle area where dictation of spoken word within the sound sample (if any) or a description of the actual sound can be displayed. Graphic display of the waveform output. Animated graphic equalisers showing intensity for various wavelength intervals at any specific time.
The BITMAP form	Each bitmap file is to be associated with auditory information describing the picture. The facility to enlarge an image will be provided to aid those persons with partial sight.
The HYPERTEXT form	A form which allows HTML files to be loaded and displayed as well as providing methods by which disabled users may interact and direct the information flow. Speech outputs to achieved by utilising Microsoft's text to speech engine.
The OPTIONS form	A form that enables the interface and its facilities to be further adapted to suit the user, for example by allowing default file locations to be specified or volume levels for specific devices to be altered.
OTHER tools	Using '***stereo sound'*** to provide a guidance mechanism to the spatial location of audicons. Providing '***hot keys***' to enable the user to be transported to a specific location within the interface. The use of Microsoft's ***Speech Recognition Engine*** to eliminate the use of mouse or keyboard input.

6. TESTING

Testing and user evaluation is essential for a system that relies extensively on the success of its interface. With this in mind feedback has been obtained from a profoundly deaf subject whose job role is to use video and signing communication strategies to help individuals who have recently lost their sense of hearing. Feedback has been extremely positive and the person in question is interested in using the software within their line of work. Other visually and aurally impaired users have also been consulted during the project. However, it is acknowledged that to date, due to the constraints of the project, testing of the system has been of a qualitative rather than quantitative nature. More extensive and formal evaluations of the MIND system, with a number of sensory impaired individuals, have been planned and will be conducted within the next couple of months. This form of testing, with large numbers of individuals and controlled experiments will provide more statistically significant results. However, in the meantime all qualitative feedback obtained will continue to impact on the development of future prototype versions.

7. SUMMARY

This paper has briefly introduced the work associated with the MIND project. Substantial progress has been made in a very short space of time and implementation is ongoing. A demonstrable prototype, successfully incorporating the features described in Section 4 is available, and is stimulating discussion and feedback into the needs and expectation of users with sensory impairments. It is expected that the results and ideas stemming from this project, which is being conducted with very limited resources, will act as a pump-priming activity for a more detailed and funded study into ways to adapt multimedia software products to meet the needs of sensory impaired users. The sensory impairments considered within the MIND project, the possible physical & psychological limitations of individuals possessing these impairments and the compensatory methods provided by the MIND system prototype are summarised in Table 4.

Table 4. *Summary of sensory impairments, possible limitations and compensatory methods*

Sensory Impairment	Possible Physical & Psychological Limitations	Compensatory Computing Methods Implemented Within MIND
Blind	Unable to see the VDU.	The provision of audicons, sound guidance mechanisms, TTS & SR. Removal of Windows 95 taskbar.
	Complete screen contents may have to be stored in memory for fluent interface navigation.	The provision of a 2-tiered hierarchical design. The 'modal' presentation of forms focusing attention. Keeping all form contents to the bare minimum without compromising on functionality.
	May feel unconfident using the mouse.	The provision of audicons that cover the whole vertical screen definition. The facility to utilise a small portion of the keyboard as an input device.
Partial Sight	May be only able to see large objects present on the VDU.	The ability to scale the screen contents.
	May be only able to focus on one particular area of the screen at any particular time.	Audicons and controls are located in close proximity to one another in 'clusters'. Forms are represented and displayed 'modally', eliminating the need to see the whole interface and subordinate child forms at the same time.
Deaf	Unable to hear sound emitted from PC soundcard.	Provision of compensatory methods such as subtitle presentation.
Hearing Impaired	May be unable to hear certain frequency ranges.	No solution at present. Real-time processing of digital sound samples could be performed to evade the frequency range the sensory impaired individual cannot perceive.
	May only have monaural hearing with consequent loss of stereo sound perception.	No solution at present. Real time processing could convert stereo sound to mono. More simply, the line-out wires from the soundcard could be fused to produce mono sound.
Co-ordination Difficulties	May be unable to utilise the mouse. Therefore such individuals may be unable to accurately locate and select controls within a two dimensional screen definition.	The provision of audicons. The ability to scale audicons to cover the complete horizontal and vertical screen definition. The ability to use the keyboard as an input device.
	May be unable to utilise more than a few control keys.	The MIND system only requires two control keys; a navigation and select key. Speech Recognition.

8. REFERENCES

A Sloane (1996), Multimedia Communication, McGraw Hill, London.

Alva Technologies Netherlands (1998), Home Page, *http://www.alva-bv.nl/.*

C Barbosa (1998), DWave Delphi Component, *http://www.digiserve.com/cmab/.*

Borland International (1997), *Borland Delphi 3 Object Pascal Language Guide*, Borland International.

J Connel, L Shafer (1995), *Object-Oriented Prototyping*, Yourdon Press.

Delphi Super Page (1998), Components For Delphi 1, 2 & 3, *http://www.delphideli.com.*

Dolphin Computer Access (1998), Home Page, *http://www.dolphinaccess.com*

Dtalk, Speech Components For Delphi, *http://www.o2a.com/*

W K Edwards, E D.Mynatt, K Stockton (1995), Access to Graphical Interfaces for Blind Users, *Interactions,* pp. 54-67, January.

M W Eysenck, M T Keane, *Cognitive Psychology*, Psychology Press (3rd Edition).

E B Goldstein, *Sensation & Perception*, Brooks/Cole Publishing (3rd Edition).

H O Istance, C Spinner, P A Howarth (1996), Providing Motor Impaired Users with Access to Standard Graphical User Interface (GUI) Software via Eye-Based Interaction, *Proceedings 1st European Conference on Disability, Virtual Reality & Associated Technologies*, pp. 109-116.

S Jarol, D Haygood, C D Coppola (1996), *Delphi 2 Multimedia Adventure Set*, The Coriolis Group.

M Lumbreras, M Barcia, J Sanchez (1996), A 3D Sound Hypermedial System for the Blind, *Proceedings 1st European Conference on Disability, Virtual Reality & Associated Technologies*, pp 187-191.

C McKnight, A Dillon, J Richardson (1993), *Hypertext: A Psychological Perspective*, Ellis Horwood.

Microsoft, Speech Development Kit, *http://research.microsoft.com/research/srg/*

MIDAS, Sound System For Windows, *http://hornet.org/*

L A Mothe (1993), *Games Programming In 21 Days*, Sams Publishing.

D Osier, S Grobman, S Batson (1997), *Teach Yourself Delphi 3 In 14 Days*, SAMS Publishing.

Jenny Preece et al, *Human Computer Interaction*, Addison-Wesley (1994)

M Rothberg, T Wlodkowski(1998), Multimedia: Making it Accessible to Blind Users, *http://www.dinf.org/csun_98/csun98_04.htm*

I Sommerville (1995), *Software Engineering*, Addison-Wesley (5th Edition).

D M Tanner (1998), Designing Accessible Computer Systems for the Blind Without Expensive Adaptive Hardware, *http://www.dinf.org/csubn_98/csun98_109.htm/*

Design of a non-contact head-control mouse emulator for use by a quadriplegic operator

E D Coyle, M Farrell, R White and B Stewart

School of Control Systems and Electrical Engineering, Dublin Institute of Technology,
Kevin Street, Dublin 8, IRELAND

ecoyle@dit.ie

ABSTRACT

A non-invasive scanning mechanism has been designed which is microprocessor controlled to locate and follow head movement in a defined zone about a persons head. The resulting head movements may then be mapped onto a computer screen and can further be used to control the cursor, thus replacing the necessity of using a standard PC mouse. To demonstrate the concept, a Graphic User Interface (GUI) of a push button telephone has been designed, and a procedure is outlined below by which one may select and highlight a digit or group of digits, and in turn dial the chosen number.

1. INTRODUCTION

Applied research is ongoing at the Department of Control Systems and Electrical Engineering, focusing on development of non-invasive computer input/output mechanisms. The intention is that the devices will be of benefit to everyday PC users, but in particular the aim is to create innovative devices which will enable quadriplegic or other persons with upper limb disability operate a computer.

Research to date has centred on a number of specific areas of application. The original application focused on developing a *non-contact head-movement electric wheelchair control device.* The objective was to design an interface which could replicate and replace a standard joystick wheelchair controller. Ultrasonic transducers positioned on a specially designed frame attaching to the wheelchair were microprocessor controlled to transmit and receive pulses which resulted in the wheelchair travelling in the desired direction of the occupant. This work was carried out by the principal researcher in part fulfilment of studies leading to his PhD degree (Coyle, 1993). The concept of control by this method was developed following an appraisal of related work by researchers at the Palo Alto Veterans Administration Medical Centre (Jaffe, 1983).

A second ongoing application is that of design of a *non-contact PC virtual keyboard emulator*, with the intention of enabling a person with upper limb disability work in a Microsoft Windows environment by simply moving the head in a defined area about the central relaxed head position. The virtual keyboard will take the form of a graphic display on the computer screen allowing the user to type without the requirement of a standard keyboard. A scholarship has been awarded by the Irish Health Research Board in support of this project.

A number of design routes have been investigated with a view to optimising on both hardware and software design. In the current system two sets of piezoelectric ultrasonic transmit/receive transducers are strategically placed on a mounting frame to the side and rear of the user's head. The transmit transducers are microprocessor controlled to cyclically transmit pulsed signals, which upon reflection from the surface of the head are picked up by the matching receive transducers. The microprocessor users a trigonometric algorithm to determine the current head location within a mapped zone about the central relaxed head position. Head movement is used as an alternative input device, replicating keyboard typing and hand-mouse action. The microcontroller is programmed for both the monitoring of head movement and as a PC Serial Mouse emulator.

This paper will outline the development of a prototype system in which a Graphic User Interface of a push button telephone has been designed (using Microsoft C in a DOS environment). This application was developed to enable critical assessment of the proposed ultrasonic head-control mechanism. Further work has resulted in development of a Windows based GUI virtual keyboard (C++), while in yet a separate project an approach has been taken in which the appropriate signals in response to ultrasonic information are inputted directly to a hardware modified mouse. A brief on these latter two applications will be given in Section 4, further project developments.

2. ULTRASONIC HEAD LOCATOR

2.1 Overview

Upon energisation with an appropriate electronic gating circuit, an ultrasonic transmit transducer emits inaudible sound waves which propagate through the air until they strike an object. An attenuated portion of the transmitted signal is then returned and picked up by an adjacent receive transducer. By application of Pulse Mode transmission the time taken for a "burst" of pulses to travel from transmitter to receiver is determined, providing a measure of distance travelled. The velocity of sound in a homogenous medium is finite (331 m/s in air) and the delay is proportional to the distance travelled by the signal.

The ultrasonic head-control unit used comprises ultrasonic transmit/receive transducers positioned on a semi-circular tubular steel headset. Two sets of *transmit* and *receive* transducers are utilised, located at angles of approximately 45 degrees in the plane to the rear of the operator's head. Head position may be estimated by obtaining the time delay for each set of transducers consecutively and applying a trigonometric co-ordinate calculation. Measurements are made relative to a central position (axial point (0,0)), located mid way between the left and right side transmitter pairs (figure 1). The system is designed that position and height of each transducer pair and of the supporting frame is possible, thus enabling best fit to meet individual user requirement. Inexpensive piezoelectric transmit and receive transducers with resonant frequency 40 KHz have been utilised.

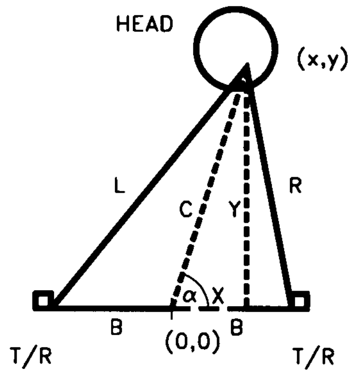


Figure 1. *Arial view depicting head position relative to fixed left and right transducer sets*

2.2 Distance measurement

Prototype transmit and receive circuits have been designed and tests performed to determine the relationship between distance travelled by the ultrasonic signal between transmitter and receiver and the time delay calculated by the microprocessor. Table 1 summaries these tests. The value of speed of ultrasound used to calculate distance travelled was 331 m/s.

Table 1. *Distance-Time measurements of echo ultrasound signal*

Distance between transmitter and receiver (cm)	Time delay to receipt of echo signal (ms)	Calculated distance between transmitter and receiver (cm)
62	1.95	64.55
57	1.80	59.58
52	1.65	54.60
47	1.50	49.60
42	1.35	44.69
37	1.20	39.72
32	1.05	34.76
27	0.90	29.79
22	0.74	24.49
17	0.59	19.53
12	0.45	14.90
9.5	0.38	12.58
7	0.31	10.26
4.5	0.22	7.20

Upon plotting a graph of distance travelled versus time delay, based on the results listed in table 1, a linear relationship is obtained.

2.3 System Control

A Motorola 68HC11 microprocessor has been adapted as the system control unit. Two control signals are required from the microprocessor to enable generation of the transmission signals, one for each of the left and right transducer sets, and the transmission is to occur periodically so that the position of the head may be regularly updated. This has been achieved by utilising the microprocessors output Compares (OC) facility . The timing system of the microprocessor is set at 2 MHz, corresponding to a cycle time of 500 ns (clock increment time). OC2 is used as the control signal for the left transducer set and OC3 for the right. OC2 and OC3 occur for set values of the free running clock, setting respective pins high. A further output compare OC1 has been configured to further clear OC2 and OC3. The returning ultrasonic echo signal causes a rising edge on either Input Capture IC2 or IC3, resulting in generation of appropriate interrupts. A block diagram representation of the system is shown in figure 2.

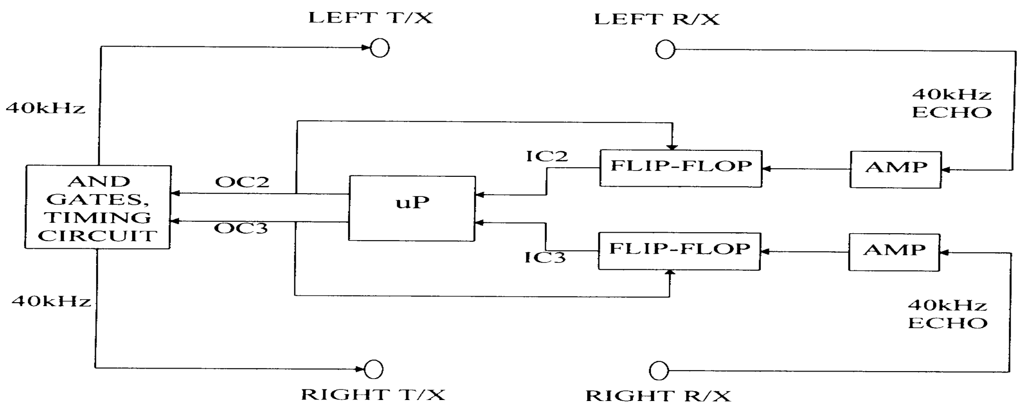


Figure 2. *Microprocessor control of ultrasonic transmission/reception process*

2.4 Averaging

The microprocessor samples head position once for every complete cycle of the free-running timer, corresponding to FFFFhex increments at 500 ns for each increment. This will result in new values for head position at 32 ms intervals. To help reduce 'jitter' associated with slight changes in head position should the user choose to remain still, the timer values obtained have been averaged in groups of 16. The averaged value is then made available for transmission to the PC.

2.5 Co-ordinate Mapping of Head Position

The two ultrasonic delay time values obtained from IC routines are processed to enable X-Y co-ordinate values be determined, providing information on user head location. The two ultrasonic distance values **L** and **R** and the distance between the two transducer pairs, 2**B**, are required. By application of the Cosine Rule the X and Y co-ordinate values may be determined. The origin reference point, (0,0), is located mid-way between the two transducer pairs. Taking **C** and **α** to represent the co-ordinates in polar-form,

$$x = C*Cos(\alpha), \qquad y = C*Sin(\alpha),$$

Cosine Rule: $$a^2 = b^2 + c^2 - 2bcCos(\alpha)$$

hence, $$L^2 = B^2 + C^2 - 2BCCos(180 - \alpha)$$

$$R^2 = B^2 + C^2 - 2BCCos(\alpha)$$

$$L^2 - R^2 = 4BCCos(\alpha)$$

$$\Rightarrow x = (L^2 - R^2)/4B.$$

The height of the triangle, i.e. the Y ordinate is found by application of Pythagoras Theorem,

if R < L: $$y = \sqrt{[L^2 - (B + X)^2]}$$

if R > L: $$y = \sqrt{[R^2 - (B - X)^2]}$$

3. TELEPHONE INTERFACE

In the current application a system has been designed which enables a disabled operator make a telephone call. This is achieved by observing a Graphic User Interface of a telephone pad on the PC screen and by making head movements which allows a number be selected and dialled.

A block diagram of the component elements making up the system is shown in figure 3. The ultrasonic transducers have been mounted on a semi-circular tubular headset, with the user head acting as a non-invasive movement control mechanism.

3.1 Graphic User Interface

A GUI is provided on the PC screen which takes the form of a telephone keypad. Upon activation of the GUI, positional movements of the head will result in placement of the cursor over a desired number. Once this number has been highlighted, a movement to the **S**elect button will result in a highlighted digit being placed in a telephone number array. A complete telephone number can be built up in this fashion and the number dialled by choosing the **D**ial function. A **De**lete button was also incorporated to allow the user remove digits from the array. In addition to these keys a number of function buttons were also provided, giving redial and directory facilities to the user. The directory facility enables the user 'fast dial' a number from a personalised directory, hence minimising upon required number of head movements in achieving a successful dial. **H**elp and **999** buttons provide the user with a quick-dial facility in emergency situations (figure 4).

3.2 Adapted Telephone

The internal configuration of a push-button telephone, the principal of operation of which is based upon the Bell Laboratories Dual Tone Multi-Frequency (DTMF) design concept, was analysed, and an imitation telephone was designed, suitable for usage in demonstrating the project concept. "Dual Tone" infers the addition of two individual tones, resulting in a single tone. The adapted telephone has an attached output speaker to enable one hear the dialled tones. The adapted telephone contains a small speaker, thus enabling one hear the generated tones.

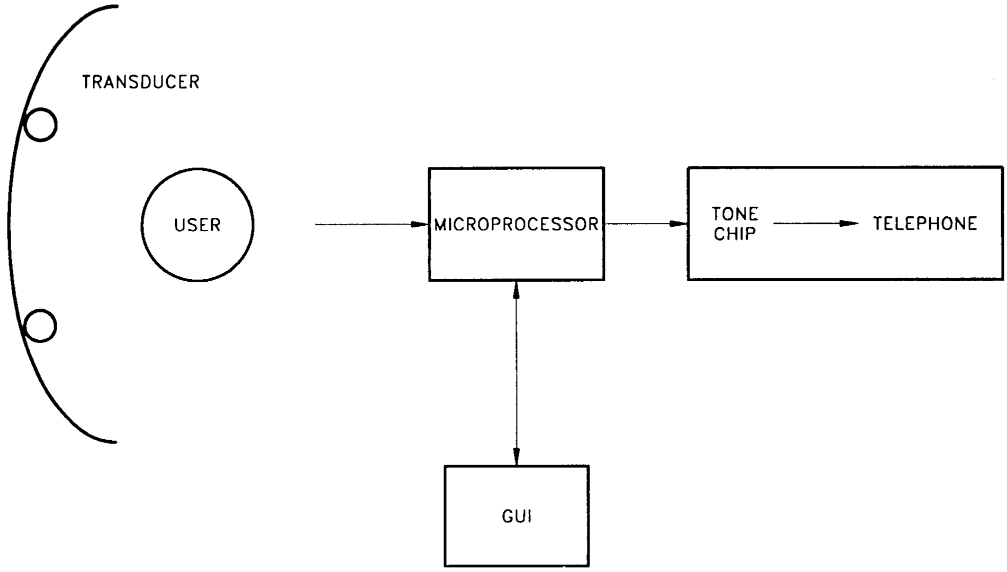


Figure 3. *Block diagram of the telephone interface*

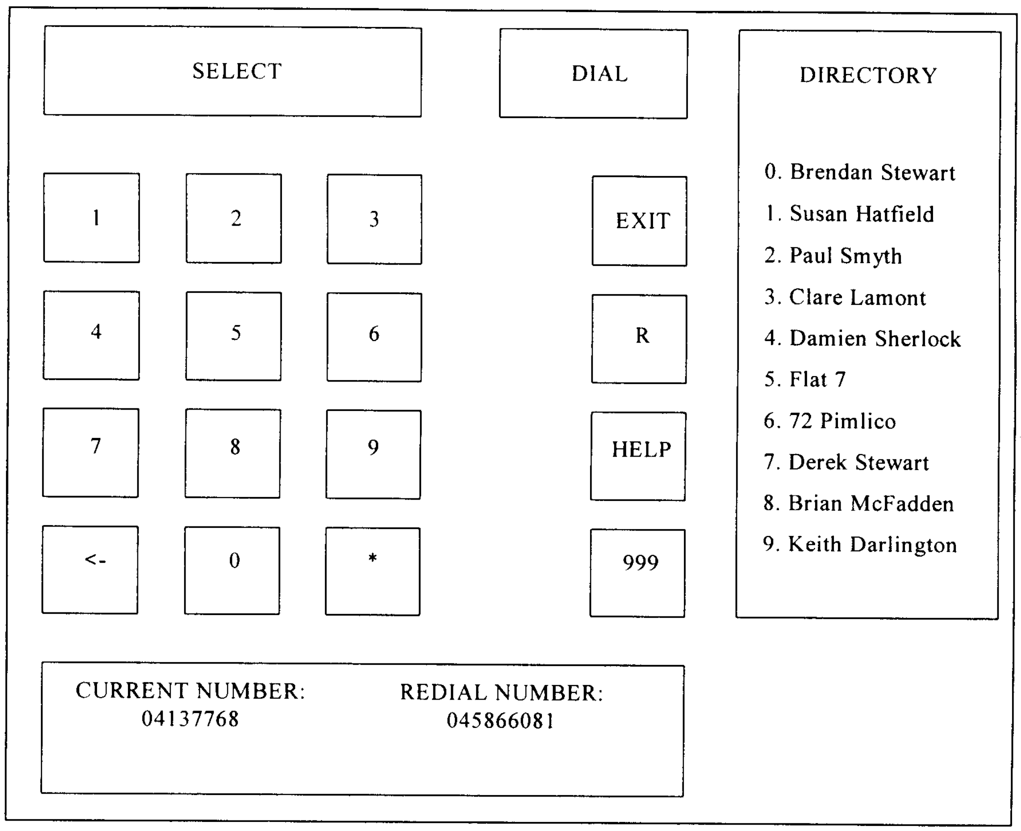


Figure 4. *Sample user interface*

3.3 System Testing

Various programmes have been designed, one of which provides the facility of cursor movement in all directions, as when using a standard PC mouse. However, for the purposes of providing a demonstration of the system, an x and y axial plane was found to be most suitable. The GUI on display was designed in a DOS environment using C language programming.

The prototype system as developed is user friendly and is relatively easy to use. Upon experimenting with the system it is a requirement that the user have reasonably good head movement control, otherwise successful operation of the system will prove difficult. The majority of people who have been invited to test the system have successfully managed to move and locate the cursor at all destination locations and to dial a desired number.

3.4 Serial Communication

The PC and microprocessor are arranged in a master-slave configuration. The PC decides when communication should take place by sending a byte of data to the microprocessor and causing a serial interrupt. The value of the byte determines the service required by the PC. There are three alternatives for the service required

1. The PC requires new time values from the µP to update the screen cursor
2. The PC wishes the µP to store a digit of the phone number
3. The PC wishes the µP to dial the number currently stored.

4. FURTHER PROJECT DEVELOPMENTS

4.1 Microsoft Windows Virtual Keyboard GUI

A further project is underway in development of an ultrasonically controlled system with the intention of enabling a user with upper limb disability interact fully in a Microsoft Windows environment. To enable the user enter text a GUI has been designed to emulate a standard keyboard. The *virtual keyboard*, taking the form of a graphic keyboard on the computer screen, is designed that a selected key results in text entry to a desired Windows application. The *virtual* GUI keyboard has been developed using Windows 'Hooks' to intercept the Windows message system. Although developed to operate by signal input command from the ultrasonic control unit, the virtual graphical keyboard has been designed as an independent entity. It is conceivable therefore that it may be applied in other areas of Medical Rehabilitation employing Personal Computers using Microsoft Windows.

To make the virtual keyboard easier to view a magnified view of each key on the keyboard has been designed into the system, the current key being highlighted in tandem with cursor mouse movement across the screen. It was important that the virtual keyboard be sufficiently large on the computer screen to be easily viewable but not too large that no other applications could be viewed at the same time. Using the magnified views of the keys reduces the size of the keyboard while maintaining the visibility of each key.

A further design feature includes a facility of updating the more commonly used words by the operator. A pop-up menu as designed to allow the user select a previously typed word with minimum mouse movement. A sample screen is shown in figure 5.

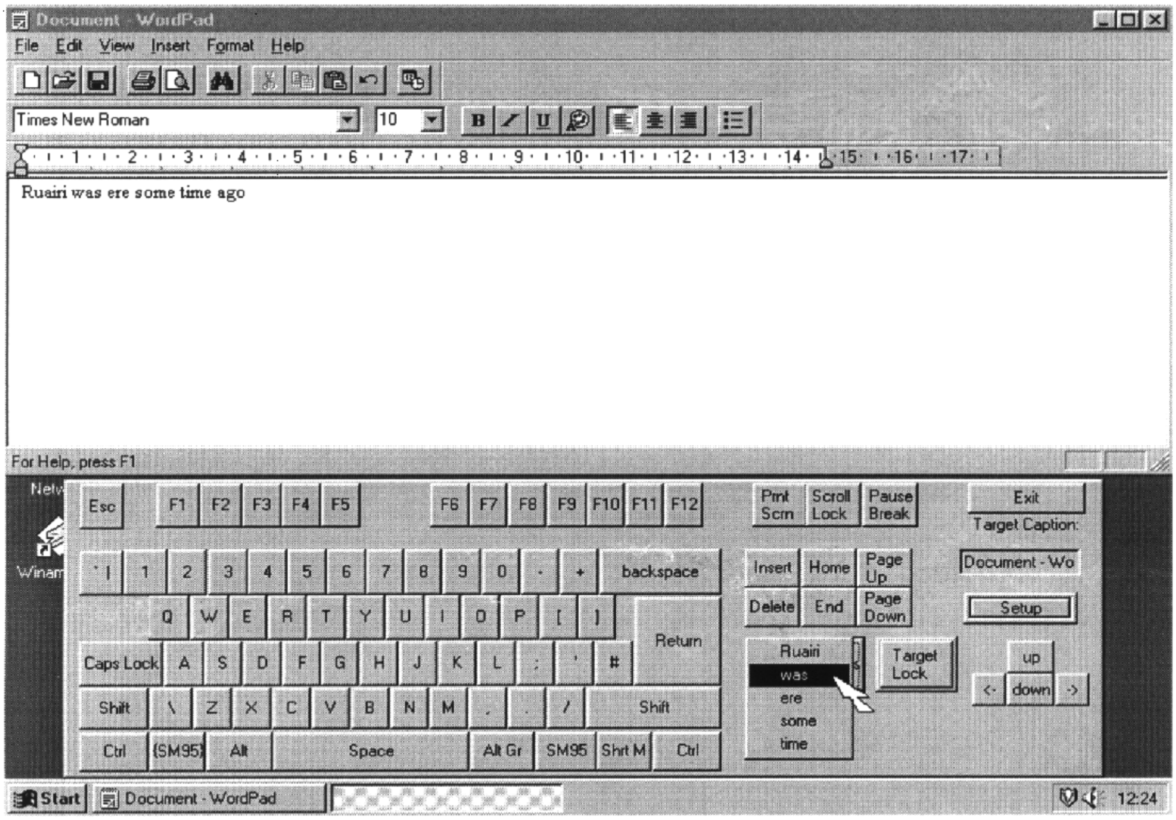


Figure 5. *Sample virtual keyboard with demonstration of 'quick selection' feature*

4.2 Hardware Mouse Modification

A second project has been initiated in which an investigation is underway into design and implementation of a hardware mouse interface. Initial results from this work are proving promising. The internal operation of a mouse may be appreciated by inspection of the diagram shown in figure 6. The push buttons connect to micro switches. This means that if they are pressed and closed a voltage level is supplied to an input pin of the mouse IC. The moving sensor system has mechanical and optical parts. The ball rotates due to the friction between the ball and mouse pad. This rotation is transmitted to two rollers, one for the x and y axes. Each roller is connected to one end of a shaft. The other end consists of a small disc with a small circle of holes. Roller rotation determines up/down and left/right motion.

A study has been made of the digital signals resulting from roller movement and based upon observation circuitry has been designed to enable replication and replacement of the sensor input. A MC68HC11 microprocessor together with the ultrasonic set have been configured to replace the direction sensors and ball. Development is at an early stage, however a prototype system has been developed and has been tested (not as yet by a person with upper limb disability). Some initial difficulties have been encountered and are to receive will further design consideration. Nevertheless the initial objective has been achieved and this presents a way forward which will require less software development than in the earlier described projects.

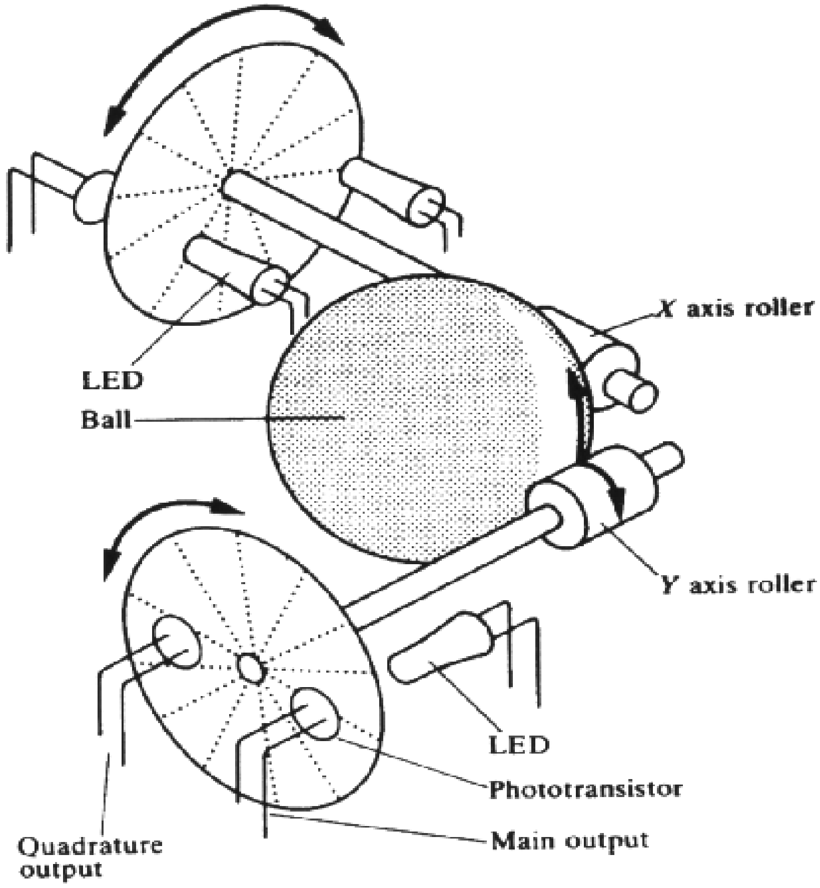


Figure 6. *The mechanical system of a mouse*

5. CONCLUSIONS

Experience has been gained in development of circuitry which has enabled ultrasonic transducers be used as an alternative non-invasive PC input device. Much effort has been made in development of a reliable ultrasonic instrumentation measuring device. A printed circuit board design incorporating a microcontroller, transmit, receive and associated circuitry components, has been manufactured.

The intention is that the measurement device may be used as the basis of and in association with one or more applications which may be of assistance to people with upper limb disability.

The primary prototype application described in section 3 of this paper, *non-invasive head control of a GUI telephone PC interface*, although software designed in a DOS environment, has been retained as an important show-piece of the developed research. The GUI screen as developed is non cluttered and pleasing to the eye and upon experimentation the user quickly gains confidence in controlling cursor movement by head movement alone. A dial tone is heard upon successful movement of the cursor to the dial button location.

Further research has been conducted in development of a small select word vocabulary *Speech Recognition* system (Newsome, 1995) and in *video Input Capture of the Eye* (Bourke, 1998), and it is intended that these additional alternative device input strategies be considered in parallel with non-invasive ultrasonic head/limb movement monitoring. It is envisaged that while results culminating from the three individual research branches will have useful application as stand alone entities, aspects from each may also be incorporated into a single system, designed to accommodate particular individual requirement.

6. REFERENCES

E D Coyle (1993), PhD Thesis - “Control. Energy Efficiency and Mechanical Redesign of Electric Wheelchairs”, National University of Ireland.

E D Coyle, (Jan 1994), journal article, The Engineers Journal of the Institution of Engineers of Ireland, “Innovative Design of Electric Wheelchairs”, vol 46, no 10, pp 49-51.

D L Jaffe, (1983), “Technical manual for Ultrasonic Head-controlled Wheelchair”, Palo Alto Veteran’s Administration Medical Centre, California, USA.

K Newsome, E D Coyle, W Brien, (1995), “ Comparison of Zero Crossing Rate and Higher Order Crossing Techniques for Improving Discrimination in Speech Signals”, Irish Colloquium in DSP and Control, pp 255-262, Queen’s University Belfast.

P Bourke, E D Coyle, (1998), “Computer Applications Driven by Eye Movement”, Irish Signals and Systems Conference ISSC’98, pp 203-209, Dublin Institute of Technology.

Virtual keyboard with scanning and augmented by prediction

Peter E Jones

Electrical and Electronic Engineering, The University of Western Australia,
Nedlands WA 6907, Western AUSTRALIA

peterj@ee.uwa.edu.au

ABSTRACT

All our teenage users are confined to electric wheelchairs and are unable to speak or make any voluntary movements much beyond either moving their head against one of three switches mounted in the chair's headrest or to hit a large "banger" switch. Real-world devices are beyond reach, only devices in a virtual world are attainable. This virtual keyboard project was designed to meet their needs for interacting with commercial off-the-shelf software such as word processors, spreadsheets, electronic mail and Internet tools. The virtual keyboard uses scanning augmented by character prediction and word completion.

1. INTRODUCTION

Children who have disabilities have reduced opportunities in all areas, more especially in education and in social settings (Alm et al 1992). The group we are working with are physically dependent and cannot control their environment. Often this can cause a problem referred to as "learned helplessness." Virtual Reality (Singh et al 1996) offers the user who, because of their impairment, cannot control or operate in the real world but can do so in a virtual one. Ellis (1995) defines virtualisation as the process by which a human viewer interprets a patterned sensory impression to represent an extended object in an environment other than that in which it physically exists. In his most abstract version it requires users to interpret an image on a flat screen with many of the usual cues missing as representing a constructed object. While this project does not provide an immersive Virtual Reality, it does provide a virtual device with which the children can interact. For them, it is a virtual world where they can type on a keyboard. Having the ability to produce text opens further doors in communication and information access without the need for one-on-one intervention from a teacher or therapist.

2. PROJECT'S AIMS

From our observations (Jones, 1997) of the users with our earlier projects we realised that activating one switch was very time consuming and error prone. To interact with a word processor effectively we would need to find some means of speeding up the interaction. We are not alone in tackling this problem, for example the Reactive Keyboard (Darragh & Witten, 1992), work at the University of Dundee in the UK (Hine et al, 1994) and a number of papers in Edwards (1995) provide solutions. Commercial systems exist as well, for example SofType (Origin Instruments, 1998). However, they don't address issues beyond those concerned with prediction or are not adaptable enough or cannot meet our users' needs.

Our prototype, developed in Microsoft Visual Basic, is shown in Figure 1. If this were to be shown in animation and colour, then you would notice the moving cursor down the left-hand column indicating the current scan row. The vowels are shown in red, the numbers in a purple colour and special keys shown in blue. Other characters are shown in black. At the end of each row is our innovative "go back" symbol (shown as a "<") which, if selected, repeats the scan of that row.

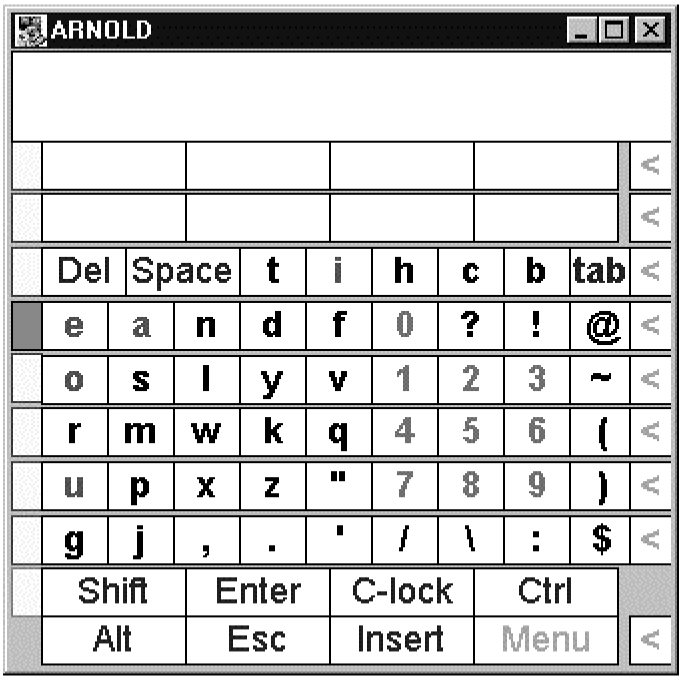


Figure 1. *Virtual Keyboard Window*

Even with our small number of users we were surprised to find significant variations in their abilities. Not only that, but there were changes in a single user throughout the day, for example through boredom or fatigue. The virtual keyboard would need to be able to be readily adaptable to these differences. Another variable is the application itself. The layout would need to be changed when moving from say a word processor to a spreadsheet. With limited funds it had to be kept to a low cost, as many commercial offerings are expensive. To this end we decided on a virtual keyboard implemented in software with a very simple electronic interface to whatever single switch the user could operate. For example, one might use a head switch while another would use a suck-blow tube. As the users could not control their eye-gaze, alternatives are needed for feedback to confirm when choices are made. Similarly, when the child is with a therapist or teacher, this helper also needs to be provided with some feedback. For example, to confirm that the switch was indeed pressed so that they realise that the child has indeed made contact and can offer appropriate encouragement. For this feedback we chose to have a light-emitting diode on the switch box as feedback to the helper. We also chose to use sound from the computer and to flash the appropriate area in the window. Finally, it showed the last input in addition to building up the complete word.

3. INPUT METHODS

The commonly used methods for input selection in augmentative assistive devices are direct selection, encoding and scanning. For a keyboard, direct selection is much the same as the ordinary method of pressing a key. Someone with a minor disability can employ a keyguard in addition to software changes to compensate for poor motor control or single-handed operation. A keyguard is a template that fits over a conventional keyboard with openings above the key tops. The user can then place their hand(s) on the template and push a finger through one of the holes to activate the key. The finger resting in the hole then avoids minor errors from tremors. Suitable software can then ignore errors such as brief key presses or repeated activations within a short time frame or allow the use of a “sticky” meta-key. For example, in one-handed operation, the shift key is first pressed and the software interprets the next key press as if the shift key was still held down. The template also assists the user in aiming at the desired key by providing protection against erroneous positioning. Although this is a useful input method for many users, it is beyond the motor skills of our group of users. They are only able to operate a single switch and a real keyboard is too complex and the key tops too small.

Another method is one of encoding. Here the user makes a sequence of actions with different sequences providing the coded input. An example of this is the equivalent of the radio transmissions using Morse Code. Morse code is produced as a sequence of short transmissions (dots), longer transmissions (dashes) and a pause to indicate intersymbol spacing. For English, the Morse Code has additionally used short sequences for the most frequently occurring letters. For the user with a disability, a similar coding scheme can be used. It is often used with a sip switch. The users have what looks like a straw. To produce an input, the user sips on the tube activating the sensor. Short and long sips make the two forms of input with pauses breaking up the sequences. This offers a number of advantages to someone with motor impairment. Once a user has learnt the code they no longer need to concentrate on the input device. It is a compact interface and does not need attention to be distracted from the task to the actual input device. It does take training before a user becomes competent. For first time users it is slow and unreliable as it totally depends on the user internalising the meaning of the sequences. Of course a visible representation of the codes can be used during the learning phase. Again, the level of control is beyond our users.

A scanning keyboard consists of an array of characters just like an ordinary keyboard. The actual device could be a real one or, as in our case, a virtual one. In a conventional keyboard any key may be selected in much the same time, i.e. it has random access. The scanning keyboard differs in both layout and method of selecting which key to activate.

Scanning is a sequential access system. That is, each key in turn has the potential to be selected. This focus moves from key to key until the required one is reached. The user then activates their switch and the character at that position is sent to the application or in some cases stored for later transmission to an application. Imagine a linear line of the 26 lowercase letters and a dwell time for each letter to allow the user to realise that this is the required letter and then a further time to activate their switch. This is clearly too slow to be acceptable. One way of speeding up the selection process is to move away from a linear array of choices, to one of grouping. Each group in turn is focussed upon, and on selecting that group the scan is then among the keys in that group. For those who can hold a switch down, some form of dual scanning is possible, though not for our particular users. Further improvements can then be made by predicting what the next most likely set of characters are and offering that always in the first group. Additionally, word prediction can be used and even the prediction of complete phrases such as "Dear Sir/Madam." This can also be applied to the use of symbols, for example the Compic (Australian) or Bliss (Canadian) set. A common way of displaying this option is as a series of rows. The system scans down the rows and on reaching the appropriate row the user activates their switch. The system then scans along the row until the desired character is reached and the user once more clicks their switch. By placing the more frequently used letters earlier in the scan sequence the overall performance can be further improved. Finally, the redundancy of natural language allows lengthy scans to be avoided if, at each selection, a set of predictions is offered based on the history.

Scanning keyboards have the advantage of being visible and use recognition rather than recall. This is less cognitive effort when compared to the systems using some form of coding. It offers those users who can operate only a single switch the simplicity of operation. The disadvantage is the waiting time for the scan to reach the desired letter. As we have also observed, it is very frustrating to have just missed your selection and have to wait for a complete repeat of the whole scan. Some users with more motor control than our users can have the means to change the direction of scan to overcome this problem. Another possibility, again requiring better motor control, is to be able to hold down a switch to mean "cycle through the groups", while a release would initiate the scan of that row. We feel that our solution, described below, offers almost as good a capability without imposing additional dexterity requirements on the users.

4. DEVELOPMENT OF THE VIRTUAL KEYBOARD

Our users were all comfortable with the concept of the scanning selection. We have used this before in some of our earlier projects at the school. Alternatives, such as direct selection or encoding, were not suitable as they require higher levels of control than most of our users have and also these other methods have longer training times.

Our particular users are unlikely to be familiar with the standard QWERTY layout for keyboards. This will not be the case for their therapists and teachers or for people who have acquired a motor control impairment later in life. However, it is the children who will be doing the virtual typing and we considered it a design advantage not to be constrained by any particular layout. For the able-bodied, alternative layouts to the conventional one of QWERTY have not shown overwhelming advantage either in ergonomics or speed. The alphabetic layout (see Figure 2), although intuitively appealing has also not shown to perform any better than the QWERTY layout. We decided that as we had a virtual keyboard it could be completely soft in the sense that it could be readily changed according to both the users and its use.

The system uses a separate template file for the layout of the keyboard. This uses a simple syntax and is easy to change with any text editor. We have provided different layouts for word processors and spreadsheets as well as an alphabetic layout in case that should be needed. Our template for spreadsheets (see Figure 3) emphasises numeric input and cell navigation. To accommodate the use of symbols such as Compic or Bliss would not be difficult, and prediction could be still be used. As you can see from the figure, the layout is not a conventional one. This will cause some minor problems with able-bodied assistants who are familiar with more conventional layouts. For our users, who have no prior experience and are never likely to need to use a standard layout, the speed advantage with a layout optimised for English more than compensates for the unusual layout and only needs a short initial learning time.

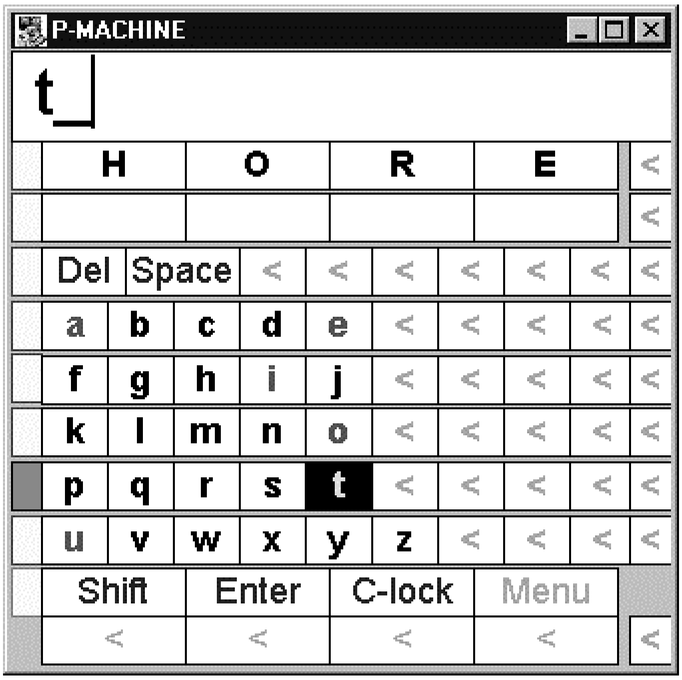


Figure 2. *Alphabetic Template Virtual Keyboard Window*

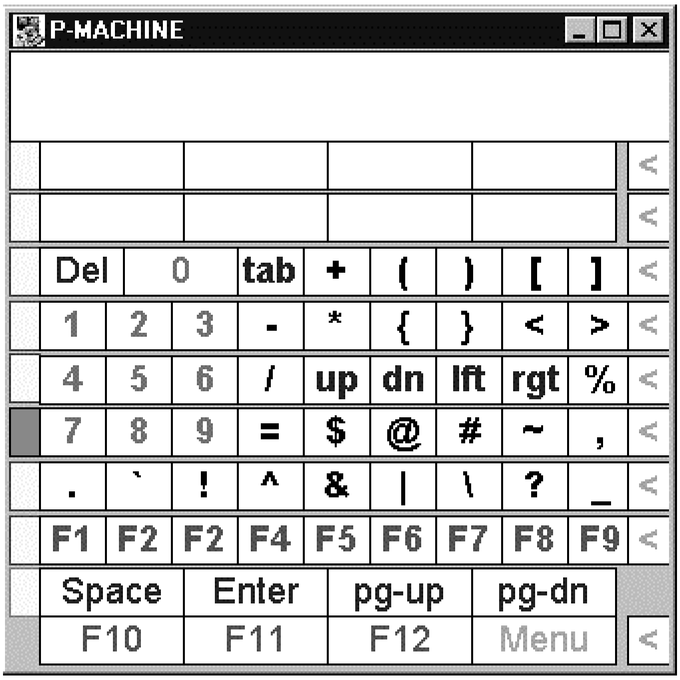


Figure 3. *Numeric Template Virtual Keyboard Window*

For the predictive keyboard we decided to provide both character and word completion. Examining the virtual keyboard (see Figure 4, with a “t” selected), the top part of the window shows (in large letters) what has been typed so far. The next row down is the character prediction row. From our analysis of English texts the system is predicting the four most likely following letters. The row below that contains four predictions to complete the current word. The left-hand column is used to highlight row selections. The layout shown is for word processing and is organised according to the expected letter frequencies found in English. Keys are colour coded to aid searching. The last row shows other keys, for example, selecting “Shift” actually changes all displayed characters to their shifted form. This is an advantage of a soft keyboard in that the key tops actually show what character will be entered when selected. We have found that the use of CAPITAL letters on conventional keyboards is very confusing for the younger child, when in fact it is only if the SHIFT key is held down do you get the capital letter. Finally, the last row has a range of keys that users may need, for instance you can send alt-sequences that can manipulate operating system functions without the need to go through ordinary menu selections.

Figure 4. *Letter “t” Chosen, with Frequency Prediction*

“Menu” allows for configuration changes (see Figure 5). Once selected the prediction rows contain the options, which may be selected in the same way as normal characters. For instance the figure shows the result of selecting “Menu”. The top row is then a set of options. The first is for the frequency ordered layout suitable for word processor applications or Internet tools such as email or URL entry. The next option is to select the alphabetic layout and the final option is aimed at a spreadsheet application. The next row allows other parameters to be altered. Each child will require different settings such as speed of dwell, scanning time etc. The window itself may be resized should the user need larger letters. All the constituent parts are scaled automatically. There are tradeoffs between sharing screen space with the applications (that is, you need as small a floating window as practical) and coping with users who have poor eyesight (where a larger window would help). For younger users it is likely that a reduced set of keys is required, as few will want the full range available on normal 101-style keyboards. Limiting the predictions to only four means that visual search times are kept brief as they are constantly updated and so cannot be learnt. The fixed layout of the remainder of the keyboard is one that the user will gradually learn and spend less time visually searching.

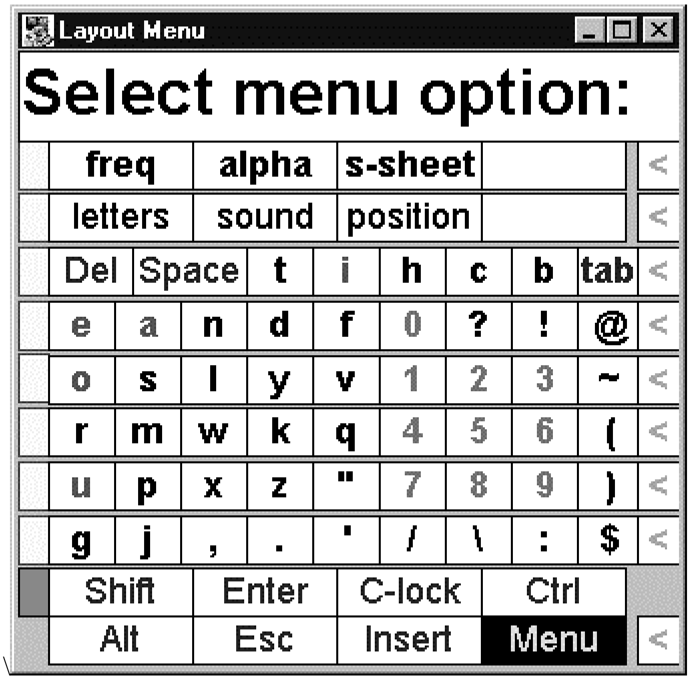


Figure 5. *Menu for Virtual Keyboard Window*

One major improvement that we have made to the scanning method is to have a "go back" symbol on the end of each scan row. A frequent cause of frustration is to miss a selection on a row, and have to wait to cycle around through all the rows again. With this simple addition the user can select the "go back" as they reach the end of the row in which they missed making the selection. The system then re-scans that row and the user has another chance at making their selection. If the row was chosen in error then the only penalty is one additional dwell time on that "go back" symbol.

Character prediction is based on n-grams. We experimented with different training texts and values for n. Performance was best with tri-grams for the English texts that we used. We found that it didn't matter whether the text was from the user or an unrelated author. This meant that we could pre-load our frequency tables for new users without waiting for a large personal file to be built. Word prediction is possible using dictionaries with some fast form of searching. Instead, we opted for a simpler one of word completion rather than prediction, using a pre-loaded dictionary that is added to as the user types. This dictionary contains an expected frequency of use. This is updated as the user enters text. For both character and word predictions we deliberately kept the choice to four in order to avoid long search times. We found that our predictions were more likely to be correct for our users than commercial systems. For instance, if a "t" is entered the predictions are "the to that this." One popular commercial system would offer a very long list such as "tab table tablet tabloid tabulate tachometer …". Most of these would not be part of our users' vocabulary and it also takes a long time to visually scan through this list to see if the prediction row contains what is wanted.

5. SYSTEM TRIALS

Our initial trials have been restricted to several able-bodied students and ourselves from our department. We wanted to evaluate the performance without confounding variables and to carry out changes without being concerned with any backward compatibility. Further, we did not want to offer our users something that was still very much an experimental prototype. We have also consulted with staff at the Cerebral Palsy Association of Western Australia who have considerable experience with a wide range of users. With the able-bodied users we found that learning was very rapid with improvements being shown on even the second trial. Errors also decreased quite rapidly due to learning of the layout and how to optimise selections from the predictions. We all appreciated the convenience of the "go back" key. Over the next few months we will be improving the prototype and will then be offering it for evaluation with a user already familiar with the type of system as he has been using a commercial system. Although our objective is to design a system for children with severe motor impairment through Cerebral Palsy, it is likely to be useable by anyone with similar levels of motor impairment.

Acknowledgements: It is a pleasure to acknowledge the efforts of Martin Masek in implementing and contributing to this project. Staff at the Sir David Brand School and the Cerebral Palsy Association were very helpful in developing the project. Finally, our thanks go to the children and our student testers in the department. The department internally funded the costs of its development.

6. REFERENCES

N. Alm, J.L. Arnott and A.F. Newell (1992), Prediction and Conversational Momentum in an Augmentative Communication System, Commn. ACM, **35**, *5*, pp. 46-56.

J.J. Darragh. and I.H. Witten (1992), *The Reactive Keyboard*, Cambridge University Press, ISBN 0-521-49375-8.

A.D.N. Edwards (1995), *Extra-Ordinary Human-Computer Interaction – Interfaces for Users with Disabilities*, Cambridge University Press, ISBN 0-521-43413-0.

S.R. Ellis (1995), Origins and Elements of Virtual Environments, in Virtual Environments and Advanced Interface Design, eds. W. Barfield & T.A. Furness III, Oxford University Press, pp. 14-57

N.A. Hine, W. Beattie, A. McKinlay and J.L. Arnott (1994), Consideration of Scanning Keyboard and Text Prediction in the Context of Access to Telecommunications Services, ISAAC: 6th Biennial Conference of the International Society for Augmentative and Alternative Communication, Maastricht, The Netherlands, 9-13 October 1994.

P.E. Jones (1997), Computer-Assisted Remote Control for the User with Motor Impairment, Proceedings of the XII BCS Conference on Human Computer Interaction ’97, Springer-Verlag ISBN 3-540-76172-1, Bristol, 12-15th August, 1997 pp. 205-221.

G. Singh, S.K. Feiner and D. Thalmann (1996), Virtual Reality — Software and Technology, Commn. of ACM, **39**, *5*, pp. 35-36, May 1996.

7. URL

Origin Instruments Corporation (1998) — SofType keyboard http://www.augmentative.com/acs-styp.htm (July 1998)

Making 3D models: A challenge in man-machine communication

Angel Osorio

Directeur de Recherche CNRS, LIMSI-CNRS
B.P. 133, 91403 ORSAY CEDEX, FRANCE

osorio@limsi.fr, http://www.limsi.fr/~osorio

ABSTRACT

The purpose of this publication is to present a graphic 3D modeler with interactive interaction capabilities operating in real time. It may be used on both in Unix workstations and PCs. The modeler can be used to make and edit 3D reconstruction of articulated objects, a facility that is particularly useful in processing medical images generated by a CT scanner or magnetic resonance units. The modeler takes account of the physical characteristics of the objects manipulated and, in developing it, we have assumed that the prospective user will have the necessary expertise to be able to interact with and interpret the model.

1. INTRODUCTION

Though there are many modeling software packages on the market, most of them are aimed at geometrical modeling: they use the full geometrical approach and contain only vertex, surface and texture characteristics in the kernel. Furthermore, and mostly for marketing reasons, only a few graphics software packages take into account the very fast evolution of computers, and even these do not make full use of the specific characteristics of today's computers. The problem of making a realistic 3D model is very complicated because of the protocols and interactions involved. A computer has, in general, a 2D screen, a 2D mouse and a keyboard; making a 3D model thus requires the operator to use 2D devices to make a 3D construction. This paper offers a summary account of the realistic 3D modeling program (AOM) developed by my team, and presents two very different applications of it: first, the real time manipulation of articulated mechanical structures, and second, the 3D construction of solids from a set of 2D sections used in medical studies, scans for example.

2. THE CONCEPT OF THE "WORKING MODEL"

Workstations and even some PCs available today are capable of modeling complex realistic structures. While CPUs are still in general too slow for many realistic displays, the addition of specific peripheral cards provides the capability required for the advanced 3D display of objects, including textures, shadows and articulations. The architecture of most computers is still of the von Neumann type, but, by using a programmer-hidden approach and exploiting the sophisticated characteristics of graphics devices, it is possible to achieve fast multi-tasking. This is the case with most PC games.

Since the beginning of the 90s, our objective has been to realize a software program that can adapt to different machine configurations including a variety of geometrical standards. It seemed to us that such a program could comprehend the idea of a "working model": if a 3D model has a set of joints, it is possible to animate them in real time via a human operator. This animation works in a realistic way, taking into account such physical characteristics as inertia, weight, limits of degrees of freedom, etc. It is achieved using standard peripheral units, i.e. mouse or keyboard, and also non-standard devices, e.g. button boxes, knob boxes, numerical data glove, etc. This approach is useful, for example, for the off-line programming of robots, which was realized ten years ago, and also for validating articulated mechanical structures, and/or the physical behavior of articulated objets, on a low cost computer.

The approach described is also useful for the advanced examination of 3D objects where both the general aspect and the inner characteristics are important. When a doctor looks at the 3D model of an organ, he needs to see all round it and also inside it. This is true for 3D images of the lung, liver, brain, breast etc., all of them subject to many different diseases which can be very difficult to diagnose. In addition, this second application of our program has proven very useful in the computerized measurement of the extent of disease. In many cases, the percentage of extent of disease is the only parameter required for measuring chemical dosage.

Our observation of doctors using 3D modeling software has shown that the real problem lies not so much with the computerized construction of 3D models as with the man-machine interface. All the cases studied to date show that the main problem is the analysis of images, in particular the detection of lesions, which involves locating and identifying minute details within a very complicated picture. Figure 1 shows a liver-scan. Even a qualified person has difficulty interpreting the 2D section of an organ like this, let alone detecting the lesions. The idea that a computer program could extract the relevant information still seems far-fetched, but the approach of the graphics system developed is to provide the doctor-user with the tools necessary for this task. The system builds up the 3D model from a set of given program-parts (modules).

3. 3D PICTURES

People often speak about "3D pictures", when they really mean 2D pictures which have a 3-dimensional connotation for an human audience, yet, as everyone knows, the most sophisticated Hollywood movie is, in fact, no more than a set of 2D pictures projected on a 2D screen. A physical modeler uses real 3D data : it manipulates functions which have values in terms of x, y and z. Few sensors currently available yield real 3D data ; the principal sensors are CT scanners and magnetic resonance units used to generate medical pictures, and laser measurement units used in cartography. One of the most difficult problems is to design a modeler which can introduce into the graphics system data "which we know and which we see" but which is not known in terms of x, y and z data, and to do this with the precision required by both the computer and the user. The modeler presented here is able to manipulate directly the date of very large 3D pictures (more than 2 giga bytes), and to meet all the user's display requirements in terms of graphics definition.

4. AOM: A REAL TIME ORIENTED MODELING SYSTEM

In the first instance, the graphics system was designed on an Unix workstation, using X11R6 and Motif1.2 as window interfaces. The graphics part of the software was developed using OpenGL and Starbase as software interfaces, the goal being to operate very close to the electronic characteristics of the computers. However, over the last two years, the world has probably witnessed the third revolution in computing with the unprecedented improvement in the performance of PCs accompanied simultaneously by substantial reductions in their cost. Indeed, for the last year, it has been possible to obtain low cost machines with the same overall performance level as that of a workstation. The practical problem for us, then, has been to write code capable of being recognized by a wide range of computers. The central idea has been to define a meta-language which conceals the code of the modeler and the interfaces. In terms of pure graphics, the task was easy because most graphics software interface work within same terms (openGL is not too far removed from Starbase or Directx). The major issue was the use of a windows system. The X11 system is quite different of windows 95 or NT, and these, in turn,; are very different from Openlook and others well known graphics interfaces where there is a lot of interaction between the windows manager and the user interface. For the present, the solution which we have adopted is to consider a window as an object (in C++ terms), which can accept and ascribe graphics characteristics. The task was easy in the case of PC development packages but it has been much harder in the case of X11. In input terms, we consider the user as an object giving answers to the system's questions.

The actual modeler is composed of the following parts:

4.1 The Viewer.

This module manages the graphics of the model to be displayed. It manages both lights and camera, and the scale and initial set of values for the articulations. It also offers the possibility of network use. The specific characteristics of workstations and of the graphics cards have been exploited, essentially in terms of Z-buffer and multiple graphics plan use. Using the primitives of OpenGL and Starbase, the characteristics of the graphics units are dumped and used for fast visualization. These capabilities are dynamically adapted to enable their use in a network display. Of course, the package is able to display the model under various aspects including initializing, lighting, shadows etc. The user has two kinds of interactions at his or her disposal : graphics scales managed by a pointer (mouse or other), and an interface for the use of specific input devices (knobs, buttons, etc.).

4.2 The file and database manager.

This enables disc input and output, and makes a logic-tree of the model, using a graph representation. A model is considered to be composed of a set of parts which are geometrically independent. Each part is a solid with specific geometrical characteristics such as scale, initial position, links, joints, deformations, etc. Each part is defined as a set of surfaces with a variable number of vertices. A vertex is a set of floating values: co-ordinate, normal, graphic and mechanical properties, etc. The model is self contained and each piece of information is present once, and only once, in the database. If a user adds a surface, again, each piece of information concerning it is included only once.

A model is represented in the database as:

np = x // number of parts
nf[i], i = 1, np // number of faces for each part
nv[i,j], j = 1, nf[i] //number of vertex for each face
O, s, Ox, Oy, Oz, Tx, Ty, Tz, Rx, Ry, Rz // origin, type, global position, translation and rotation
L, t, Ox, Oy, Oz, Dx, Dy, Dz, g // light, type, position, orientation, gamma
X, t, n, Ox, Oy, Oz, Tx, Ty, Tz, Rx, Ry, Rz // part scale, type, part, part position, translation and rotation
A, t, n, Ox, Oy, Oz, Tx, Ty, Tz, Rx, Ry, Rz // joint, type, part, joint position, translation and rotation
C, t, n, f, ..., s, ... // chain, type, father(s), son(s)
Q, t, v, ..., // physical properties, type, value(s)
P, t, x, y, z, px, py, pz, nx, ny, nz, ... // point, data type, position, property, normal

There are two ways of defining a point : by its co-ordinates (the normal and property values of each point being computed from the others elements of the model) or by giving all the properties of the points. Even when the information is redundant, the data used to define the point is retained.

4.3 The Model Editor

This allows the user to edit a model in geometrical as well as functional (physical) terms. It is possible to edit vertex co-ordinates and characteristics, and also to define and modify the logical representation of a model in terms of links between parts, definition of joints, limits of degree of freedom etc. It calls up a conventional graphics module for the geometrical management of parts.

4.4 The Matrix Processor

The rapid manipulation of an articulated system requires fast operations between matrices. A conventional “push - pull” software engine has been designed. Using a stack of 4 x 4 elements, the operation between matrices is achieved by means of an inverse polish notation: “push” introduces a matrix into the stack, while the entry of an operator ensures the operation and “pull” of the stack. These functions have been programmed on two types of workstation (SGI and HP), using the specific capabilities of the machines for the purpose. The new PC graphics cards include many improved graphics characteristics. The driver provided by the manufacturer controls matrix manipulation in PC graphics cards. For example, in a particular case, we were able to use the graphics memory of a card to save directly the geometrical vertex of the model, after which it was possible to avail ourselves of the very fast capabilities of the device.

4.5 The Man-Machine Interface

Two types of protocol (interaction) have been programmed in terms of the specifications of the workstation used. They are :

4.5.1 The Software Interface. If the workstation has only a mouse and keyboard as peripherals, then a window displays a set of scale-bars. The operator can assign each one of these to a function including: degrees of freedom, rotation and translation axes between parts, limiting conditions, logic-tree of parts, etc. A major problem today is the ability of point to a 3D point in a 2D representation. A “3D mouse” sufficiently precise to point at a 3D point in a2D representation is currently not available on the market. For the time being, therefore, we have defined, as a solution to this problem, a virtual 3D widget which includes three planes and a simplified representation of the 3D body. This enables us to point in terms of x, y and z using a standard mouse and the modeler then adjusts dynamically the precise characteristics of the area thus marked.

4.5.2 Hardware Interface. Three kinds of peripherals have also been considered : (1) a button box (an electrical box with 24 buttons); (2) a knob box (a box with 9 rotary knobs scanned at a thousand times per second); and (3) a numerical data glove. This interface lets the user assign each peripheral unit to a degree of freedom, and also accepts data into the database for a driver to take account of a new peripheral. Of course, each peripheral unit can be assigned to such general characteristics of the viewer as light, scale, shadow, modes, etc. Furthermore, a complete CT scanner interface has been built into the modeler so that the system may be used in a medical context. Figure 1 shows an aspect of the software interface.

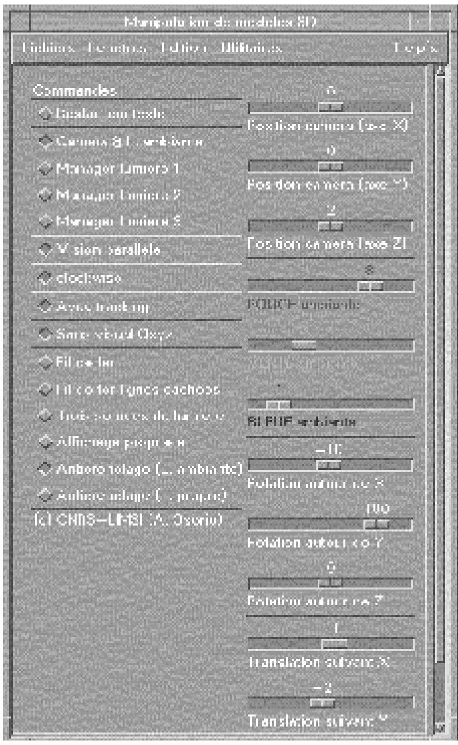


Figure 1

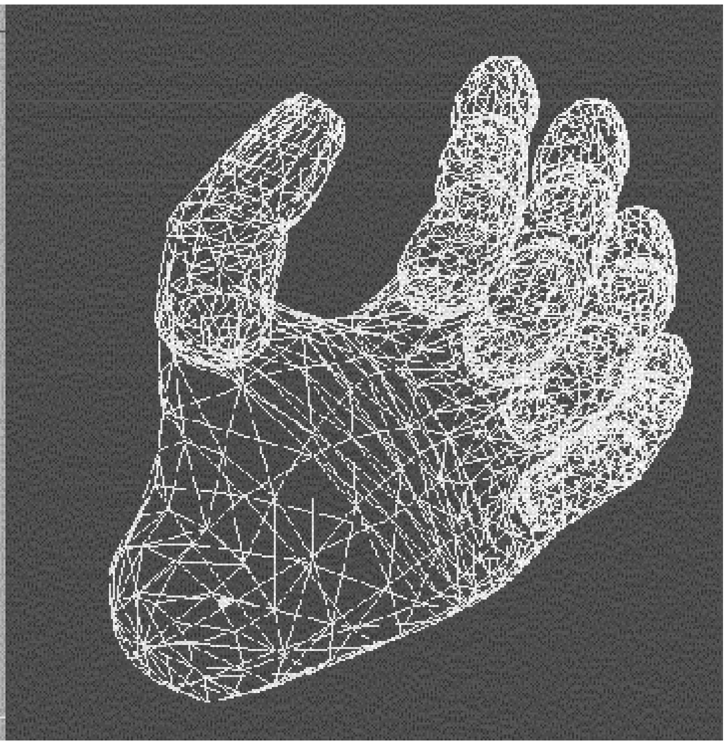


Figure 2

5. APPLICATIONS

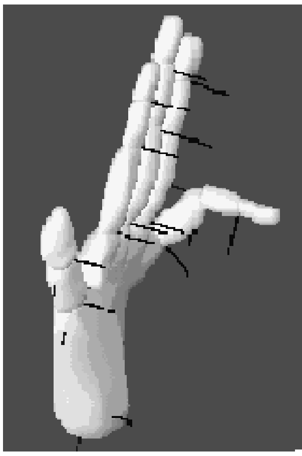
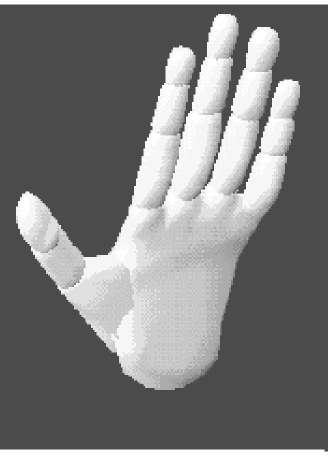
Over the past 20 years, there has been a great deal of research into off-line programming of robots. The goal has been to create realistic programs for manipulators, and also to take account of most possible real-life situations, including errors, accidents, hidden information and so on. Many developments of the past can today be run in real time on modern workstations. However, the general economic situation has led recently to a fall off in this kind of approach.

Recently, "virtual reality" has become the buzz word in computers and the number of applications is growing fast. With this approach, the general problem is the same as with off-line programming of robots. The computerized world has a set of objects with mobile parts between the elements. We have considered this idea of "joints" from a very general point of view. For instance, we can set the scene with a table, an overhead projector on the table, a transparency film over the projector and a picture projected onto a screen. If the table is held by a robot, or even placed on a vehicle, the general model must take into account the particular positions of each part in order to display the picture on the screen. This example suggests the idea of an hierarchical tree where the joints between parts can be fixed, slipped or be deformed.

6. THE REALISTIC MODEL OF A HUMAN HAND

For validation purposes, and in order to create a "real-world" experience, we have used the AOM system to make the model of a very popular, articulated manipulator: the human hand. The hand has a great number of degrees of freedom : for example, picking up a ball, one of the easiest of movements, for the purposes of this model, requires us to control at least 15 degrees of freedom (21 dof in our case). Three different phases have been used to construct a model of the hand: (1) the geometrical model of each part, using conventional software; (2) the editing of the parts in order to set the specific properties; (3) the editing of the joints and their limits.

Figure 2 shows a still of the realistic virtual hand and figure 3 includes the articulation axis. Figures 4, 5 and 6 shows three different positions of the same hand.

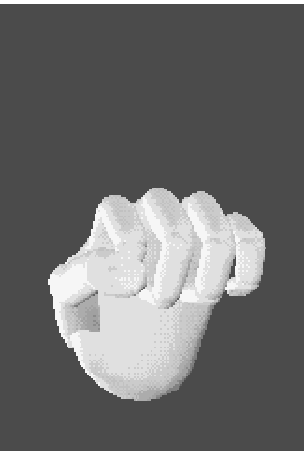


Figure 3 **Figure 4** **Figure 5** **Figure 6**

7. THE 3D CONSTRUCTION OF MEDICAL IMAGES

In spite of recent progress in computing, medical imaging has advanced little over the last 50 years. Doctors still study black and white images of extremely poor quality, whether these are body or ultrasonic scans or radiographs. In many cases, the use of realistic 3D models would be really useful, for example, in gaining knowledge of the extent of lesions in order to determine the quantity and kind of chemicals to be used in a treatment (cancer pathology for instance).

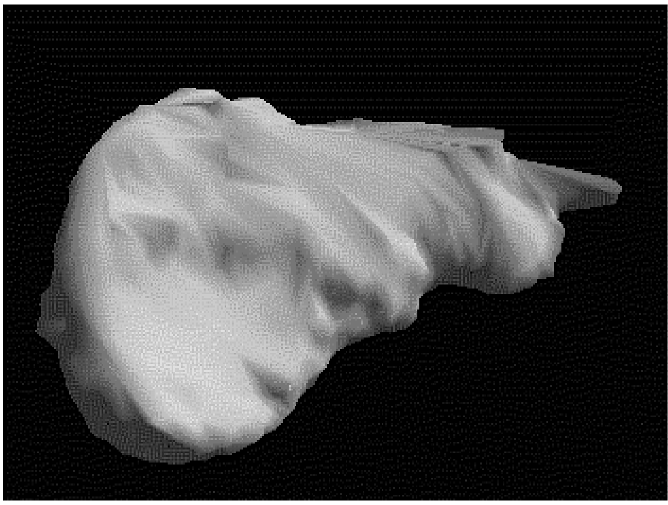
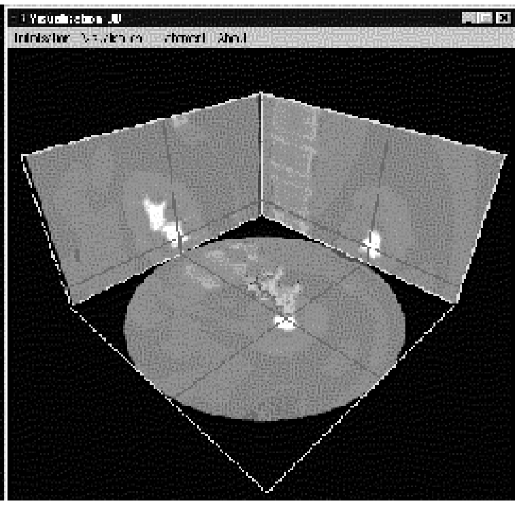



Figure 7 **Figure 8**

Making a 3D model from a set of 2D images is a very difficult task which requires a significant and expert contribution from the doctor. The problem of distinguishing between healthy and abnormal tissue is still too difficult and specialized for today's computer technology. The AOM system includes a 3D reconstruction model capable of helping the doctor in constructing a 3D image of an organ. This work includes an advanced man-machine interface which helps the doctor to: (1) visualize and edit 2D images in order to identify and extract important areas of tissue; (2) scale 2D images in order to set an homogeneous scale and obtain an original point of reference; (3) manage the model's transparency in order to visualize lesions and take their relative measurements.

So far the software has been used for the 3D reconstruction of a lung with major inner lesions, and for a liver with many secondary lesions. Figure 7 includes a 3D view of the liver which has been realized from a set of 25 scanned images. Figure 8 shows a 3D reconstruction of a tumor in a kidney.

8. CONCLUSIONS

The high performance of the low cost computers available today permit the manipulation of realistic 3D models. Many field, in which this revolutionary technology might be useful, do no benefit mainly because the interfaces are not well adapted for use by professionals. This situation is particularly acute in the field of medicine where images are still studied and exploited as they were 50 years ago. A crucial factor in the widespread use of computers must be the overall cost of equipment. The software developed here is currently being used by two hospitals in the Paris region and, in each case, this has been achieved by adding only one per cent to the overall running costs of the Medical Images Center. As a result, doctors in both places now have access to an advanced system for processing and studying images, which yields clearer, more accurate and detailed information than hitherto and in less time. In this way, the new system also indirectly helps doctors cope with their principal problem, namely the shortage of man-hours. In the light of this and also considering the situation from the patients'point of view, we can only hope that, in the very near future, this technology becomes universally available to the medical profession.

Acknowledgements: For the last 4 years my research team has been working on the manipulation of 3D models. This paper is a summary of our work. The principal student collaborators were Olivier Bedelet, Laurence Chomarat, Olivier Chomarat, Hervé Marechal. The CT scan images were made at the Medical Image Center of the Clinique du Bois de Verrieres (Antony), in Paris area. Thanks are due to Eleanor Gregh and Jim Johnsen for their help in the English version of this paper.

9. REFERENCES

[1] M. Krus, P. Bourdot, A. Osorio (1998), Primitive-based management of virtual scenes for a human interactive navigation, *Complex Systems Intelligent Systems & Interfaces*, Nîmes, may 1998

[2] O. Bedelet (1997), Un système interactif de description d'objets 3d. *PhD stage. LIMSI-CNRS*, septembre 1997.

[3] H. Maréchal (1995), Un langage de description d'objets articulés. *PhD stage. LIMSI-CNRS*, septembre 1995.

[4] O. Chomarat (1994), Interface temps réel pour la manipulation d'objets 3d articulés . *PhD stage. LIMSI-CNRS*, juin 1994.

[5] L. Chomarat (1993), Représentation 3d d'objets articulés. *PhD stage. LIMSI-CNRS*, juin 1993.

A practical example using virtual reality in the assessment of brain injury

R C Davies[1], G Johansson[1], K Boschian[2], A Lindén[2], U Minör[2], B Sonesson[2]

[1]Department of Ergonomics, University of Lund,
Box 118, SE 22100 Lund, SWEDEN

[2]Department of Rehabilitation, Lund University Hospital,
Orupsjukhuset, SE 24385 Höör, SWEDEN

roy.davies@amt.lth.se, gerd.johansson@amt.lth.se
kerstin.boschian@skane.se, ulf.minor@skane.se, bengt.sonesson@skane.se

ABSTRACT

Virtual Reality (VR) as a complementary tool for medical practitioners in the assessment and rehabilitation of people who have suffered a brain injury is discussed. A pilot-study has been undertaken on a prototype VR assessment tool. The design involved nine occupational therapists with expertise in the care of traumatic brain-injured patients and one (computer experienced) patient. The aim was to begin a dialogue and to ascertain the potential of a VR system. A common method for occupational therapists to assess function and ability is to ask a patient to brew coffee. From the performance of such a task, an individual's "functional signature" can be determined. The prototype was built using Superscape®, a personal computer based VR system, to be close to the real coffee making task, including effects of making mistakes, realistic graphics and sound effects. The world was designed to be as easy to use and intuitive as possible, though problems of mental abstraction level, transfer of training and realistic interaction have yet to be resolved. The comments from the test participants have highlighted problem areas, given positive insight and pointed out other scenarios where VR may be of use in the rehabilitation of people with a traumatic brain injury.

1. INTRODUCTION

Virtual Reality (VR) has shown great potential in various training and visualisation fields. In the rehabilitation of people with traumatic brain injury, there has been much discussion but little tangible work. A few groups, notably; Andrews *et al* (1995), Rose (1996), Pugnetti *et al* (1995) and Rushton *et al* (1996) have, however, shown that VR technology can provide new and useful techniques for rehabilitation. For example, training using VR could help to alleviate the problems of under-stimulation during recuperation (so called fourth level damage - degeneration of brain tissue due to an impoverished environment) by providing continuous and early sensory stimulation (Rose, 1996). Similarly, early assessment of cognitive ability can allow a rehabilitation programme to be instigated quickly (Rose, 1996).

VR is an attractive training medium since:

- It allows control over which details are to be included in an environment, leaving out those that may be distracting or irrelevant.
- The environment can be structured so that exactly the same situation can be repeated as many times as the trainee (or trainer) desires.
- The trainee receives positive reinforcement in the form of direct feedback, rewards for correct responses and help from the computer.
- The trainee can work independently, with the training sequences being recorded for later appraisal by a human teacher.
- Dangerous or costly situations can be trained without risk.
- Training can occur despite physical disability.

A common problem with brain injury is a reduction in motivation. Given that computer programmes can be very motivational for some people (particularly younger people), VR based training may allow patients to learn despite themselves. This will, of course, require careful consideration of behavioural training methods and make use of direct feedback and rewards from the computer.

One issue that is not clear is what sort of information can be successfully transferred to the real-world situation from training using VR. This is entirely dependent on the cognitive abilities of the trainee. There have been a number of successful instances of training people with mental disabilities using VR. For example Standen and Cromby (1995) have used desktop VR for education and training of people with learning difficulties, and VIRART at the university of Nottingham have shown that VR has a positive effect in special needs education (eg. Brown and Stewart, 1996). It is hypothesised that since people with mental disabilities can be taught using VR technology, it may also work for people with a brain injury.

Nevertheless, it would be presumptuous to assume that VR will provide the perfect tool, indeed, research has shown that care must be taken where VR is applied - not all such training is beneficial. Kozak *et al*, (1993), for example, have found no evidence of transfer of training using VR for physical tasks due mainly to the lack of tactile feedback. This is backed up by Wilson (1997), where he warns against seeing VR (or indeed any new technology) as a panacea and to critically compare new methods with existing techniques.

Furthermore, side effects of immersive VR usage can be uncomfortable, if not damaging. Such effects include overloading of neck muscles, nausea (Regan, 1995) and optical disturbances (Mon-Williams *et al*, 1992) which would be augmented and unpredictable on people with brain injury. These effects are absent in desktop VR systems.

Finally, the ethics of trying a new technology on people who have been through a traumatic experience must be considered. The accepted rehabilitation methods cannot be delayed in order to test a new procedure if it is not guaranteed to provide at least the same level of help. Therefore, experiments must be carefully and ethically designed.

1.1 Uses of VR in rehabilitation after brain injury.

VR has the potential to be used in three main areas: assessment of the extent of a person's remaining cognitive function; training weak cognitive areas; and as a tool for everyday living. The former two are discussed below.

1.1.1 Assessment. There are two methods of assessment currently used that can be roughly classified as theoretical and practical.

The theoretical method entails a one to one discussion between a medical practitioner and the patient and includes pen-and-paper tests designed to highlight particular cognitive functions. These tests have been developed and validated over many years and are only recently being augmented by computerised versions (Bradley *et al*, 1993). It is not expected that VR can take the place of any of these tests in the short term, mainly due to the rigorous and time-consuming validations required. However, the practitioner not only uses tests, but also experience and subtle clues in behaviour to pinpoint the precise nature of the disability. It is in this area that VR could be useful.

Practical assessment occurs in the rehabilitation hospital by the nurses and occupational therapists caring for and working with the patient.

The responsibilities of an occupational therapist include evaluation and treatment of patients who, due to some functional limitation, find it difficult to contend with daily activities. A brain injury can mean that a person is hampered in the performance of daily activities by lack of motor or cognitive proficiency or both. Reduction in motor proficiency can be seen as difficulties in walking, bending down, lifting, gripping and co-ordinating. Reduction in cognitive proficiency often has implications for how independent a person can be. Examples of such are; taking initiative, being attentive and able to set goals, choosing and managing artefacts, beginning and ending an activity, proceeding through a sequence in logical order, and learning from previous experience.

Tests are carried out in this environment as well, though these tend to be more of a practical nature, based on daily activities. VR could also be of help here, particularly in the early assessment of people who have other physical injuries that restrict movement.

1.1.2 Training. Rehabilitation of people with brain injury entails preparing them for re-entry into daily life, and retraining them in everyday activities necessary to partake in society and work and to look after themselves. This is a time-consuming process involving practice in real situations and can be stressful since the patient often has to go into the outside (unprotected) world, for example to learn how to catch a train. Some of these situations may be possible to represent in VR, such as using the train ticketing machine or an automatic bank teller.

The cognitive processes that VR can help with are: following instructions; sequencing; attention to periphery information (audible and visual); increasing the brain's working speed; problem solving; and navigation.

1.2 Physical and cognitive burdens imposed by computerised tools.

A cognitive (and physical) burden is placed on a person to manage a tool to complete a task. Most tools are, of course, designed to simplify the completion of the task, but often require training to be used effectively. In the extreme case, using the tool is too complex and completely detracts attention from the task (for example, a non-typist trying to use a keyboard for writing). This is particularly relevant for computerised tools, where it is often not known beforehand if the tool will actually simplify the process. If the issue is further complicated by involving people who may well have abstraction difficulties with everyday objects, then extreme care must be taken in feeling a way forward in the design of a computerised tool.

In perceiving depth in the real world, both monocular and binocular cues are used (Eysenck and Keane, 1995). Many of the monocular cues such as linear perspective, aerial perspective (haziness of far-off objects), interposition (nearer objects obscure farther ones), shading, texture, familiarity with object's size and motion parallax can be implemented in VR. This is indeed vital if we are not to tax the patient's abstraction understanding too much, for all the binocular cues (convergence of the eyes, stereopsis and accommodation) are lost when we put the picture on a computer screen. To complicate the issue, the patient also has to understand that the image on the screen is a two dimensional rendition of a three dimensional model which in turn represents a real environment. It is expected that some brain injured patients will not be able to overcome these abstraction barriers, particularly if their three dimensional perception is damaged.

Clearly, one of the first tasks is to identify exactly which group of patients can benefit from the technology.

2. A STUDY IN ASSESSMENT USING VR

It was decided to conduct an exploratory study as a lead-in to further work to assess where VR could in practice be applied to brain injury rehabilitation and to which groups of patients. A collaborative project was initiated between the Lund University Ergonomics department and the Orup Rehabilitation Centre to build and test a simple desktop VR world as an assessment tool. The task chosen was a reproduction of one of the tests used for practical assessment - that of brewing coffee.

Figure 1: *A scene from the VR coffee brewing environment.*

Table 1. *Activity Analysis of brewing coffee*

Idea / verbal or visual stimuli	Clarification
Walk	Between coffee machine and sink
Position correctly	Place oneself at a suitable distance with respect to the object
Gather together	Take filter, coffee etc together from a cupboard
Grip	The handle of the coffee pot, the tap, the coffee container and the coffee scoop
Manipulate	Local hand movements, for example the coffee filter
Co-ordinate	For example, take the top off the coffee container
Lift	Coffee pot, coffee container
Transport	Coffee pot between machine and sink
Navigate	Don't crash into other objects
Manage	Get through the task, don't rest, break off, show tiredness etc
Adapt environment	Place the coffee machine in a suitable place, turn on/off machine, tap
Bend over	Stretch, rotate, take up something from the floor
Choose	Agreed material or tool
Use	Use the tool in a logical way
Attentiveness	Focus on the task
Manage tools	How one, for example, copes with the filter
Search	The objects one needs
Do in logical order	For example, filter, then coffee grounds, turning on the power
Continue	Do the task without interruption
Goal driven	Do what one thought or agreed to do
Return	Put back objects, clean up
Finish	Stop putting in water and coffee grounds when sufficient amount
Adapt behaviour	Compensate for unexpected problems
Learn	Use previous experience, don't repeat mistakes
Organise	The workplace in a suitable way
Stabilise	Balance without support
Question	Ask relevant questions

Cognitive component – suitable for testing with desktop VR environment
Physical component – not suitable for testing with desktop VR environment

Brewing coffee is a frequent and common task in the Swedish culture. A task such as this is usually subdivided into both physical and cognitive components (Table 1) which are used by the occupational therapist in the assessment process. Whilst the physical components cannot be assessed using desktop VR technology, over half are cognitive, suggesting that VR could indeed have a role in early assessment using this task. The cognitive aspects can be further divided into operational and tactical proficiencies. The operational aspects depend on variables such as ability to visually scan the locale, orientation in the environment and action readiness or confusion with complex motions. Such behaviours or deficiencies are recognisable in an individual regardless of the task's character.

Tactical proficiency is more problem or task orientated and tied to the completion of an activity and the specific requirements of the task. In this area are qualities such as problem solving and judgement of the actual situation, risk assessment, simultaneous capacity, planning of procedures and behaviour control.

The current coffee brewing test is described in a document used in the rehabilitation field which details the exact process to be carried out and what the therapist should be on the look out for (Fischer, 1995). This was used as a basis for the design of the VR world which was contrived to look like a simplified kitchenette (fig 1). The following aspects were taken into account – some of which come from Fischer (1995) whilst others are standard human-machine interaction considerations:

1. *Task concentration.* The user has to be able to concentrate on the task, not on using the tool. This is particularly important in this case as the users may have no computer experience at all. Furthermore, any assessment using the tool is going to be coloured by the effects of the tool itself. This leads to an important research question – can VR be used for assessment, or is the weight of usage too great to be overcome by brain-injured patients? Methods of reducing tool distraction are discussed below.
2. *Sound.* Not only vision is important – sound, too, plays a role in determining what is occurring in an environment. For example, the sound of running water reminds us that the tap is still on. Some brain-

injured people have trouble incorporating such peripheral information into their task performance. Therefore, realistic and directional sound effects were included for all objects.

3. *Controlling the world:* One method of reducing distraction by the tool is to simplify the input procedure. This is particularly important in this case as a 3D mouse or dataglove is both expensive and may be hard to use. All interactions are by single mouse click with automatic positioning of the viewpoint. It was hoped that by restricting in this way, the user would quickly learn the tool. Similarly, all interaction was via objects in the environment – no abstract symbols, icons or hidden keypresses.
4. *Abstraction level, visibility of objects and affordances.* To decrease the cognitive load, the level of abstraction from the image on the screen to the real world had to be minimised. In essence, this entailed making the VR world as realistic as possible within the hardware and software limitations. For example, the textures on the objects were scanned from the real objects to give them added realism. There were some problems, however, in keeping items of interest always visible and of a sufficient size to be clicked upon without drawing undue attention to the object. An interesting question would be whether a split-screen method is better. A further issue was whether the objects gave sufficient affordance to being clicked upon.
5. *Making mistakes.* A vital part of the assessment process is noticing what sort of mistakes the patient makes. For example, not being able to follow a procedure might manifest in putting the coffee grounds in the machine before the filter. Therefore, the VR world had to allow for mistakes to be made and produce a sensible result, so forgetting to put the coffee grounds in should result in hot water in the coffee pot, not coffee. This differs from the usual situation where mistaken avenues of action cannot be taken (for programming efficiency) since the objects are not made active until the correct prerequisites have been fulfilled. Users also had to have the freedom to take alternative routes since not everybody makes coffee in the same way.

An important aspect to consider was whether the patient would be able to manage the VR tool, and whether he or she could overcome the cognitive (and physical) burdens imposed by the tool (as opposed to those demanded by the coffee brewing process). These are estimated to be as follows in this prototype:

- Being able to move a mouse with enough precision to point and click.
- Perceiving the coupling of mouse movement with cursor on the screen.
- Understanding that clicking with the left button can cause an event to occur.
- Realising that clicking an object means “pick up this object and carry it”.
- Knowing that clicking another object whilst holding the first means to “perform an operation using the first object on the second”.
- Comprehending that the flat two dimensional picture is a representation of a three dimensional world
- and that this, in turn, is a representation of a real environment.

It has not yet been investigated whether the patient is able to understand these concepts, nor the detriment of these on the performance of the task and the usability of the tool for assessment of brain injury – this is currently the subject of further research.

3. METHOD

Superscape® was chosen as the VR world building environment since it works on ordinary personal computers, such as those found in a rehabilitation centre and allows quick and easy prototyping and programming of complex object behaviours. Desktop VR was used rather than immersive, again due to the availability of hardware, and also because of the aforementioned side effects.

3.1 Subjects

The subjects were nine occupational therapists (aged from 31 to 49, 6 women, 3 men) who work with brain injured patients on a daily basis as well as one brain injured patient (aged 19, male). Only the patient had significant computer experience, the others were ‘word-processing’ level users. Since the intention was to gather ideas and show the potential of the system, it was considered too early to include more patients, the one included was comfortable with technology and was included as an expert rather than a ‘patient’.

3.2 Procedure

The subjects were made aware of the purpose of the test and asked beforehand to consider the potential of the system. They were told to brew coffee as they usually do, that the program was controlled by single mouse clicks and performed the task between one and five times. The time taken to perform the task was recorded. The subjects were then asked for criticisms about the graphics, animation sequences and tempo as well as to give suggestions about other activities or areas where VR technology could be used. Note that this was not a real user-test, rather a preliminary test to gather in expertise from the experts in brain injury care and assessment.

4. RESULTS

With 28 tests by 10 people, time varied from 1.17 minutes to 12 minutes to complete the task, with a tendency to reduction on subsequent tests by an individual. The average time taken was 3.9 minutes with a standard deviation of 2.6 minutes. Two attempts were stopped prematurely due to computer problems, all others finished when the coffee was successfully brewed. The following observations and comments from the users were gathered and divided into three areas – graphics, animation sequences and tempo:

4.1 Graphics

- Eight of the ten participants had problems reading the packets and identifying what the objects actually were.
- Some considered that simply clicking with the mouse was confusing and restricting and that dragging objects should be possible.
- Some technical problems (bugs) became apparent, such as milk cartons left hanging in the air and a disappearing coffee-pot.
- Some participants thought it disturbing the way objects suddenly came towards them and just hung in the air when being carried – this could be frightening for patients.
- It was not possible to see the coffee grounds going into the filter.
- More objects should be put onto the shelves and a clock on the wall.
- Colouring of the objects was questioned (such as the water in the coffee-pot). Superscape® does not allow shadowing, and in this model, all the objects were flat-shaded to increase speed. This gave a cartoon-like quality to the pictures.

4.2 Animation sequences

- Missing being able to open the coffee packet.
- Being able to see the sink, coffee machine and bench together was judged desirable.
- Need an overview of the whole area.
- Water runs too fast into the coffee-pot.
- Some “reality problems” were discovered such as water not seen to be pouring into the machine, though ending up there anyway.
- One person commented that, due to the hopping backwards and forwards of the viewpoint and automatic movements of objects, that it was difficult to know what was going to happen next.
- The one person that commented on sound said it was suitable.
- Not possible to have two things in the hands at once.

4.3 Tempo

- The tempo was thought to be reasonable, though some parts were considered too fast (such as filling the coffee-pot with water), whilst others too slow and jerky.

Some of the comments can be traced back to the computer experience of the participants that, though limited, nevertheless gave expectations of how a graphical interface should work. Completely computer untrained people would probably not have commented in the same way.

The users were also asked for suggestions for other applications and came up with: automatic bank teller, automatic ticket machine, making dinner, laying the breakfast table, performing a transaction at the bank or post office, training in planning, training in traffic and finding one’s way. Suggestions covered both assessment and training. The general feeling was positive and that such technology is exciting and could be more motivating for younger patients than household tasks. It was also felt that adaptability to an individual's needs would be easier and the ability to repeat difficult situations a definite bonus, particularly where the consequences of a wrong action can be expensive or dangerous (such as with a ticket machine or in traffic).

5. DISCUSSION

This preliminary test and participative design has identified some areas where VR technology can be used in the rehabilitation of people who have suffered a brain injury and indeed desirable in the following circumstances: as an early assessment tool for patients who are also physically injured; as a training tool for difficult or dangerous situations; and as a positive training motivator for technology interested (perhaps younger) patients. However, the disturbance effect of using a VR tool on task performance and assessment potential has yet to be measured. Actual user tests under real conditions are planned using a more advanced prototype to determine the suitability of the tool and whether any transfer of training actually occurs. Results will be compared with existing assessment and training methods.

This prototype design posed some problems and forced often arbitrary choices of interface. Comparative studies are warranted to test alternative methods of control and the appropriate trade-off between level of realism and computer ability. These are essential due to the lack of information on the ability of brain injured patients in coping with the abstraction required in using a computerised tool. Such studies could be:

- Comparison of the current point-and-click mouse control with drag-and-drop and other input devices such as a dataglove, touch screen or force-feedback device.
- Comparing the effect of picture realism (realistic lighting, shadows and realistic object behaviours) with the current more primitive version (no lighting effects or shadows, flat, unshaded surfaces and somewhat unreal water behaviour) on performance and abstractability.
- Determining the effect of sound effects - how much sound is required and how realistic should it be?

It was also difficult at times to find a suitable viewpoint where the user could both see and click on the objects that should be available at a particular stage in the brewing process. Alternative organisations of the objects could be tested against, perhaps, a split screen showing the coffee machine and a roving viewpoint.

Other VR technologies could also be tested but a careful analysis of cost vs. benefit must be made concerning the availability of suitable computer equipment in the rehabilitation centres and the possible dangers of such technology. This includes stereoscopic glasses to give depth vision to the screen, to immersive VR using a head mounted display or even projected immersive VR as well as alternative input devices as mentioned above.

During the test, many comments were gathered on both the negative and positive aspects of the tool. Despite problems such as unpredictable behaviour of the viewpoint and occasional bugs, the average time taken by the participants to brew the coffee was less than four minutes with no prior training and they showed a tendency for improvement over successive trials. This is a significant achievement and suggests that the usability level of the tool is indeed at an appropriate level, at least for non brain injured people, though how it will be received by patients cannot be concluded from this test. The negative comments could be mostly tied back to hardware or software deficiencies, programming bugs or limitations in the prototype. Furthermore, there was little mention made of reality problems such as not-quite-realistic sound effects, flat shading of objects and totally unrealistic water behaviour. This may be explained by considering the expectations of the users – if the scene is too realistic, people tend to notice the deficiencies, if it looks like a model, they tend to accept reality deficiencies more easily. Whether people with brain injury will do the same should be investigated in the context of the trade-off between reality level and hardware costs. The positive comments suggested that the tool does indeed have potential, though it is obvious that clear and indisputable proof will be required before the medical practitioners and occupation therapists put any trust in a tool based on technology that many see as solely for games or researcher's ramblings.

Acknowledgements: We would like to thank the people at Orup Rehabilitation centre who participated in the evaluation of the prototype and who have given us many good ideas. Thanks also go to Vårdal stiftelsen (The Swedish foundation for treatment and allergy research) for financial support of this pre-project.

6. REFERENCES

T K Andrews, F D Rose, A G Leadbetter, E A Attree and J Painter (1995), The use of Virtual Reality in the assessment of cognitive ability, In *The European context for Assistive Technology: Proceedings of the 2nd TIDE* Congress (P Porrero and R Puig de la Bellacasa, Eds) IOS Press, Amsterdam, pp 276-279.

V A Bradley, J L Welch and C E Skillbeck (1993), In *Cognitive Retraining using Microcomputers*, Lawrence Erlbaum Associates, Hillside, USA.

D J Brown and D S Stewart (1996), An emergent methodology for the development of virtual learning environments, In *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technology* (P M Sharkey, Ed) University of Reading, Maidenhead, Berkshire, UK, pp 75-84.

M W Eysenck and M T Keane (1995), *Cognitive Psychology – A Student's Handbook*, Psychology Press, UK.

A G Fischer (1995), AMPS – *Estimation of motor and process abilities (bedömning av motoriska färdigheter och processfärdigheter)* (in Swedish) – contact Orup Rehab Centre.

J J Kozak, P A Hancock, E J Arthur and S T Chrysler (1993), Transfer of training from Virtual Reality, *Ergonomics*, **36**, pp. 77-784.

M Mon-Williams, J P Wann and S Rushton (1993), Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display, *Ophthalmic Physiol Opt* , **13**, pp. 387-391.

L Pugnetti, L Mendozzi, A Motta, A Cattaneo, E Barbieri and A Brancotti (1995), Evaluation and retraining of adults' cognitive impairments: which role for virtual reality technology?, *Computers in Biology and Medicine*, **25**, pp. 213-227.

C Regan (1995), An Investigation into Nausea and Other Side-effects of Head-coupled Immersive Virtual Reality, *Virtual* Reality, **1**, pp. 17-32.

F D Rose (1996), Virtual Reality in rehabilitation following traumatic brain injury, In *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technology* (P M Sharkey, Ed), University of Reading, Maidenhead, UK, pp 5-12.

S K Rushton, K L Coles and J P Wann (1996), Virtual Reality technology in the assessment and rehabilitation of unilateral visual neglect, *Proc. 1st European Conference on Disability, Virtual Reality and Associated Technology* (P M Sharkey, Ed), University of Reading, Maidenhead, UK, pp 227-231.

P J Standen, and J J Cromby (1995), Can students with developmental disability use virtual reality to learn skills which will transfer to the real world?, *Proc of the 1995 VR and persons with disabilities conference*, Center on Disabilities, California State University, Northridge, http://www.csun.edu/cod/virt95.html.

J R Wilson (1997), Ergonomics' contributions to virtual environments, *Proc. of the 13th Triennial Congress of the International Ergonomics Association*, Tampere, Finland, **2**, pp 8-10.

Transfer of training from virtual to real environments

F D Rose[1], E A Attree, B M Brooks, D M Parslow, P R Penn and N Ambihaipahan

Department of Psychology, University of East London,
London E15 4LZ, UK

[1]*f.d.rose@uel.ac.uk*

ABSTRACT

Training is one of the most rapidly expanding areas of application of the technology of Virtual Reality (VR) with virtual training being developed in industry, commerce, the military, medical and other areas of education and in a variety of types of rehabilitation. In all cases such training rests upon the assumption that what is learned in the virtual environment transfers to the equivalent real world task. Whilst there is much anecdotal evidence there have been few systematic empirical studies and those that have been carried out do not lead to clear conclusions. This paper reports preliminary findings from a study, using a simple sensorimotor task, which seeks to establish not only the extent of transfer, but also the reliability and robustness of whatever transfers. The findings demonstrate a clear positive transfer effect from virtual to real training and suggest that the cognitive strategy elements and cognitive loads of the two types of training are broadly equivalent. However, caution is advised in the interpretation of these findings. The results are discussed in the wider context of models of transfer of training.

1. INTRODUCTION

One of the most rapidly developing applications of virtual environments (VEs) is in the field of training. Virtual training regimes have been devised not only for pilots but car, bus and train drivers, divers, firefighters, surgeons, quality control operators and space mission controllers (Durlach & Mavor, 1995). Moving closer to the disability focus of this conference, there has been much interest recently in using VEs in the training of people with learning disabilities (Cromby *et al.*, 1996; Mowafy & Pollack, 1995; Standen *et al.*, 1997; Stanton *et al.*, 1996; Strickland, 1997), in rehabilitation following brain damage caused by traumatic brain injury (Rose, 1996; Rose *et al.*, 1996, Rose *et al.*, in press), stroke (Rose *et al.*, submitted) and neurodegenerative diseases (Pugnetti *et al.*, in press), and in desensitisation training for people with phobias (Carlin *et al.*, 1997, North *et al.*, 1997). Several authors have argued that, because they can be so comprehensively controlled and trainees' responses to them so meticulously monitored, VEs represent an almost ideal training medium (Darrow, 1995; Rose, 1996; Schroeder, 1995; Rizzo, 1998). Seidel and Chatelier (1997) have even suggested that the use of VEs may be "training's future".

Crucial to these training applications of VEs is the issue of transfer of training. Does training carried out in a VE transfer to the equivalent real world situation? Within the VE training literature there is a wealth of more or less anecdotal evidence that transfer does occur. However, there have been relatively few attempts to investigate empirically the virtual to real transfer process in terms of what sort of training shows transfer, in what conditions, to what extent, and how robust the transferred training proves to be. Where the transfer process has been the focus of the investigation findings have been mixed. Regian (1997) has reported positive transfer in both a console training task and a spatial task. However, Kozak *et al.* (1993) in a much cited "pick and place" task failed to find transfer from virtual to real, although the methodology in this study has been questioned by Durlach and Mavor (1995) and the results disputed in a follow up investigation by Kenyon and Afenya (1995).

Clearly there is a need for further systematic investigation of transfer from virtual to real environments. However, it is important to recognise that all studies involving transfer cannot be taken together for the purposes of analysis and review. For example, within the studies referred to above the intended outcomes of the training process are very varied, including simple sensorimotor performance, complex sensorimotor skills, spatial knowledge of an environment, vigilance, memory, and complex problem solving. It would be surprising to find equivalence between them in terms of the extent and type of transfer which occurs. This brings us to an additional point.

A serious criticism which can be levelled at studies of transfer of training from virtual to real situations is that few authors have sought to analyse the process in terms of the well established literature on the transfer or training (Cormier & Hagman, 1987) which forms part of the more extensive literature on the psychology of learning. Modern psychological thought about transfer of training can be seen as having developed from concerns about the theory of formal discipline, dominant within education in the early part of the 20th century, which held that core mental skills embedded in learning disciplines such as Latin and Mathematics would automatically transfer to other subjects. Thorndike and Woodworth (1901) took issue with this assumption, suggesting that transfer of training between two sequential tasks would occur only to the extent that the two tasks shared identical elements. A similar view was espoused by Wylie (1919) and Bruce (1930), later known as the Bruce-Wylie laws. Later still Osgood (1949) sought to generate a predictive model (called a transfer surface) from which one could estimate the extent of transfer between two tasks on the basis of the degree of overlap between them in terms of stimulus and response elements.

The theories of transfer so far described were firmly rooted within the Behaviourist tradition within psychology and, as cognitive psychology gained greater influence within the discipline so interpretations of transfer took on a more cognitive slant (e.g. Newall, 1980). In particular the emphasis was now placed upon the extent to which two tasks were similar in terms of cognitive processing demands (i.e. using the same knowledge in a similar way) in predicting how much transfer will occur between them.

In terms of predicting transfer of training effects within training and retraining programmes for those with disabilities, especially for those with brain damage, one attempt to combine the best of both the identical elements models and the more recent cognitive models is that proposed by Parenté, Anderson-Parenté and DiCesare (Parenté & Anderson-Parenté, 1990; Parenté & Dicesare, 1991). According to this model a good rehabilitation programme must have significant similarity with the real world situation in which those it is designed for will be operating, both in terms of stimulus and response elements but also the cognitive strategies which need to be employed.

In the present paper we report the preliminary stages of an attempt to systematically investigate the nature of the transfer process occurring between a virtual and real training environment in terms of the extent and robustness of what transfers, within the theoretical framework of the model proposed by Parenté *et al.* Using a simple steadiness tester wire loop task we have sought to produce a high level of overlap in both stimulus and response elements between the virtual and real training conditions. In this way we have sought to focus on the cognitive strategies involved in virtual and real training (i.e. if there is ample overlap in terms of sensorimotor elements any failure to transfer is likely to be due to a lack of overlap in terms of the cognitive strategies needed for virtual and real training). In addition to investigating the extent of transfer of training in this case (experiment 1) we are interested in the possibility that the cognitive loads associated with training in virtual and real situations may be different (experiment 2). In other words, even if superficially similar in terms of conventional measures of transfer, virtual training may be in some respects less robust than real training. We have investigated this by introducing different types of interference into the real loop task after participants have been trained either in the real or virtual versions of the task. Our specific hypothesis is that motor and cognitive interference will have differential effects on real and virtually trained performance.

2. EXPERIMENT 1

2.1 Method

2.1.1 Participants. 150 university staff and students. (mean age =37.6, SD=4.97, 97 women and 53 men). All were unpaid volunteers recruited through poster announcements.

2.1.2 Tasks

Real-world task: The real world version of this test consisted of a curved wire, 450mm in length and 2mm in diameter, suspended between two vertical side supports at a height of 140mm above the table. Using the non-preferred hand, the participant held a rod on the end of which was a circular wire loop (80mm diameter) and was required to guide the loop along the wire as quickly as possible but without touching it. Contact between the loop and the wire (an error) produced feedback in that the background screen lit up

Virtual reality task: The virtual environment was created using dVISE, and was run via a HP 715 workstation, using dVS. In the virtual version of the task the participant viewed a computer generated three dimensional simulation of the wire and its supports via a head mounted display. Participants controlled their movement along the wire by moving a 3-D mouse, and feedback was produced by lighting up the background in the VE.

2.1.3 Procedure. Participants were randomly allocated to one of three equal sized groups. All three groups were tested on the real world wire loop task before and after training but differed in terms of the type of training given in between. For Group 1 training consisted of eight trials on the real-world wire loop task. Each trial consisted of moving the loop along the curved wire from left to right and then returning along the wire to the start position. Between each trial each participant had a one minute rest. Group 2 training consisted of eight trials on the virtual version of the task. As in the real-world task, participants completed eight trials interspersed with one minute rest periods. Participants in the Group 3 no-training control spent the period between pre and post training measures on a non-related task (this time period was based upon pilot data which showed that 15 minutes was the average time taken to complete either real-world or virtual training).

2.2 Results

As the pre-test error scores (baseline) had a wide between participants variation (range 10 to 127) the baseline scores were partialled out from the analyses.

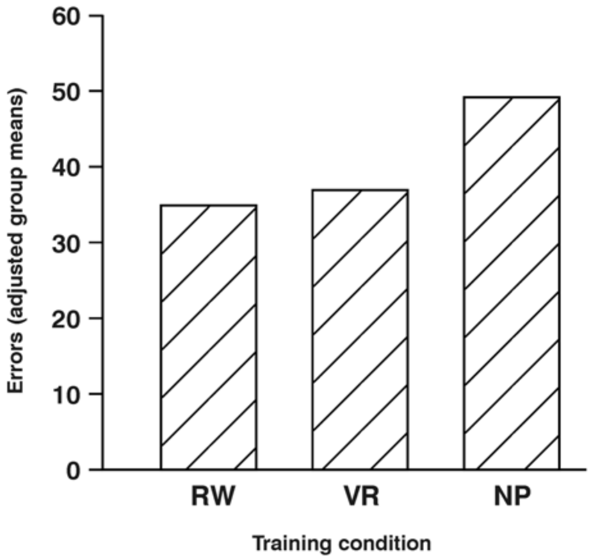


Figure 1. Adjusted group mean error performance scores (after baseline error scores were partialled out) for real, virtual and no training groups.

As can be seen from the adjusted group means in Figure 1, more errors were made in the no practice training condition than in the other two conditions. A one-way analysis of covariance, using baseline error scores as the covariate, showed that there was a significant difference between training conditions $F(2,147)=17.00$, $p>0.0001$. Planned comparisons showed that significant differences existed between the real-world practice condition and the no practice condition ($p>0.01$), and between the VR condition and the no practice condition (p>0.0.01). There was no significant difference between the real-world practice and VR conditions ($p<0.05$).

3. EXPERIMENT 2

3.1 Method

3.1.1 Participants. 100 university staff and students (mean age=30.9, SD=6.3, 60 women and 40 men). These participants had taken part in Experiment 1.

3.1.2 Procedure. Participants were randomly allocated to one of two groups, a concurrent motor task condition or a concurrent cognitive task condition. The task was concurrent with the carrying out of a single trial on the real-world loop task. As before, the trial consisted of moving the loop along the curved wire from left to right and then returning along the wire to the start position with their non-preferred hand.

Participants in the concurrent motor task condition tapped a Morse-code key with the middle finger of their preferred hand. The key had to be tapped at the same tempo (two per second) as that heard on a pre-recorded audio-cassette tape.

Participants in the cognitive concurrent task condition listened to a pre-recorded audio-cassette tape which presented 40 words at three-second intervals. Interspersed with these words were names of fruit. Each time the participant heard the name of a fruit they had to say yes.

3.2 Results

For the purposes of analyses, participants' post-test error scores following real-world or VR training undergone in study 1 were used as the baseline to measure the effect of the concurrent task variable. As this baseline also had a wide between participants variation (range 3 to 79) the baseline scores were partialled out from the analyses.

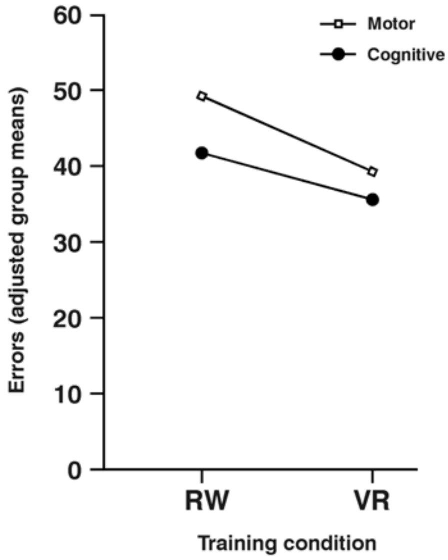


Figure 2. Adjusted group mean error performance scores (after baseline error scores were partialled out) for real and virtual training groups when carrying out either a motor or cognitive concurrent task.

The adjusted group means in Figure 2 indicate that carrying out a concurrent motor task led to a greater number of errors in the loop task than carrying out a concurrent cognitive task. It also appeared that introducing a concurrent task had a greater effect on the performance of participants previously trained on the real task than on the performance of those previously trained on the virtual task. However, this effect was not statistically significant. A two by two analysis of covariance, using post-test error scores as the covariate, showed that there was a main effect of type of concurrent task (motor vs. cognitive) $F(1,95)=4.22$, $p=0.043$, but no main effect of training condition (real vs. virtual) $F(1,95)=0.734$, $p=0.394$. No significant interaction occurred between the training condition and concurrent task $F(1,95)=1.313$, $p=0.255$.

4. DISCUSSION

In predicting transfer of training from a rehabilitation programme to subsequent real world performance Parenté and Hermann (1996) have drawn a distinction between "task elements" (broadly speaking the sensory and motor elements of a task referred to in the introduction), and "organisational set" (the cognitive processing demands of a task previously referred to). Thus the best possible transfer is predicted when both the task elements (A) and the organisational set (B) in the rehabilitation programme and in the real world environment in which the patient must subsequently operate are identical. An example would be retraining a stroke patient with prospective memory problems in the sequence of actions needed to cook a meal in his/her own kitchen. For resource reasons it is rarely possible for rehabilitation to be individualised in this way. The nearest approximation which is usually possible is when the real world task elements and the organisational set are similar but not identical to those rehabilitation situation (A' and B' rather than A and B). These two situations are represented below:

Rehabilitation Programme		Real World		Transfer
Task elements	Organisational set	Task elements	Organisational set	
A	B	A	B	Very high
A	B	A'	B'	High

In this study we have made the task elements in the virtual and real training situations, not identical, but as similar as possible. For example, as well as making real and virtual visual displays almost identical we modelled the handle of the metal ring the participants had to move along the wire on the handle of the 3-D mouse they used in the virtual task. In this way we have sought to focus on the organisational sets required in the virtual and real tasks. Since we obtained very high levels of transfer it is reasonable to conclude that, within the constraints of the investigation, the organisational set (or cognitive processing strategy) required to learn the virtual and real versions of the steadiness tester are very similar.

This finding does not exhaust the questions we need to answer regarding virtual to real transfer if virtual training is to become a reliable tool. Although the organisational set involved in learning in the virtual and real worlds may be similar enough to support high levels of transfer of training, the cognitive loads associated with operating the organisational set may differ between the two situations. More specifically we predicted that the "cognitive cost" associated with virtual training would be greater than real training and that this might be reflected in post training real world performance being less reliable or "robust" in virtual trained participants than in those trained throughout on the real task. If true this would clearly be an important factor to take into account in deciding when to use virtual training. We sought to investigate this cognitive load hypothesis by introducing both motor and cognitive interference into a post-training test trial.

In the event we found that the motor interference was more disruptive than the cognitive interference. This is perhaps unsurprising given that the steadiness tester task has a high sensorimotor component. With a more obviously cognitive task the effect may have been reversed. There was no evidence that either type of interference was more disruptive for those participants trained in the virtual task than those trained in the real task. The results do not support our "differential cognitive load" hypothesis, therefore, and taken at face value would lead us to be reassured that virtual training is as reliable as real training. We recommend caution, however. Firstly, and as we have just observed, the task employed in this investigation is predominantly a sensorimotor task. With a task which could be considered more demanding in terms of cognitive load, a different result might have been obtained. Secondly, within the constraints of this study we were only able to investigate the disruptive effects of one motor and one cognitive interferer. Just as training task nature and difficulty need to be varied, so it is necessary to examine a range of levels of intrusiveness of the interference. In the absence of a more complete, parametric, investigation we believe firm conclusions would be premature.

Currently we are extending our investigations to take account of these considerations in healthy volunteers as well as studying transfer of training from virtual to real environments in people with definable cognitive impairments.

5. REFERENCES

R.W. Bruce (1933), Conditions of transfer of training. *J. Exp. Psychol.*, **16**, 343-361.

A.S. Carlin, H.G. Hoffman and S. Weghurst (1997), Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. Behav. Res. Ther., **35**, 153-158.

S.M. Cormier and J.D. Hagman (1987), Transfer of Learning. Contemporary Research and Applications. Academic Press, London.

J.J. Cromby, P.J. Standen, J. Newman and H. Tasker (1996), Successful transfer to the real world of skills practiced in a virtual environment by students with severe learning difficulties. In: *Proc.: First European Conference on Disability, Virtual Reality and Associated Technology.* (P Sharkey Ed.) p103-107.

M.S. Darrow (1995), Virtual reality's increasing potential for meeting needs of persons with disabilities: What about cognitive impairments? In *Proc.: Third Intnl. Conf. Virtual Reality and Persons With Disabilities*, (H J Murphy, Ed.), San Francisco.

N.I. Durlach and A.S. Mavor (Eds.) (1995), *Virtual Reality. Scientific and Technical Challenges.* National Academy Press, Washington, D.C.

R.V. Kenyon and M.B. Afenya (1995), Training in virtual and real environments. *Ann Biomed Eng*, **23**, 4, pp445-455.

J.J. Kozac, P.A. Hancock, E.J. Arthur and S.T. Chrysler (1993), Transfer of training from virtual reality. *Ergonomics*, **36**, 7, pp777-784.

L. Mowafy and J. Pollack (1995), Train to travel, *Ability*, **15**, pp18-20.

A. Newall (1980), Reasoning, problem solving and decision processes: the problem space as a fundamental category. In *Attention and Performance 8*. (R. Nickerson Ed.), Lawrence Erlbaum, Hillsdale, NJ.

M.M. North, S.M. North and J.R. Coble (1997), Virtual reality: An effective treatment for psychological disorders. In: *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation*. (G Riva Ed), Netherlands: IOS Press, 59-70.

C.E. Osgood (1949), The similarity paradox in human learning: a resolution. *Psychol. Rev*., **56**, 132-143.

R. Parenté and J.K. Anderson-Parenté (1990). Vocational memory training. In *Community Integration Following Traumatic Brain Injury: A Functional Approach*. (J. Kreutzer & P. Wehman Eds.), Paul H. Brookes, Baltimore pp 157-169.

R. Parenté and A. DiCesare (1990), Retraining memory: Theory, evaluation and applications. In *Community Integration Following Traumatic Brain Injury: A Functional Approach*. (J. Kreutzer & P. Wehman Eds.), Paul H. Brookes, Baltimore pp 157-169.

R. Parenté and D. Hermann (1996), In *Transfer and generalisation of learning. Retraining Cognition. Techniques and Applications*. (R. Parenté & D. Hermann Eds.) Aspen Publications, Gaithersburg. pp 53-62.

L. Pugnetti, L. Mendozzi, E.A. Attree, E. Barbieri, B.M. Brooks, C.L. Cazzullo, A. Motta and F.D. Rose (in press), Probing memory and executive functions with virtual reality. Past and present studies. *Cyberpsychology and Behavior*.

J.W. Regian (1997), Virtual reality for training: evaluating transfer. In *Community Integration Following Traumatic Brain Injury: A Functional Approach*. J. Kreutzer & P. Wehman (Eds.), Paul H. Brookes, Baltimore pp157-169.

A.A. Rizzo, J.G. Buckwalter, U. Neumann, C. Kesselman and M. Thiebaux (1998), Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities. *Cyberpsychology and Behaviour*, **1**, 59-78.

F D Rose (1996), Virtual reality in rehabilitation following traumatic brain injury. In: *Proc. of the First European Conference on Disability, Virtual Reality and Associated Technologies.* (P Sharkey, Ed), 5-12.

F.D. Rose, E.A. Attree, B.M. Brooks and D.A. Johnson. Virtual Reality in Brain Damage: A Rationale From Basic Neuroscience. In: *Virtual Environments in Clinical Psychology: Scientific and Technological Challenges in Advanced Patient-Therapist Interaction*. (G. Riva Ed.) Netherlands: IOS Press, (in press).

F.D. Rose, E.A. Attree and D.A. Johnson (1996), Virtual reality: An assistive technology in neurological rehabilitation. *Curr. Op. Neurol.*, **9**, 461-467.

F.D. Rose, B.M. Brooks, E.A. Attree, D.M. Parslow, A.G. Leadbetter, J.E. McNeil, S. Jayawardena, R. Greenwood and J.A. Potter (submitted), A preliminary investigation into the use of virtual environments in memory retraining of stroke patients: Indications for future strategy?

R. Schroeder (1995), Learning from virtual reality applications in education. *Virtual Reality*, **1**, **1**, 33-40.

R.J. Seidel and P.R. Chatelier (Eds) (1997), *Virtual Reality, Training's Future? Perspectives on Virtual reality and Related Emerging Technologies*. Plenum Press, New York.

D. Stanton, P. Wilson, N. Foreman (1996), Using virtual environments to aid spatial awareness in disabled children. In: *Proc. of the European Conference on Disability, Virtual Reality and Associated Technology*, (P. Sharkey P, Ed), 75-84.

D. Strickland (1997), Virtual reality for the treatment of autism. In: *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical and Methodological Issues in Assessment and Rehabilitation*, (G. Riva, Ed) IOS Press, Netherlands, 81-86

E.L. Thorndike and R.S. Woodworth (1901), The influence of improvement in one mental function upon the efficiency of other functions. *Psychol. Rev.*, **8**, 247-261.

H.H. Wylie (1919). An experimental study of transfer of response in the white rat. *Behavioural Monographs*, **3**, (No. 16).

Developments of a collaborative research on VR applications for mental health

L Pugnetti[1], L Mendozzi[1], E Barbieri[1], A Motta[2], D Alpini[1],
E A Attree[3], B M Brooks[3] and F D Rose[3]

[1] Scientific Institute S. Maria Nascente, Don Gnocchi Foundation, via Capecelatro 66, 20148 Milan, ITALY
[2]Psychiatric Unit, Casa di Cura "Policlinico di Monza", Monza, ITALY

[3]Dept. of Psychology, University of East London, Romford Road, London E15 4LZ, UK

[1]*melau@tin.it*, [3]*E.A.Attree@uel.ac.uk*

ABSTRACT

The collaboration between or two scientific institutions is giving significant contributions to VR research into several fields of clinical application. Concerning the important issue of side-effects, future studies will clarify whether the encouraging results obtained in the recent past on patients with neurological diseases can be confirmed, and whether specific recommendations for the use of immersive VR in selected clinical populations can be made. Recent collaborative studies on the application of non-immersive VR to improve clinical testing of spatial memory provided evidence of good replicability of results in both healthy and neurologically affected groups. The development of retraining applications for spatial memory impairments and future studies aimed at assessing the impact of ambulatory disability on spatial cognitive abilities will be based on these findings. Finally, a newly approved transnational project will lead our groups into the field of the assistive technology to improve working skills and opportunities for employment of subjects with mental disabilities who seek a job.

1. INTRODUCTION

Though it has been promoted as a revolutionary tool, VR has not yet entered the practice and changed the methods of cognitive testing and rehabilitation. In the last years, our two groups in Milan (FDG) and London (UEL) have been developing the rationale to propose VR as a clinical tool (Rose et al., 1996), along with software programs and testing protocols to assess whether this technology can be safely used to produce meaningful results in both research and clinical contexts. This report deals with some the most recent developments of our collaborative projects.

The ARCANA system has been originally developed by the FDG group on a proprietary platform as a research prototype for testing so-called "strategy application disorders" (Pugnetti et al., 1995). An improved version is now being developed in collaboration with UEL which has an open configuration and runs on a standard fast PC workstation in both immersive and non-immersive modes. The paradigm has been conceived as an analog of the Wisconsin Card Sorting test (WCST) and, as such, has a well developed rationale and a known field of application. The psychometric characteristics of the preceiding immersive system have been recently published (Pugnetti et al., 1998). It has also proved useful to expand our investigations into aspects of VR-induced exploratory behavior and their psychophysiological correlates (Alpini et al., 1996; Pugnetti et al., 1996b; see also Alpini et al., this volume).

While the results of studies with the new versions are not yet available, here we discuss unpublished findings concerning side-effects reported by subjects performing with a previous version of the ARCANA system. Data are pertinent to an immersive VR system equipped as described in a previous publication (Pugnetti et al., 1995).

2. METHODS

A group of 36 outpatients with neurological impairments as a result of multiple sclerosis (MS; n.25), cerebrovascular disease (n.4), traumatic brain injuries (n. 5) and normal-pressure hydrocephalus (n.2), as well as 32 healthy subjects matched for age and level of education gave informed consent to participate in VR test sessions of 30 min. (max 45 min.) duration. Patients were recruited among those with stable neurological conditions, good bilateral visual acuity, preserved dominant hand dexterity, no history of epilepsy, psychiatric, and vestibular disorders, and no severe cognitive impairments. Before any test session patients and healthy subjects were carefully explained the aims of the research, the potential risks of immersive VR and the way they could interact with the virtual environment (VE). They were also interviewed about previous experiences with immersive VR, their susceptibility to common kinetosis and were given the opportunity to wear the headset (HMD) and practice with a VE for up to 15 min. while instructions were repeated to assure full comprehension. All practice and test sessions were carried out with the subjects seated comfortably on a revolving chair; this was necessary to equate controls' posture to that of some of the patients who could not stand easily because of their illness. They were instructed to rotate smoothly and slowly, and to avoid the combination of forward movements in the VE and real head rotations that produces a visuo-vestibular sensorial conflict. Subjects who reported any side-effect during or after the practice trial were reassured about the benign nature of the symptoms and let reexamine their decision to participate. Both training and test sessions were carried out in a small but quiet facility in the hospital. Test sessions were run usually the day after practice. The recording of side-effects was done according to the methodology reported by Regan and Price (1993, 1994). The Motion Sickness History Questionnaire was used to rate each individual's susceptibility to common kinetoses. A Malaise Questionnaire (MQ) was used to rate every 5 min. the presence of physical discomfort of the cybersickness type. A Simulator Sickness Questionnaire (SSQ, 26 items version) was compiled by the subjects prior and immediately after the VR session to rate the presence and severity of symptoms other than malaise. In addition, structured interviews based on the Equipment and Display Questionnaire (E&DQ) and a 6 items Immersion Questionnaire were carried out at the end of each test session to investigate ergonomic factors and subjective feelings. The reader is referred to the original papers by Regan and Price (1993, 1994) for details on the rating instruments.

3. RESULTS

No instance of severe malaise was observed during the practice sessions. Five subjects (7%) - 3 healthy controls and 2 MS patients - asked to discontinue the test because of severe nausea which always occurred between 10 and 20 min. from the start. Their condition did not require any treatment besides removal of the HMD and a brief rest. Their data were not included in the other statistics.

The prevalence of VR-induced side-effects of any type was measured by the pre- vs post-VR change on the total sickness score of the SSQ, and was around 7.5 points for both patients and healthy controls who completed 30 min. of testing. This figure was neither related to the severity of neurological condition (in the patients) nor to an unspecific sensitivity to kinetoses (for both groups). The maximum change was observed on Kennedy's (Kennedy et al., 1992) disorientation factor (mean of 12 points), followed by the oculomotion and nausea factors (< 6 points each). Patients' change scores did not differ from those of healthy controls, but patients reported more symptoms than controls on the pre-immersion questionnaire (Fig. 1). The time course of the malaise ratings (any symptom) followed the pattern already described by Regan and Price; it increased to reach a maximum of 40% of the subjects reporting at least one symptom after 25 min. of exposure, decreasing thereafter to return at baseline levels 10 min. after the end of the session (Fig. 2). According to the Malaise Questionnaire the mean prevalence of VR-induced symptoms across a 30 min. session was 16% in the total sample. Again, patients did not report more symptomatology than healthy controls. No clinically significant aftereffects were noticed on standard clinical balancing tests performed only on healthy subjects after the session. Since it was felt that clinical measurements could not pick up subtle changes, a pilot study using a computerized stabilometric platform was carried out on a subsample of 8 healthy subjects and 6 MS patients who performed on ARCANA1 and were matched to 10 controls and 7 MS patients who did not (Pugnetti et al., 1996). Results showed that immersive VR did not worsen static balance of either healthy controls or neurological patients, though the latter showed significant absolute impairments due to their illness. On average, 38% percent of the total sample (n.63) reported some negative rating concerning ergonomic factors of the HMD (8 items on the E&DQ) ; over 60% rated it as uncomfortable for a prolonged use, while 48% found it too heavy. On average, the efficiency of the display was rated somewhat negatively by 25% of the subjects, mainly due to the difficulty to adjust the focus and the interpupillary distance or to keep regulations steady, so that the images were felt to be unstable. The response of the tracker was also criticized by 25% of the users; they felt it too slow or imprecise for an optimal interaction with the VE. Fourty-five percent of the users were not satisfied with the pointer we have adopted (a hand-made light-weight woodden key incorporating the tracker). Only 21%, however, reported that the above ergonomic factors may have interfered with their performance. Interestingly, patients tended to complain less than healthy controls. From the analysis of the Immersion Questionnaire it emerged that 34% of the subjects did not felt "immersed" in the virtual environment (VE) during the session, whereas 43% did; in the remaining 22% the effect was modest. Only 14% felt that the virtual experience had changed their affective state toward excitation; 58% reported no change, whereas 27% reported only slight changes. Again, no significant differences between patients and controls were noticed.

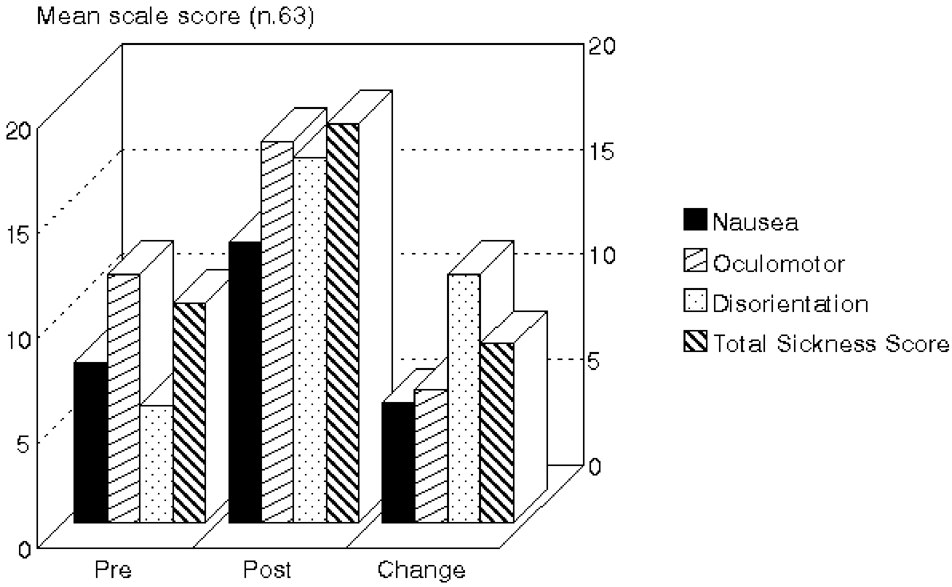


Figure 1. *Pre-immersion, post-immersion and change profiles on the SSQ for all subjects*

4. DISCUSSION

These findings are consistent with those reporting that immersive VR causes symptoms incompatible with the continuation of the experience in 5 to 30% of the users (Stanney et al., 1998). Since hardware factors are important determinants of the type and prevalence of VR-induced side-effects, our data should be compared only to those obtained with comparable systems and methodologies. Regan and Price's data (1993) seem to fit that criterion. These authors reported higher overall ratings of nausea both on the MQ and on the SSQ which were aministered to a larger sample (n.150) of healthy volunteers. Eight (5%) of their subjects withdrew from the experiment because of severe side-effects. The differences may be explained by a number of concurrent factors such as the weight and technical characteristics of the HMDs, but perhaps the most notable differences were that our subjects, unlike those in the largest Regan's study, were seated on a revolving chair during the experiment and that they were instructed to avoid movements that could exacerbate symptoms. In a further experiment comparing 44 subjects who sat while using VR with 24 subjects who stood, the same authors did not find significant differences in malaise ratings (results reported by Kolasinsky, 1995). Further studies are needed to clarify this issue.

The time-course of malaise ratings was also different; in Regan's study symptoms were reported to last longer after reaching their maximum at the end of the VR period (20 min.). In our study, symptom reports tended to decrease after 25 min., when subjects were still interacting with the VE, and returned to baseline levels at the 10 min. post-VR rating point. Therefore, our findings do not totally support the conclusion that malaise ratings increase steadily as a function of immersion time. The occurrence of a within-session adaptation may also be considered. Our hypothesis is that adaptation may be associated to the changing pattern of interaction that we have shown to occur with our VR task; i.e. the permanence in each successive virtual room decreased as subjects routinized their strategy and reduced exploratory movements (head and body rotations) in the second half of the session (Pugnetti et al., paper presented at MMVR6, S. Diego, CA, January 21, 1998).

Perhaps the most relevant finding of our study is that neurological patients - who may be considered at greater risk to develop side-effects from immersive VR - are, in fact, no more susceptible than matched healthy subjects. It should be noticed, however, that this conclusion is based mostly on subjective reports, and that they do not necessarily imply that immersive VR can be already safely introduced in a clinical setting. Our findings on the maintenence of static balance after VR exposure confirm Regan and Price's results (1993, 1994), but appear at variance with other studies (Di Zio and Lackner, 1997) reporting significant - albeit transient - impairments of static balance after immersive VR. Admittedly, our data need confirmation on larger patient samples. It seems, however, that comparisons across studies employing such diverse hardware, software, time of exposure, postures during exposures, measurement devices and criteria must be considered very cautiously. Replication studies are needed in this important area of VR research. The development of the new immersive and non-immersive versions of the ARCANA paradigm in collaboration with the UEL group will allow the planning of studies aimed at revisiting the main questions raised by our former experiences: whether immersive VR can really be safely proposed for clinical use, and whether immersive and non-immersive versions differ in terms of psychometric yield.

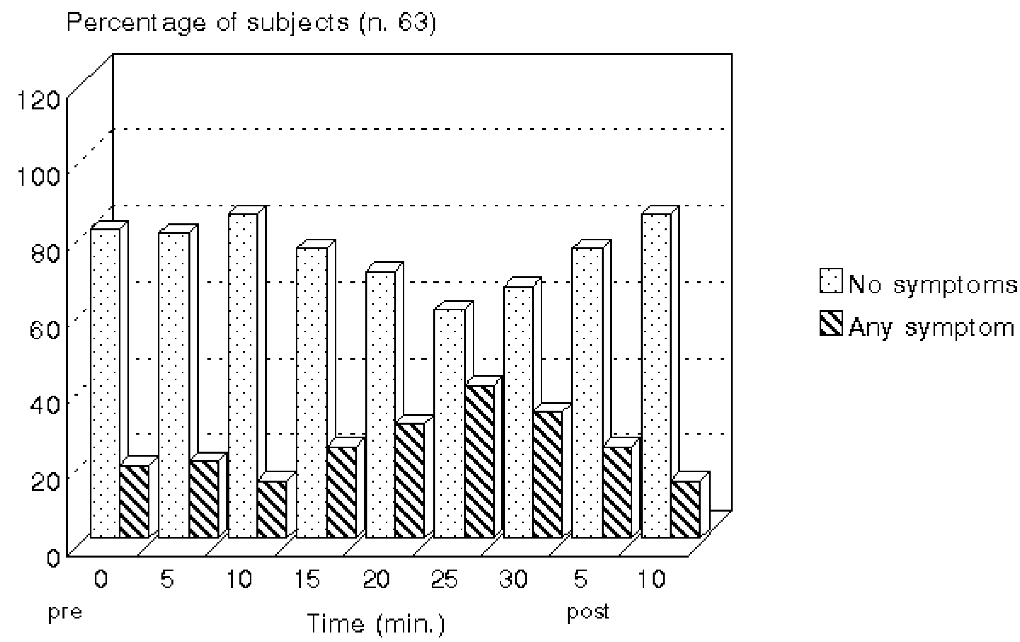


Figure 2. *Percent of subjects reporting at least one symptom on the MQ as a function of time*

5. VR FOR MEMORY RESEARCH

The demonstration that VR is an efficient tool for diagnosing memory deficits and their retraining is also actively pursued by our groups. Most, if not all, the evidence concerning visuospatial memory deficits in neurological patients has been collected using two-dimensional, non-interactive stimuli (e.g. pictures) in typical "paper-and-pencil" tests. These studies have generally found impairments on tests of egocentric spatial orientation and on tests of anterograde memory for visuospatial (topographical) information, all of which exclude locomotion and/or extensive exploration of large-scale spaces. This is unfortunate, because the integration of sensorimotor information has been shown to be important in the development of a cognitive representation of space and, specifically, of the external environment (Kirasic, 1991). The most essential use of spatial knowledge is that of assisting the interaction between an individual and his/her surrounding space. For example, memory for a layout should facilitate route finding whereas memory for objects should serve other purposes. In fact, there is growing neurobiological evidence that cortical areas and pathways involved in processing spatial coordinates (*where?*) are rather distinct from those processing shapes or patterns that mediate objects recognition (*what?*) (Epstein and Kanwisher, 1998). VR lends itself to devise experiments employing true 3D visual stimulation and exploration of large-scale spaces while assessing the differential effect of being an active or passive participant on different memory systems. Such studies have been pioneered by the UEL group in London, and have recently been replicated - for the first time - in Milan using the original yoked-control methodology described by Andrews et al. (1995) and Attree et al. (1996) in a non-immersive VR setup. The replication study investigated whether exploration of computer-generated environments can selectively enhance spatial memory in patients with MS (Pugnetti et al., 1998b). The hypothesis was that active subjects would show a better recall of the spatial layout of the environments they explored, whereas passive subjects would show a better recall of the contents of the VEs. Consistent with the hypothesis and results of the previous studies, 15 patients and 15 healthy subjects who controlled their movements in a virtual house using a joystick recalled the spatial layout of the environments better than 15 patients and 15 controls who merely watched the active participants' progress (Fig. 3). Among passive subjects, only healthy controls did significantly better than active participants in the recall of virtual objects. There were no significant differences between active and passive participants' recall of correct object locations in the virtual environments. MS patients' recall of the spatial layout and of the virtual objects was significantly worse than that of healthy subjects, but patients' data did not correlate with traditional neuropsychological measures of spatial memory on which MS patients have been shown to be impaired (Beatty et al., 1988). We concluded that VR can be used to test aspects of spatial memory that are not measured by traditional tests. We have also reported an enhanced effect of being an observer in a second experiment in which a different group of 26 MS patients were asked to recognize pictures of the objects they had incidentally memorized while exploring the same VE (Pugnetti et al., 1998). In that study we found the time dedicated to VE exploration to be directly related to recognition memory in patients but not in healthy controls, whereas the ability to handle the joystick to navigate the VE was slightly impaired in some of the patients and may have selectively influenced their ability to freely recall the objects they have seen. These results have implications also for the design of future clinical VR applications based on standard PC platforms.

Though the precise reasons why an enhanced spatial memory occurs after active exploration are uncertain (see Brooks et al.,1998 and Pugnetti et al., 1998b for discussions on this issue), it appears that the use of VR-based simulations can help at least the identification of the conditions and factors - other than known disease-related variables - that favor or mitigate the expression of memory deficits in neurological disorders such as MS and stroke, as recently reported by Rose et al. (1997). This appears to be a genuine instance of added value of a VR application to a clinical diagnostic problem. We have shown that a form of spatial memory which has not been tested so far in MS patients because of the lack of adequate means is defective, but can be modulated by direct interaction with the environments, whereas object memory does not seem to benefit by not being involved in an active exploration. Future research could investigate whether there is any transfer to the real world of spatial knowledge trained with VR in selected MS patients, and whether memory for objects would be improved by active interaction with them. VR applications could also contribute to understand if severe motor disability *per se* can be a factor in the development of visuospatial deficits due to the restriction imposed on self-controlled exploratory activity.

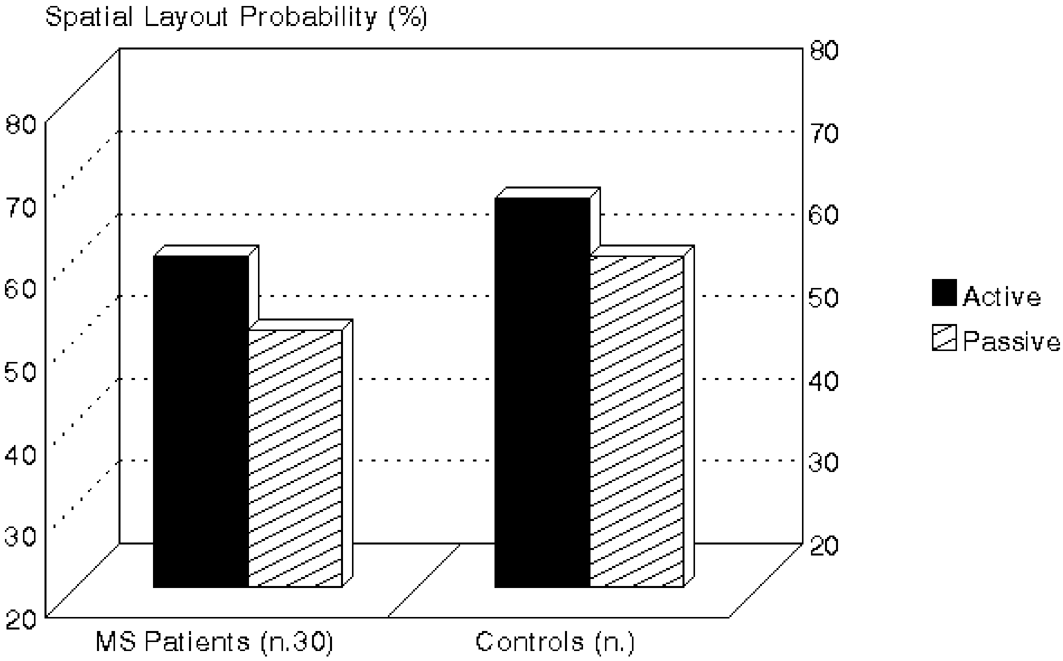


Figure 3. *Mean probability for recall of spatial layout in MS patients and healthy controls.*

6. THE VIRT PROJECT

On winter 1997 a project aimed at the development of a training tool based on VR to assist subjects with mental disabilities who are seeking employment, has been approved by the Italian Ministry of Labour and Social Affairs. The project, named VIRT (for Virtual Reality Training), is led by CIRAH, a non-profit association based in Milan that supports initiatives for the disabled who are eligible for a job application. Our two institutes - Fondazione Don Gnocchi and University of East London - participate in the VIRT consortium as partners along with national cooperatives (IL MELOGRANO and CSLS) and national associations from Spain (FEPROAMI) and France (UNAPEI and QUARTZ) that promote education, training, social and work integration of people with disabilities. The project is supported by the Horizon - Employment Initiative european funding program and has links with other projects, such as TIME, that also offer training opportunities to disables using new information technologies.

The main aim of the VIRT project is to assess whether low-cost VR technology is suitable to develop models that will supplement and improve current training procedures for subjects whose mental impairments do not totally preclude their integration in a productive activity. All too frequently subjects with a mental disability have limited access to new and more effective training tools based on latest technologies. As the latter becomes more and more available to non-disabled in most educational, training and work places, it will inevitably create more and more discrimination if educators, trainers and employers of the disabled are denied the opportunity to participate in the development of specific applications for their clients. This is a critical issue as far as equal access to resources for both disables and their tutors are concerned. The project will make educators, trainers and sensible employers aware of the potential of VR and will also make them responsible for a number of critical choices concerning the development and the use of the new training tool. In particular, their advice and decision will be important in the planning and refining phases of the development of the VR applications, while they will have major responsibility for the testing phase. We started from an analysis of the training curriculum and work experiences that are currently offered to employable mental disables in the participating countries. It emerged that the offer is far less than what is needed to cope with the number of potential trainees and the variety of working experiences that are necessary to make them aware of their role in a productive process and, more broadly, in a working organization. A flexible tool to simulate a wider range of working tasks than those available at the actual place will have a precise role in the training curriculum of disables. In general, it will serve to broaden their working experience, to foster their decisional autonomy, and to be more aware of their skills. More specifically, it will serve as an additional way to interact with tutors and workmates, to share experiences with them, to get reliable feedbacks, to apprehend difficult attitudes such as self-monitoring and self-correction trough exercise by trial-and-error, free of personal and material risks. The applications will combine VR scenarios and multimedial presentations in order to maximize the access to the different contents (e.g. instructions in different formats, examples of real tasks sequences, testing of transfer between a virtual and a real representation, etc.) and their flexible use. Tasks will be organized in coherent "modules" according to a topological specificity, which is currently maintained in every work training organization participating to the project. Accordingly, a variety of warehouse, workshop and office tasks are planned to be simulated, each with its own variants and difficulty levels. The trainers' role will be to configure each application and select the appropriate training schedule and methodology of approaching the VR system for each subject. A special effort will be made to provide the system with the wider possible capability to personalize the training in order to mimick what is currently done with real job training and insure that the psychological peculiarity of the relationship between the trainer and the trainee be maintained. As many as 60 disabled will be trained by the participant cooperatives during the two and a half years duration of the VIRT project. Data concerning feasibility, acceptance, side effects, impact on psychological and work organization factors, ways of interaction, transfer of learning, perception of utility and efficacy, data analysis and definition of performance will be obtained.

7. REFERENCES

D Alpini, L Pugnetti, A Cesarani, L Mendozzi and E Barbieri (1996), Evaluation of vestibular strategies in human exploration of the environment by means of immersive virtual reality, In *Giddiness and Vestibulo-Spinal Investigations* (CF Claussen, MV Kirtane, L Constantinescu & D Schneider, Eds), Elsevier Science, Amsterdam, pp. 315-323

TK Andrews, FD Rose, AG Leadbetter, EA Attree and J Painter (1995), The use of virtual reality in the assessment of cognitive ability, In *The European Context for Assistive Technology: Proceedings of the 2nd TIDE Congress*, IOS Press, Amsterdam, pp. 276-279

EA Attree, BM Brooks, FD Rose and AG Leadbetter (1996), Memory processes and virtual environments: I can't remember what was there, but I can remember how I got there. Implications for people with disabilities. In *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technologies* (PM Sharkey, Ed.), Maidenhead, UK, pp. 117-122.

WW Beatty, DE Goodkin and N Monson (1988), Anterograde and retrograde amnesia in patients with chronic progressive multiple sclerosis, *Arch.Neurol.*, **45**, pp. 611-619

BM Brooks, EA Attree, FD Rose and AG Leadbetter (1998), The specificity of memory enhancement during interaction with a virtual environment, *Memory*, in press.

P DiZio and JR Lackner (1997), Circumventing side effects of immersive virtual environments, In *Design of Computing Systems: Social and Ergonomic Considerations*, (M Smith, G Salvendy & R Koubek, Eds.), Elsevier Science, Amsterdam, pp. 893-896

R Epstein and N Kanwisher (1998), A cortical representation of the local visual environment, *Nature*, **392**, pp. 598-601

RS Kennedy, NE Lane, MG Lilienthal, KS Berbaum and LJ Hettiger (1992), Profile analysis of simulator sickness symptoms: application to virtual environment systems, *Presence*, **1**, *3*, pp. 295-301

KC Kirasic (1991), Spatial cognition and behavior in young and elderly adults: implications for learning new environments, *Psychol. Aging*, **6**, *1*, pp. 10-18.

EM Kolasinski, SL Goldberg and JJ Hiller (1995), Simulator sickness in virtual environments, Technical Report 1027, U.S. Army Research Institute for Behavioral and Social Sciences, Alexandria, VA.

L Mendozzi, A Motta, E Barbieri, D Alpini and L Pugnetti (1998), The application of virtual reality to document coping deficits after a stroke: report of a case. *Cyberpsychol.Behav.* **1**, *1,* pp. 79-91.

L Pugnetti, L Mendozzi, A Motta, A Cattaneo, E Barbieri and A Brancotti (1995), The evaluation and retraining of adults' cognitive impairments: which role for virtual reality technology? *Comp.Biol.Med.,* **25,** *11*, pp. 213-228

L Pugnetti, D Alpini, A Cattaneo, E Barbieri and L Mendozzi (1996), Computerized stabilometry in the assessment of some physiological after-effects of immersive virtual reality: a pilot study, In *Giddiness and Vestibulo-Spinal Investigations* (CF Claussen, MV Kirtane, L Constantinescu & D Schneider, Eds), Elsevier Science, Amsterdam, pp. 47-52

L Pugnetti, L Mendozzi, E Barbieri, FD Rose and EA Attree (1996b), Nervous system correlates of virtual reality experience, In *Proceedings of the 1st European Conference on Disability, Virtual Reality and Associated Technologies* (PM Sharkey, Ed.), Maidenhead, UK, pp. 239-246.

L Pugnetti, L Mendozzi, EA Attree, E Barbieri, BM Brooks, CL Cazzullo, A Motta and DF Rose (1998), Probing memory and executive functions with virtual reality: past and present studies, *Cyberpsychol.Behav.* **1**, *2,* pp. 151-161.

L Pugnetti, L Mendozzi, BM Brooks, EA Attree, E Barbieri, D Alpini, A Motta and DF Rose (1998b), Active versus passive exploration of virtual environments modulates spatial memory in MS patients: a yoked control study, *It.J.Neurol.Sci*, in press.

EC Regan and KR Price (1993), Some side-effects of immersion virtual reality, Report 93RO10, Army Personnel Research Establishment, Farnborough, Hampshire, UK.

EC Regan and KR Price (1994), The frequency of occurrence and severity of side-effects of immersion virtual reality, *Aviat.Space Environ.Med ,* **65,** pp. 527-530

FD Rose, EA Attree and DA Johnson (1996), Virtual reality: an assistive technology in neurological rehabilitation. *Curr.Op.Neurol.* **9**, pp. 461-467

FD Rose, EA Attree and BM Brooks (1997), Virtual environments in neuropsychological assessment and rehabilitation, In *Virtual Reality in Neuro-Psycho-Physiology,* (G Riva, Ed.), IOS Press, Amsterdam, pp. 147-156

K Stanney, G Salvendy, J Deisinger et al. (1998), Aftereffects and sense of presence in virtual environments: development of an R&D agenda, *Int.J.Hum.Comp.Interact.*, **10**, *2*, pp. 135-187

Virtual professional networks between speech pathologists

Magnus Magnusson

University College of Karlstad, Handikapp & Språk/ Utbildningsvetenskap
S-651 88 Karlstad, SWEDEN

Magnus.Magnusson@hks.se

ABSTRACT

All speech pathologists in one Swedish county will develop methods for supervision, therapy and development of professional methods, using ISDN-based videotelephony and Internet. The project is the first one of its kind and has initiated follow-up projects. It is primarily based upon research developed at the University of Karlstad, the Department of Disability and Language.

1. BACKGROUND

Speech Pathology is a fairly young discipline. The medical area of phoniatrics has existed for one hundred years but the professional group of speech therapists with its different subgroups has existed for a few decades, at least based upon formal and general education. Today, most countries have colleges for academical basic and further education for speech therapists. In Sweden there are four Universities which have four year courses for speech pathologists and also a master-course. It is also possible to go on to a PhD-exam. There have also been colleges for special teachers, specializing into the field of speech pathology, However, these colleges have been postponed and at the moment there is only one professional basic education whereby you can become a speech therapist in Sweden.

In the field of speech pathology there have existed several computer-based applications for therapeutical use since the late 1960's. However, the field did not really start to develop until the beginning of the 1990's when personal computers had started to be common tools at clinics, schools and hospitals. The development in Sweden has probably been a little more rapid since several speech pathologists started to develop language training software, mainly for Aphasia therapy already in the early 1980's. Especially one program called Lexia and its earlier version Afasi very rapidly became a household word for Swedish speech therapists (Gunnilstam, 1989). By 1990, most speech therapists in Sweden had access to that program in their daily work. There were also special 5-week courses available concentrating on computer usage for speech pathologists from the college in Stockholm, By 1996 about one hundred out of the 800 of Sweden's speech pathologists had passed that course and about 10 of the Speech pathologists had established themselves as part-time programmers.

However, most of the applications were directed towards use in the individual clinic, in the everyday use with patients or clients and not over a distance. Telematical applications were very rare among the speech therapists. A few professionals tried the text-telephone and even fewer tried out electronic BBS:s in the late 80's and the early 90's. There were a few pioneering activities in Norway, Sweden and the US. For instance, the journal CUSH (Computer Users among Speech and Hearing Therapists) was published in the US for almost 10 years and included many articles on computer applications in that field and a few articles on telematics as well. The earliest mention of telematical networking for professionals was published already in 1985 (Blache, 1985). Probably due to the relatively high cost of long-distance calls in the US, usage of BBS:es and fax and that sort of technology was a local thing for a relatively long time and no real results of investigations were published until the end of the 80's and the beginning of the 90's when a few practicians published reports almost simultaneously at the ASHA-conferences or in CUSH. Most of these presentations introduced special national BBS-systems as especially suited for professional exchange (Bull, Cochran, 1987/a, 1987/b, 1988/a-c; Dean, Pickering Jr, 1989, 1991; Krupke, 1990; Bull, La Pine, 1990).

A few years later in the US a few papers on the possibilities to make diagnostics (and therapy) using telematics were published (Wertz, 1992; Wertz & al, 1992).Most of their suggestions were based on assumptions and not so much on real work. The real breakthrough in the US came when the Internet was established and quite suddenly a lot of sites have been established, concerning speech pathology, including courseware and professional information (Goldberg, 1997).

In Sweden the development has been similar in many ways although the experiments with BBS-communication have been more widespread over the country. However, the professionals have waited to use electronic telecommunication systems until Internet came into use, but not disabled users. People with communication disabilities have been using BBS-.systems regularly since the mid-80's in Sweden. Today, about 10 professionals have their own private sites outside of any general site or server and probably at least 15% of all the professionals in the country have home-pages in connection with their employer's server or site (Johansson, 1995). Most professionals also have e-mail addresses. However, there are no networks in existence in that field yet.

In Norway the development has been somewhat diferent since the company Tandberg decided to develop a special videophone. The first pieces of equipment were available in the late 1980's and some of them were used in field experiments to evaluate professional supervision and therapeutical work in the field of speech pathology and other fields (Kristiansen, 1991). Today there are at least 100 videophones in use in Norway at special schools and clinics which means that there are possibilities to establish a network for synchronous multimedia distance communication in Norway, covering the whole of the country. The very first systematic language training trials over a distance were made in Norway (Holand, 1991).

Speech pathologists as a professional group in Sweden are greatly understaffed. Certain of the more unpopulated counties of Sweden have a few speech pathologists in one or two places only. This means that the distances are great between provider and patients to be. A new computer application could then be to try and give support to the patients over a greater distance, using different types of distance communication applications. So far, trials have been made using Internet, e-mail and fax communication, in other words more or less indirect communication means for people with language disabilities (Lifvergren, Lundell, Magnusson, 1997). However, during the last few years a few trials have been made, trying to establish more direct real-time communication over a distance, using videotelephony. The preliminary results show that this medium is accepted by the users and their spouses (Johansson, Magnusson, Wallin,1997).

From these brief pioneering examples we can make the following summary. The technology for distance communicaiton which has been most in use is a technology which uses text. It is also indirect or asynchronous or used for non-interactive information retrieval. BBS, Fax and Internet offer this type of communication with the promise of pictures and sound integrated during the last few years. However, very little technology offer synchronous multimedial real-time communication. Still the most common synchronous distance communication today is the auditive telephone call. The most recent distance technology, however, is the multimodal videophone technology which in its most recent form can be used as a computer-integrated technology. This type of technology seems to offer the best possibilities to simulate a real meeting betwen people and to offer optimal multichannel forms of communication. Unfortunately, the networks which are available today offer limited resources in transmission rate so that it is impossible to offer an optimal quality in the transmission.

A few projects have been created during the very last few years to start to evalutate the new and not wholly developed technology. Several European projects have evaluated videotelephony for deaf and/or hearing-impaired people. A few European or national projects have studied the medium for language impaired people. The most wellknown Swedish project is called VITSI and studies language training and social networks for people with mental retardation or Aphasia who are using videophone technology (Magnusson, Gunnilstam, 1995; Johansson, Magnusson, Wallin, 1997; Brodin, Alemdar, 1996). This report tells about a few projects which are trying to develop professional and virtual networks for speech pathologists, using this multimodal technology.

2. GOAL

Four projects are cooperating at the moment from different parts of Sweden – TELELOG 1, TELELOG2, REGLOG & DISTANCE COURSE. A preliminary and general goal is to establish a few local or regional networks with speech pathologists who learn how to use the videophone technology as a natural tool in their everyday work. The projects have separate subgoals:

- TELELOG 1: to teach the speech pathologists from one county how to use the technology
- TELELOG 2: to evaluate the technology as texttelephones for the same group and their patients
- REGLOG: to support part of the education at the speech pathology college of Umeå with videophones
- DISTANCE COURSE: the speech pathology college at the Karolinska Institute in Stockholm will give a videophone-based course to 12 speech pathologists from all over Sweden

There are more projects in the planning but they are not presented in this context.

3. METHOD

In 1997 the county council of Värmland in the eastern part of Sweden (population 260.000 inh) started the project TELELOG 1. The project was aiming to connect the speech pathologists working in the county through a network built on desktop videoconferencing equipment. During one year the speech pathologists would be using the systems in the network to see what type of situations that could be handled using videotelephony. During the last decade the computer has become an everyday tool in the work of speech pathologists, mostly since there has been a lot of training software developed. The computer is used both at the clinic and in the home of patients and it is quite common that a patient gets a training disk to take home and install it on the private computer (Magnusson, Gunnilstam, 1995).

Building on these results, the speech pathologists of Värmland intend to learn how to use this medium during one year. After that, in the autumn of 1998, they will begin to use the medium in regular trials with patients in different places in the county and also in the country, since colleagues all over the country have expressed interest in the project. Today, ten systems are installed, including one system in the home of the project-leader. All the clinics can use 20% of their work-time for the project. They have also received basic training in the usage. This means that all speech pathology cliniques are connected with videophones. After the summer five systems will be installed in the homes or workplaces of disabled people. One of the systems has already been installed in the office of the local Aphasia Society. That will be the main content of TELELOG 2.

The common goal of the two projects is to study possible emerging new methodology within this new medium, and also the reactions from the therapists. The study will mainly be qualitative, to see whether the medium offers similar experiences to the users as the so called ”real worlds” meeting, where similarities and differences can be found.

In the autumn of 1998, the project will be connecting to a distance course from the College of speech therapists at the Karolinska Institute in Stockholm. That course will be worth five academic credits and direct itself towards 12 participating speech pathologists, from all over the country. Lessons will be held, using the videophone, possibly combined with the use of videorecordings. To be able to fulfill that project, videophones have been installed in the offices of at least two of the colleges for speech pathologists.

Finally, the REGLOG project will install nine systems in each of the nine clinics that are involved in the speech pathology college of Umeå. Three of the systems are already installed and will be in use in August 1998.

All the projects are using the same type of computer system: Pentium computer 200 Mhz or more; 32 – 64 MB RAM; 2 – 8 GB HD; 17” – 19” VGA Monitor; 24 – 32 CD-ROM; 1200 dpi scanner; Colour ink-jet printer; Still picture digital camera; Headset; Loudspeakers; Windows 95; WORD 7; and Photo Plus 4. Five of the systems use Picture Tel 200 videophone card and the rest use Intel Proshare. Every system is using ISDN Duo, that is 2x64 kb/s double channel transmission rate. Most of the systems also use Internet and E-mail over ISDN.

4. RESULTS

Most of the projects are young which means that there are few analyzed data available yet. At the time for this presentation there will be interview data available about the experiences from TELELOG 1 and REGLOG. Preliminary quick interviews have indicated that the participants are aware of the positive possibilities and that in TELELOG they are cooperating between the clinics in the county in formal as well as in informal ways and that they experience the technology as fairly easy to handle. It is also evident that some of the participants are experiencing possible new methodologies which they see as a result of their usage. The University of Karlstad will analyze the results and include them in their growing knowledge base on the usage of videoteleohony in distance education and therapy

5. REFERENCES

Blache, S. (1985) Computer Networking, Information Transfer and the Communication Disorders Specialist CUSH 1(1985)1, sid 17-28

Brodin, Jane, Alemdar, Ingegerd (1996) Videotelephones. A Tool for Facilitating Commu-nication and Social Integration for Persons with Moderate Mental Retardation, Stockholm University, Dept of Education, Technology, Communication and Disability Report nr 13

Bull, Glen L., Cochran, P. (1987/a) SPECIAL TALK- An Electronic Network for Clinicians SYNERGY '87, ASHA/American Speech-Language-Hearing Foundation Computer Conference, Houston, TX

Bull, Glen L., Cochran, P. (1987/b) Peer Support for Clinicians in the Field: AnElectronicConference SYNERGY '87, ASHA/American Speech-Language-HearingAssociation New Orleans, FA

Bull, Glen L., Cochran, P. (1988/a) An Electronic Conference for Special Educators and Clinicians Third National Conference on the Technology and Media Division of the Council for Exceptional Children, Baltimore, MD

Bull, Glen L., Cochran, P. (1988/b) Use of BITNET by Communication Disorders Program Paper presented at the ASHA/American Speech-Language-Hearing Annual convention, Boston, MA

Cochran, P., Bull, Glen L. (1988/c) Electronic Conversations among Special Educators Council for Exceptional Children, Washington D.C.

Dean, C. R., Pickering Jr, J. E. (1989) Accessability to BITNET for Communication Disorders Institutions Paper presented at the ASHA/American Speech-Language-Hearing Annual convention, St Louis, MO

Dean, C. R., Pickering Jr, J. E. (1991) Telecomputing in Communication Disorders CUSH 7(1991)1, sid 35-43

Goldberg, Barbara (1997) Linking up with Telehealth ASHA Fall sid 26-31

Gunnilstam, Olle (1989) Datorprogram för logopeder (Computer Software for Speech Pathologists), Nord T log fon 14(1989)2, sid 29-32

Holand, Unni (1991) Use of Graphic Communication in Distance Training of Patients with Aphasia Brussels, In: "Issues in Telecommunication and Disability", COST 219, sid 289-95

Johansson, J-E (1995) Logopeden på Internet, Stockholm, Logopeden och Datorn, unpublished

Johansson, Irene, Magnusson, Magnus, Wallin, Eva (1997) Videotelephony as a Training Tool for People with Aphasia and Mental Retardation, Aphasia In: Murphy, Harry (Ed): Proceedings from Technology for Disabled People, Los Angeles, 16-22.3

Kristiansen, Tove (Ed) (1991) A Window to the Future. The Videotelephone Experience in Norway., Oslo, Teledirektoratets Forskningsenhet

Krupke, David L. (1990) Making Connections - Networking: An Iowa Effort CUSH 6(1990)1, sid 54-56

LaPine, Peter R., Bull, Glen L. (1990) Electronic Conferencing: An Open Invitation to CUSH Members CUSH 4(1988)1, sid 6-13

Lifvergren, Johan, Lundell, Jonny, Magnusson, Magnus (1997) Internet for People with Aphasia In: Murphy, Harry (Ed): Proceedings from Technology for Disabled People, Los Angeles, 16-22.3

Magnusson, Magnus, Gunnilstam, Olle (1995) Remote therapy for Persons with Aphasia In:Murphy, Harry (Ed): Proceedings from 3rd conference on Virtual Reality and Persons with Disabilities, (1995), University of Karlstad

Wertz Robert T., Dronkers, Nina F., Bernstein-Ellis, Ellen, Sterling, Lisa K., Shubitowski, Yvonne, Elman, Roberta, Sheanut, Gregory K., Knight, Jon L., Deal, Robert T. Jon L. (1992) Potential of Television and Telephonic Technology for Appraising and Diagnosing Neurogenic Communication Disorders in Remote Settings Aphasiology, 6(1992)2, sid 195-202

Wertz, Robert T.(1992) Potential of Television and Telephonic Technology for Appraising and Diagnosing Neurogenic Communication Disorders in Remote Settings Aphasiology, 6(1992)2, sid 195-202

Virtual reality for blind users

Roman Berka and Pavel Slavík

Department of Computer Science and Engineering, Czech Technical University Prague,
Karlovo nám. 13, 121 35 Prague, CZECH REPUBLIC

berka@sgi.felk.cvut.cz, slavik@cslab.felk.cvut.cz

ABSTRACT

The paper describes a solution to the problem how blind users can work with the three dimensional information available in the computer environment. This serious problem gains more and more importance as the use of three dimensional information is permanently growing. An experimental system that allows such communication has been designed and implemented. This system can be used as a sort of kernel for applications of various kinds that deal with the problem of communication between a blind user and a three dimensional model.

1. INTRODUCTION

Virtual reality has gained more and more importance in recent years. This term covers a large spectrum of meanings which means that its interpretation is rather complicated in some cases. In this paper we will deal with the meaning that is used in the web environment and in similar applications. These applications cover for example virtual walkthroughs in virtual scenes. Both these applications and the web varieties use the 3D model of a scene the user works with. Unfortunately there is a large group of users that can take no benefit the possibilities offered by virtual reality. This group of users consists of blind users [1 , William A. Barry, 1994],[3 ,D. Hermsdorf, www]. As the share of graphical information (and 3D information in particular) in various media(especially on the web [2 , R. Berka, 1996]) steadily increases, this group of users is more and more limited in the usage of information resources, the importance of which is permanently growing. This situation requires an immediate solution. It is necessary to develop tools that will allow the blind user to obtain information about a 3D scene (or its model) in a form that will be suitable for such a user. Such a user can not perceive the scene in a global way as normal users can. The solution to this problem is to develop a special type of communication that will make it possible to ask questions of various kinds. The answers to these questions will provide the user with information both structural (e.g. which object is to the left of the current object) and geometrical (e.g. how far from the current position is the object B). Moreover, it is necessary to have some statements at hand by means of which the movement of the user from one position to another could be described. The system that would satisfy all these needs will be based on a special language, the statements of which will allow communication between the user and the system working with a 3D scene.

The communication with the 3D model can have various frameworks. Let us discuss some typical situations:

1. Navigation in a 3D virtual environment. As an example we can give the situation when a blind user tries to learn about some building he/she is going to visit the next day (bank, hospital, railway station etc.). By means of the system the user can obtain information about the structure of the building and thus some idea how to get from point A (let us say the entrance) to point B (let us say the wash-room). The user can train his/her movement in an unknown environment. The communication between the user and the system can be performed by means of standard input and output devices.

2. The extension of the previous example can be done in such a way that the user will use the model for real navigation in unknown environment. This will require some additional equipment that will allow him/her to establish their exact position in space. As the user moves in a building, where a lot of obstacles can be, it is necessary to establish the position with very high accuracy. One solution would be to have some signal sources (ultrasound, infrared, radio etc.) that could be received by the appropriate equipment attached to the computer carried by the user. Processing such a signal would establish the exact position of the user. Then the user can either ask questions about his/her position or can be automatically navigated to the final position. The situation described corresponds to the use of wearable

computers.The information obtained has a very close link to reality and the information from the 3D model can be projected to the real environment.

3. A modification of the above given approach can lead to the development of a special 3D browser for web application. This browser will be oriented to blind users. The idea is that the 3D scene description will be in VRML and the user can ask questions about the situation in the immediate vinicity of the current position, etc. In such a way the blind user can get information about the 3D scene. This information was not available to him/her when using the web in a standard way (the user could work mostly with textual information only).

2. THE TRAINING AND NAVIGATION SYSTEM EYE

Within the framework of our research, we have implemented the EYE system. This system satisfies the requirements stated above. Using the current implementation of this system a blind user can train his/her movement in an unknown environment (e.g. railway station halls, airport halls and other public accessible areas).

Based on a 2D map, the EYE builds a virtual scene and locates the user in this scene. The user can freely walk through the scene and the system informs him/her about important places, objects and obstacles. The system has two data sources: end-user interaction and the 2D map created by user-author. This approach is based on three components:

- a digital model of the scene
- the way the user communicates with the system
- the way the system communicates with the user

The digital model of the scene is reduced to a 2D form since it contains all information necessary to solve our task. In addition, the 2D solution is much easier then the use of the 3D algorithms used in virtual reality. The map represents necessary information including obstacles, significant objects and important paths among these objects.

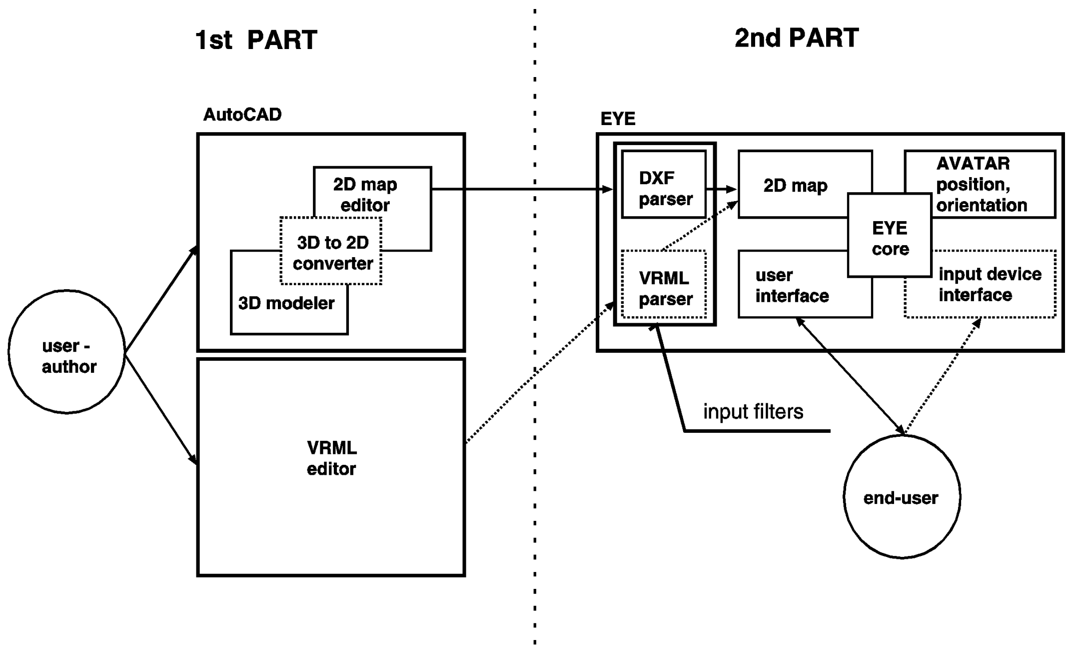


Figure 1: *The data flow through the system*

One of the most important parts of the system is the data preparation module. Since the data structure imported as a 2D map has to meet special requirements it is impossible to work with a general 3D model. There are two ways to obtain a suitable 2D data. First, the 2D map is generated by conversion from any 3Dmodel and certain additional information is then appended. Second, the 2D map is generated using a special software tool. We used the second way in our approach in which a special application for map creation was developed.

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

The communication between the system and the user and vice versa represents a set of problems due to the fact that the information will be presented to the blind end-user in a non-visual form using some screen-reading tools or a synthetic voice. In addition, it is expected that the set of potential users will be wide and that every user will have to learn to work with the system. Therefore, there are demands on the communication protocol between the user and the EYE. First, the data flow coming in direction from the system to the user has to be minimal as the user must be able to absorb all information given by the EYE. Second, the system is controlled by a set of commands in which the minimal necessary subset needed to work with the system is reduced to 3commands. All commands have to be very simple as they must be easy to memorise.

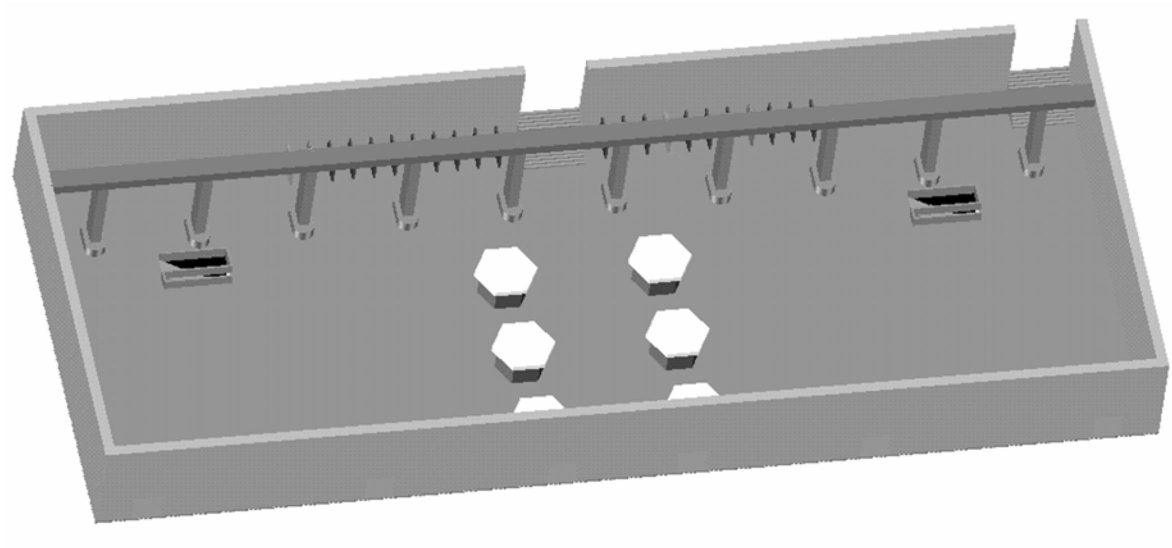


Figure 2: *3D model of Wilson railway station hall in Prague*

2.1 Two Parts of the System.

From the view of an author who generates input data, the system consists of two parts (see Fig. 1). The first is a data preparation process during which the scene description is created using text and geometrical data. The second part consists of importing a scene description to the system EYE and of the training process. Both parts are independent and are not necessarily parts of one program. The end-user then works with the second part only.

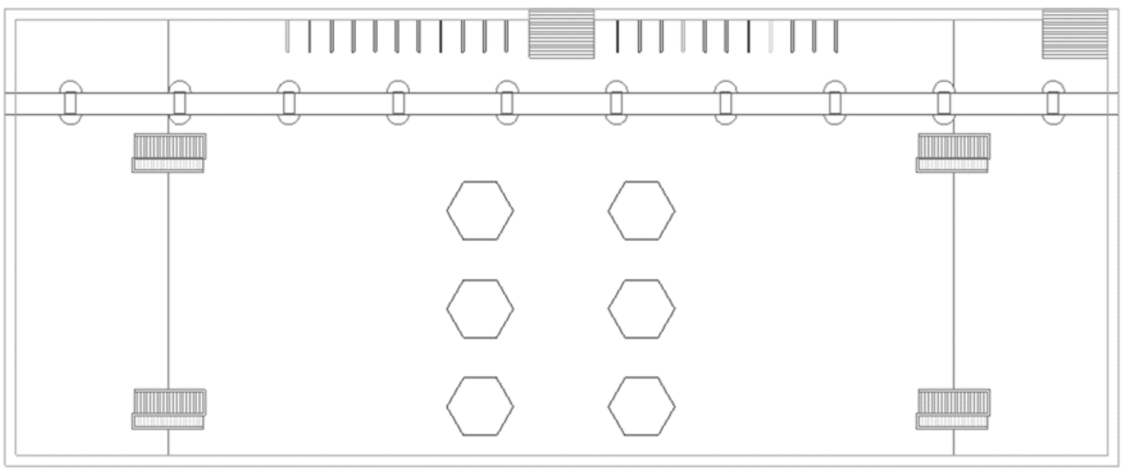


Figure 3: *Top view to model of Wilson railway station*

A special application called a "map editor" (see Fig. 1 - the first part) is used to create 2D maps according to rules which make the data readable by the EYE system. Generally, the map editor loads any geometrical description of the scene, converts it to the acceptable form and appends additional information. This information includes possible paths among important places of the scene and also the special attributes of all objects. These attributes have a text form. Every object in the scene has a set of predefined attributes with the meaning of names and short descriptions of the object (object functionality - e.g. "newspaper stand").

The above mentioned fact means that some part of the input information is removed as e.g. the third dimension of objects in the scene and the other information is added using a map editor. This editor makes it possible to add some attributes in text form. These attributes, in most cases describing the objects, are later included into messages sent to the user during conversation. Next, using the map editor the author of the map defines all possible paths between every two potential goals of the user. These paths are generated automatically by the map editor. Later, the author can edit generated paths by hand and correct them.

Fig. 2 shows a 3D view of part of Prague's Wilson railway station hall and Fig. 3 shows a view of the same scene from the top. Fig. 4 shows a 2D map corresponding to the previous view. This 2D map already includes added information.

All these pictures demonstrate the two individual steps of the data preparation process. During the first step, 3D input data are converted into the 2D description according to the special rules. In the next step, additional information is appended as described in the previous two paragraphs.

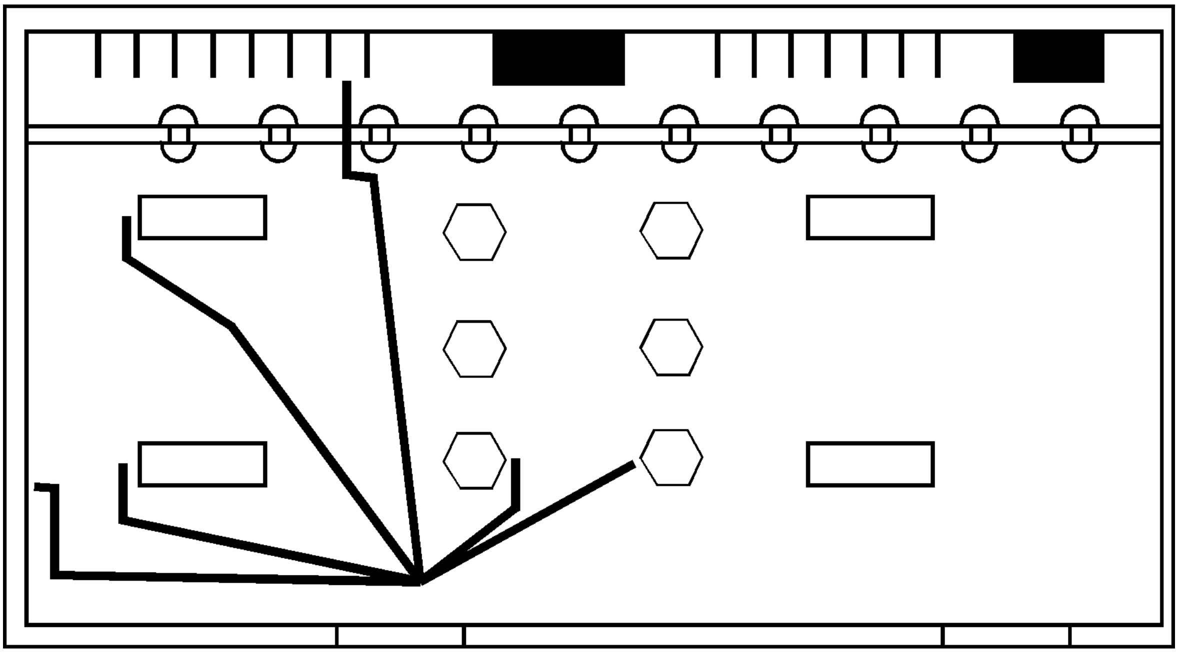


Figure 4: *Processed 2D map of Wilson railway station hall in Prague*

The second part is represented directly by the EYE system as shown in Fig. 1. It consists of modules communicating via a core which processes all data flows in the system. The first activity of the EYE after start up is reading the map. The input data representing the map is entered through input filters which transform this data into internal data structures. These filters should be specialised for any specific input data format. Our current version includes only one input filter for DXF (AutoCAD Drawing eXchange Format) used usually in area of CAD systems. To extend the possibility of processing another format (e.g. VRML), a new filter has to be implemented and linked to the system as shown in Fig. 1.

When the map is stored in the desired structure communication with the user can start. First of all, the initial localisation of the user in the scene should be made. Information about the position and orientation of the user are stored in an object representing the user in the scene. This object will be called AVATAR (according to terminology used in the area of VR). The AVATAR has a circular shape with a size corresponding to the projection of the human body to the ground. The core uses this representation to compute distances of the AVATAR from the goal position and to detect AVATAR's collisions with other objects in the scene.

The user interface module communicates with the user by means of natural language using short formulations in text form. The user's communication with the system consists of a small set of commands. These commands are transformed into internal instructions, in most cases changing position or orientation of the AVATAR. While the localisation of the AVATAR is based on information given by the user it is also possible to use a special input device interface which gets this information from devices measuring orientation and distances. Such a solution could also be used as navigation system in the real environment.

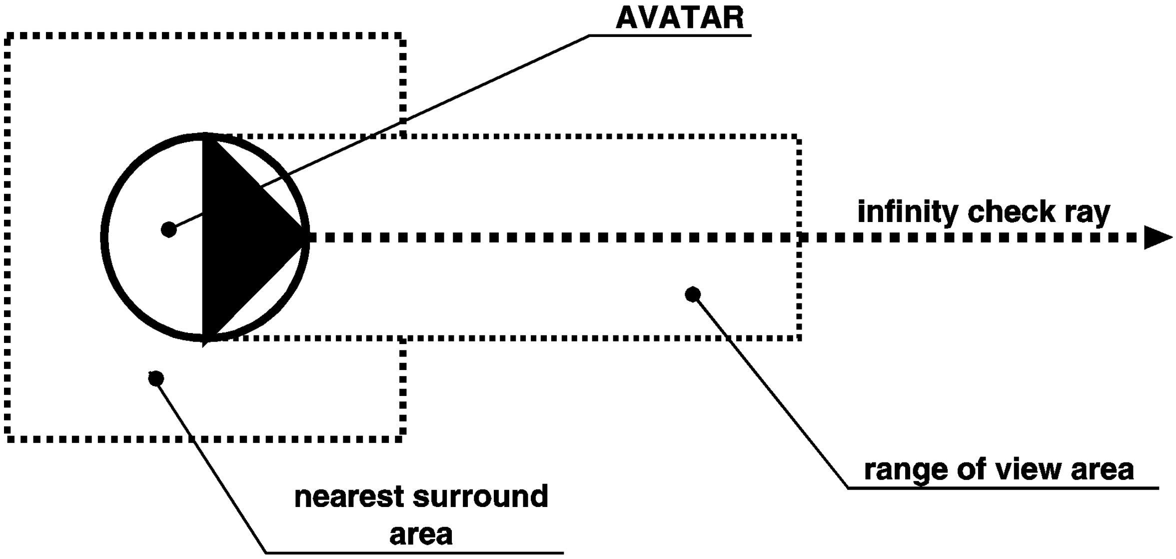


Figure 5: *The AVATAR and its sensitive areas*

2.2 The System - User Communication

All objects in the scene are potential obstacles for the user. Since the system is proposed for use by blind people the EYE recognises six classes of objects with respect to the danger level of collision with them (e.g. a door represents another class than a hole).

PORTAL - represents doors
ACCESSIBLE - represents objects accessible without danger (e.g. grass plot along the path)
WALL - represents general obstacles
HANGING - represents obstacles hanging in given height over the ground and user's head could be hit by the obstacle
HOLE - represents dangerous obstacles as holes in the floor
UNKNOWN - represents obstacles without specific description

All classes differ in reaction from the side of the system when the user collides with the obstacle (e.g. a collision with the door will evoke entering to an other room but on the other hand a collision with a wall evokes a warning message).

The AVATAR is equipped with sensors which are sensitive to the existence of obstacles in specified areas around him. As the AVATAR is walking through the virtual scene the EYE system informs him/her about obstacles which appear in his/her surroundings. Fig. 5 shows the AVATAR and his/her sensitive areas. The AVATAR has three active sensors. The first one checks the area in front of the AVATAR and corresponds to a healthy human's range of view. The second sensor works as a ray sent by the AVATAR into the space. The system then returns information about objects intersected by the ray. This sensor is equivalent to the vision to large distances needed for orientation in large space. The last sensor informs the AVATAR about objects in the nearest area surrounding him. The two last sensors are needed for orientation and they are evoked on request of the user. The first sensor is activated every time the AVATAR changes its position or orientation and gives information concerning which objects would be potential obstacles for the walking user/AVATAR.

As the AVATAR moves in the scene the system checks his/her collisions with all other objects and informs the user about objects in the scene and about the possibility of colliding with them. Since the potential user will communicate with the system using special peripheries for blind people the messages returned by the EYE should be as short as possible. This requirement is important for the data preparation process where most of the attributes have a text form and they are used as parts of messages sent to the user.

The user gives the system very simple commands from a small set (see this complete set in APPENDIX A). The basic two commands are used to change the position or orientation of the AVATAR. Using the other two commands the user can check either the immediate surroundings or distant area in front of him/her. The other commands are used to control the EYE system where the user can set the length of his/her step or the range of the area checked by the "range of view" sensor.

For easier communication the system uses special units which correspond to terminology used by blind people. The distances are measured in steps in which the step length is optional. The direction is given in a natural form in which the direction is derived from a clock-arm position relative to the AVATAR. Fig. 6 shows all possible directions with respect to AVATAR's orientation. Here, for example 3 hours equals east.

The initial communication of the user in the virtual scene starts with a list of possible goals in the current scene (see APPENDIX B). This list is offered by the system. After the user answers he/she enters the room and the system gives him/her basic information about the room, about his/her immediate surroundings and gives him/her directions to the chosen goal.

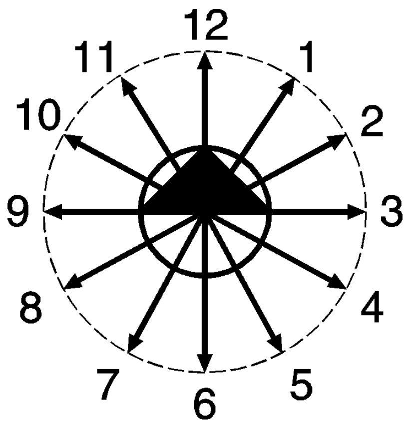


Figure 6: *Possible directions measured in clock-arm positions*

The user can freely walk through the scene and the system navigates him/her to the right path. When the user reaches the goal the system stops the navigation process and gives him/her a new list of possible goals.

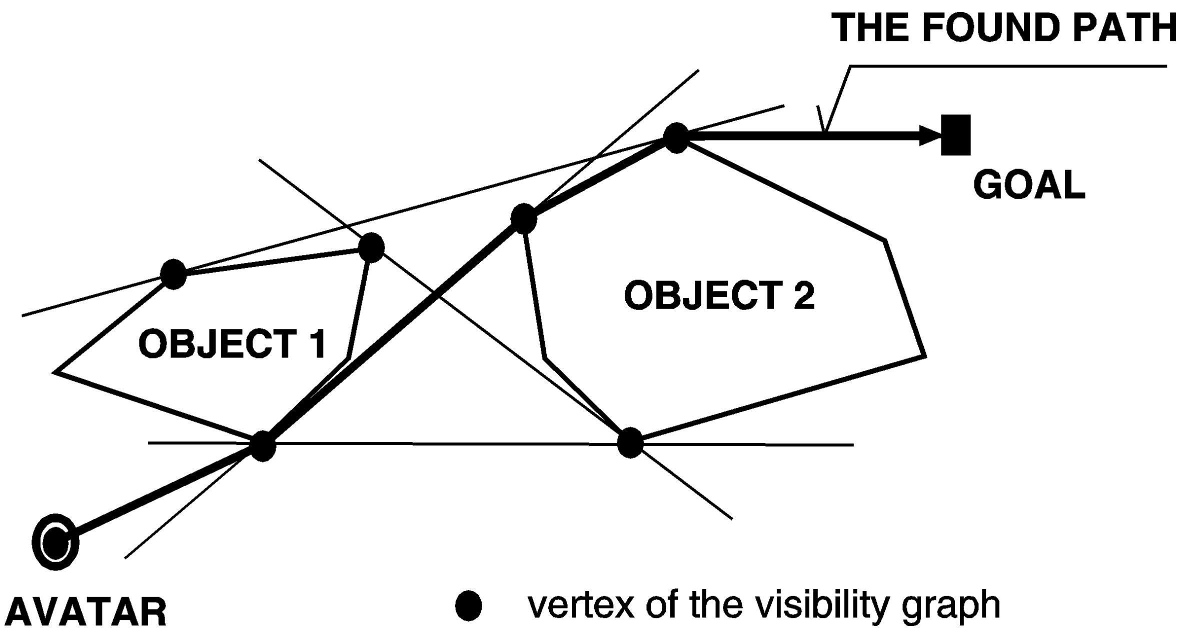


Figure 7: *Example using the tangent algorithm*

3. IMPLEMENTATION DETAILS

The system is realised in C++ programming language using object oriented technology where individual objects represent real objects in the system as described above. This approach is suitable for most similar problems solved in virtual reality. The following three main problems had to be solved in order to implement the EYE.

- integration of graphical and non-graphical data in one data format
- navigation algorithms
- communication peripherals

The first is not only a file format problem. It is a problem of data import from standard graphical systems. Since the data read by the EYE is required to include special information it was necessary to implement a special application under any standard graphical system which is able to produce data directly in the required format.

We have chosen AutoCAD as the drawing system and the application is now implemented using an ADS development extension. The DXF format meets the requirements specified above and therefore we used it as a format for our data. See Fig. 1 which shows the data flow through both parts of the system and between the system and peripherals.

The navigation algorithms use polylines representing the paths included in the 2D maps during data prepa-ration process. Every polyline has a set of attributes. Two of these attributes are labels at both end-points of every path. These labels correspond to the names of places or objects where the path is terminated. They are also used for user orientation in the scene. The user is navigated by the system using vertices of the polyline. At any time the user is informed about the direction to the visible vertex lying nearest the goal on the specified path.

During the data preparation process, the path can be created by hand or automatically using the shortest path searching algorithm implemented as a subpart of the first part in Fig. 1. The path searching algorith muses a visibility graph method [4 , F. Kuo-Chin, 1994] where the graph is created by a "tangent algorithm". The basic idea of the tangent algorithm assumes that there exist exactly four tangents between two non-overlapping 2D objects in a plane. Creating tangents between every two objects we obtain sets of vertices and edges (including edges from original objects and tangents) which define the desired visibility graph. The path between every two given points in the scene is then found as the shortest path in the visibility graph. The Fig.7 demonstrates the tangent algorithms on two general objects. The marked vertices of both objects are also vertices of the visibility graph. The some edges of the objects and parts of tangents are the edges of the graph.

4. FUTURE WORK

The system is still under development. In the near future all parts that are not yet finished will be completed and the system will be fully operational. In addition there are many possibilities for extending the system. These extensions would allow us to use the system and the results of the research performed in specific applications. The extensions will cover:

- the use of voice input and output. This will significantly ease the use of the system for blind users - especially in the case when navigation in a real environment takes place. The overall cost of the system will not increase significantly as the voice cards and speech synthesizers can be purchased at a reasonable price.
- the use of auxiliary sound navigation. This feature is like navigation in real 3D space when blind persons generate sounds when walking (e.g. the sound of their steps, clicks with their white stick etc..). Such a sound typically generates an echo that gives an idea about the size of the room in which the person is and also about its basic configuration. There are currently methods under development that will generate 3D sound in a virtual model described by means of VRML (or in another format). The development also takes into account the materials which the walls, ceiling, objects in the room, etc. are made from. This feature is of interest in case the user trains his/her movement in the training environment.
- the research performed up to now a was also targeted to the development of real 3D browser for the blind users. Besides system modules already existing it would be possible to include both "sound" modules described above. Such an approach will allow blind users to work to a certain extent with 3D information available on the web. This kind of information is not currently available for them to a satisfactory extent. Such a development will ease the use of the web considerably. This activity is linked with other research performed in our department where pictorial information is transformed into structured textual information what will allow the user to work with pictorial information as well. Integration of these two approaches will allow the development of special modules that will enhance

current web browsers. These modules will compensate in certain ways for the handicap of the given class of users.

- a very important field of application will be the use of modules developed in the environment of networks. Currently there exists a lot of applications where virtual worlds have been created (e.g. WorldChat). The users virtually meet in these worlds and communicate in a different way. As the 3D environment is described by means of standard formats (like VRML) it would be possible to enhance existing software with modules that will allow blind users to use these systems as well. Another aspect of communication in networks will be the navigation in actual space.

When speaking about the application of the system developed in the framework of wearable computers there is an important feature that could substantially extend the field of use of such a system. Up to now we have assumed that the 3D model is stored permanently in the computer (on the hard disk) or on a floppy disk with the model which will be used. The direction in which the system can be developed is the use of the system in the network environment. In the basic version of such a system the situation will be characterised by having the possibility of acquiring the 3D model via network. This will give an opportunity to dynamically obtain information about the environment where the user currently is. This mode of use lies in fact in the field of mobile computing where the connection between the user and database (where 3D models are stored) can be accomplished by means of GSM or similar technologies. This approach will allow the user to get very detailed information about the environment even in very complicated cases (multi-storey buildings, etc.).

5. CONCLUSION

The first two versions of our new training and navigation system EYE were developed with the goal of making a large amount of 3D data accessible to blind people. This data can be used by blind users learning to move in unknown environments. Every user can communicate with the system using a few simple basic commands and the EYE system answers with short simple sentences. Therefore, the communication should be easy to learn for most potential users. We tested the system with a blind user who found all the desired objects in the scene. The first results of our experiment indicate the strategy of navigation in the virtual scene is correct and can be used in the future development of the system.

The system developed can be used as a kernel for various applications (3D browser, navigation in a real environment, virtual walk-throughs etc.). Some of the results achieved were obtained in the framework of national project GACR 201/96/0706: Graphical Communication Between Human and Computer for Blind People [5 , P. Slavík, 1997].

6. REFERENCES

[1] William A. Barry and John A. Gardner. Accessibility to Scientific Information by the Blind. http://dots.physics.orst.edu/csun94.html, 1994. Department of Computer Science, Cornell University.

[2] R. Berka, M. Brachtl, A. Holecek, M. Novotny, J. Prikryl, P. Slavik, and J. aZra. Three-dimensional Information System on the World Wide Web. In Multimedia '96, pages 70-78, Wien, 1996. Eurographics, Springer-Verlag.

[3] D. Hermsdorf. Teleworking for Disabled People. http://www.yuri.org/webable/tedis.html. The German National Research Center for Computer Science: TEDIS, Sankt Austin, Germany.

[4] F. Kuo-Chin and L. Po-Chang. Solving Find Path Problem in Mapped Environments Using Modified A* Algorithm. IEEE Transactions on Systems, Man, and Cybernetics, 24(9):1390-1396, september 1994.

[5] P. Slavik and R. Berka et al. Graphical Communication Between Human and Computer for Blind People. Technical report, Czech Technical University Prague, 1997. national project GACR 201/96/0706, pp. 69.

APPENDIX A. THE COMMAND SET

The end-user can communicate with the EYE using ten simple commands. These commands are divided into three categories. The first category includes commands changing the position or orientation of the

AVATAR. The second category consists of commands requiring any information and the last class represents a group of system commands. Here is the list of all commands:

1. step [number of steps] - move a number of steps forward
2. turn [clock] - turn on the clock
3. back [number of iterations] - return a number of iterations back
4. look - give information about objects in the "range of view" area
5. describe - give summary information about the current room
6. infinity - give information about objects positioned far from the AVATAR
7. check - give information about objects in the immediate surrounding
8. range [number of steps] - sets the length of strip representing the "range of view" area
9. preset [distance in meters] - sets the user's step length
10. quit - terminates the program
11. ? – help

The user can type only the first character of every command for faster communication.

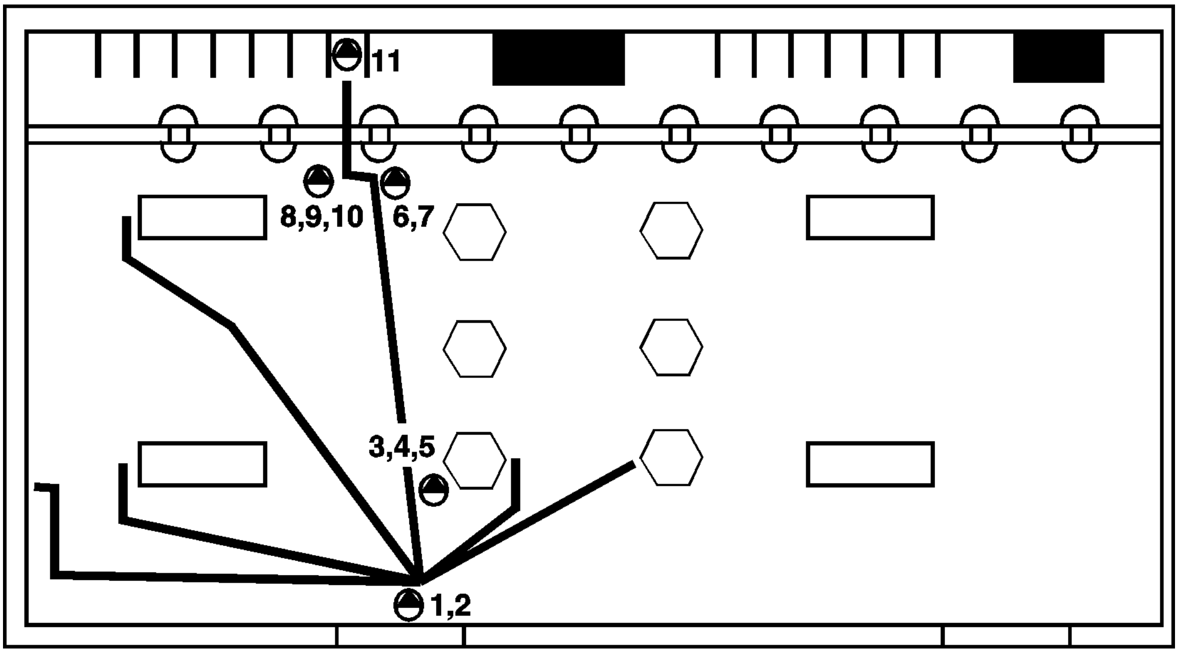


Figure 8: *The dialog corresponding with corresponding iterations between the user and the system*

APPENDIX B. EXAMPLE OF TYPICAL DIALOG

A typical dialog between an end-user and the system starts, as mentioned above, through the selection of a goal. Then the user gives the system commands changing its position until the goal is reached.

Fig. 8 shows the individual phases (iterations) of AVATAR's motion, and the corresponding sequence of commands follows. Fig. 8 describes situations after commands given by the user. The AVATAR is represented by symbols as on Fig. 5. Every symbol is marked by a set of numbers and every mark means a situation after a command with the corresponding number in parentheses.

The text of an example is written according to rules such that all messages given by the EYE are typesetby *italics* font and for all commands received from the user, the **bold** font is used.

This is EYE the Virtual Reality navigator for blind people version 2.0

Description of the room:
The Wilson railway
station hall in Prague.
The door: north left entrance from park

Choose one path from the following list, type the corresponding number and confirm by pressing of ENTER, please.

0 ... ticket offices
1 ... underground toward Holesovice railway station
2 ... underground toward the center and South Town
3 ... post office
4 ... newspaper kiosk
5 ... fast food

path number: **0**
from: north left entranceto: ticket offices

command (1): **l**

No objects detected in your range of view.The direction to your goal is 0 o'clock.The distance to the goal is 100 steps.

command (2): **t 2**

No objects detected in your range of view.The direction to your goal is 10 o'clock.The distance to the goal is 100 steps.

command (3): **s 20**

objects in your range of view:fast food
The direction to your goal is 11 o'clock.The distance to the goal is 82 steps.

command (4): **c**

objects in your immediate surroundings:fast food IN FRONT of you on yourRIGHT hand.

*command (5):***t 11**

No objects detected in your range of view.The direction to your goal is 0 o'clock.The distance to the goal is 82 steps.

command (6): **s 65**

objects in your range of view:concrete column
The direction to your goal is 9 o'clock.The distance to the goal is 15 steps.

command (7): **t 9**

No objects detected in your range of view.The direction to your goal is 0 o'clock.The distance to the goal is 15 steps.

command (8): **s 5**

No objects detected in your range of view.The direction to your goal is 3 o'clock.The distance to the goal is 13 steps.

command (9): **t 3**
No objects detected in your range of view.The direction to your goal is 0 o'clock.The distance to the goal is 13 steps.

command (10): **i**

far objects in infinity:ticket office

command (11): **s 13**

Your goal has been achieved,do you wish to continueto next possible goal?

command (12): **n**

Thank you for usingthe EYE. Goodbye.

The use of haptic virtual environments by blind people

C Colwell[1], H Petrie[1], D Kornbrot[1], A Hardwick[2] and S Furner[2]

[1]Sensory Disabilities Research Unit, University of Hertfordshire,
College Lane, Hatfield AL10 9AB, UNITED KINGDOM

[2] British Telecommunications plc, Research Laboratories,
Martlesham Heath, Ipswich IP5 7RE, UNITED KINGDOM

[1]*{c.g.colwell|h.l.petrie|d.e.kornbrot}@herts.ac.ukc*, [2]*{andrew.hardwick|stephen.furner}@bt-sys.bt.co.uk*

ABSTRACT

This paper describes two studies concerning the use of haptic virtual environments for blind people. The studies investigated the perception of virtual textures and the perception of the size and angles of virtual objects. In both studies, differences in perception by blind and sighted people were also explored. The results have implications for the future design of VEs in that it cannot be assumed that virtual textures and objects will feel to users, whether blind or sighted, as the designer intends.

1. INTRODUCTION

The research presented here concerns the perceptual aspects of the development of haptic VEs for blind people. It is important to know how blind users haptically perceive virtual objects, so that such objects can be incorporated appropriately into large scale VEs. Haptic perception incorporates both kinaesthetic sensing, (i.e. of the position and movement of joints and limbs), and tactile sensing, (i.e. through the skin) (Loomis and Lederman, 1986). At present, most VEs use visual displays, with some use of auditory and very little haptic information. The development of haptic, kinaesthetic and tactile devices offer a new dimension of realism to VEs for sighted users and has particular potential for blind users.

1.1 The Impulse Engine 3000 ™

The device used in the current studies was the Impulse Engine 3000 (Figure 1). This force-feedback device was developed by the Immersion Corporation and was used with software written by Andrew Hardwick. The device can display virtual textures and objects which users can feel using a probe. The probe is the length and diameter of a thick pen and has 3 degrees of freedom of motion, i.e. it can move in 3 spatial dimensions: forwards and backwards, up and down, and left and right. The system provides force-feedback to users by monitoring the position of their hand and altering the force accordingly (Hardwick, Furner and Rush, 1997). The force is created by three motors which exert resistance against the probe. This gives users the impression that a texture or object is present.

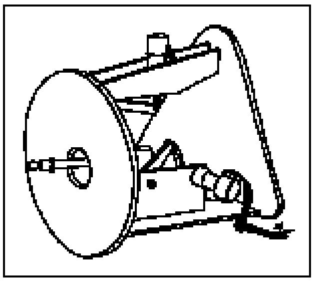


Figure 1. *The Impulse Engine 3000 ™.*

Two types of virtual stimuli were used in the current studies: textured surfaces and simple 3-dimensional objects such as cubes and spheres. The cubes and spheres could be felt from both the inside and the outside of the object. When exploring the inside of an object, it is as if the user is inside the object and they cannot feel the outside of the object. An example from the real world might be exploring the outside of a closed box but not being able to explore inside it and then getting inside the box, closing it, and exploring the inside of it.

Twenty-two participants took part in both studies, 9 were blind and 13 were sighted. Six of the sighted participants were female and all the other participants were male. The sighted participants were all university students, from different disciplines. The blind participants were all employed in computer-related jobs or on a computer science course except one, who was a retired audio engineer. Six of the 9 blind participants were either born without sight or lost their sight by the age of 30 months. The other 3 lost their sight between 8 and 26 years of age. The participants ranged in ages from 18 to 65; the average age being 32.

2. THE PERCEPTION OF VIRTUAL TEXTURES

The first study involved virtual textures with varying groove widths. The dimensions of the virtual textures were as close as possible to those used in the investigations of the perception of real textures by Lederman (1974; 1981; Lederman and Taylor, 1972), the only difference being that those real textures involved grooves with a rectangular profile whereas the textures used in the current study involved sinusoidal shaped grooves. This difference was unavoidable, as Lederman was unable to produce grooves with a sinusoidal profile (although she would have preferred to use such a form) whereas with the Impulse Engine it was not possible to simulate usable rectangular profile textures (preliminary simulations showed that the probe of the device became “caught” in the corners of the grooves). The widths of the grooves varied from 0.375 mm to 1.5 mm in steps of 0.125 mm and had a fixed amplitude of 0.0625 mm (Figure 2). There were no visual representations of the virtual textures. The magnitude estimation technique (Snodgrass, Levy-Berger and Haydon, 1985), widely accepted in studying psychophysics, was used to assess the roughness of ten textures. Initially, participants were given a standard stimulus (one of the textures from the middle of the range), to which they assigned for themselves an easily remembered number (e.g. 10), the “modulus”. They were then presented with a random sequence of 60 textures (the ten textures, each presented six times). For each presentation, they were asked to give a number which represented the texture of the new presentation relative to the modulus. So, if the texture seemed twice as rough, they would give the number 20, if it seemed half as rough, they would give the number 5. Participants find this method difficult, but for many physical stimuli it has been shown to reveal the relationships between physical parameters and psychological sensation.

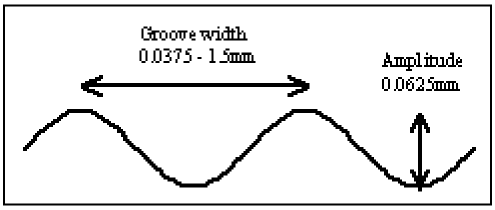


Figure 2. Dimensions of the sinusoidal grooves used in the virtual textures study.

The data were analysed by calculating the power function between physical texture parameter (i.e. groove width) and psychological sensation (i.e. the magnitude estimates) for each participant. Regression analyses were also conducted to determine how much of the variation in the sensation of the textures could be accounted for by the variations in the groove width. Regression analyses were conducted for each participant individually and on the massed data which allowed a comparison of the performance of blind and sighted people.

Overall, there was a highly significant relationship between the perception of virtual texture and its simulated physical characteristics ($F_{1,216} = 12.09$, $p < 0.001$). All nine blind participants also individually showed a significant relationship between perception of virtual texture and its simulated physical characteristics. For three of these participants the exponent was positive, meaning that they perceived the narrower grooves to be rougher than the wider grooves. This was in contrast to the other six participants for whom the exponent was negative, meaning that they perceived the wider grooves to be rougher than the narrower grooves. Only five of the thirteen sighted participants showed a significant relationship between perception of virtual texture and its simulated physical characteristics. For all the sighted participants the exponent was negative. The magnitude of the exponents ranged from 0.51 to 0.84, making them higher than those obtained by Lederman for the closely corresponding real textures.

The results from this study showed that more blind people were more discriminating than sighted people in their assessment of the roughness of the textures. Most of the twenty-two participants perceived the wider groove widths to be more rough than the narrower groove widths, although three participants perceived the narrower grooves to be rougher.

 ISBN 0 7049 1141 8

3. THE RECOGNITION OF VIRTUAL OBJECTS

For the second study a range of virtual objects was presented in various ways. The Impulse Engine 3000 allows virtual objects to be explored from both inside and outside the object, so for some of the virtual objects used, both inside and outside presentation were given to investigate any differences this factor produced. The virtual objects used were: cubes (inside and outside presentation), spheres (inside and outside presentation), rotated cubes (outside presentation only), sheared cubes (inside presentation only). Three sizes of each type of virtual object were presented: cubes with edges ranging from 1.0 cm to 2.5 cm (Table 1), spheres with diameters ranging from 1.5 cm to 2.5 cm. The amount of rotation of the cubes varied between 30 degrees and 70 degrees and the amount of shear between 18 degrees and 64 degrees. Since this was an initial exploratory study, a full factorial design was not used. Each type of virtual object was presented three times, with a range of different sizes and angles of rotation and shear. A multiple choice matching response method was used. Participants were asked to feel an object and then choose from a set of four objects of varying size the one they thought they had felt. Sighted participants were shown scale drawings and blind participants were shown scale tactile 2-D representations.

Table 1. *Mean perceived size/angle of virtual objects with percent over- and under-estimation (data from blind and sighted participants combined).*

Object Type	**Actual Size/Angle (cm/degrees)**	**Perceived Size/Angle Inside presentation**		**Perceived Size/Angle Outside presentation**	
		Mean and standard deviation (cm/degrees)	**Over/Under estimation (Percent of actual)**	**Mean and standard deviation (cm/degrees)**	**Over/Under estimation (Percent of actual)**
Cube	1.0	1.8 (0.40)	+ 80%	See Note 1.	
	1.5	1.7 (0.30)	+ 13	1.6 (0.50)	+ 7%
	2.0	2.4 (0.20)	+ 20	2.0 (0.50)	0
	2.5	See Note 2.	-	2.4 (0.20)	- 7
Sphere	1.5	2.1 (0.1)	+ 27	1.2 (0.40)	- 20
	2.0	2.3 (0.1)	+15	1.8 (0.50)	- 10
	2.5	2.5 (0.1)	0	2.3 (0.30)	- 8
Rotated	30°	-	-	40° (12.0)	+ 33
Cube	50°	-	-	52° (12.0)	+ 4
	70°	-	-	45° (18.0)	- 36
Sheared	18°	20° (11.0)	+ 11%	-	-
Cube	41°	37° (11.0)	- 10%	-	-
	64°	59° (9.7)	- 8%	-	-

Since a full factorial design was not employed, a series of analyses of variance were used to analyse different components of the data. Mean perceived sizes/angles for the various objects used are shown in Table 1. No significant difference was found between the perceptions of sighted and blind participants, except that the sighted participants judged the sheared cubes more accurately than the blind participants. Both groups were significantly more accurate in their perception of larger objects than of smaller objects. For example, the 1.0 cm edge cube was perceived on average to have a 1.8 cm edge when explored from the inside, an overestimate of 80%, whereas the 2.0 cm cube was perceived on average to have a 2.4 cm edge, an overestimate of only 20%. The size of the objects felt from the inside tended to be overestimated, the mean overestimation across all sizes of cubes and spheres being 25.8%. However, the size of objects felt from the outside tended to be underestimated, with a corresponding mean underestimation of 6.3%. Finally, the angles of the rotated cubes seemed to be difficult to judge, although this may have been due to the lack of a reference point for judging the rotation in the VE.

4. DISCUSSION

The way in which a user of the Impulse Engine 3000™ can explore virtual objects differs from the way in which real objects are felt in several ways. An example is that the device currently requires the user to feel textures and objects with the probe. This is not a particularly intuitive way of interacting with objects and several participants said they would rather use their hands because they are more used to feeling their environment in this way. A further example is that if the user pushes hard enough they can have the sensation of pushing through the surface of an object. This is because the Impulse Engine 3000 motors are capable of withstanding only 8 Newtons (approximately 2 lbf) of force from the user.

Hardwick, Rush, Furner & Seton (1996) observed an interesting phenomenon associated with the Impulse Engine 3000, whereby people differ in terms of where they think the virtual space is located in real space. Some people have a mental image of the virtual space being outside the device, so that virtual objects are felt to be near the hand and are touched by the end of the probe that they hold (Figure 3a). In contrast, others imagine the virtual space to be within the device, so that virtual objects are touched by the other end of the probe (Figure 3b).

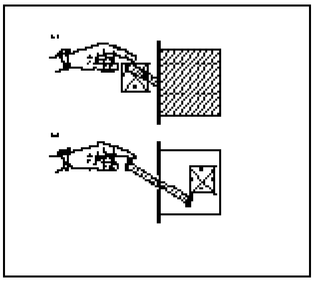


Figure 3. Representation of different mental models of the location of a virtual object. (a) outside the device. (b) inside the device.

This phenomenon was explored further during the current studies asking each participant where in real space they thought the object was located, and to point to this location. Data on this phenomenon were collected from 19 of the participants. 14 (74%) imagined the objects to be located inside the device, 4 (21%) imagined the objects to be outside, and 1 (5%) imagined them to be half-way. Three (33%) of the blind participants imagined the objects to be located outside of the device, compared to only 1 (8%) of the sighted participants. Of the participants who imagined the objects to be outside the device, 3 were blind and 1 was sighted. Therefore, this phenomenon may be more prevalent amongst blind people than sighted people, but is worthy of further investigation.

5. CONCLUSIONS

This paper has presented two studies exploring the perception of virtual textures and objects using the Impulse Engine 3000 haptic device. These studies have illustrated both the potential and some of the problems of using current haptic technology to simulate real world objects or to create totally virtual objects. In designing haptic interfaces, designers need to exercise care and not assume that the virtual world will be perceived in exactly the same ways as the real world, particularly given the current limitations of haptic devices which use probes and joysticks. However, the current devices do provide realistic feeling textures and objects which replicate the psychophysical properties of real textures and can be judged like real objects. These virtual objects and textures have enormous potential for enhancing VEs for both sighted and blind people.

Acknowledgements: We would like to thank all the people who participated in these studies. Chetz Colwell is a PhD candidate supported by the Economic and Social Research Council, U.K. and M.A. Systems and Control Ltd, U.K.

5. REFERENCES

Boyd, C. D. & Darken, R. Psychological issues in virtual environment interfaces, in Proceedings of CHI '96 (Vancouver, BC, Canada, April, 1996) ACM Press, 49-53.

Hardwick, A. Furner, S. & Rush, J. Tactile access to virtual reality on the World Wide Web for visually impaired users. Submitted for special issue of IEE Displays: Developments in tactile displays.

Hardwick, A. Rush, J. Furner, S. & Seton, J. Feeling it as well as seeing it - haptic display within gestural HCI for multimedia telematics services, in P.A. Harling & A.D.N. Edwards (Eds.) Progress in Gestural Interaction (Proceedings of Gesture Workshop '96) Springer-Verlag, 1996, 105-116.

Hardwick, A. Furner, S. & Rush, J. Tactile access for blind people to virtual reality on the World Wide Web. IEE Digest No. 96/012 (Colloquium on Developments in Tactile Displays) 1997, 9/1-9/3.

Lederman, S. J. Tactile roughness of grooved surfaces: the touching process and effects of macro- and microsurface structure. Perception & Psychophysics 16, 2 (1974) 385-395.

Lederman, S. J. Perception of surface roughness by active and passive touch. Bulletin of the Psychonomic Society 18, 5 (1981) 253-255.

Lederman, S. J. & Taylor, M.M. Fingertip force, surface geometry, and the perception of roughness by active touch. Perception and Psychophysics 12, 5 (1972) 401-408.

Loomis, J. M. & Lederman, S. J. Tactual perception. In K. K. Boff, L. & Thomas, JP. (Eds.) Handbook of perception and human performance. Wiley/Interscience, New York, 1986, 31.31-31.41.

Snodgrass, J.G, Levy-Berger, G. and Haydon, M. Human experimental psychology. New York: Oxford University Press, 1985.

Can a haptic force feedback display provide visually impaired people with useful information about texture roughness and 3D form of virtual objects?

Gunnar Jansson

Department of Psychology, Uppsala University
Box 1225, SE-751 42, Uppsala, SWEDEN

gunnar.jansson@psyk.uu.se

ABSTRACT

The aim was to investigate the usefulness of a haptic force feedback device (the PHANToM) for information without visual guidance. Blind-folded sighted observers judged the roughness of real and virtual sandpapers to be closely the same. The 3D forms of virtual objects could be judged accurately and with short exploration times down to a size of 5 mm. It is concluded that the haptic device can present useful information without vision under the conditions of the experiments. The result can be expected to be similar when observers are severely visually impaired, but this will be controlled in a separate experiment.

1. INTRODUCTION

The most common function expected to be fulfilled by haptic force feedback displays is to enhance the perception of virtual reality scenes rendered by visual and/or auditory displays in medical, entertainment, telerobotic, and military applications (Burdea, 1996). When a haptic display is considered for people with severe visual impairment, the situation is quite different; the lack of visual guidance may decrease the effectivity in utilising the haptic information as vision and haptics normally cooperate (see, for instance, Heller, 1982).

The problems include getting an overview of the display and locate the relevant parts of it, as well as picking up 3D aspects of the display depicted in 2D (cf. Jansson, 1988). If overview and location of parts have to be obtained by haptics alone, very long exploration times may be needed. Concerning getting 3D aspects from 2D haptic depictions, it has been suggested that it is an impossible task (Révész, 1950), and in the applied work with tactile pictures (for an overview, see Edman 1992), 3D aspects of the pictures are not emphasised. On the other hand, there are reports indicating that perspective information may be useful for haptic pick up of 3D aspects (Heller et al., 1996; Holmes et al., in press; Kennedy, 1993). 2D cutaneous information may thus contribute to the perception of 3D objects, but there are many problems left for research on what conditions favour 3D percepts from cutaneous information.

In addition to cutaneous information, information from the movements is available for the observer. The information is provided by sensors in the muscles, tendons and joints. The importance of this information was strongly emphasised in seminal papers by Katz (1925/1989) and Gibson (1962), but the relative contribution of cutaneous and movement information to haptic perception has since then been much discussed. Many authors state that relative motion between skin and object is the important factor pointing to experiments where the performance is the same whether the hand or the object is moving (see, e.g., Lamb, 1983, and Lederman, 1981, 1983). Hughes & Jansson (1994) noted that most of studies of this problem concerned texture perception and that the applicability of the equivalence of movement of observer and of object can not without further evidence be generalised to other types of haptic perception. Vega-Bermudez et al. (1991) got the same result, however, for tactile letter recognition but they studied only patterns smaller than the finger pad. In contrast, Jansson (in press) found significant differences between active exploration and passive reception of cutaneous information when studying larger 2D virtual geometric forms.

The problem of the relative contribution of cutaneous and movement information to haptic perception has direct relevance for the usefulness of force feedback displays for people with severe visual impairment. Force feedback displays emphasise movement information and are much less concerned with cutaneous information. In many cases, including the one to be studied here, the equipment defines only one point at a time for contact between observer and virtual object which is much less than cutaneous information in natural contexts. If movement information can be sufficient this does not decrease the effectivity of the display, but if the cutaneous information also is important the restriction to one point decreases the effectivity. Lederman & Klatzky (in press) made a series of experiment indicating that restriction of cutaneous information to one point substantially impaired the performance. When they applied these results to the design of haptic displays they suggested that there may be significant costs of not providing the fingertips with spatially distributed force patterns, at least for novice operators.

As discussed above, there are both pros and cons concerning the usefulness of presently available haptic force feedback displays. The general aim of the present project, part of which is reported here, is to find to what extent devices of this kind can be useful in spite of their limitations. The most positive aspect in favour of their usefulness is their offer of free exploratory movements in 3D space, the most negative aspect the restricted cutaneous information. A reasonable hypothesis is that the importance of movements increases with the complexity of the depiction, especially when 3D aspects are included. The aspects of the virtual objects to be studied here are texture roughness and 3D form.

2. EQUIPMENT

2.1 Haptic Display

A PHANToM 1.5A from Sensable Technologies, Inc., Cambridge, MA, USA, was used as haptic display. It is a robot driving a two-linked arm the tip of which is freely movable within a 19.5 x 27 x 37.5 cm workspace with a nominal position resolution of .03 mm, maximum exertable force of 8.5 N and enertia (apparent mass at tip) of < 75 g according to Sensable specifications (for more details, see the site http://www.sensable.com). The device was driven by a Scandic Computer equipped with a Pentium Pro 200 MHz and with Windows NT Workstation 4.0.

2.2 Exploration Styluses

The tip of the PHANToM arm was provided with the standard stylus, which means that the point of contact with the virtual object was at the end of this stylus. For the exploration of the real sandpapers a stylus was constructed which was a copy of the one used by the PHANToM but with an additional 50 mm long and 3 mm thick steel tip with a pointed tip corresponding to the stylus used when collecting physical sandpaper data.

3. PERCEIVED ROUGHNESS OF REAL AND VIRTUAL SANDPAPERS

3.1 Problem

Texture is one of the most important properties of an object and a property that haptics readily can pick up. Sandpapers have been used in many studies about texture, one reason being that the physical properties of their texture can be clearly defined and ordered. In order to study how well virtual sandpapers reproduce the texture of sandpapers in a form that is useful for observers, the perception of real and virtual sandpapers were compared. As exploration method may effect the result, the same method was used in both conditions, namely exploration with a stylus.

The experimental problem was thus the following. How well do blind-folded observers' perception of the roughness of real and virtual sandpapers agree when they are explored with a stylus?

3.2 Method

3.2.1 Real Sandpapers. Four Norton Metalite sandpapers with 50, 80, 120 and 220 grit, respectively, were used. (For standard specifications, see http://www.wirecloth.com/howto/convert/ussueve.html.)

3.2.2 Virtual Sandpapers. Virtual sandpapers were presented by the PHANToM with a method developed by Green & Salisbury (1997). The PHANToM is first used to acquire data from a sample of respective sandpapers. A vertical probe with one end attached to the PHANToM arm and the other end resting on the horizontal sandpaper is made to follow a trajectory in the form of a straight line at a constant speed and exerting a constant force. Lateral forces and the z position of the endpoint during the movement are recorded. The collected data are used to calculate the vector of z values and means and standard deviations of the static friction coefficients which are used for the simulation. The virtual surfaces are not exact copies of the real surfaces but the properties used in the simulation are intended to be sufficient for allowing accurate perception of the roughness of the sandpapers.

3.2.3 Procedure. Before using the PHANToM participants were informed about the device and safety aspects of its use (a standard head protective device common in industry was placed on the participant's head) and were allowed to acquaint themselves with it. They were instructed to hold the stylus as vertically as possible close to its lower end and to move it approximately in a straight line back and force applying the same constant force during the whole experiment.

Before exploring the real sandpapers the participants were instructed to hold the specially made stylus as vertically as possible and closely above the steel part of the stylus (about 5 cm from the pointed tip) and keep the hand such that it did not touch the sandpaper. The instructions about movements were the same as those for the virtual sandpapers.

When the participants were ready to start, they were asked to choose hand for the exploration the same hand used for the whole experiment. They were equipped with eye cover and earphones playing white noise and the experiment proper began.

3.2.4 Phychophysical Method. The roughness of the sandpapers were judged with magnitude estimation with 120 grit defined as standard to be given the value of 100; no roughness was defined as 0 and there was no maximum limit. In half the trials the real 120 grit sandpaper was the standard, in the other half corresponding virtual sandpaper. The task of the participants was to judge each presented sandpaper to have a roughness value such that it was related to 100 in the same way as its perceived roughness was related to that of the standard sandpaper.

3.2.5 Design. All participants took part in all the experimental conditions and the trials were arranged in four main blocks: virtual texture with virtual standard, virtual texture with real standard, real texture with virtual standard and real texture with real standard. Each main block consisted of six blocks each containing a presentation of the standard sandpaper followed by the four experimental sandpapers. All orders were randomised for half the participants, and the reverse orders were used for the remaining participants.

3.2.6 Participants. Twelve paid sighted university students (seven women and five men) with a mean age of 25 years (SD = 2.4 years) participated. All with the exception of one man worked with their right hand.

3.3 Results

The data were collected and analysed by Billberger (1998). A fourway ANOVA demonstrated significant effects of roughness and replication ($p < .001$) and interaction between standard and replication ($p < .05$), but no significant effects for stimulus type (real/virtual) and standard ($p > .05$), nor for any other interactions. Fig. 1 demonstrates the effects of physical roughness on perceived roughness for real and virtual sandpapers.

3.4 Discussion

The result indicates that the real and virtual sandpapers are perceived in very much the same way, at least when they are similarly explored. The simulation of these textures can thus be considered as successful. It should be noted, however, that there was a tendency at all levels of roughness of virtual sandpapers to be perceived as somewhat rougher than corresponding real sandpapers. This may mean that a significant difference would show up if the number of participants were larger, but the smallness of the difference means that such a result can not be expected to have any practical importance.

There was hardly any difference between the results when the standard was a real sandpaper and when it was a virtual sandpaper. Any of them can be used in future experiments.

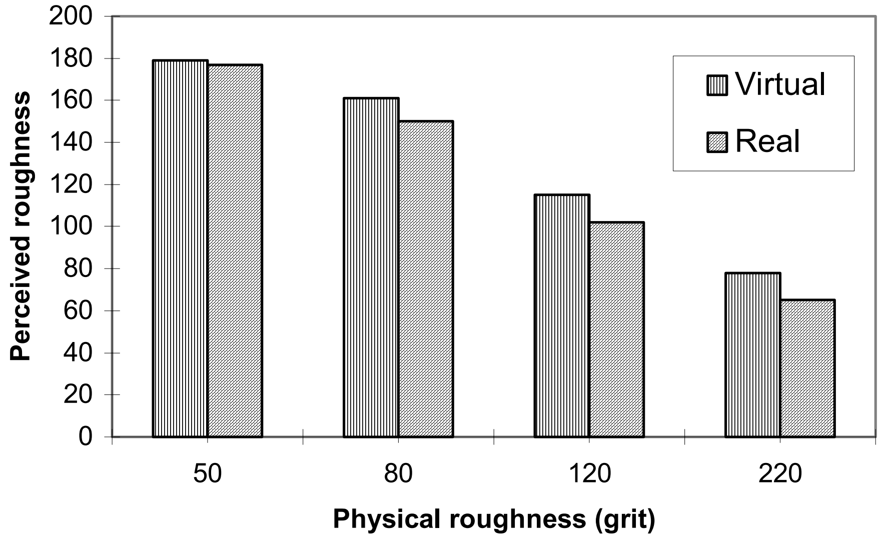


Figure 1. *Perceived roughness as a function of physical roughness and stimulus type.*

4. IDENTIFICATION OF 3D VIRTUAL GEOMETRIC FORMS

4.1 Problem

If any method of rendering virtual objects would be successful, it is important that their 3D form can be identified by the observers. The experimental problem in this part of the investigation was to get a first idea about how well observers can identify differently sized 3D geometric forms rendered by the PHANToM and explored with the stylus.

4.2 Method

4.2.1 Rendering of 3D Geometric Forms. The software, called ENCHANTER, for rendering the experimental forms was developed by Fänger and König (1998) in co-operation with the author. It is based on the software GHOST™ SDK and provides the user with possibilities of easy rendering of 3D geometric forms with several different properties for presentation in experiments.

4.2.2 3D Geometric Forms Studied. Four 3D forms were used, cube, sphere, cylinder and cone, in three different sizes, maximum width and height being 5, 25 and 50 mm, respectively. In order for the 3D forms to be easily localised they were positioned in the middle of a cubical enclosure with dimensions twice those of each 3D form, and for the 3D forms and their enclosure to be certainly discriminated the 3D form surface had no static friction while the inside surfaces of the enclosure had a high such friction.

4.2.3 Procedure. The participants were informed about the PHANToM and the safety aspects and they were allowed to acquaint themselves with the device. The 3D geometric forms to be used were explained for the (sighted) participants with the help of drawings. There were no restrictions on how to use the stylus, but the participants usually kept the stylus similar to a pen. The head protective device and eye cover were applied, and the experiment proper began. (As the PHANToM made very little noise during the exploration of these 3D forms the sound was not masked.)

The participants were presented with the 3D virtual forms one by one and asked to judge their form as fast and accurately as possible (with equal emphasis on both aspects). Maximum 1 min was allowed per 3D form. The verbal responses and the time used for each 3D form was recorded.

4.2.4 Design. All participants took part in all conditions. Each participant was presented three blocks with the 12 3D forms in random order, thus altogether 36 3D forms.

4.2.5 Participants. Ten paid sighted university students (seven women and three men) with a mean age of 22 years (SD = 2 years) took part. All used their right hand for exploration.

4.3 Results

The percentages of correct responses and mean explorations times, both parameters over all participants, are presented in Figs. 2 and 3.

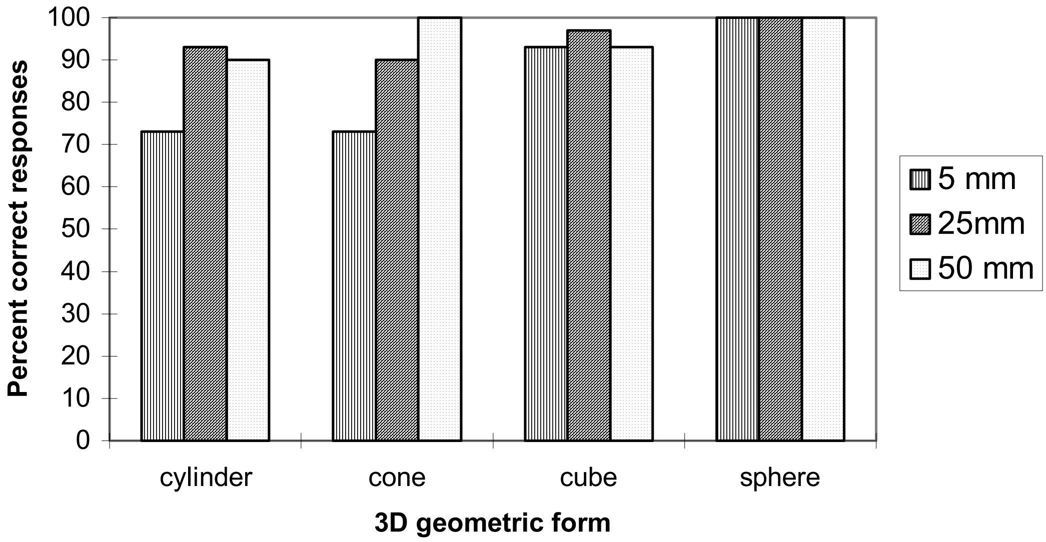


Figure 2. *Percent correct responses for each of the four 3D forms and three sizes.*

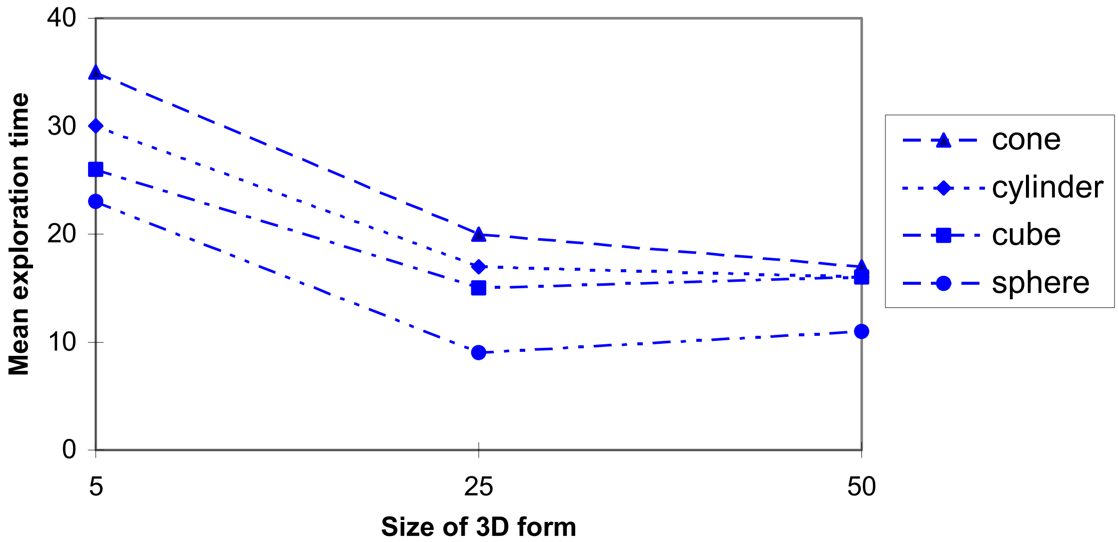


Figure 3. *Mean exploration time (sec.) as a function of 3D form and size (mm).*

4.4 Discussion

The results show clearly that the force feedback device used can provide observers with useful information without vision under the conditions of the experiment. The percent of correct responses is highly above chance level (25 %). In fact, the sphere was correctly identified every time, even in its smallest size. A majority (52 %) of the mistakes for the other 3D forms were made during the first replication. If only the second and third replication had been included the percent correct responses would have been 95 % over all 3D forms and sizes. This demonstrates a quite rapid learning to identify the 3D forms when the identification is not perfect from the start. The size threshold for correct identification is apparently smaller than 5 mm.

Fig. 3 indicates differences in exploration time between the 3D form and sizes. For all the 3D forms the time for the 5 mm size in longer than for the larger sizes. The sphere is not only always correctly identified, but also the time to explore it is shortest for all sizes.

A note should be made about potential effects of sound not being masked. It was assumed before the experiment that the sound could not be used for identification of the 3D forms. Spontaneous comments by some participants indicated, however, that it may have contributed. It can not be excluded that it was used by some participants for the detection of edges. If this was the case, it is not a problem from an applied point of view, as the auditory information is available also for visually impaired people (without hearing loss), but for future experiments about haptics alone it is recommended that sound is always masked.

5. CONCLUSIONS

The investigation demonstrates that the force feedback device studied can present useful information to observers for whom vision is not available. Even if the aspects involved are quite limited they are basic for haptic perception of objects. It is an important result that texture and 3D form can be judged with such accuracy and speed. However, it is evident that the extent to which this can be generalised to other contexts remains to be studied.

In the present study the observers explored the objects via a stylus. That this may not mean a disadvantage compared with other exploration methods is indicated by a theoretical and experimental investigation by Klatzky and Lederman (in press).

The observers were blind-folded sighted people. A study with visually impaired observers would probably not show very large differences, especially not concerning the relations between experimental conditions. Even if the basic haptic capability can thus be expected to be the same for sighted and for visually impaired observers, it is necessary to make specific experiments with visually impaired observers to make sure that the usefulness of the device is similar for them. They have probably more training than sighted people in using haptics which may mean generally better results. On the other hand, especially people with early appearance of severe visual impairment may have less experience of spatial aspects of the environment which may lead to not as good results in general. Therefore, an investigation on related problems with observers having severe visual impairments has started.

Acknowledgements: This study was made possible by grants from the Swedish Council for Research in Humanities and Social Sciences and from the Swedish Transport and Communications Research Board. The author is indebted to Donald Green, Massachusetts Institute of Technology, Cambridge, MA, for making his computer software PALPATE available for this study, to Jens Fänger and Henry König, Universität Magdeburg, Germany, for developing, in co-operation with the author, the computer software ENCHANTER, to Lars-Erik Larsson for technical assistance, and to Katarina Billberger for running the experiments.

6. REFERENCES

K. Billberger (1998), *Haptisk perception av verkliga och virtuella texturer* (Haptic perception of real and virtual textures), Undergraduate thesis, Department of Psychology, Uppsala University, Uppsala, Sweden.

G C Burdea (1996), *Force and touch feedback for virtual reality*, Wiley, New York.

P K Edman (1992), *Tactile graphics*, American Foundation for the Blind, New York.

J Fänger and Henry König (1998), *Entwicklung einer Modellierungs- und Experimentierumgebung für eine Kraftrückskopplungsgerät* (Development of a form production and experiment environment for a force feedback device), Praktikumsdokumentation, Institute for Simulation and Graphics, Magdeburg University, Magdeburg, Germany.

J J Gibson (1962), Observations on active touch, *Psychological Review*, **69**, pp. 477-491.

D F Green and J K Salisbury (1997), Texture sensing and simulation using the PHANToM: Towards remote sensing of soil properties, In *Proceedings of the Second PHANToM Users Group Workshop* (J K Salisbury & N A Srinivasan, Eds), A.I. Technical Report No. 1617 and R.L.E. Technical Report No. 618, Massachusetts Institute of Technology, Cambridge, MA.

M A Heller (1982), Visual and tactual texture perception: Intersensory cooperation, *Perception & Psychophysics*, **31**, pp. 339-344.

M A Heller, J A Calcaterra, L A Tyler and L L Burson (1996) Production and interpretation of perspective drawings by blind and sighted people, *Perception*, **25**, pp. 321-334.

E Holmes, B Hughes and G Jansson (in press), Haptic perception of texture gradients, *Perception*.

B Hughes and G Jansson (1994), Texture perception via active touch, *Human Movement Science*, **13**, pp. 301-333.

G Jansson (1988). What are the problems with tactual pictures, and what can we do to solve them? In *Man-Machine Interfaces, Graphics and Practical Applications* (C W M Magne'e, F J M Vlaskamp, M Soede, & G Butcher, Eds.), Royal National Institute for the Blind, London, pp. 18-24.

G Jansson (in press). Haptic perception of outline 2D shape: The contributions of information via the skin, the joints and the muscles. In *Proceedings of the Fifth European Workshop on Ecological Psychology (EWEP5),* Pont-à-Mousson, France, July 7-10, 1998.

D Katz (1989), *The world of touch* (L E Krueger, Transl.), Erlbaum, Hillsdale, NJ. Original work published 1925.

J M Kennedy (1993), *Drawing and the Blind*, Yale University Press, New Haven, CT.

G D Lamb (1983), Tactile discrimination of textured surfaces: Psychophysical performance measurements in humans, *Journal of Physiology (London),* **338***,* pp. 551-565.

R L Klatzky and S J Lederman (in press), Tactile roughness perception with a rigid link interposed between skin and surface, *Perception & Psychophysics*.

S J Lederman (1981), The perception of surface roughness by active and passive touch, *Bulletin of the Psychonomic Society,* **18***,* pp. 253-255.

S J Lederman (1983), Tactile roughness perception: Spatial and temporal determinants, *Canadian Journal of Psychology,* **37**, pp. 498-511.

S J Lederman and R L Klatzky (in press), Sensing and displaying spatially distributed fingertip forces in haptic interfaces for teleoperater and virtual environment systems, *Presence*.

G Révész (1950), *Psychology and Art of the Blind*, Longmans, Green and Co, London.

F Vega-Bermudez, K O Johnson & S S Hsiao (1991), Human tactile pattern recognition: Active versus passive touch, velocity effects, and patterns of confusion, *Journal of Neurophysiology*, **65**, pp. 531-546.

Ambisonic sound in virtual environments and applications for blind people

M Cooper and M E Taylor

Institute of Educational Technology/Knowledge Media Institute/Multimedia Enabling Technology Group, Open University, Berrill Building, Walton Hall, Milton Keynes, MK7 6 AA, UK

m.cooper@open.ac.uk, *m.e.taylor@open.ac.uk*

ABSTRACT

To date there has been much more effort directed to generating credible presentations of virtual worlds in a visual medium than in reproducing corresponding or even self contained worlds with synthesised or recorded sound. There is thus today a disparity in the relative fidelity of these 2 modes in most VR systems. While much work has been done in hi-fidelity and 3 dimensional sound reproduction in psycho-acoustic research and applications for audiophiles this has rarely been taken onboard by the VR community.

This paper describes work ongoing to apply Ambisonic techniques to the generation of audio virtual worlds and environments. Firstly Ambisonics is briefly outlined then principles behind its implementation in this context described. The design of the implementations to date is described and the results of trials discussed. The strengths and limitations of the approach are discussed in the light of this practical experience.

There is a range of applications envisaged for this technique that would be particularly of benefit to disabled people. The 2 principal areas under consideration by the authors is the extended use of the audio mode in HCI for people with disabilities and VR applications for blind and partially sighted people.

Both of these areas are described and specific examples of applications given in each. The limitations given available low cost technology are highlighted and the technology evolution required to make such applications widespread commented on.

The results of the development work and subsequent user trials undertaken to date are given and discussed. Lines of further exploration of this technique and its application are outlined.

1. INTRODUCTION

Ambisonics is a method of recording and reproducing sound in 3 dimensions. Advances in digital audio and computing now hold the possibility to exploit this approach to facilitate computer synthesis of 3 dimensional audio environments. Applications of this approach have been envisaged offering particular benefits for visually impaired people. Detailed feasibility studies, user needs capture and pilot implementations are now being undertaken to prove the validity of the approach and that it will meet identified needs. This is the subject of this paper, which is a description of work in progress together with references to related work of others and notes on the aspirations of potential users that are directing it.

2. AMBISONICS AN OVERVIEW

Ambisonics, originally developed in the 1970s, is a method of capturing, encoding for recording and reproducing sound in 3 dimensions. A very brief, largely non-technical, overview of Ambisonics is given here to set the rest of the paper in context. Full descriptions of the technique with its psycho-acoustic and mathematical justifications can be found in the key papers of the field's founding fathers [Fellgett 72, Gerzon 74, and Gerzon 92]. A paper outlining its advantages in audio virtual reality was given at the first ECDVART conference [Keating 96].

In the capture of sound Ambisonics detects not only the nature of the sound but the direction at from which it arrives at the point the recording is made. The term "soundfield" is coined to describe the sound impinging on an imaginary sphere around the listening position. The goal of the technique is to be able to record a soundfield at one time and place and then reproduce it at another.

In first order Ambisonics, the only degree of complexity implemented to date, the sound is ideally captured by a set of 3 dipole microphones (i.e. with figure of eight response patterns) coincidentally located and lying along the axes of a Cartesian co-ordinate system. This is physically unrealisable, however the commercially available Soundfield Microphone effectively accomplishes this by having 4 microphone capsules arranged in a tetrahedron then combining the outputs of these in such a way as to yield the required response patterns. The fact that the capsules are not truly coincident is compensated for electronically. The microphones output is 4 separate signals conventionally labelled X, Y, Z and W. The X signal represents the sound components from the front minus those from the rear and similarly Y left minus right, and Z up minus down. The W signal is a non-directional reference signal generated from the combined outputs of all of the microphone's capsules. Within Ambisonics these W, X, Y, and Z signals are known collectively as "B-Format". In the applications discussed in this paper, a microphone is not used to capture a "live" soundfield but the same B-Format signals are generated from a computer model of the sound sources and their environment.

3. TECHNOLOGY CONTEXT

A key reason for Ambisonics failing to gain popularity within the audio industries in the 1970s, when it emerged, is that they could not at that time foresee the required 4 channels being available for broadcasts or recording media for the domestic market. However in the last year or so this has become readily possible with the arrival of digital broadcasting and the DVD (Digital Versatile Disc) recording format.

The "digital revolution" that has occurred in sound recording and processing, combined with the increase of computing power available in standard desktop PCs, means it is now practical to use Ambisonic techniques for the creation and reproduction of audio virtual environments. Software tools have been developed, which enable a standard PC equipped with a suitable multi-channel sound card to be used both for the creation of the B-Format as digital signals, and the processing needed for their playback over a particular loudspeaker arrangement [Farina 1998]. There is a commercially available and very powerful system, from Lake DSP of Australia, which facilitates the same (potentially alongside the real-time generation of the virtual environment) by using dedicated digital audio signal processors (DSPs), connected to a standard PC. Both these approaches are based on digital algorithms for the fast implementation of a mathematical process called "convolution". The response of any system to a given input is the convolution of its impulse response with that input. Acoustically the impulse response can be viewed as the result of a gunshot within the acoustic space.

Other systems have emerged for the creation of a 3 dimensional sound effect (e.g. Dolby MP encoding, and Dolby Surround and Pro Logic decoding, DTS, THX, etc.). Theses are all devised to give an impression of sound surrounding the listener but are unable to precisely locate a sound source from the side or the rear. These other techniques are mainly used in cinema where the focus of attention is towards the front. Further, the sound is combined with a wide screen high quality picture, thus the effect of these sound processes on the overall perception of the experience is significant but the brain effectively ignores any spatial imprecision in the sound heard. Thus Ambisonics remains the best available technique for generating of the soundfields for the applications principally directed toward people with a visual impairments described in this proposal.

4. APPLICATIONS FOR BLIND COMPUTER USERS

4.1 Virtual reality for blind people

Virtual reality (VR), in its many guises, has been held up as a potentially powerful tool in education and training. Indeed that potential has now been demonstrated in various commercially available and in house software packages. Much of the work to date in developing virtual reality systems has concentrated on the visual modality. However if the potential of virtual reality is to be extended to blind users then the computer generation of credible audio worlds is required. Further this would be of benefit in a wide range of VR applications for users without a significant visual impairment. There is some evidence to the effect that the perceived veracity of a virtual world is more dependent of the fidelity of the audio rather than visual representation of that world. We appear to be able to much more readily "suspend disbelief" in what we see than what we hear.

4.2 Sound Environments as access to GUIs

The move in personal computing over the last 10 years, almost universally, to Graphical User Interfaces (GUIs) is an example of advances in usability for the majority creating barriers for one user group; blind people. Various approaches have been have been developed to address this problem and have met with mixed results. These include the use of tactile displays, extended speech synthesiser systems and the modal transformation of icons to "audicons", and approaches that seek to combine these. In the audicon approach a sound representative of the icons function is given to the user when the cursor within the GUI is over a particular icon. It is suggested that the usefulness of this approach could be greatly enhanced if the position that this audicon appears to emanate from directly maps to the position of the corresponding icon within the visual display.

Some work has been done to investigate this within the EU TIDE programme sponsored GUIB project [Crispen and. Petrie 1993]. The approach taken in their work was based on the direct calculation of the signals for feeding into the headphones of the user based on a modelling the Head Response Transfer Function (HRTF). If movements of the head are need to be accounted for as would be the case in many practical applications this approach becomes very computationally intensive [See Keating 96]. In terms of using sound to access a GUI the approach taken within the GUIB project was to us a sound cue for the cursor position and a separate one for the icon location. The user was then required to control the cursor through a mouse to bring the first sound to the other. The evaluations undertaken indicated that there were still significant challenges in taking this forward to a practical system. They found that the acuity in the vertical dimension was very much less than the horizontal as would be expected from the theoretical understanding of auditory location. It is not possible to determine from the published results to what degree this was effected beyond the human perceptual limits by their implementation. Certainly inaccuracies in the HRTFs used would have particularly affected the vertical acuity.

The authors of this paper would like to suggest that an alternative model for audio interaction with a GUI than that adopted in the GUIB project would be more powerful and less effected by the limitations of the technology and human perception. The sound synthesised by the computer could be done from the perspective of the user being at the cursor position. Thus the directional information from the available audicons would be in direct relation to their position with respect to the cursor. Thus the relative position of the sound rather than an absolute location becomes the important factor. The scale and mapping of the perceived sound space could be such that accounts for the different spatial perception of individuals and in the different aspects. This is a key direction in the ongoing work and many perceptual and technical factors need further research to validate this approach.

As well as facing particular challenges in the use of GUIs, blind computer users face increased challenges when seeking to learn to use a new operating system or application software. A significant disadvantage for blind computer users compared with their sighted colleagues is the time required for them to learn to such software, particularly when it is GUI based. This is principally due to the difficulty of learning by exploration and experiment. The following indicates a learning strategy typical of sighted computer users learning to use a new piece of software:

"Oh, what does this icon do?"

"Let's select it and see."

The challenge for the blind computer user even given currently available assistive technologies is firstly how to identify the existence of an icon that potentially offers a useful function and then how to evaluate its action. An extension of the spatially located audicon approach could significantly address the first of these and assist with the second.

One can envisage an exploratory mode for interacting with a GUI through audio. In this mode the audicon would be triggered when the cursor was within a given radius of the icon. The volume of the audicon would then increase as the cursor moves toward the icon. The advantage of this is that the user would have a greater awareness of the location of the cursors position within the arrangement of icons and of the existence of the available icons. The potential for this mode resulting in a meaningless noise when the cursor is in the vicinity of multiple icons is largely obviated by the spatial location of the audicons. Human audio perception is very good at focusing its attention at a sound from a given direction even given the presence of a high level of other sounds from other directions (the "cocktail party effect"). The phase “mumbling icons” has been coined to describe this approach. Given the successful implementation of spatial audicons the extension to this exploratory mode is technically relatively trivial but will need extensive user trials to arrive at optimal settings of the various system parameters.

5. WORK IN PROGRESS

5.1 Generation of Simple Audio Virtual Worlds

To date only a basic system for synthesising Ambisonic sound due to a single sound source in a user defined rectilinear world has been developed. This uses simple inverse ray tracing techniques to calculate the resulting B-Format signals at the “listening” position due to a stationary or moving source within the world. Any Wave Format (.WAV) file can be used as the source and the resulting B-Format signals are calculated and stored as 4 individual .WAV files. This is an off line calculation of the audio virtual world. The 4 B-Format signals are then played out through a multi-channel PC sound card (Gadget Labs™ Wave/4) using commercially available sound studio software (Cool Edit Pro). It should be noted that it was not possible to use standard SoundBlaster™ compatible PC sound cards because although it was possible to install 2 cards in a single PC it was not possible to exactly synchronise the outputs from both cards.

The playback facility is a regular array of speakers at the corners of 2 orthogonally bisecting rectangles, one in the horizontal plane, at the level of the listening position, and the second in the vertical plane passing through the listening position from front to back. A key feature of Ambisonics is that the signals recorded, or synthesised, are independent of the configuration of the playback speakers (unlike Dolby 5.1, etc.). Thus the particular array chosen was mainly determined by the to use 2 of the speakers for stereo work at other times and the physical constraints of the sound booth. An 8-speaker array was selected as previous work with Ambisonics had shown this was a practical minimum for a stable reconstruction of a 3 dimensional soundfield. The system was installed in a sound booth that had some acoustic treatment but was only a "semi-dead" acoustically and in no way could be considered anechoic. Currently a speaker decoder, that takes the analogue B-Format signals and derives the individual speaker feeds has been constructed in-house with the parameters as calculated from those given in Gerzon 1980 for the particular speaker array installed. Others, [Farina and Ugolotti 1998] have developed and demonstrated the use of software decoders working on the digital B-Format signals but this for the configuration described then requires 8 soundcard audio outputs. This is perfectly possible with the above listed software by installing a second Wave/4 soundcard but was not selected as the initial route.

The choice of the sound source needs, if fidelity is the objective, to be recorded by close microphone techniques and not subject to any artificial reverberation treatment and the same applies for synthesised sources. That way it is the acoustic of the virtual world that determines the sound as calculated for the listening position. The current version of the software enables any rectilinear space to be modelled, with any number of rectilinear features, of different acoustic properties, on any of its surfaces. The sound source and the listener can be placed anywhere within the room and the sound source moved through any path that can be described as a timed series of Cartesian co-ordinates. Acoustically the modelling is very simple with only surface absorption and inverse square law effects being taken into account.

Trials with multiple users with and without a visual impairment of this system described begin this autumn. The initial trials to date have confirmed that for most people a believable sound image is created but there is some variability between subjects as to the perceived location of a sound within the world. What has been demonstrated to date is the feasibility of the basic approach and useful information gleaned as to the computation levels required. The typical time of calculation of the B-Format signals due to a 1-second source signal on a 133MHz Pentium PC is about 20 seconds. This is of course subject to a lot of variation with programme and virtual world parameters. The most dominant here being the ray step and acceptance angle within the inverse ray tracing algorithm the above is with 0.5 degrees set for each.

There is now an ongoing programme of work seeking both to increase the complexity of the audio worlds modelled and to arrive at optimal calculating of these so that interaction with them can be achieved in real time. This will be based on the use of a Digital Audio Convolution Processor (a multi-channel DSP board and development environment) from Lake DSP (see http://www.lakedsp.com/products/index.html for further information.)

5.2 *Justification for Approach*

There is an issue here of why apply this high level of, currently expensive, processing power in researching and developing a technology when the vision is for it to become widely used in education and the home? The answer is principally in the lead-time of the proposed work. The research is directed towards applications that may only come into widespread use in 3 to 5 years time. With the current speed of evolution in computer technology it is highly likely that the necessary processing power will be readily available within that time scale on standard PCs. Further developments in computer related technology such as DVDs and PC Sound Card technology will mean the that peripherals that facilitate the implementations that require less processing power will also become available. A significant part of the research and development work outlined will be in the ensuring that the applications will run efficiently on platforms readily available to the target users. Much can be achieved towards making the software more efficient once the effects of the various software parameters on the perception of the users is fully understood but greater levels of processing power are required to fully investigate this in the first place.

5.3 *Workshop of Blind Computer Users*

A workshop was held with a small group of blind people, on 4 July 1998, to introduce them to the basic concept of 3 dimensional audio and encourage them to brainstorm on potential applications that could be of use to them or other blind and partially sighted people. The majority of the visually impaired participants were current or former OU students and this gave an educational bias to their perspective. Some of the key potential applications to arise from this are listed here but to set them in context the make-up of the workshop is tabulated:

Total number attendees (including facilitators)	14
Number of attendees having a significant visual impairment?	7
Self-identifying as blind	4
Sighted assistants	2
Involved in the development of computer or audio applications for visually impaired people	6
Having no visual memory	1
Number of those with a significant visual impairment describing there current computer usage as at least once per week	5
Of those with a significant visual impairment current usage of assistive technology when using a computer:	
Speech output	4
Braille Displays	-
Enlarged VDU Displays	1
Sighted Assistant	3

Key suggestions for the application of 3D audio virtual environments that emerged from these discussions were:

- Modelling of physical relationships of objects etc. - E.g. Model of the Solar system; Electro Magnetic Fields and their interaction; positions of fielders in a cricket match.
- Use in GUI-controlled Desk Top Publishing and Spreadsheets, etc - Use of 3D sound to give a better indication of where objects were on a page and for moving objects around, flowing text into boxes, etc. Use of sound cues to indicate colour; text attributes such as **Bold** and *Italic*.
- General improvements to the accessibility of a GUI by enhanced sound cues (a need for a common protocol for these cues was identified)
- Simulation of a work or social situations - e.g. in social science study of group situations
- In tele-conferencing position of speaker indicated by position their voice emanates from - turn-taking cues.
- VR - Suggestions of walkthroughs of buildings, London Underground - The idea of "hearing" walls, i.e. introduction or training in the use of echolocation. Navigational training in general.
- Sound maps

5.4 Investigations in Audio Perception and Practical Ambisonics

The limitations of both human audio perception and the practical implementation of Ambisonic theory both need to be further investigated to confirm that the envisaged applications for blind people are in fact viable and practical and to inform further development work.

Key questions to be addressed are summarised in the table below:

In human audio perception	In practical Ambisonic implementations
• What is the resolution and variability of human spatial discrimination in sound	• What spatial precision is required in the computer modelling of the soundfield to meet the limits of human perception
• Is the expected sense of "being present" in an audio virtual environment achieved with the implementations made and what are the key success factors for this	• Can virtual environments of sufficient complexity to be believable be created "real-time" given the available technology
• What are the particular requirements for visually impaired users of audio virtual environments where others use the visual modality. (e.g. GUIs)	• What is the most appropriate mapping of a 2D graphical display into 3D audio environment (e.g. GUIs)
• What is the effect of a 3D audio representations on exploration of a GUI by blind computer users	• In an implementation of "mumbling icons" what are the optimal system parameters for different tasks

These questions will be addressed in a series of simple psychophysical experiments undertaken with sighted, blind and partially sighted users. The visually impaired subjects are being recruited from OU students in the regions around Milton Keynes and others involved in local blind and visually impaired groups in the area (e.g. British Computer Association of the Blind). A series of inter-related experiments is being planned over the next year with 20 subjects attending at the laboratory on 3 separate occasions each. It is judged that valid detailed methodologies can be constructed with such sample sizes to meet the research objectives. A challenge for the detailed design of the experimental method is to isolate those factors due to the technology and those due to human perception.

6. REFERENCES

K. Crispen, H. Petrie (1993) Providing access to GUIs for blind people using a multimedia system based on spatial audio presentation. *Pre-prints of the 95th AES (Audio Engineering Society) Convention*, New York

A. Farina, E. Ugolotti (1998) Software Implementation of B-Format Encoding and Decoding *Pre-prints of the 104th AES Convention*, Amsterdam

P.B. Fellgett, (1972 Sept.) Directional Information in Reproduced Sound, *Wireless World*, vol. 78, pp. 413-417

M.A. Gerzon, (1974 Dec.) Surround Sound Psychoacoustics, *Wireless World*, vol. 80, pp. 483-486

M.A. Gerzon, (1980 Feb.) Practical Periphony, The reproduction of full sphere sound. *Pre-print of the 65th Audio Engineering Society Convention*, London

M.A. Gerzon, (1992 Mar.) General Metatheory of Auditory Localisation, *Pre-print 3306 of the 92nd Audio Engineering Society Convention*, Vienna

D.A. Keating, (1996) The generation of virtual acoustic environments for blind people, *Proc. 1st European conference on Disability, Virtual Reality and Associated Technologies,* pp. 201-207, Maidenhead UK

3D aural interactive hyperstories for blind children

Mauricio Lumbreras and Jaime Sánchez

Department of Computer Science, University of CHILE
Blanco Encalada 2120, Santiago, CHILE

mlumbrer@dcc.uchile.cl, jsanchez@dcc.uchile.cl

ABSTRACT

Interactive stories are commonly used for learning and entertaining purposes enhancing the development of several perceptual and cognitive skills. These experiences are not very common among blind children because most computer games and electronics toys do not have appropriate interfaces to be accessible by them.

This study introduces the idea of interactive Hyperstories performed in a 3D acoustic virtual world. The hyperstory model enables us to build an application to help blind children to enrich their early world experiences through exploration of interactive virtual worlds by using 3D aural representations of the space. We have produced AudioDoom, interactive model-based software for blind children. The prototype was qualitative and quantitatively field-tested with several blind children in a Chilean school setting.

Our preliminary results indicate that when acoustic-based entertainment applications are careful applied with an appropriate methodology can stimulate diminished cognitive skills. We also found that spatial sound experiences can create spatial navigable structures in the mind of blind children. Methodology and usability evaluation procedures and results appeared to be critical to the efectiveness of interactive Hyperstories performed in a 3D acoustic virtual world.

1. INTRODUCTION

Interactive computer games have been used for entertainment purposes for some time. However, it is during the last years that games have been available to a wider population of children. Today, most kids worldwide have had some type of experience with computer games delivered preferably through video-based devices (Druin et al, 1996). This scenario is not the case for children with disabilities (BCC 1993). They do not have interactive entertainment software available in quantity and variety. The case turns more critical with blind children, because they cannot take advantage of visual games.

This study reacts to this growing need by introducing an application based on highly interactive hypermedia stories for blind children. Hyperstories for blind children intend to assist the enrichment of early world experiences through free exploration of interactive virtual worlds by using a 3D aural representation of the space and surrounding entities such as the ones explored in (Mereu et al., 1996; Savidis et al, 1996; Lumbreras et al., 1996). The application introduced here stems from a design model for generic hyperstories. A hyperstory is defined by the combination of a navigable virtual world, a character manipulated by the user, a set of dynamics objects and other characters, and the traces of interaction among these entities given by the plot of the story (Lumbreras and Sánchez, 1997).

Our main research issue was to explore a model to describe an acoustic navigable environment. We also explored how spatialized acoustic modality combined with haptic manipulation of the environment, allow kids to construct mental associated structures such as haptic/acoustic correlation, spatial navigation without visual cues, object permanence in time through the hyperstory methapor, and interface and usability issues related to interactive software for blind children.

2. WHAT IS A HYPERSTORY ?

There are several applications, disciplines and computer environments that come together to the concept of hyperstory. One type of these environments are MUDs (Multi-User Dungeons) and their variations (MOOs, etc.). In the original version, these text-based systems allow many users to connect simultaneously to virtual "worlds" composed by rooms, objects, and people. Depending upon the design of a particular system, themes vary from fantasy environments with dragons and wizards, to futuristic exploration with spaceships and aliens.

Our model extends these ideas by including the elements of a story. These elements are: plot, roles, and characters. The main idea is to capture these elements in the representation (Hayes-Roth et al., 1996). Plot is a temporal sequence of actions involving a set of individuals. A plot and its constituent actions may be quite abstract, i.e.: A meets B, A loves B, A loses B, A wins B. Role is the class of individuals whose prototypical behaviors, relationships, and interactions are known by both actors and audience. For example, the plot outlined above ordinarily is instantiated with alternative roles, for instance: the boy in love and the girl he loves. Character is a personality defined as a coherent configuration of psychological trait, for instance, any of the characters in the present scenario might be: shy and sensitive, silly and affectionate.

We first introduce the definition of a Hypermedia Virtual Environment (HVE) as:

$$HVE := hypermedia_{I} + dynamic\ objects_{II} + characters_{III} \tag{1}$$

Where:

I. Hypermedia is:

- In charge of modeling the virtual world composed by several navigable environments connected between them by links. This is a special case of hypertext, each node basically represents a container of objects and a potential scenario of the hyperstory. Physical gates, portals and doors represented as links render the connectivity. Thus hypermedia models the spatial relationship and connectivity of environments. The concept of associated hypertext as underlying model to describe spatial navigable metaphors is commented in (Dieberger 1996).
- A modeling technique to provide basically the branching in the course of the story. But while this definition might be sufficient, it fails to convey a significant semantic aspect of the structures embedded in the hyperstories because of the complexity to model interaction patterns among entities by using just the node-link model.

II. Dynamic Objects are:

- In charge to represent the objects of the virtual world. They are entities that have behavior in time and react to the events produced by the user and other entities.

III. Characters are:

- The entities that carry on the main course of events involving a complex behavior. There is a distinguished character called the protagonist. This is manipulated by the user and represents the user-system connection. If the protagonist is third person viewed, an avatar will be in charge of this representation. Characters are special cases of dynamic objects and they are very important to the story level. Characters represent the main plot and elicit the content of the story. For example, in a film the most interesting events happen to the characters and develop the actions that emotionally impact to the audience.

But at this point a MUD or an adventure computer game could be in some way similar to the previous definition. A Hyperstory (HS) is an extension of this concept and structurally is:

$$HS = HVE + narrative \tag{2}$$

For this reason our model extends the idea of HVE by introducing the idea of an intentional sequence of events, based on plot, roles and characters. Other differences from MUDs arise from the idea of closure or explicit final, described as a good feature in narrative (Landow 1992).

2.1 The added value of hyperstories

A hyperstory is an interactive story guided by an intentional argumentative structure in a greater degree than a casual scenario. The plot in a hyperstory is not linear, is a hyper-plot. Here, action, object activation and dialog can trigger a change in the flow of the story. Thus, we borrow ideas from hypertext/hypermedia technology by including narrative in a virtual environment context (Bernstein 1996). Hyperstories have improved conventional literary stories by allowing a "dynamic binding" between characters, the world in which they move, and the objects they act on (Sánchez et al., 1996). The learner performs this binding through a greater flexibility in the learning process. In other words, a hyperstory is a combination of a virtual world where the learner can navigate, a set of objects operated by the learner, and the pattern of interaction between entities (Lumbreras and Sánchez, 1997).

In a particular execution instance of a hyperstory, two children may experience different views of the same virtual world, extending the ideas of (Joiner 1994). Slight changes introduced by the child to the object's behavior can produce different hyperstories in the same world. Children when manipulating a character can also interact with other characters to solve a given problem. Familiar environments such as schools, neighborhoods, squares, parks, and supermarkets can be interesting metaphors for building virtual worlds. It is interesting to notice that conventional computer authoring tools do not provide an adequate set of facilities for building acoustic hyperstories as we have conceptualized them.

3. THE MODEL

Our model is a design model based on object oriented concepts, providing a framework to describe the diverse building blocks of the hyperstory. The model supplies a framework composed by three foundational classes as described in OOD techniques (Rumbaugh et al., 1991). These classes are: context, link, and entity. In addition to these classes, there is an associated constructor called channel, similar to the route constructor in VRML environments. Contexts model the static world, links model the connectivity between contexts, entity is the abstract class that captures any object or character definition, and channels work as a broadcast media of events in a fan-in or fan-out fashion to the subscribed entities.

Each base class has a predefined behavior and a set of attributes that makes them different from each other (e.g. a link knows about the transportation of entities between contexts). Another example of specialized behavior arises from contexts: if an entity sends an event to a context, it sends the event to all contained objects. Thus a context works as a diffuser of events. All these base classes have behavior, based on a modal programming. The Objectcharts is the formalism to specify behavior (Coleman et al., 1992).

3.1 Main conceptual design building blocks

A hyperstory specification can be splitted in two interrelated conceptual parts by using the following classes:

- static scenarios (contexts and links),
- objects (entities) and the explicit routing mechanisms (channels).

3.1.1 The static world. Hyperstories with several scenarios organize them according to their physical connectivity (linking). For this purpose, we can describe the virtual world as a kind of nested context model. A virtual world is defined as a set of contexts that represent different environments. Each context contains an internal state, a set of contained contexts, a set of objects, links to other contexts, and a specific behavior. Different relationships may be held between two different contexts, such as:

- neighborhood (there is a link from one context to the other),
- inclusion (one context is included in the other),
- none (contexts are "disjoints").

The idea of a context is the same as in standard hypertext technology: a node or container. Different "real world" metaphors can be implemented easily with this simple model, such as a town, a house, a room -or houses within a town and rooms in a house-. All these metaphors are built in such a way that can be freely navigated. Another important concept about context is perception: a context is a spatial container that can be perceived as a whole rendered as a unity at the interface level. In this stage of the modeling, the base classes context and link are used by creating new ones through inheritance. At this point of the design, we are dealing with the first term of the Eq. (1).

3.1.2 Populating the world. In order to bring life to the hyperstory, we populate the environments with objects, some active, and some passive, orthogonal-composed by a navigational dimension. To avoid misunderstandings we briefly define some terms concerning objects in this context.

- **Passive**: the object answers only to simple events like "Who am I?"
- **Active**: the object has a noticeable behavior while the time progresses -continuous or discrete- or they respond to events with some algorithm that reflects some behavior.
- **Static**: the object always belongs to the same context.
- **Dynamic**: the object can be carried to the contexts by some entity or travel autonomously.

Any object or character (even the protagonist) will be a subclass of an entity. Therefore we need to extend the basic attributes and behavior of an entity. Basically, an entity can be viewed as an object that has a set of attributes that define an internal state and a behavior. Using a special made state-based scripts we describe the object behavior. In each state, there are a set of rules containing a triggering event, a pre-condition, and a list of actions that must be performed when the event arrives and the pre-condition holds. Each rule plays the role of a method in OOD jargon. But if we try to capture the nature of the narrative and the diverse branches of a hyperstory, the model must consider this requirement. Certain entities in a story can respond to the same event (message in OOD jargon) in a different way according to the story stage. For example, according to the stages of the hyperstory a person can respond: "fine" or "tired" related to the question "How are you?". In short, an object can behave differently to the same message received in its life stage. This concept is called programming with modes (Taivalsaari 1993) or state-based programming. To capture this feature the rules are not specified in a flat way, they are blocked and grouped according to the entity life stage. In short, we use state-based scripts in order to deal with this feature. By embedding narrative in the behavior of the entities we are satisfying the Eq. (2).

3. AUDIODOOM: A HYPERSTORY FOR VISUALLY IMPAIRED CHILDREN

Our aim was to test the hypothesis that a highly interactive and immersive aural environment can serve as a tool to stimulate and reinforce some processes and skills in blind children such as space representation. Sound serves as the output media of the system, but the transient nature of the sound imposes a bias in the interface design, leaving this tightly linked to temporal constraints. For this reason, the conceptual idea of interactive narrative combined with game challenging must be organized and rendered in a very simple way to model our target user, blind children aged 8-12 years old.

AudioDoom is the prototype that enables us to test our ideas about interactive Hyperstories for visually impaired children. This software is based on the idea of user-navigation in a set of corridors where the child gets and interacts with virtual objects, resembling in some way the classic Doom game. AudioDoom is based on a fantastic story about an extraterrestrial invasion to the earth, developing the action inside an extraterrestrial-flying source. The child must save our planet in order to get a successful story end. In the course of the hyperstory, the child encounters characters, objects, and challenges that may change the flow of the plot of the story.

The structure of the flying source is presented as a set of perpendicular corridors with different lengths (see Fig.1). These corridors are connected by means of doors that can appear at the end or at the side as an optional exit to other corridor. In each case the user can activate the desired door in order to access to the right corridor. Related to the physical navigation inside a corridor, the user is allowed to move in forward direction step-by-step. Certain entities can appear suddenly after a current step has finished. If this happens the user must solve a challenge depending on the type of entity found. For example, the monster is simple to destroy -three shoots- but the mutant moves in the space jumping between neighborhood voxels. For this reason, the child must localize this entity as soon as possible and then shoot immediately. It must be clear that each user action or entity appearance is rendered with spatialized sound.

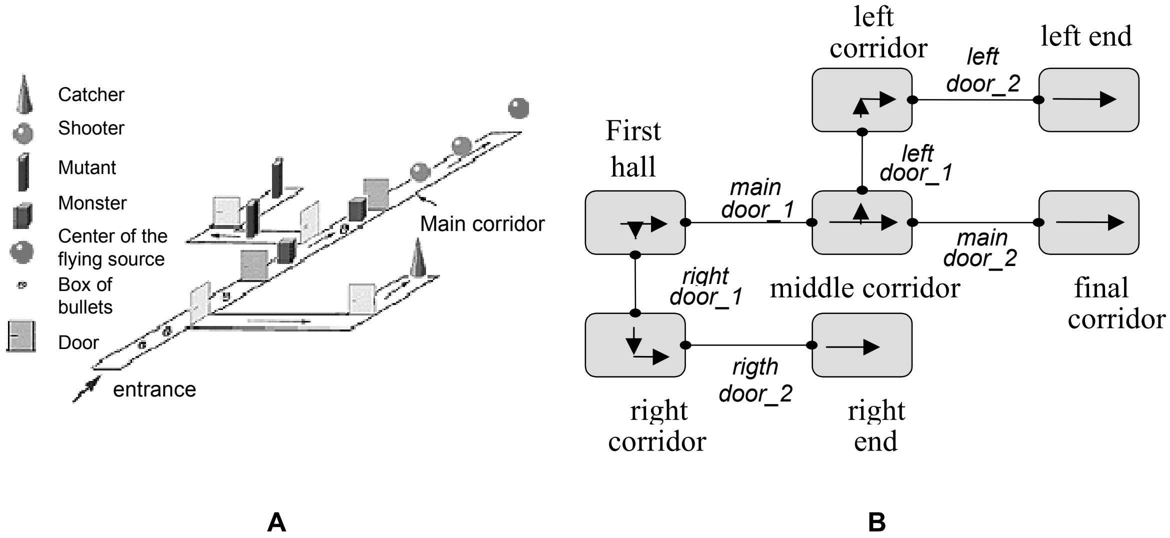


Fig 1 *Artistic version of the AudioDoom environment (A). Related to the hyperstory model, we have modeled AudioDoom as a set of context (grey rounded rectangles) linked by doors (lines with rounded ends) as the diagram (B) shows.*

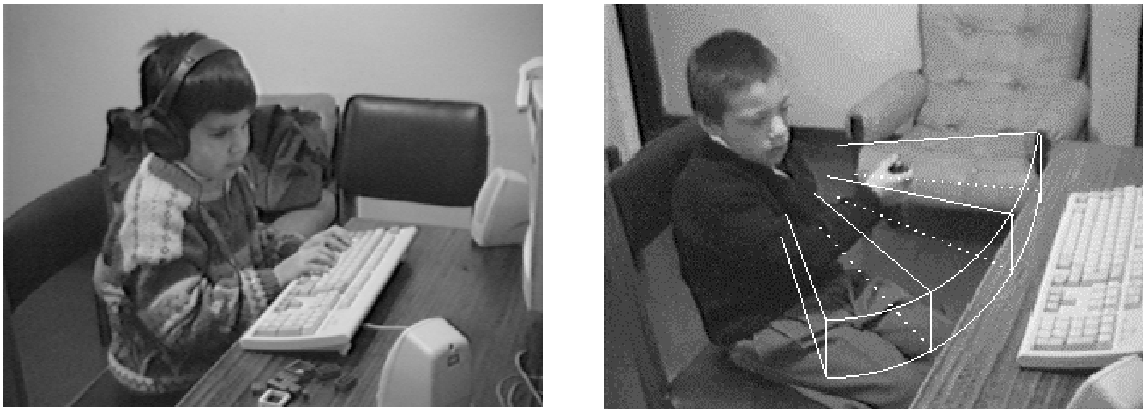
Typical actions involve to get objects (box of bullets), shoot to an entity (monster and mutant), or localize and interact with a determined character (the catcher) in a position of the space. The soul of the story presents multiple branches, but some of them are not deterministic, because some story entities may or may not be encountered, depending on the casual user-entity encounter. This scenario brings new alternatives in each session with AudioDoom. This spatial sequencing of the space enables the user to be involved in the story -resolving challenges *in crescendo*- increasing the level of complexity.

The added value of AudioDoom comes from the fact that we have used the hyperstory methapor to evaluate how a virtual acoustic representation can build a mental spatial representation in blind children. For this reason, we have built some tasks where the child interacts several times with AudioDoom and then tries to describe the taxonomy, organization, hyperstory entity location, and space organization of the environment by using LEGO blocks. In short, the hyperstory serves as an engagement device to test our hypothesis.

3.1 Interacting with AudioDoom

To interact with the virtual environment the child operates over the surrounding space, acting on voxels or minimal discrete units of volume. The voxel concept determines a discreteness of the space, simplifying the surrounding positions of interaction and creating a concrete repository for a certain entity. For example, in a moment a voxel can be empty or contain an entity. This entity usually is a virtual object represented acoustically, a door, a box, a character, etc. This entity can receive some events from the child depending on the entity: take, activate, and open. AudioDoom presents a modal interface where the same physical event can be interpreted according to the context, mode, and entity located in the target voxel. We must take into account that an entity can have a kinetic behavior, a movement in space along the time. This activity involves several voxels because a voxel is an atomic space container. This approach may appear a little restrictive, but we can divide the environments into the desired quantity of voxels until we obtain the desired granularity.

From the child point of view, AudioDoom is manipulated by using a wireless ultrasonic joystick called The Owl (Pegasus 1997). Through this device, the child can interact and move in the environment by clicking in different voxels of the surrounding space (see fig. 2). According to the position of the sound, the child must coordinate the haptic/kinestesic device with the perceived sound position. This scheme of action-reaction is strongly stimulated in the child, because of the strong haptic-acoustic correlation embedded. To deal with this topic we design AudioDoom to be mainly used with a ultrasonic joystick with 3 degree of freedom (X,Y, Z) and the use of 3D sound, but the child can use AudioDoom either by interacting with the standard keyboard or the mouse.



A **B**

Fig 2 *To introduce AudioDoom we test different and simple ways of interaction. For example we test 3D sound presented with headphones and an interaction mechanism keyboard-based (A). In other cases, the blind child interacts with an ultrasonic joystick and hear through external speakers. The voxels drawn over the photograph show three different volumes of interaction (B).*

3.2 The dynamic of interaction

The basic idea of AudioDoom is to split the navigable space into small atomic environments, the minimal scenario of action in a given moment. In this environment the child can interact with entities in different voxels. The linear connection of the atomic environments renders a corridor. This structure organizes the space into several corridors, giving a semantic and argumentative connection of the hyperstory and the space. These corridors are modeled as contexts and the doors as links.

The child can produce different types of activities in an atomic environment such as:

- To move forward the next atomic environment by giving a step
- To open a door
- To take a turn. This action has sense if a door appears in a different direction to the advanced.
- To interact with an entity in a certain way

If we consider the type of presentation media and the method of interaction of this hyperstory with a strong physical metaphor, we must consider three key points at the interface: the structuring of elements at a moment, the pointing of objects, and the dynamic of selection and interaction. In general, the system presents one or several entities at the time, each localized in a voxel. Then, the child after the acoustic localization, tries to point the entity and issue some events. According to the type of the entity, the interaction can be reduced to a discrete event -take a bullet box or to hit a door to be opened- or could be a chain of events with a given purpose: i.e. to shoot three times to destroy an alien, to shoot several times to destroy a mutant moving randomly between contiguous voxels.

3.3 Inside AudioDoom

AudioDoom was conceptualized with the following constraints:

- Stimulate spatial relations, by exploiting the surrounding child physical environment
- Capability to present disjoint and distinguishable acoustic entities, located in some point of the space
- Clear isolation between the input-output media in order to test various concepts according to each device
- Reflect a real time response related to the child action

Moreover, we choose to produce a software used by a wide population. In South America, schools and institutes for disabled children have little resources. For this reason, our software must run in a minimal platform. All these restrictions must not degrade the chance to render a virtual acoustic environment. In this version of AudioDoom, the sounds are only presented at ear level. It means that we do not include elevational cues.

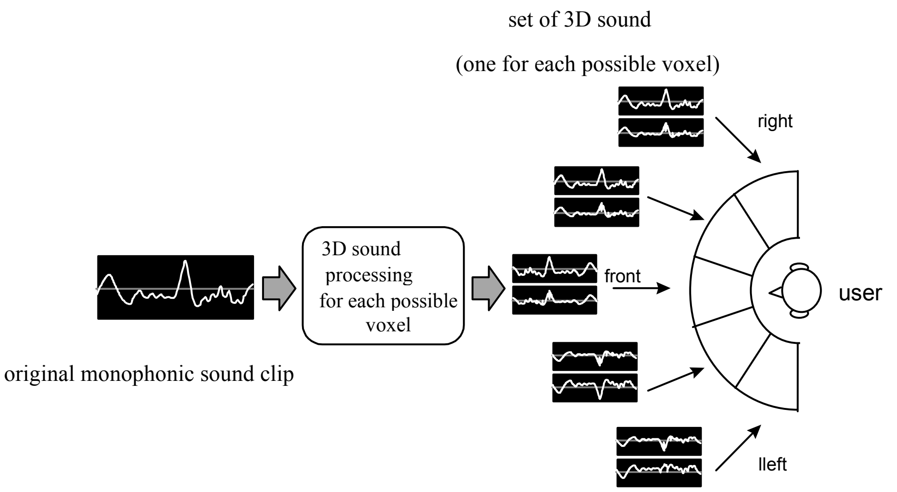


Fig 3 *Each sound clip is processed in an off-line way to create a 3D sound clip for each possible voxel. If v is the number of voxels and c is the number of sounds then n is the number of sound clips managed by the system, with n:= v * c .*

3.4 Implementation issues

Our approach followed the idea that if some entity can move between *n* possible voxels, we take the monophonic sound of this entity. Then by convoluting different sets of HRTF -one pair of each position- to the monophonic sound, we obtain n clips of 3D sound. This processing was done off-line. The result is a big set of 3D sounds requiring only a cheap sound board to be played (see Fig. 3). To deal with the real time mixing -background music, sound effects, entity sound, etc.- we use the Dynamic Link Library *wavemix.dll* - included in MS Windows-. Thus the execution hardware platform needs only a PC, Windows 3.1, and a stereo soundboard.

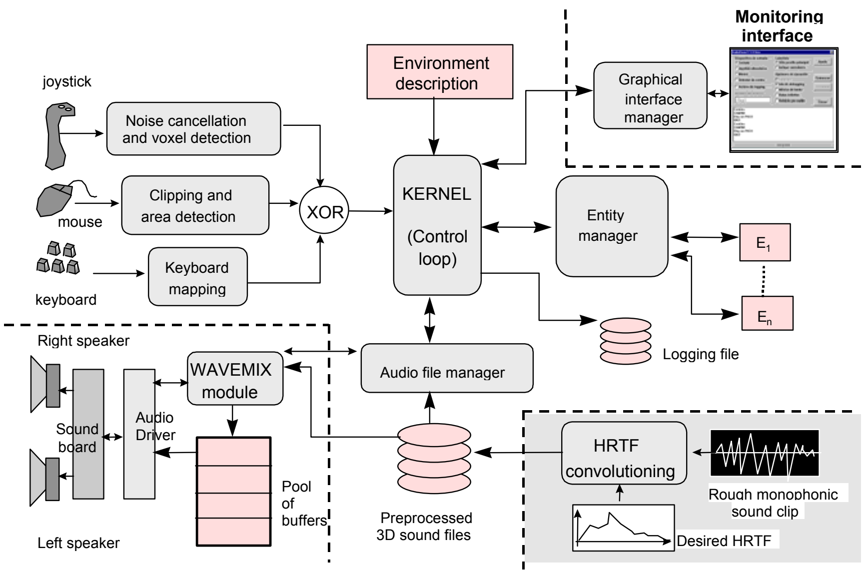


Fig 4 *AudioDoom is composed by several modules: the more important are the input processing, the entity management, and the sound playing. The gray area shows an off-line processing to reneder the preprocessed 3D sound clips.*

4. THE EVALUATION OF AUDIODOOM

At the beginning of our work there were several unclear topics: could this study allow to make strong inferences between causal relations?, how well defined were the theoretical ideas?, how confident we can predict that our findings are true and reliable, and the limits beyond will not?. Thus, we begin our evaluation with an exploratory approach to identify the mechanism related to perception and externalization of psychologically quantified variables. In this context, the evaluation of AudioDoom was an exploratory experiment, where by using several strategies we intend to know clearly the domain. The evaluation of AudioDoom was qualitative in a sense that we try to establish relevant elements about usability of interactive applications to be used without visual cues and to determine if our hypothesis was well grounded.

4.1 The testing scenario

AudioDoom was tested with seven Chilean blind children aged 8-11 ranging from totally blind since birth to other children with light and dark discrimination. In the first session the child interacts with AudioDoom by using the keyboard. The keys F, J, Enter, and the Space Bar are used to serve as milestones to orient the child in the keyboard. After a short oral explanation the child explores the interface and begins the hyperstory. The child interacts with AudioDoom during five hyperstory sessions and then we set the first evaluation. By using LEGO blocks the child tries to represent the environment structure of AudioDoom as he imagines and perceives. To accomplish this task, each type of LEGO block has an assigned semantic: long blocks represent a part of one corridor, cubes represent mutants, small cubes represent box of bullets, etc. Small plastic doors represent the perceived doors. Table 1 shows the testing progress.

5. DISCUSSION

After a preliminary user evaluation we have demonstrated that it is possible to render a spatial navigable structure by using only spatialized sound. This mechanism preserve in a notably degree the structure, topology, structural relationships, meaningful orientation, navigation, and mobility elements. The result is preliminary because we have not included free navigation in open places within the virtual environment due to the restriction of the navigation in straight corridors with divergent branches connected to 90°. Some children show some difficulties, especially with mapping transversal corridors. This problem apparently arises from the fact that the turn disorientates the user, because the real surrounding space is fixed -chair, table, etc.- For this reason, we face as a key issue the representation of distinguishable milestones in the environment to facilitate the orientation. The use of some artificial auditory beacon can improve the orientation.

Even though we use 3D sound with several limitations -no head tracking,limited quantity of voxels- children usually prefer external speakers. Children are not so clear about this fact but one reason could be that the headphones impose the isolation, limiting the oral interaction with the evaluator. The discomfort imposed by the headphone used (a Sony MDR CD30) appears to be another reason. We detected this pattern of preference at the beginning, so we adapted the HRTFs to be used to external speakers by reprocessing the amplitude of the signal of each channel. This result motivates the use and study of transaural audio, which enables to spatialize sound with external speakers (Gardner 1997).

We carefully observe the mechanism of interaction in AudioDoom. The keyboard offers better confidence because there is no ambiguity of the selected voxel. The ultrasonic joystick reflects some problems due to erroneous voxel selection and undetected clicking because misalignment of the joystick related to the ultrasonic sensors. But children report more level of satisfaction with the joystick. It seems that the movement of the child arm increases the level of immersion. Furthermore the haptic-acoustic correlation is an excellent mechanism to stimulate the available skills in visually impaired children.

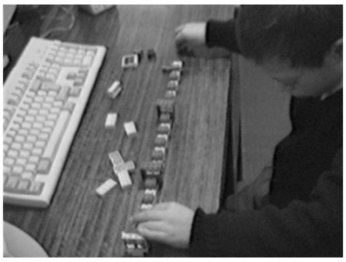
One key element in the further improvement of AudioDoom is the construction of an editor, because right now AudioDoom is a hardwired solution.

Table 1 *Photograph sequence showing how the child build the perceived navigable structure*

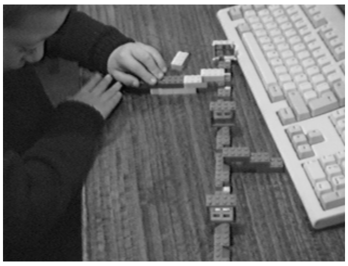
1. The child begins the first hyperstory sessions by using the keyboard. After some practical training, the child test the joystick and external speakers (see Fig. 1).

2. Some interactions with AudioDoom are sufficient for the child to begin the construction of the main corridor by using the LEGO blocks.

3. After some questions about each part of his model, the child continues the building of the main corridor



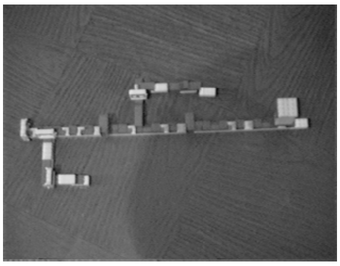
4. The main corridor practically is finished, representing the path from the beginning to the center of the flying source. The child locates each entity at the perceveid position in his traveling.



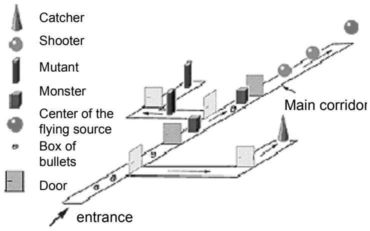
5. With confidence about the main structure, the child interacts again with AudioDoom, navigating the divergent corridors. To accomplish this task the child comes back to the previous built model and then extends it with the new perceived acoustic structure.

6. In this case the child locates a door found in the navigation of the new corridor. At this moment the child reflects difficulty because the door is actually located aside instead of the front direction of advance.

7. While the child progresses in the construction of the model, he is orally inquired about the current activity. In this case the evaluator asks about the advance direction in the model and the child answers the perceived advance direction, relating adequately to the LEGO construction.



8. This is the model built by the child. You can observe the similarity with the artistic graphical version of the AudioDoom environment. This result reflects that an acoustic virtual environment creates a mental image. The most interesting thing is that the child never saw the graphical representation of the AudioDoom navigable structure.



9. This diagram represents the topology and distribution of entities of AudioDoom. With a high degree of precision the child expressed the perceived structure. This is easy by comparing the last photo and this graphic.

6. FINAL REMARKS

One of the promisoriest results derived from AudioDoom comes from the fact that not only virtual acoustic environment can serve as entertainment environment. Moreover these ideas can be used to deliver educational material. It is well known that blind children need help to know and mentally map their neighborhood, school, downtown, etc. In this way we are exploring the possibility to render not only fantastic environments, but virtual representations of real and familiar places. These representations can be modeled with the hyperstory model by including motivating elements to capture the attention of the children.

We have presented a conceptual model for building highly interactive stories. Learners have control over stories, access to diverse tools and materials to construct with, in order to develop strategies and test hypothesis with the implicit idea of fostering the development and use of the skills to determine spatial relationships and laterality. We believe that 3D aural hyperstories can contribute to make the interaction with computers much more enjoyable and learnable to learners. Children like stories and remember them easily. When children get engaged in a story, they can identify, retrieve and use relevant data to solve a challenge by having rapid and flexible access to the story sequence. From our experience with AudioDoom we have learned that hyperstories highly motivate learners, facilitate free navigation, and promote active constructivist learning by providing powerful materials and tools to construct with.

Acknowledgement: This work was partially supported by the Chilean Science and Tecnology Fund (FONDECYT) Grant 1950584 and CYTED Project SIMHI, VII-8.

7. REFERENCES

BCC (1993), *First Steps, a Handbook for teaching young children who are visually impaired,* Blind Children Center, Los Angeles, California, USA.

M Bernstein (1996), Conversation with friends: hypertext with characters, *Lectures on Computing,* Springer Verlag.

D Coleman, F Hayes, S Bear (1992), Introduciong object charts or how to use state charts in object-oriented design, *IEEE Transactions on Software Engineering*, Vol.18, No.1: 9-18.

A Dieberger (1996), Browsing the WWW by interacting with a textual virtual environment - A framework for experimenting with navigational metaphors, *Proceedings of ACM Hypertext '96*, pp.170-179.

A Druin , C Solomon (1996), *Designing Multimedia Environments for Children,* John Wiley & Sons Inc.,USA.

B Hayes-Roth, R Gent, D. Huber (1996),.Acting in character, *Technical report KSL-96-13*, Knowledge Systems Laboratory, USA.

D Joiner (1994), Real Interactivity in Interactive Entertainment, *Computer Graphics*, Vol 28, number 2: 97-99.

G Landow (1992), *Hypertext: the convergence of contemporary critical theory and technology,* Baltimore: The John Hopkins University Press.

M Lumbreras, J Sánchez, M Barcia (1996), A 3D Sound Hypermedial System for the Blind, in *Proceedings of the First European Conference on Disability, Virtual Reality and Associated Technologies*, pp.187-191, Maidenhead, UK.

M Lumbreras, J Sánchez (1997), Hyperstories: A Model to Specify and Design Interactive Educational Stories, in *Proceedings of the XVII International Conference of the Chilean Computer Science Society*, Valparaiso, Chile. Published by IEEE.

S Mereu., R Kazman (1996), Audio Enhanced 3D Interfaces for Visually Impaired Users, in *Proceedings of ACM CHI 96*, pp. 72-78.

Pegasus, *Pegasus Web Site*, http://www.pegatech.com.

J Rumbaugh, M Blaha, W Premerlani, F Eddy, W Lorensen (1991), *Object-Oriented Modeling and Design,* Englewood Cliffs, NJ:Prentice Hall.

J Sánchez, M Lumbreras (1996), Interfaces for learning. To be Published in *Advances in Human/Factors: Human-Computer Interaction*. Elsevier Publishers

A Savidis, C Stephanidis, A Korte, K Crisipie, K Fellbaum (1996), A Generic Direct-Manipulation 3D-Auditory Environment for Hierarchical Navigation in Non-visual Interaction, in *Proceedings of ACM ASSETS 96,* pp. 117-123

D Schwabe, G Rossi, S Barbosa (1996), Systematic Hypermedia Application Design with OOHDM. *Proceedings of ACM Hypertext ´96* , Washington DC, pp. 116-128.

A Taivalsaari (1993), Object-Oriented programming with modes, *Journal of Object Oriented Programming*, Vol. 6, No. 3: 25-32.

Sign language formal description and synthesis

O Losson[1] and J-M Vannobel[2]

[1,2]Laboratoire I3D (CEP) - IRRH, Université des Sciences et Technologies de Lille
Bâtiment P2, Cité Scientifique, 59655 Villeneuve d'Ascq, FRANCE

[1]*lo@cal.univ-lille1.fr*, [2]*jmv@cal.univ-lille1.fr*

ABSTRACT

Special needs of deaf people appeal henceforward to sign language synthesis. The system presented here is based on a hierarchical description of sign, trying to take the different grammatical processes into account. Stress is laid on hand configurations specification thanks to finger shapes primitives and hand global properties, and on location and orientation computation issues. We then expose the results achieved from the corresponding written form of signs, leading to their computer virtual animation.

1. INTRODUCTION

Sign language appears to be the main means of communication within the deaf community. In spite of a repression cycle it has suffered from in several countries until recent past, it remains a living language, in the full acceptance of the term, with communication capacities similar to those of oral languages.

Our project of synthesizing sign language is based on the established facts that half of deaf people encounter difficulties in reading and, as a result, suffer from subeducation. Being able to synthesize their mother tongue would undoubtedly be of major interest for those persons. In education as well as in daily life – such as in an emergency -, the generation process of synthetic signs offers many advantanges in comparison with video : facility and swiftness of generation and transmission, availability of the virtual signer and smoothness of transitions between lexical items.

The first three-dimensional synthesis of sign has been achieved in the early eighties (Shantz and Poizner, 1982). But it suffered from the lack of power of computers at the time, which entailed low-level (joint angles) specification of body postures. Surprisingly enough, very few attempts have been made since then to carry out the same goal.

This study is in keeping with the global aim of translating written French into signs. But the starting point considered lies in a textual representation of the signed sentence which is to be animated; as a consequence, we will not deal with sign language syntax here. The main core of the system is a sign formal description, grounded on a set of primitives isolated beforehand, stemming from both linguistic works and more synthesis-oriented research.

2. FORMAL DESCRIPTION

2.1. Grammatical Bases and Modulation Processes

Following Stokoe's footsteps, linguists have proved that sign languages, like oral ones, were doubly articulated into phonemic and morphemic (or monemic) levels. This decomposition has definitely given evidence for deaf people's signs system to be counted among all languages in the world. Second-level articulation units have been identified on the minimal pairs criterion (Nève, 1997), i.e. two signs differ from one another in only one of these *cheremes*. Belonging to one of four spatial types - hand configuration, manual location and orientation, and movement -, they combine to form morphemes, lexemes (signs) and whole statements; their psycholinguistic actuality has been besides fully validated.

Sign language has its own grammar, based on spatial, physiological and temporal dimensions. For instance, strongly iconic classifiers and size-and-shape specifiers, but also pronouns and index references, appeal to mechanisms involving the signer's space. Non-manual expressions both play paralinguistic (kinetic

 ISBN 0 7049 1141 8

stress, face expression) and grammatical roles (marking syntactic clauses). At last, it has been shown that special modulation processes (Klima and Bellugi, 1979; Namir and Schlesinger, 1978) affecting repetition, shape, amplitude and kinematics of movement (Loomis et al, 1983), were used to express subtle variations in meaning.

Liddell and Johnson (1989) have proposed a highly detailed description of sign based on the partition between segmental (holds and movements) tier and articulatory bundle tier (containing features of the hand). One of the major interest of this approach is to tackle with particular phonological (hold deletion, assimilation) but also morphological processes of sign language, including the one – of fundamental importance – of subject and object agreement.

2.2. A Sign Language Formal Description

The proposed sign language formal description system tries to take the widest range of such processes into account. To achieve that purpose, it has been split into two levels. Sentence, on the one hand, has its own parameters (localization and indexic references, grammatical clause descriptor, …). On the other hand, signs description may inherit from some of those discourse parameters. We will focus here on sign specification.

In order to provide the most general possible synthesis, with abilities to describe not only French- but every Sign Language, one of the basic underlying principles was to identify primitives at all levels, aimed at gradual combination into more and more complex structures (see figure 1).

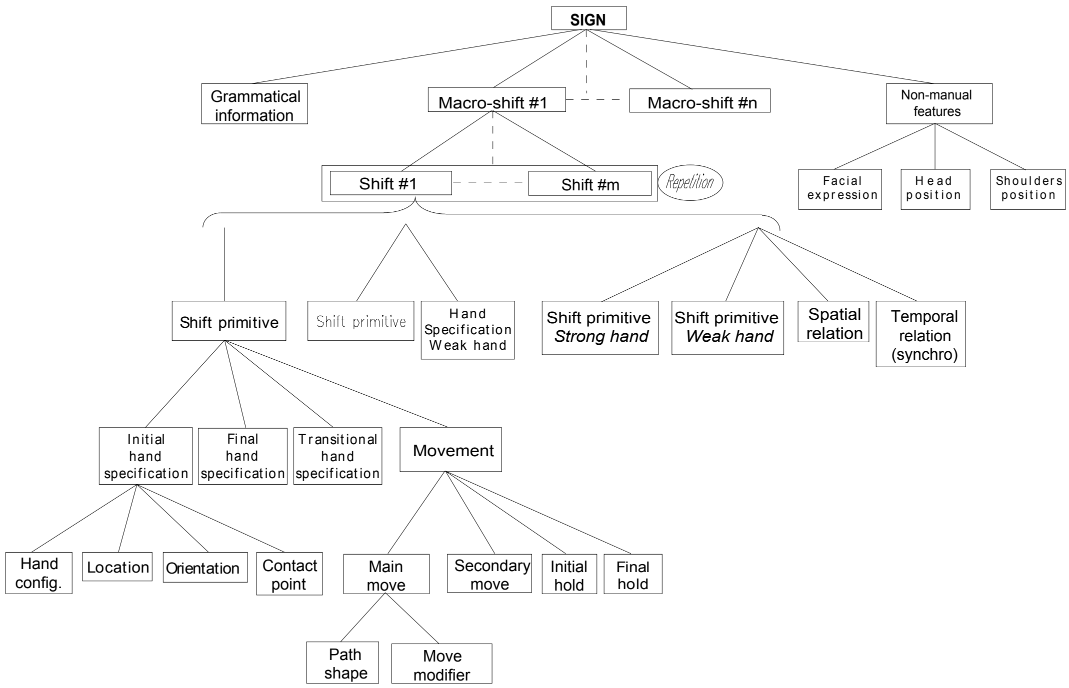


Figure 1. *General hierarchic structure of sign specification.*

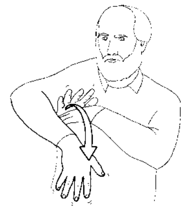
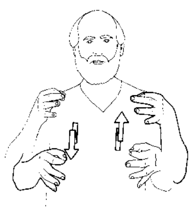
In that way, a hand configuration is described in terms of digital primitives and global hand properties (see details in section 3). Together with orientation, location and an optional manual point, it makes up what we call *hand specification*. Specifying a contact point is convenient in those many cases where the latter, rather than the wrist, must be located at the given position. Movement, as for it, is composed of one main move, with specific path and dynamics. Optional preceding and following holds, as well as some superimposed secondary movement (waving of fingers for instance), may complete movement specification.

A *shift primitive* is built as a bundle of hand specifications at the beginning and end of the sign, and of the movement itself. A transitional hand specification may be added too if necessary. Depending on the number and activity of articulators implied in a sign, a *shift* can be described

- either as a single shift primitive (if the sole strong hand is present),
- or a (dominant hand) shift primitive plus a hand specification for the weak hand,
- or two shift primitives if both hands are active. The description is completed with a spatial relationship

and a descriptor for synchronization between the two movements in this case (see table 1).

Table 1. *Two examples of signs with main features description.*

Sign	Meaning	Description
	French Sign Language (FSL) : ***Waterfall***	*Hand specification* for the static weak hand : Hand configuration [Flat], close to [torso], facing [down] *Shift primitive* for the strong hand : Initial hand specification : [5] at [weak hand back] facing [down] Final hand specification : [5] at [lower torso] facing [back] [Arc-shaped] movement while [wiggling] of fingers
	American Sign Language (ASL) and FSL : ***Judge***	*Shift primitive* for the strong hand (implicit for the weak hand) : *Hand spec.* : [clip] at [ipsilateral upper torso] (initial) or [ipsilateral lower torso] (final) facing [contralateral] [Straight] movement *Spatial relationship* : [symmetrical] hands, [alternate] move repeated shift according to [simple repetition]

We introduce the notion of *macro-shift* in order to be able to describe quite complex signs (such as compounds). This is made up of one or more shifts, optionally followed by a repetition. Isolation of the different kinds of repetitions, temporal and spatial relations, has been made by a systematic census on a french signs corpus (Moody, 1986).

At last, we have endowed our sign description with grammatical and non-manual information. Facial expression is of particularly salient importance when considering grammatical parameters inherited from the sentence level.

3. HAND CONFIGURATIONS

Hand configurations in sign languages have given rise to several coding systems (Kurokawa, 1992; Lee, 1994). Among them, HamNoSys (Prillwitz and Zienert, 1990) seems to be able to encode the widest diversity of hand shapes. We have set up a new descriptive method, based on the examination of different inventories, and split into three levels : finger shape primitives, hand global properties and constraints on joints.

3.1. Finger Configuration Primitives

All fingers except thumb have the same behaviour in terms of movement : they may flex at their interphalangeal joints, and both flexion and abduction (spreading of fingers) are permitted at the metacarpophalangeal (MCP) joint. Thumb has a larger reachable workspace due to greater articulatory complexity. Especially, it may contact other fingers in various ways. In those cases, its shape will be considered here as constrained in a hand global property (primitives specifying only free thumb configurations).

Figure 2. *Example of* #Flat *finger configuration primitive.*

We have isolated seven finger shape primitives, and five ones for the thumb, as being used in sign languages. For instance, the configuration called '#Flat' (see figure 2) has flexed MCP joint and extended proximal and distal interphalangeal joints. The very peculiar '#E' configuration has been taken into consideration at the present level, although fingers interaction is involved. As a matter of fact, it almost merely appears in the corresponding manual alphabet entry.

3.2. Hand Properties and Constraints on Joints

Hand properties imply more than one digit. Thumb and finger contact (or opposition), and abduction or crossing of fingers are here taken into account. Many kinds of such relationships were found in sign languages :

- Contact, or opposition without contact, between thumb and another finger :
  - *Ventral contact* : thumb tip is in touch with the middle phalanx of the finger in question, that may have any configuration (#Hook in most cases).
  - *Flat contact* between pads of thumb and of another finger, which should have #Flat configuration.
  - *Contact between tips* of thumb and of another finger, which should have #O configuration.
- *Crossing* of index and middle fingers. Other kinds of crossing are far more constraining for joints and absent from sign language.
- *Intercalation of thumb*, mainly between index and middle fingers. Thumb may also functionally come between middle and ring fingers, or between ring and middle fingers; nevertheless, those situations are seldom used in sign language.
- *Dorsal contact* : thumb pad covers a finger on the back of its middle phalanx.
- *Abduction / adduction* is also considered here, insofar as it generally involves several fingers, if not the whole hand.

Figure 2 above shows an example of a flat contact between the thumb and index finger.

Final step for hand configurations specification is the application of constraints on joints. The metacarpophalangeal structure of ligaments implies indeed more or less strong interdependence between flexions of neighbouring fingers at that joint. Inequalities proposed in (Lee and Kunii, 1993) have been applied on MCP joints as flexion limits, in order to achieve realistic hand shapes synthesis. Moreover, in a clenched hand, fingers converge towards the scaphoid point. That phenomenon has been included in the synthesis process by setting artificial abduction, proportional to flexion, when the latter exceeds two third of its static maximal value.

3.3. Hand Configurations Synthesis

The kinematic skeleton of the thumb, as for the whole body, is mathematically modeled by ideal joints and flat segments, each one being defined in the local coordinate system attached to the proximal joint. Axes are chosen such that x is the main axis of each segment, oriented from the proximal to the distal joint, and that (xz) is the plane in which segment points are given (palm for instance in figure 3), y being defined by the right-hand rule. Rotations about those axes will be considered in the following order : adduction-abduction (yaw) θ about the y-axis, then flexion-extension (pitch) φ about the x-axis, and finally axial rotation (roll) ψ about the z-axis.

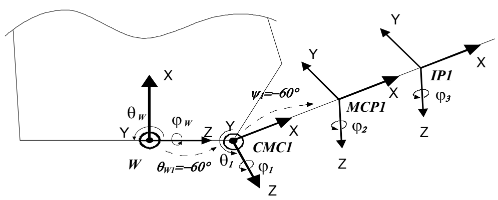


Figure 3. *Kinematic model of the thumb.*

The major issue in hand configurations synthesis is thumb posture computation. As called to mind above, this articulator has more degrees of freedom than other fingers. Several methods have been described to position such articulated structures, even when multiple goals are to be achieved (Badler et al., 1987). We have rather rationally tried to simplify our initial model so as to obtain a non-redundant system and solve it. Axial rotation takes place at the carpometacarpal (CMC) joint, especially during opposition movement. This one will here be undertaken by a fixed initial rotation, about the axis of the first metacarpal (see figure 3). Neglecting the low adduction-abduction angle, only flexion will be considered at the MCP joint, which is moreover assumed not to vary in a wide range of values around ten degrees. Thumb tip may therefore be expressed in the CMC local coordinate system as

$$x_T = \begin{pmatrix} c\theta_1 c\varphi_1 & -c\theta_1 s\varphi_1 c\psi_1 + s\theta_1 s\psi_1 & c\theta_1 s\varphi_1 s\psi_1 + s\theta_1 c\psi_1 & p_1 c\theta_1 c\varphi_1 \\ s\varphi_1 & c\varphi_1 c\psi_1 & -c\varphi_1 s\psi_1 & p_1 s\varphi_1 \\ -s\theta_1 c\varphi_1 & s\theta_1 s\varphi_1 c\psi_1 + c\theta_1 s\psi_1 & -s\theta_1 s\varphi_1 s\psi_1 + c\theta_1 c\psi_1 & -p_1 s\theta_1 c\varphi_1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\varphi_2 & -s\varphi_2 & 0 & p_2 c\varphi_2 \\ s\varphi_2 & c\varphi_2 & 0 & p_2 s\varphi_2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_3 c\varphi_3 \\ p_3 s\varphi_3 \\ 0 \\ 1 \end{pmatrix}$$

where *c* means cosine, *s* means sine, and p_1, p_2, p_3 are segment lengths, from the proximal to the distal one. Solving of those inverse kinematics equations is achieved thanks to a method presented in (Kobrinski and Kobrinski, 1989), assuming that the distance from x_T to the target position x expresses as a function of only one of the unknowns, others remaining fixed. Let α_i be the i-th rotation angle and let us determine its optimal value α_i^*. Rewriting x_T as

$$x_T = a_\alpha \sin \alpha_i + b_\alpha \cos \alpha_i + d_\alpha ,$$

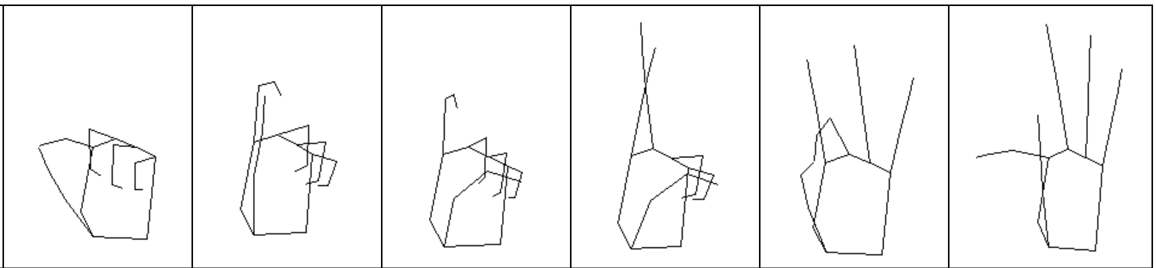
where vectors a_α, b_α, and d_α do not depend upon α_i, the minimum of distance $|x - x_T|^2$ is reached when

$$tg\alpha_i^* = \frac{{}^t a_\alpha (d_\alpha - x)}{{}^t b_\alpha (d_\alpha - x)}$$

From the two possible values of α_i^*, we must keep the one leading to a positive value for the second-order derivative, or for the quantity ${}^t(a_\alpha \sin\alpha_i + b_\alpha \cos\alpha_i)(d_\alpha - x)$.

This method provides realistic shapes for the thumb, as shown in figure 4 below.

Synthesized hand configurations



Hand Prop. / Parameter(s) / Config. name	#Pad	#Ventral	#Covered	#Covered	#Tip	#ThumbBetween
Parameter(s)	2	2	3, 4, 5	4, 5	3	2, 3
Config. name	[Beak]	[Key]	[Hook]	[R]	[8]	[T]

Figure 4. *Wireframe 3D synthesis of hand configurations using different hand properties.*

4. LOCATION AND ORIENTATION

4.1 Location

Ranges of locations in sign language are limited by the reachable workspace of the hand (Lenarcic and Umek, 1994) : almost always above the waist, ahead of the front plane, roughly within half spheres placed in front of the torso, in front and both sides of the head. In our system, hand may be assigned a location in two ways :

- in space, with main points on a grid defined as intersections of three planes (horizontal, parallel to the frontal plane, parallel to the sagittal plane);
- on the signer's body (including the weak hand), at various locations inspired by linguistic studies (Liddel and Johnson, 1989).

Points on the body can be expressed in terms of spatial coordinates in a 3D synthesis. Therefore, what we have to do here is, given a target location for the wrist, compute the four angles of the arm (three at the shoulder - θ_s, φ_s, ψ_s – and one at the elbow - φ_e). Redundancy of that positionning problem has been solved as explained below.

- First, one natural configuration of the arm is selected for each main point in space, among a collection of possible quadruplets obtained by successive iterations on θ_s in the following equations :

$$s\varphi_s = \frac{BC \pm A\sqrt{A^2 + B^2 - C^2}}{A^2 + B^2} \quad , \quad with$$

$$A = 2l_1 xc\theta_s - 2l_1 zs\theta_s \quad , \quad B = 2l_1 y \quad , \quad C = {l_1}^2 - {l_2}^2 + x^2 + y^2 + z^2$$

$$l_2 c\varphi_e = xc\theta_s c\varphi_s + ys\varphi_s - zs\theta_s c\varphi_s - l_1$$
$$l_2 s\psi_s s\varphi_e = xs\theta_s + zc\theta_s$$

- In the second time, we can get the four arm angles for any point in space from those configurations of the neighbouring points N. The starting point is computing the average angles weighted by the inverse distance to each of them :

$$\bar{q} = \left\{\ \bar{\theta}_s, \bar{\varphi}_s, \bar{\psi}_s, \bar{\varphi}_e\ \right\} \quad with \quad \bar{\alpha} = \frac{\sum_{i \in N} \frac{1}{d(P,i)} \alpha}{\sum_{i \in N} \frac{1}{d(P,i)}} \quad , \quad \alpha \in \left\{\ \theta_s, \varphi_s, \psi_s, \varphi_e\ \right\}$$

  We then keep the solution, among the set *Poss* of possible ones, which is the nearest from that mean value :

$$q^* = \min_{q \in Poss} \left(\left| \ q - \bar{q}\ \right| \right)$$

The above stratagem keeps us out of generating unnatural configurations for the arm, while avoiding greedy algorithms of inverse kinematics.

4.2 Orientation

According to linguistic studies, palm orientation seems to be sufficient to encode hand orientation. Six main directions only (up, down, front, back, ipsilateral and contralateral) are generally isolated as a result of phonological substitution. We cannot release from specifying hand orientation more precisely in our synthesis perspective. This is done by defining both a forward vector ($\vec{n}$) normal to the palm, and a vector ($\vec{i}$) oriented according to the (virtually) pointing index finger, or to the axis of the third metacarpal, from CMC3 to MCP3.

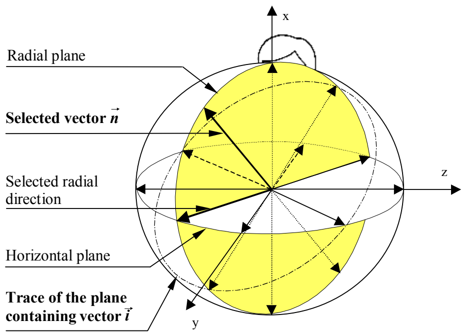


Figure 5. *Directions specifying absolute orientation (FSL credit).*

The problem of hand global orientation consists in computing :

- the radio-ulnar pronation-supination angle ψ_e,
- two angles at the wrist joint : adduction-abduction θ_w and flexion-extension φ_w.

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998
©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

In mathematical terms, let ${}^{R_0}A_{R_2'} = (s \quad n \quad a)$ be the current orientation of the local coordinate system at the elbow, after shoulder rotations and elbow flexion have been applied. The required global orientation of the hand thus expresses as :

$${}^{R_0}A_{R_3''} = (i \quad n \quad i \wedge n) = (\sigma \quad \nu \quad \alpha) = {}^{R_0}A_{R_2'}\,A(x_2',\psi_e)\,A(y_3,\theta_w)\,A(z_3',\varphi_w)$$

Identification leads to the desired angles :

$$\psi_e = -Arctan\left(\frac{{}^t n\alpha}{{}^t a\alpha}\right) \;,\; \theta_w = Arctan\left(\frac{{}^t s\alpha.c\psi}{{}^t a\alpha}\right) \;,\; \varphi_w = -Arctan\left(\frac{{}^t s\nu}{{}^t s\sigma}\right)$$

It often appears easier and convenient to specify a particular point M on the hand that is to be settled at the given place, rather than setting location of the wrist W. In sign *I-give-you* for example (see section 5), the index finger tip should be located on the upper torso at the beginning of the movement. Since coordinates of M are known in the local coordinate system R_3'' attached to the palm, the wrist location can be found easily as

$$\overrightarrow{OW}_{R_0} = {}^t(x \quad y \quad z) = \overrightarrow{OM}_{R_0} - {}^{R_0}A_{R_3''}\,\overrightarrow{WM}_{R_3''}$$

In the last place, specifying relative orientation has been contemplated by using a set of predefined values for each angle ψ_e, θ_w, and φ_w. In a number of cases besides, only pronation-supination of the forearm is required to describe the orientation of the hand. For the moment, this method is solely used when the wrist joint – no hand point – is concerned by the target location.

5. COMPUTER SYNTHESIS OF SIGN

5.1 Synopsis

To specify any sign according to the structure presented in section 2, we have prefered a textual extensive description of the sign, rather than a symbol-based coding system (on grounds of legibility and easiness of data exchange). The sign compiler implements a lexical analyser, a parser, and the evaluator itself. It builds the hierarchic sign specification structure in memory from predefined entities like hand configurations, while inheriting from grammatical information from the sentence level through sign parameters (see figure 6).

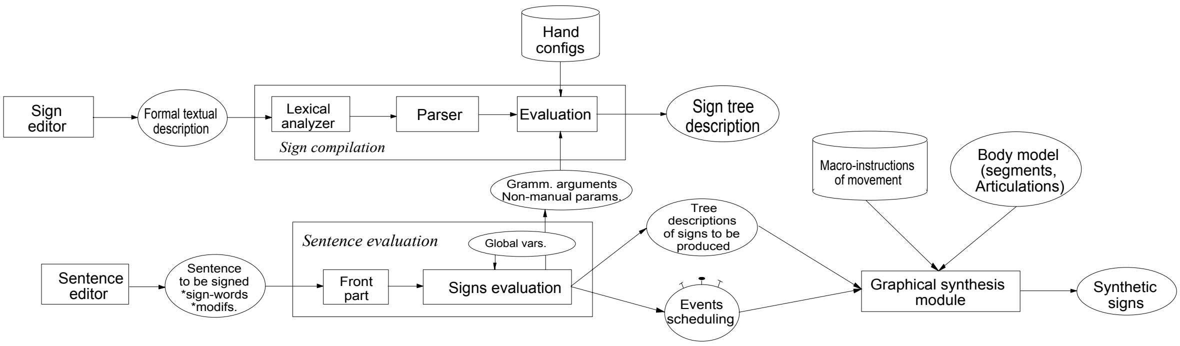


Figure 6. *Synopsis of sign synthesis.*

The graphical synthesis module generates the signs after each one has been previously evaluated in this way. The underlying model of the human body is a hierarchic tree of upper body segments and joints, from the torso to the eyebrows and distal phalanges. Each segment is described by a collection of 2D-points and has one or more children segments together with the proximal joint attached (including the local and global coordinate systems). Moreover, segments are regrouped within macro-segments (fingers, hands, head, ...) accepting high-level messages.

Input sentences look as streams of sign-words and escape codes handling grammatical features such as indexic references and role play, time setting, clause type (condition, wh-question, yes/no question, imperative), grammatical repetition, etc. Signs are described in a formal language with special syntax close to the Smalltalk object-oriented programming language.

5.2 Results

In the current progress of our work, the system computes arm postures as well as hand configuration and orientation from the given sign textual description. Symbols used to specify hand shapes, contact points and spatial locations and orientations, are evaluated so as to transmit suitable objects to the human body structure. The graphical synthesis has been achieved by means of connected deformable polygons, with distance-dependent lighting and removal of hidden surfaces. Primitives of facial expressions have been added too, as part of lexical items, but conveying also crucial grammatical information.

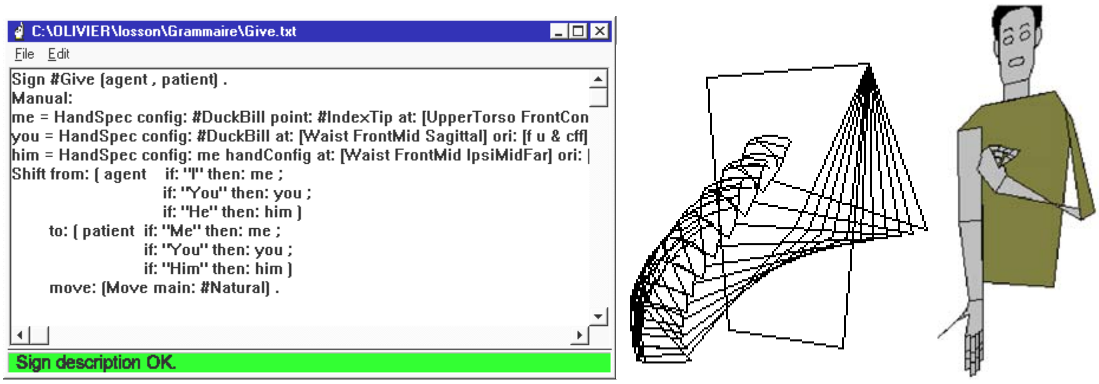


Figure 7. *Textual description, wireframe and polygonal synthesis of sign* I-give-you.

Two parameters have been inserted in the sign *give* above to take the subject and object agreement into account : thanks to local variables, we are able to determine the sign features with the different possible values (*I/me*, *you* and *he/him*) for the agent and patient parameters. The syntax used leads to a very readable description of the sign, in which the formal hierarchic sign specification clearly appears. It is built and referenced to by sending messages to objects. A graphical interface is provided with the editor in order to make sign edition easier. So far, it allows the user to add hand specifications with direct tridimensional visualization of the selected elements.

For the moment, movement remains simple with ‘natural’ shifts (through joint angles extrapolation) from one hand specification to another. But the main types (straight line, circle, arc) have been isolated and characterized. Movements may be then considered as series of goals (Lebourque and Gibet, 1997), and synchronization between the two articulators be handled by event-driven processes with Grafcet or Petri-net representations. Besides, we have recently developped a specific application intended for the analysis of sign duration and dynamics from video sequences. This should enable us to specify tense of signs better, just as pauses between signs at the sentence level.

6. CONCLUSION

One of the guidelines of the system described here was to be able to generate any kind of sign, as far as possible. The proposed formal description is generic enough to take up such a challenge. Not only does it provide primitives and more complex structures found in every sign, but it also tries to take the widest variety of sign language grammatical processes into consideration. Such an open system lets the user entirely free to adjust the encoding sharpness of signs.

Movement must still be specified precisely in terms of path dimension and dynamics, together with hands arrangement and synchronization. But the virtual signer already shows interesting results for simple signs, especially a high-level specification of hand configuration, location and orientation. In collaboration with native deaf signers, we now intend to test the generated synthetic signs in order to optimize their expressive potential.

7. REFERENCES

N I Badler, K H Manoochehri and G Walters (1987), Articulated figure positionning by multiple constraints, *IEEE Computer Graphics and Application*, **7**, *6*, pp. 28-38.

E Klima and U Bellugi (1979), *The signs of language*, Harvard University Press, Cambridge, London.

A A Kobrinski and A E Kobrinski (1989), *Bras manipulateurs des robots - Architecture et théorie*, Mir editions, Moscow [in French].

T Kurokawa (1992), Gesture coding and a gesture dictionary for a nonverbal interface, *IEICE Trans. Fundamentals*, **E75-A**, *2*, pp. 112-121.

T Lebourque and S Gibet (1997), Synthesis of hand-arm gestures, In *Progress in gestural interaction - Proc. Gesture Workshop '96* (P A Harling & A D N Edwards, Eds), Springer-Verlag, Berlin, pp. 217-225

J Lee, T L Kunii (1993), Constraint-based hand animation, In *Models and Techniques in Computer Animation World* (N Magnenat-Thalmann & D Thalmann, Eds.), Springer-Verlag, Berlin, pp. 110-127.

J Lee (1994), Notational representation of sign language: a structural description of hand configurations, In C*omputers for handicapped persons – Proc. 4th Int. Conf., ICCHP '94 Vienna, Austria* (W L Zagler, G Busby, R R Wagner, Eds.), Springer-Verlag, Berlin, pp. 38-45.

J Lenarcic and A Umek (1994), Simple model of human arm reachable workspace, *IEEE Trans. on Systems, Man and Cybernetics*, **24**, *8*, pp. 1239-1246.

S K Liddell and R E Johnson (1989), American Sign Language: the phonological base, *Sign Language Studies*, **64**, pp. 195-277.

J Loomis, H Poizner and U Bellugi, Computer graphics modeling of American Sign Language, *Computer Graphics*, **17**, *3*, pp. 105-114.

B Moody (1986), *La langue des signes – Tome 2 : dictionnaire bilingue élémentaire*, International Visual Theatre (I.V.T.), Editions Ellipses, Paris [in French].

L Namir and I M Schlesinger (1978), The grammar of Sign Language, In *Sign Language of the Deaf : Psychological, Linguistic and Sociological Perspectives* (I M Schlesinger & L NAMIR, Eds), Academic Press, New York, pp. 97-140.

F-X Nève (1997), *Essai de grammaire de la Langue des Signes Française*, Bibliothèque de la Faculté de Philosophie et Lettres de l'Université de Liège, Liège (Belgium) [in French].

S Prillwitz and H Zienert (1990), Hamburg Notation System for Sign Language: Development of a sign writing with computer application, In *Current Trends in European Sign Language Research - Proc. 3rd European Congress on Sign Language Research* (S Prillwitz & T Vollhaber, Eds), Signum, Hamburg, pp. 355-380.

M Shantz and H Poizner (1982), A computer program to synthesize American Sign Language, *Behavior Research Methods and Instrumentation*, **14**, *5*, pp. 467-474.

Segmentation and classification of hand gestures for man-machine communication

F García-Ugalde[1], D Gatica-Pérez[2] and V García-Garduño[3]

[1,3]Depto. de Ing. Eléctrica-DEP, FI-UNAM, Apdo. Postal 70-256
04510 México, D.F., MEXICO

[2]Human Interface Technology Laboratory, University of Washington,
Box 352142, Seattle, WA 98195, USA

garcia@verona.fi-p.unam.mx

ABSTRACT

Man-machine communication in a natural way, it means without cumbersome gloves, is still an open problem. Keeping in mind the need to develop some friendly tools for helping people with disabilities to use the computer as a support tool for training into reinforced methods for learning to read or any other application. In this work we have addressed the problem of communication with a computer using some recognition of very basic hand gestures. From an engineering point of view our system is based on a video camera which captures image sequences and in a first time a segmentation of hand gestures is developed in order to provide information for its posterior classification and recognition. For classifying the segmented fields named **e** of gestures, for instance *hand # 1* and *hand # 2,* see figures 5a and 6a, we have proceed first to obtain a binary version of these segmented fields comparing them with a threshold, so rendering the classification faster, then based on the Radon transform (Lim, 1990), a computation of the projected sum of the binary intensity of gestures has been done at directions 0° and 90°, see figures 1 and 2. For reducing the number of data to be processed a wavelet decomposition of the projected sum of the binary intensity for each orientation (0° and 90°) has been done using Daubechies filters: d4 (Daubechies, 1988). This projected and wavelet decomposed information has been used for classifying the gestures: training our system with our dictionary and computing the correlation coefficient between the wavelet coefficients corresponding to *trained sequences* and others captured and computed in continuous operation, the computer is able to recognize the very simple gestures. The region segmentation has been done using a dense motion vector field as the main information then each region is matched to a four-parameter motion model (Gatica-Pérez et al, 1997). Based on Markov Random Fields the segmentation model detects moving parts of the human body with different apparent displacement such as the hands (García-Ugalde et al, 1997). The motion vector field has been estimated by a Baaziz pel-recursive method (Baaziz, 1991) and considered together with others sources of information such as intensity contours, intensity values and non-compensated pixels as inputs of the Markov Random Field model. The maximum a posteriori criterion (MAP) is used for the optimization of the solution, and performed with a deterministic method: *iterated conditional modes* (ICM). The complete segmentation algorithm includes initializing, region numbering and labeling, parameter estimation of the motion model in each region, and optimization of the segmentation field. So our probabilistic approach takes into account the fact that an exact displacement field does not exist (errors usually occur at or around motion boundaries), and that better results can be attained if an indicator of the quality of the vector field is known, this indicator is obtained from the non-compensated pixels as well as the intensity contours (García-Ugalde et al, 1997).

1. INTRODUCTION

The problem of developing friendly interfaces for man-machine communication is still an open problem. However these interfaces become more and more necessary for helping people with some kind of disabilities to communicate with the computer. For instance, there exist some methods used for permitting handicapped people to learn to read: these methods consist essentially in defining one hand gesture for each of the 26 letters of the alphabet and use these hand gestures at the same time as a therapist shows the letter sign and makes the assocaited sound, this permits to deliver more information and then as a result people learning to read could learn more quickly because they assocaite to each letter three parameters: the sign of the letter, the sound and the hand gesture.

Our system is thinking as a support for the therapist, it has to recognize the hand gesture associated with each letter and then by synthetic speech generate the associated sound. So in these conditions people with disabilities could training without the constant presence of the therapist, they just need a computer with a video camera in front of them and when they are practicing to read, they are doing the hand gestures associated with letters and the computer is generating the sounds, making a close loop.

From an engineering point of view, for this system works the computer needs to recognize the hand gesture: segmenting and classifing it. The problem of segmenting moving regions was traditionally studied for coding video, accurate motion vector field estimation is crucial in this application as well as the segmentation and can be seen as interdependent problems, because one is needed to obtain the other with accuracy (estimation-segmentation ambiguity). Thus, motion-based segmentation is crucial for extracting high level information from the time-varying intensity of a sequence and for improving the motion measurement process (García-Garduño, 1995), (François, 1991), it represents a qualitative change from a local motion description to a regional one. In this paper, we present an algorithm to segment image sequences that begins with a Baaziz (1991) pel-recursive estimation of a motion vector field. Later, we model the image sequence using Markov Random Fields and pursue the optimization of the segmentation problem by a Bayesian estimation criterion (MAP) performed with a deterministic method: *iterated conditional modes* (ICM). Our probabilistic approach takes into account the fact that an exact displacement field does not exist (errors usually occur at or around motion boundaries), and that better results can be attained if an indicator of the quality of the vector field is known, this indicator is obtained from the non-compensated pixels as well as the contours. The classification of the segmented areas has been developed by computing for the *trained sequences* and others, a correlation coefficient of wavelet coefficients, of the projected sum of the intensity (at orientations 0^0 and 90^0) of the segmentation field (in its binary version).

2. GESTURE CLASSIFICATION

For classifying the segmented fields **e** of gestures, for instance *hand # 1* and *hand # 2* we have proceed first to obtain a binary version of these fields by comparing with a threshold, so rendering the classification faster. Then based on the Radon transform (Lim, 1990)

$$p(t) = \int_{u=-\infty}^{\infty} \mathbf{e}\left(t_1, t_2\right) \Big|_{t_1 = t\cos\theta - u\sin\theta,\ t_2 = t\sin\theta + u\cos\theta} du \tag{1}$$

a computation of the projected sum of the binary intensity has been done at orientations: 0^0 and 90^0 see figures 1 and 2 for *hand # 1* and *hand # 2* respectively. Then for reducing the number of data to be processed a wavelet decomposition of the projected sum of the intensity for each orientation (0^0 and 90^0) has been done using Daubechies filters: d4 (Daubechies, 1988). This projected and wavelet decomposed information has been used for classifying the gestures, training our system with our dictionary and computing the correlation coefficient between the wavelet coefficients corresponding to *trained sequences* and others obtained in continuous operation of the system.

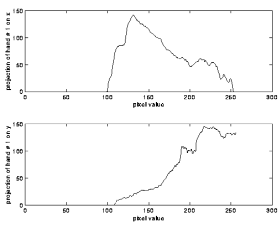


Figure 1. *Projected sum of the intensity of the binary version of the segmented field* **e** *of hand # 1 at* 0^0 *and* 90^0 *respectively.*

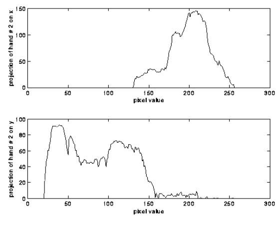


Figure 2. *Projected sum of the intensity of the binary version of the segmentation field* **e** *of hand # 2 at* 0^0 *and* 90^0 *respectively.*

3. THE SEGMENTATION ALGORITHM

From the point of view of image processing in presence of pure divergent motion, simple spatial clustering techniques for motion-based segmentation do not work well (García-Garduño, 1995). In this case, a model Θ_M of both the motion and the structure of the regions in the scene has to be introduced. Thus, the goal of the segmentation process is to assign each pixel in the image to one out of several regions, depending on the accuracy between each estimated motion vector and the assumed model. Each region is characterized by a motion parameter vector. The obtained regions can then be associated to different regions of the same object, or to different objects in the scene. The proposed segmentation algorithm is based on Markov Random Fields and estimation theory, using the *maximum a posteriori* criterion as optimality principle. Such a combined approach provides a common framework in which we can introduce information sources of distinct nature, model their interactions, and incorporate expected properties on the solution. Markov Random Fields modeling is appropriate for motion segmentation: we have a dense displacement vector field as the main information for separating an image into regions, but as we have discussed, motion information is not always correct, especially at movement discontinuities; in this case we may also include other data sources: intensity gray values, non-compensated pixels and intensity contours, as additional observations to improve the final solution. Furthermore, we add physical properties to the model: (a) a motion model Θ_M for each region in the scene to be segmented, (b) spatial continuity for the segmentation, (c) presence of motion boundaries only when strong intensity changes occur, and (d) expected geometrical shapes for the region boundaries. According to MRF theory, we will represent each information source as an *observation field* and each expected result as a *label field.* In this case, observations are:

- the estimated horizontal and vertical components of the motion field $\mathbf{d}_x$ and $\mathbf{d}_y$
- the binary non-compensated pixel field $\mathbf{p}$. As we mentioned earlier, it can be considered as a simplified way of removing motion outliers, for it represents a way of switching between displacement and more reliable information (intensity values) when the motion field is not accurately estimated
- the image intensity gray values field $\mathbf{i}$
- the binary intensity contour field $\mathbf{g}$ that favors the coincidence of motion boundaries and strong spatial gradients: 0 means no contour; 1 means contour.

On the other hand, desired *label* fields are:

- the desired segmentation label field $\mathbf{e}$ which has associated a four-parameter simplified linear motion model $\Theta_{MLS} = (t_x, t_y, k, \theta)$, that can describe combined translational, rotational, and divergent motions of planar surfaces parallel to the image plane (García-Garduño, 1995)

$$\begin{bmatrix} \mathrm{d_x} \\ \mathrm{d_y} \end{bmatrix} = \begin{bmatrix} \mathrm{t_x} \\ \mathrm{t_y} \end{bmatrix} + \begin{bmatrix} \mathrm{k} & -\theta \\ \theta & \mathrm{k} \end{bmatrix} \begin{bmatrix} \mathrm{x - x_g} \\ \mathrm{y - y_g} \end{bmatrix} \qquad (2)$$

where (x_g, y_g) is the center of gravity of each surface.

- To improve the segmentation process we introduce an auxiliary binary motion discontinuity line field $\mathbf{l}$ along with the segmentation label field: motion boundaries (0 means no motion discontinuity; 1 means motion discontinuity).

In figure 3 we show the interaction model of observations, labels and physical assumptions.

Assuming *N* pixels in the image we formulate the motion-based segmentation as an estimation problem: simultaneously find the *label* fields $(\hat{\mathbf{e}}, \hat{\mathbf{l}})$ that maximize the *a posteriori* probability density function (**pdf**) of the labels, given the *observe**d*** data :

$$(\hat{\mathbf{e}}, \hat{\mathbf{l}}) = \arg max_{\mathbf{e},\mathbf{l}}\, p\left(\mathbf{e},\mathbf{l} \middle| \mathbf{d}_x, \mathbf{d}_y, \mathbf{i}, \mathbf{p}, \mathbf{g}\right) \tag{3}$$

Reversing the problem using the Bayes rule, the last equation can be expressed as

$$(\hat{\mathbf{e}}, \hat{\mathbf{l}}) = \arg max_{\mathbf{e},\mathbf{l}}\, p\left(\mathbf{d}_x, \mathbf{d}_y, \mathbf{i} \middle| \mathbf{e}, \mathbf{l}, \mathbf{p}, \mathbf{g}\right) p\left(\mathbf{e},\mathbf{l} \middle| \mathbf{p}, \mathbf{g}\right) \tag{4}$$

In (Gatica-Pérez et al, 1997) we have shown that maximizing the *a posteriori* **pdf** is equivalent to minimize a so-called *energy function* $U(e, l, d_x, d_y, i, p, g)$ which has the form

$$U(\mathbf{e},\mathbf{l},\mathbf{d}_x,\mathbf{d}_y,\mathbf{i},\mathbf{p},\mathbf{g}) = \alpha\ U_{\mathbf{d}}(\mathbf{d}_x,\mathbf{d}_y,\mathbf{e},\mathbf{p}) + \beta\ U_{\mathbf{i}}(\mathbf{i},\mathbf{e},\mathbf{p}) + \gamma\ U_{\mathbf{e}}(\mathbf{e},\mathbf{l}) + \kappa\, U_{\mathbf{l}}(\mathbf{l},\mathbf{g}) \tag{5}$$

where α, β, γ and κ are weighting terms, all these energy terms has been also defined in (Gatica-Pérez et al, 1997).

3.1 Global optimization using iterated conditional modes method

To overcome the great computational cost required by simulated annealing, the global optimization of the solution is reached by using an iterative deterministic relaxation procedure: a modified *iterated conditional modes* (ICM) method based on an instability table (Françoise, 1991). ICM methods minimize the local energy ΔU_s in each pixel $s(x, y)$ of the image. Our minimization scheme considers two phases in each iteration (García-Garduño, 1995): one for the optimization of the segmentation field through minimizing

$$U_1 = \alpha\ U_{\mathbf{d}}(\mathbf{d}_x,\mathbf{d}_y,\mathbf{e},\mathbf{p}) + \beta\ U_{\mathbf{i}}(\mathbf{i},\mathbf{e},\mathbf{p}) + \gamma\ U_{\mathbf{e}}(\mathbf{e},\mathbf{l}) \tag{6}$$

and the other for the optimization of the motion discontinuity line field, minimizing

$$U_2 = \gamma\ U_{\mathbf{e}}(\mathbf{e},\mathbf{l}) + \kappa\ U_{\mathbf{l}}(\mathbf{l},\mathbf{g}) \tag{7}$$

The term $U_{\mathbf{e}}(\mathbf{e},\mathbf{l})$ represents a link term between the two stages of the optimization general process.

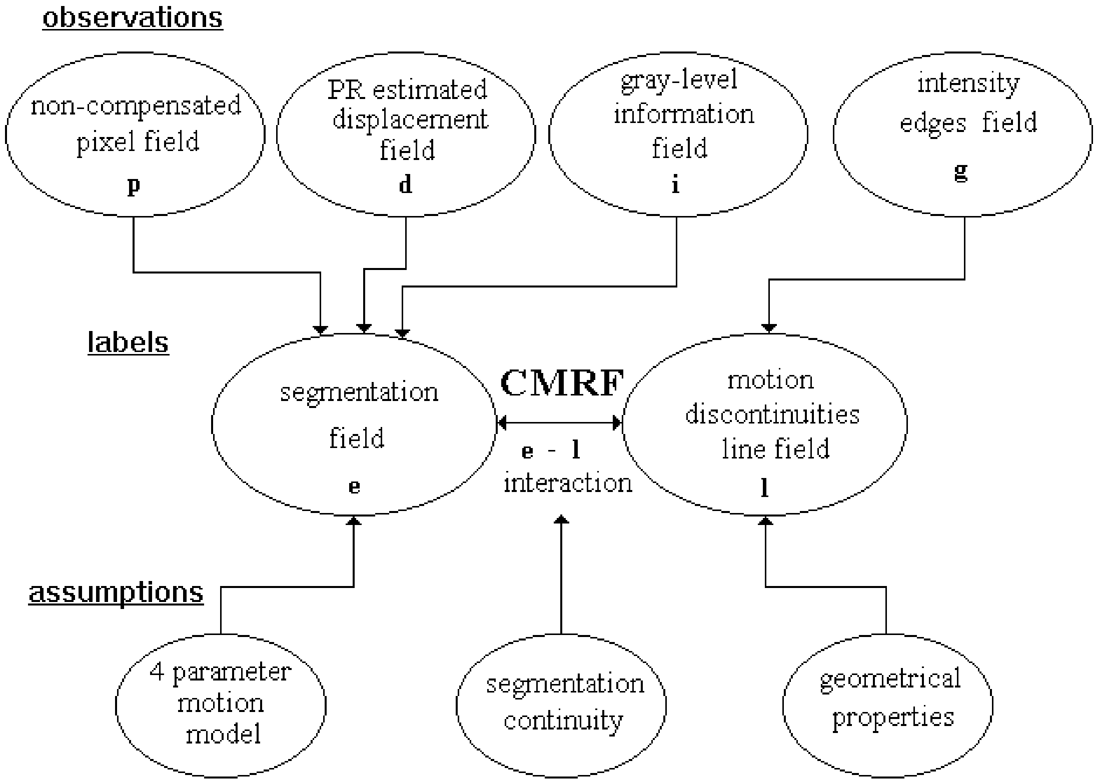


Figure 3. *Interaction model for the motion-based segmentation algorithm.*

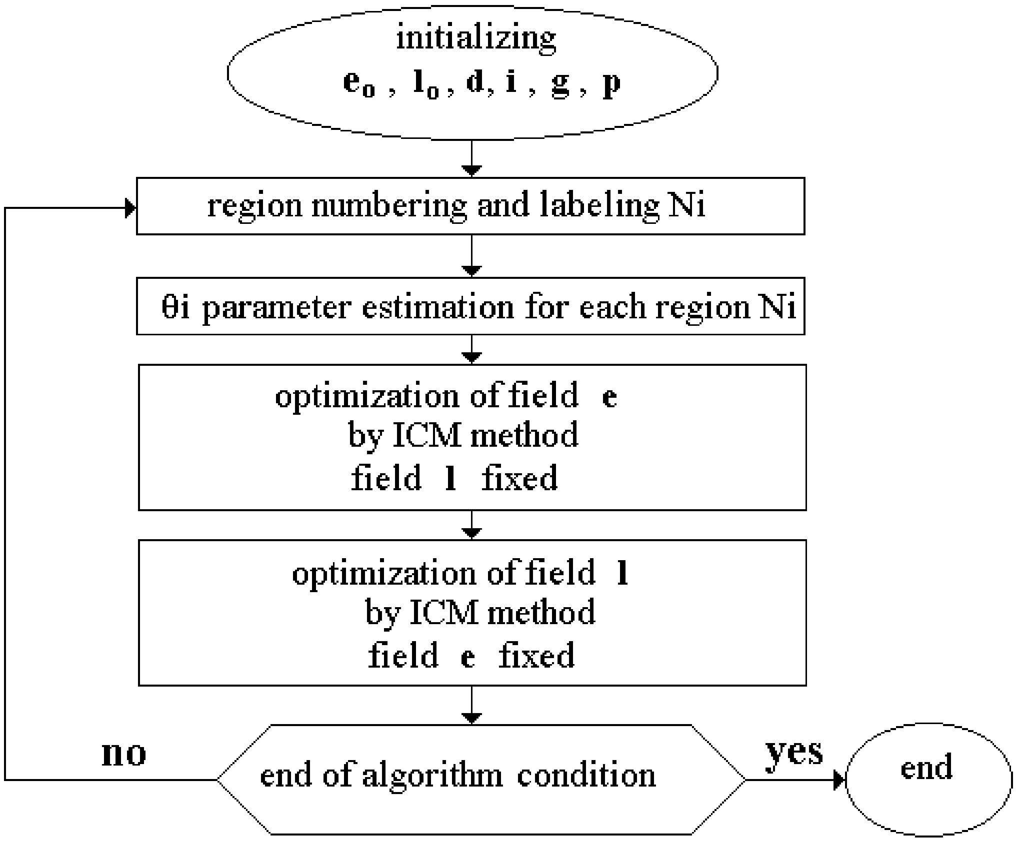


Figure 4. *General diagram of the proposed motion-based segmentation algorithm.*

3.2 The complete motion-based segmentation algorithm

The complete motion-based segmentation algorithm includes four stages (a) initializing, (b) numbering and labeling of each region in the image, (c) motion model parameter estimation in each region, and (d) optimization of the label fields. These steps are repeated until the method reaches the maximum number of iterations allowed, or until the segmentation becomes stable. An advantage of our algorithm is that the number of regions in the image is not fixed through the segmentation process, see figure 4.

4. SEGMENTATION RESULTS

Results obtained on the test sequences *hand # 1* and *hand # 2* for segmenting the moving parts are presented in figures 5 and 6 respectively. The segmentation fields obtained using the proposed algorithm are shown in figures 5a and 6a. A superposition of *hand # 1* and *hand # 2* with their respective segmentation regions can be seen in figures 5b and 6b. From the results it can be observed that the *hands* in motion have been well segmented from the rest of the scene. This result is qualitatively correct and reached only after 2 iterations of the segmentation algorithm (for each case: *hand # 1* and *hand # 2*), the tiny regions remaining in the background could be fused in posterior iterations.

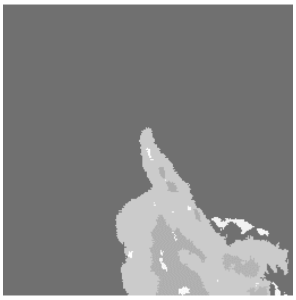


Figure 5.a. *MAP motion-based segmentation algorithm. Segmentation field* **e** *of hand # 1.*

Figure 5.b. *Superposition of frame 3 of hand # 1 and segmentation field* **e.**

Figure 6.a. *MAP motion-based segmentation algorithm. Segmentation field* **e** *of hand # 2.*

Figure 6.b. *Superposition of frame 7 of hand # 2 and segmentation field* **e**.

No further processing has been done on the segmentation frontiers. The motion vector fields obtained with the Baaziz method, figures 8a and 10a, are somewhat homogeneous and properly adjusted to the moving areas. The non-compensated pixels shown in figures 8b and 10b, represent respectively only 4.65 % and 1.99 %. A very important input to the segmentation algorithm is the initial binary segmentation regions, this initialization was obtained by thresholding the difference between, respectively frames 1 and 2 for *hand # 1* and frames 6 and 7 for *hand # 2*, and passing the resulting difference through a median filter of window 3x3. Keeping in mind the use of segmented regions for helping people with some kind of disability to communicate with a computer, when the segmentation of the *hands* has been completed, we have defined a very simple dictionary of gestures which are used to reinforce methods designed for learning to read.

5. PEL-RECURSIVE MOTION ESTIMATION

The mainly use of pel-recursive displacement estimation since it was proposed by Netravali and Robbins (1979), has been on predictive motion-compensated image sequence coding. A dense motion field based on the spatio-temporal varying intensity of a sequence is produced, such a field can be considered as a low level information source. For its computing, in this work we have selected the Baaziz method which consist on two main steps: in the first one it uses the method proposed by Biemond et al. (1987) and in a second step for reducing the number of non-compensated pixels the Walker and Rao method is used (on non-compensated sites only). We have applied it towards a higher level representation of an image sequence: a dense field will constitute the main clue to guide the pixel fusion process into regions of similar motion. Displacement is computed along the scan direction according to a prediction-updating scheme until convergence is obtained. One proper criteria for convergence is the recursive minimization of the reconstruction error; this minimization can also be iterative. Thus, in the Wiener-based algorithm, the displacement is estimated for each pixel until the DFD has been minimized using the equation

$$\hat{\mathbf{d}}^{\mathbf{i}} = \hat{\mathbf{d}}^{\mathbf{i-1}} - \begin{pmatrix} \sum_{j=1}^{N_n} \left(i_x^j\right)^2 + \mu & \sum_{j=1}^{N_n} i_x^j i_y^j \\ \sum_{j=1}^{N_n} i_x^j i_y^j & \sum_{j=1}^{N_n} \left(i_y^j\right)^2 + \mu \end{pmatrix}^{-1} \cdot \begin{pmatrix} \sum_{j=1}^{N_n} i_x^j \cdot \mathrm{DFD}\left(\mathbf{z}_j\,,\ \hat{\mathbf{d}}^{\mathbf{i-1}}\right) \\ \sum_{j=1}^{N_n} i_y^j \cdot \mathrm{DFD}\left(\mathbf{z}_j\,,\ \hat{\mathbf{d}}^{\mathbf{i-1}}\right) \end{pmatrix} \tag{8}$$

where

- $i(x,y,t)$ is the intensity of each pixel of the sequence
- $\mathbf{z} = (x,y)$ is the position of each pixel
- $\mathbf{d}(x,y,t) = (d_x(x,y,t), d_y(x,y,t))$ is the displacement vector of each pixel in the interval $(t - k\Delta t, t)$
- $\hat{\mathbf{d}}^{\mathbf{i-1}}$ is the initial displacement estimation (prediction) for each pixel. If estimation is iterative, it represents the displacement after $\mathbf{i}-1$ iterations
- $\hat{\mathbf{d}}^{\mathbf{i}}$ is the final estimation for each pixel, (or after $\mathbf{i}$ iterations)

DFD is the displaced-frame-difference (reconstruction error)

$$DFD(\mathbf{z},\mathbf{d}) = i(z,t) - i(z-d, t-k\Delta t) \tag{9}$$

- N_n represents the number of pixels in a small casual spatial neighborhood of each pixel
- i_x^j y i_y^j are the components of the intensity gradient vector ∇i on the displaced positions in frame $t - k\Delta t$

$$\begin{aligned} i_x^j &= i_x(z_j - \hat{\mathbf{d}}^{\mathbf{i-1}}, t - k\Delta t) \\ i_y^j &= i_y(z_j - \hat{\mathbf{d}}^{\mathbf{i-1}}, t - k\Delta t) \end{aligned} \tag{10}$$

$\mu = \sigma_v^2 / \sigma_u^2$ is the ratio between linearizing error variance and actualization term variance respectively.

Figure 7.a. *Sequence hand # 1. Frame 1.*

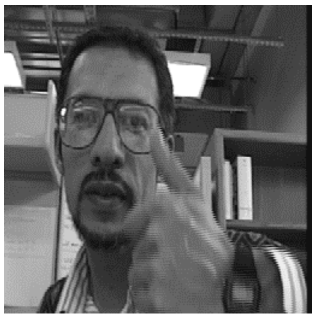


Figure 7.b. *Sequence hand # 1. Frame 4.*

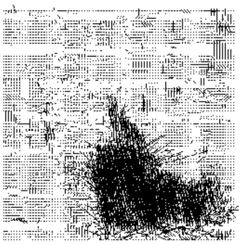


Figure 8.a. *Motion field obtained with the Baaziz method using hand # 1 , frames 1 and 2.*

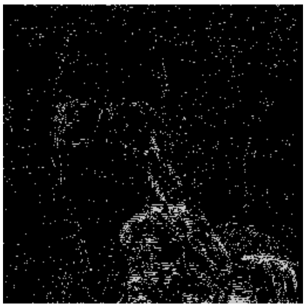


Figure 8.b. *Non-compensated pixel binary image using hand # 1 (0=compensated, 1=non-compensated)*

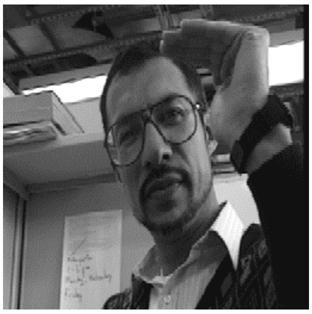


Figure 9.a. *Sequence hand # 2. Frame 1.*

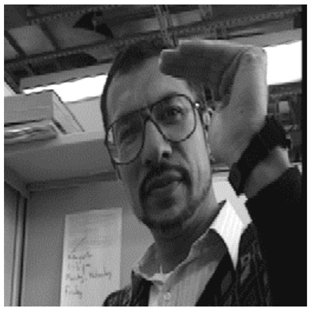


Figure 9.b. *Sequence hand # 2. Frame 9.*

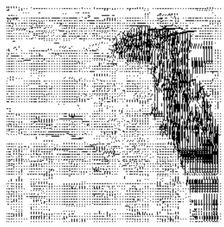


Figure 10.a. *Motion field obtained with the Baaziz method using hand # 2, frames 6 and 7.*

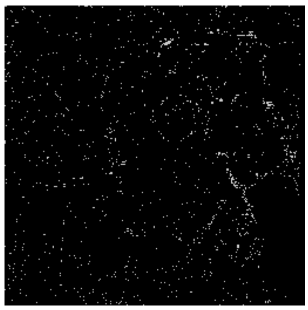


Figure 10.b. *Non-compensted pixel binary image using hand # 2 (0=compensated, 1=non-compensated).*

Pel-recursive algorithms based on linear estimation includes local context information so its motion fields are more immune to noise and quantitatively more accurate, but simple enough to compute. On those pixels in which the convergence criterion is not satisfied we had applied the Walker and Rao method and even if after this second step the pixel remains non-compensated (NC), this information will be useful during the segmentation process as a simple partial confidence measure of the motion estimation quality.

Figures 7a and 7b show frames 1 and 4 of the test sequence *hand # 1.* In figures 8a and 8b we present respectively the motion field and the non-compensated pixel image, obtained using frames 1 and 2 and the Baaziz method, the hand is moving from bottom to top. Figures 9a and 9b show frames 1 and 9 of the test sequence *hand #* 2, in which the hand is moving from top to bottom, in figures 10a and 10b we present respectively the motion field and the non-compensated pixel image, obtained using frames 6 and 7 and the Baaziz method.

6. CONCLUSIONS

The results obtained for segmenting and classifing hand gestures as a way for helping people with disabilities to communicate eassily with a computer are very encouraging, for this very simple dictionary of gestures a 100% successful has been obtained.

Our system is very simple consisting only in a computer and a video camera attached to it. However our major drawback is time processing, we are not able to operate in real-time. Comparing our work with others developped previously (Starner and Pentland, 1995), (Quek et al,), we have a more general system able to recognize more diverse hand gestures but not working in real-time jet.

Acknowledgments: This work was supported in part by ``Universidad Nacional Autónoma de México" (UNAM) and ``Consejo Nacional de Ciencia y Tecnología" (CONACyT).

7. REFERENCES

[1] A.N. Netravali and J.D. Robbins (1979), ``Motion-compensated television coding: part I," *The Bell System Technical Journal*, Vol. 58, No. 3, pp. 631-670, March.

[2] D. Gatica-Pérez, F. García-Ugalde and V. García-Garduño (1997), ``Segmentation algorithm for image sequences from a pel-recursive motion field." *Proc. of the VCIP, SPIE*, Vol. 3024, pp. 1152-1163.

[3] E. François (1991), *Interpretation qualitative du mouvement a partir d'une séquence d'images*, Ph. D. Thesis, Université de Rennes I, France, June.

[4] F. García-Ugalde, J. Savage-Carmona, T. A. Furness III, D. Gatica-Pérez and V. García-Garduño (1997), "Segmentation of moving human body parts by a modified MAP-MRF algorithm," *Proc. of the VSMM, IEEE,* pp. 197-205.

[5] F. K. H. Quek, T. Mysliwiec and M. Zhao. "Finger mouse: A freehand pointing interface." *Vision Interfaces and* Systems *Lab. (VISLab), The University of Illinois at Chicago.*

[6] I. Daubechies (1988), ``Orthonormal bases of compactly supported wavelets," *Comm. Pure and Applied Mathematics*, Vol. 41, pp. 909-996.

[7] J. Biemond, L. Looijenga, D.E. Boekee and R. Plompen (1987), ``A pel-recursive Wiener-based displacement estimation algorithm." *Signal Processing*, Vol. 13, No. 4, pp. 399-412, December.

[8] J.S. Lim (1990), *Two dimensional signal and image processing*, Englewood Cliffs, NJ: Prentice Hall.

[9] N. Baaziz (1991), *Approches d'estimation et de compensation de mouvement multiresolutions pour le codage de séquences d'images*, Ph. D. Thesis, Université de Rennes I, France, October.

[10] T. Starner and A. Pentland (1995), "Real-time american sign language recognition from video using hidden Markov models." *MIT Media Lab. Perceptual Computing Section Technical Report* No. 375.

[11] V. García-Garduño (1995), *Une approche de compression orientée-objets par suivi de segmentation basée mouvement pour le codage de séquences d'images numériques*, Ph. D. Thesis, Université de Rennes I, France, May.

Japanese sign-language recognition based on gesture primitives using acceleration sensors and datagloves

Hideyuki Sawada, Takuto Notsu and Shuji Hashimoto

Department of Applied Physics, School of Science and Engineering, WASEDA University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN

{sawa, taku, shuji}@shalab.phys.waseda.ac.jp

ABSTRACT

This paper proposes a Japanese sign-language recognition system using acceleration sensors, position sensors and datagloves, to understand human dynamic motions and finger geometry. The sensor integration method realized a robust gesture recognition comparing with a single sensor method. The sign-language recognition is done by referring to a Japanese sign-language database in which words are written as sequences of the gesture primitives. Two recognition algorithms which are the automata algorithm and the HMM are introduced and tested in the practical experiments.

1. INTRODUCTION

Gesture plays an important role not only as nonverbal media for emotional human communication [1][2], but also as a necessary method of communication among people who are hearing impaired as sign language. Sign language has been formed as a language for the verbal communication. Hearing impaired people sometimes need to have communications with able bodied people with the help of a sign-language interpreter. Since the number of interpreters is limited, they cannot always have enough communication when they demand. A computerized interpreter which understands the sign-language automatically is greatly needed as a communication tool in the welfare field. If a computerized system is able to understand the human gesticulation, it will generate a new solution not only as a machine interpreter which translates nonverbal expressions into verbal languages, but also as a flexible human-machine interface which reacts to nonverbal expressions and emotional feelings.

Many attempts to the realization of sign-language recognition have been reported so far [3]-[5]. Most of the approaches introduced in the existing researches can be classified into two categories: one employs wearable devices such as a dataglove and a position sensor, and the other uses an image processing technique. The former approach is suitable for the realtime recognition, since the finger geometry and the hand posture are directly obtained as three dimensional physical data. Constraints by wearing the device and wiring for the data transmission are, however, the problem. On the other hand, the latter approaches are actively studied these days, because the image processing technique lets the users be free from the constraints of the wearable equipment. Great deal of computational processing and the estimation of occluding objects, however, have to be solved by the computing algorithms. The common problem which remains in both approaches is how to extract the starting point of a gesture. Furthermore the dynamic characteristics of gestural motions are hard to be recognized in realtime because the physical data obtained in the both techniques has the positional dimension.

The arm motion and the finger geometry are considered to be the primitive information in understanding gesticulations. Motion can be measured directly as acceleration caused by the applying forces to the body[2][6]. In this study, acceleration sensors and a pair of datagloves are employed for the gesture acquisition. This paper presents a gesture recognition algorithm based on acceleration patterns caused by dynamic movements. The gesture patterns are extended to the classification as gesture primitives for the construction of Japanese sign-language (JSL) recognition system. Two algorithms which are the automata algorithm and the HMM (Hidden Markov Model) are introduced in the sign-language recognition. The HMM has better performance rather than the automata algorithm in the current study. JSL database is also constructed in which JSL words are described as the sequence of gesture primitives.

2. CHARACTERISTIC PARAMETERS FOR GESTURE PRIMITIVE RECOGNITION

2.1 Characteristic parameters Extracted by Acceleration sensor

An acceleration sensor (Nihon Kohden TA-513G 3D acceleration sensor) used in this study is small enough to be attached to any points of the human body. The size is 20x15x12[mm], and the weight is 12.5[g]. It can sense three dimensional accelerations by the piezo-electric devices which cause the change of voltage according to the amount of acceleration applied to the sensor. The sensitivity is about 5mV per 1G in the range between -25G and +25G. The acceleration data are amplified and fed to the computer through an A/D converter as 12 bit binary data.

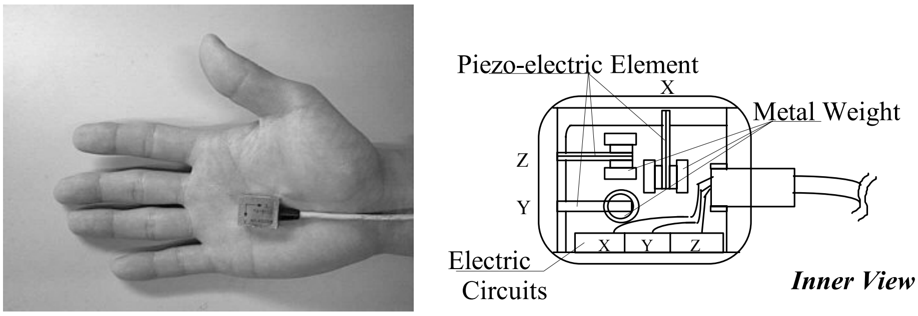


Figure 1. *3D acceleration sensor : its outer view(left) and schematic inner view(right)*

The three acceleration data $a_X(t)$, $a_Y(t)$ and $a_Z(t)$ are independently obtained in realtime, which correspond to the accelerations in x, y and z directions at time t, respectively. There is a large variation of acceleration patterns in the same kind of dynamic motion, as in the case of automatic recognition of handwritten characters. Furthermore, the sensor employed in this study does not have a good quantitative reproducibility. Consequently, a high recognition rate cannot be obtained by applying pattern matching method of acceleration vector series. The global features of the gesture motion must be extracted.

In this study, the motion features are extracted from the following three two-dimensional vectors in order to obtain intuitive characteristics of the motion:

$$\mathbf{A}_1(t)=(a_y(t),\ a_z(t)),\ \mathbf{A}_2(t)=(a_z(t),\ a_x(t)),\ \mathbf{A}_3(t)=(a_x(t),\ a_y(t)) \qquad (1)$$

$\mathbf{A}_1(t)$, $\mathbf{A}_2(t)$, $\mathbf{A}_3(t)$ are the projection vectors of the acceleration on the *y-z*, *z-x* and *x-y* planes, respectively. In the study, the data acquisition frequency f_{in} was set to 30 Hz. Although this is not sufficient as a sampling frequency in the measurement of motion which includes rapid changes of the direction or the velocity, we are focusing on the realtime recognition with the minimum acceleration data. A succession of fifteen to thirty data set which accords with the duration of one shot gesture is used for gesture recognition. Eleven characteristic parameters shown in Table 1 are extracted from one sequence of each set of projection vectors on the *y-z*, *z-x* and *x-y* planes, for the realtime gesture recognition[2][6].

Table 1. *Characteristic Parameters of Acceleration Sensor.*

Pd :	Change of Force Given as Time Differences of Vectors.
Pg :	Rotating Direction Given as Vector Products $\mathbf{A}(t)$ * $\mathbf{A}(t+1)$.
Pr :	Directional Characteristics Given by Maximum Vector.
Pa0 - 7 :	Characteristics of Directional Distribution of Vectors.

2.2 Characteristic parameters Obtained by Dataglove

The static shapes of hands also contribute to the meanings of the gestures. Therefore a pair of datagloves (Nissho Electrnics SuperGlove) are employed in this system, which allow the realtime analysis of three dimensional hand position and finger geometry. By distributing ten bend sensors along the five finger parts of the glove, the device outputs roughly proportional values to finger joint angles. Furthermore, a magnetic position sensor (Polhemus Sensor) fixed on the back side of the wrist area gives three dimensional positions x, y and z and orientations which are azimuth (θ_{az}), elevation (θ_{el}) and roll (θ_{rl}) attitudes. Fourteen characteristic parameters R_0 - R_{13} as shown in Table 2 are defined. R_3 - R_8 represent the sine and cosine of the azimuth, elevation and roll data. Cosine and sine functions eliminate the disconnection of the degree at $+\pi$ and $-\pi$, and work for the selection of the hand posture at a degree of $(n\pi)$ [n = -1, 0, 1] and $(n\pi)/2$ [n = -1,

+1], respectively. At first before the use of the system, a user inputs the limits of hand positions (up, down, right and left) and the finger curvature (grasping and straight), so that the physical factors of sensor data influenced by a user's physique are normalized. The characteristic parameters are obtained from the normalized data.

Table 2. *Characteristic Parameters of Dataglove.*

R_0 - R_2 :	3D Position of hand.
R_3, R_4 :	$sin(\theta az)$, $cos(\theta az)$.
R_5, R_6 :	$sin(\theta el)$, $cos(\theta el)$.
R_7, R_8 :	$sin(\theta rl)$, $cos(\theta rl)$.
R_9 - R_{13} :	Joint Angles of 5 Fingers.

2.3 Gesture Recognition Algorithm Based on Characteristic Parameters

The gesture recognition uses *P*'s and *R*'s. The realtime recognition is made by comparison with standard data acquired in the learning phase to make the system suited for the individual users. In order to recognize a gesture from the acceleration time series, the starting point of the gesture has to be detected first. Even if the gestures of the same kind are measured by the acceleration sensor, the force pattern may be observed differently depending on the individuals. In this study, the start of the gesture is identified from the magnitude of the acceleration, since the beginning of the motion is accompanied by the local peak of acceleration.

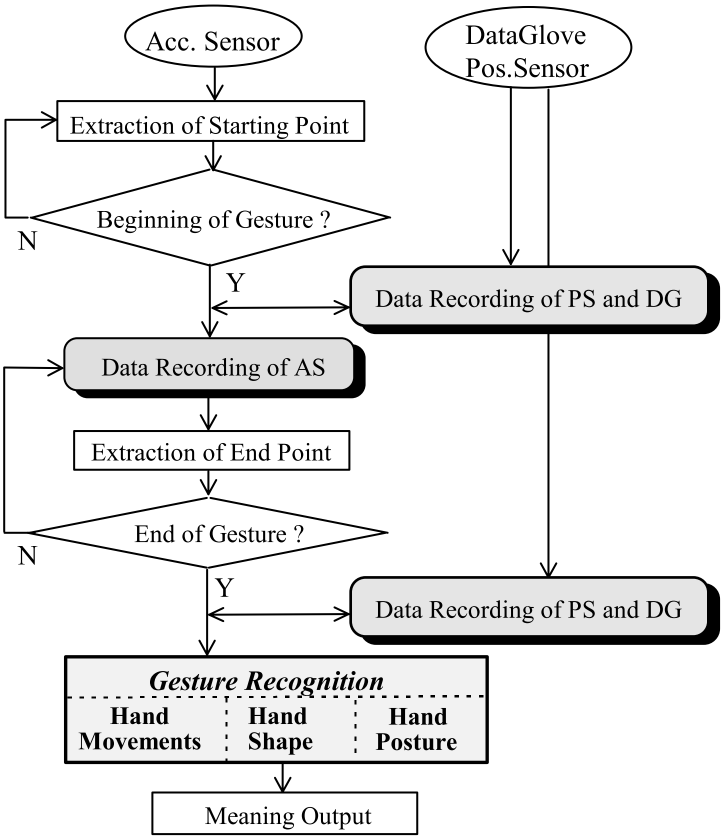


Figure 2. *Diagram of data acquisition and recognition*

The gesture recognition is executed as follows. By observing acceleration data, the recognition algorithm starts its action when acceleration value exceeds a pre-determined threshold. First, data of the datagloves and the position sensors are recorded in the computer memory, which is followed by the acceleration data recording. When the end of the gesture is detected by observing the acceleration data being recorded, data of the datagloves and the position sensors are recorded once again. Next whole recorded data are transformed into characteristic parameters which is handed to the recognition algorithm, then the meaning is determined by referring to the standard pattern data. The flow diagram of the data acquisition and the recognition is described in Figure 2.

In the learning phase, a user inputs gestures to be recognized M times each. Then the average $E_\alpha{}^G$ and the standard deviations $\mu_\alpha{}^G$ of the characteristic parameters are calculated for each gesture G as shown below, and stored as the standard pattern data.

$$E_\alpha^G = \frac{1}{M}\sum_{i=1}^{M} V_\alpha^{g_i} \quad , \quad \mu_\alpha^G = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(V_\alpha^{g_i} - E_\alpha^G)^2} \qquad \begin{array}{l} V: \textit{Parameter Values of Learning Pattern} \\ g_i: \textit{i - th Sample of Gesture g} \\ \alpha: \textit{P's or R's} \end{array} \tag{2}$$

In the recognition phase, the characteristic parameters V_α' are extracted, and the normalized distance e_G is calculated for each standard pattern data as below,

$$e_G = \sum_\alpha \varepsilon_\alpha^G = \sum_\alpha \frac{(V'_\alpha - E_\alpha^G)^2}{(\mu_\alpha^G)^2} \qquad V': \textit{Parameter Values of Input Pattern} \tag{3}$$

Then the minimum e_G is selected as a candidate. In case it is smaller than a predetermined threshold value Th shown below, the result of the gesture recognition is confirmed.

$$Th = \min_{G \neq H} \{\sum_\alpha \frac{(E_\alpha^H - E_\alpha^G)^2}{(\mu_\alpha^G)^2}\} \qquad \begin{array}{l} \textit{for all } G,H \\ G, H: \textit{Gesture } G,H \end{array} \tag{4}$$

Table 3 shows the results of recognition experiments for the 10 kinds of gestures which are:

vertical swing, horizontal swing, diagonal swing, sharp single swing,
clockwise rotation, counterclockwise rotation, star shape motion, triangular shape motion,
heart-shaped motion, pointing of direction, pause.

Gesture recognition was repeated 10 times each for two persons who are the particular individual for whom the standard pattern was constructed, and another individual. 100 percent gesture recognition is achieved for the individual for whom the standard pattern was constructed. And relatively good results were also obtained from the gestures by another individual. Mis-recognition arises for the “diagonal swing”, “counterclockwise rotation” and “triangular shape”, where the direction of motion/rotation is almost the same, as well as for “pointing of direction” and “sharp single swing”, where only the direction of the motion are different. Thus, it is obviously better that the user prepares the standard patterns of the individual before operating the system and enters the recognition phase. But the recognition is also possible to some extent even if the parameters are used to recognize the gestures of other individuals.

In the gestures used in our daily life, single operations such as vertical swing, horizontal swing, diagonal swing, rotation, sharp single swing, and pointing of direction, as well as their combinations, are commonly used. In the proposed system, “star shape”, “triangular shape” and “heart-shaped motion” can be recognized from the three dimensional acceleration patterns, even though their real-time recognition by image processing and other techniques seems quite difficult.

Table 3. *Experimental results of gesture recognition.*

	Own Gestures: Recognition Rate %	Another Person’s Gestures: Recognition Rate % (Mis-recognized as)
1. Vertical swing	100	100
2. Horizontal swing	100	100
3. Diagonal swing	100	70(Triangle)
4. Single swing	100	100
5. Clockwise rotation	100	100
6. Counterclockwise	100	80(Triangle)
7. Star shape	100	100
8. Triangular shape	100	80(Counter-clockwise)
9. Heart shape	100	100
10.Pointing of direction	100	70(Single swing)

3. SIGN-LANGUAGE RECOGNITION ALGORITHM BASED ON GESTURE PRIMITIVES

3.1 Gesture Primitives and JSL Database

JSL currently has almost 3,000 words, most of which are represented by the combination of simple motions. It means JSL words seem to be determined by observing both arm motion and hand figures. Strictly speaking, features from both motion and figures should be extracted simultaneously, changes of hand figure during arm motion, however, can be negligible, by considering the visual cognition process of human. In this study, arm motion and hand figure before and after the motion are highly paid attention to. So, gesture primitive blocks as

- Hand figure **-** Motion **-** Hand figure **-**.

are regarded as morpheme, the minimum unit of a gesture, which is used in the first step of JSL recognition procedure. Simple motion patterns are selected as 11 motion primitives and 14 hand figure primitives as listed in Figure 3 and Table 4, respectively, together with the hand posture patterns (**U**p, **D**own, **F**ront, **B**ack, **O**utside and **I**nside direction). We are constructing a JSL database in which sign-language are described by the repetition of primitive blocks.

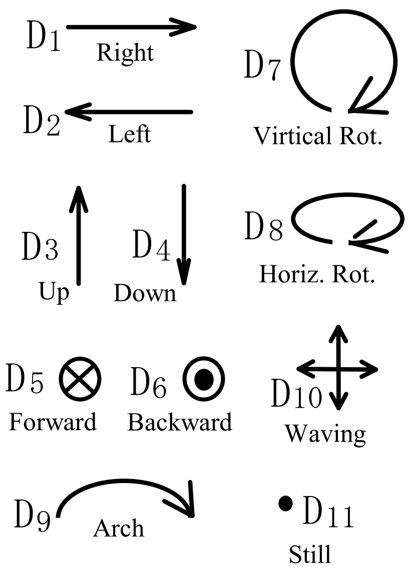


Figure 3. *Hand motion primitives*

Geometry Pattern	Finger States **s**traight/**b**end
F1	bbbbb
F2	sbbbb
F3	bsbbb
F4	bbsbb
F5	Bbbbs
F6	Ssbbb
F7	Bssbb
F8	Sbbbs
F9	Bsbbs
F10	Sssbb
F11	Bsssb
F12	Bbsss
F13	Bssss
F14	Sssss

Table 4. *Hand figure primitives*

Right	*Left*	*: Supplements*	*Right*	*Left*	*: Supplements*
Happy		: Meaning	You		: Meaning
S F3/B	F3/B	: Geometry/Posture	S F3/O	-	: Geometry/Posture
D3	D4	: Motion	D5	D11	: Motion
F3/B	F3/B	: Geometry/Posture	E F3/O	-	: Geometry/Posture
D4	D3	: Motion			
E F3/B	F3/B	: Geometry/Posture	Japanese ‘a’		: Finger Spelling
			F2/B	-	: Geometry/Posture
North		: Meaning			
S F7/O	F7/O	: Geometry/Posture	Japanese ‘ta’		: Finger Spelling
D4	D4	: Motion	F2/O	-	: Geometry/Posture
F7/O	F7/O	: Geometry/Posture			
D2	D1	: Motion	Japanese ‘ya’		: Finger Spelling
E F7/O	F7/O	: Geometry/Posture	F8/B	-	: Geometry/Posture

Figure 4. *Examples of Japanese sign-language database*

30 JSL words often used in the self introduction and in the daily life, and also Japanese 50 finger spellings are currently registered in the database and used for the realtime recognition. Three descriptions of JSL and three finger spellings are listed in Figure 4 as examples. The description of finger spellings consists of the geometry and the posture of hands, on the assumption that they are shown around the shoulder position. In case motion or hand figure has no effect on the meaning, it is marked as -.

3.2 Configuration of JSL Recognition System

The recognition system consists of two acceleration sensors, a pair of datagloves and computers. Two PCs are employed for the data acquisition from the sensors: one for the acceleration data, and the other for the datagloves and the position sensors on both hands. Every time motion and hand figure primitives are extracted in each PC, they are transmitted to SGI Indy computer, which are used for the execution of JSL recognition. By wearing a pair datagloves with the acceleration sensor and the position sensor fixed on the back side of the wrist area, user is able not only to execute the recognition of JSL but also register JSL words in the JSL database. The recognizable subjects in this system are the JSL words and finger spellings represented by the sequences of motion primitives and hand figure primitives as described in the JSL database.

Schematic diagram of the JSL recognition system is described in Figure 5. By observing acceleration data, the recognition algorithm starts its action when acceleration value exceeds a pre-determined threshold. First, hand figure primitives and hand postures of both hands are recorded on the gesture table prepared in SGI Indy, which is followed by the motion primitive extraction. Extracted motion primitives as shown in figure 3 are stored in the gesture table as a sequence in temporal order. When the changing point of motion is detected by observing the motion primitive sequence, a hand figure primitive and hand posture data are measured. By the repetition of the procedure, a primitive sequence as

Start: Hand figure - Motion - Hand figure - Motion - ..... - Motion - Hand figure : End

is obtained and stored in the gesture table until the end of gesture is detected by the acceleration data observation. Then the sequence is handed to the JSL recognition algorithm. For the finger spelling recognition, shoulder position of a user is assigned as the display spot. In case the gesture is presented at this area, it is regarded as a finger spelling and its recognition procedure is executed.

JSL recognition is made by comparing gesture primitive sequences stored in the gesture table with JSL database. For the comparison, two algorithms are currently examined. One is an automata algorithm, and the other is HMM (Hidden Markov Model) which is widely used for the time series pattern recognition.

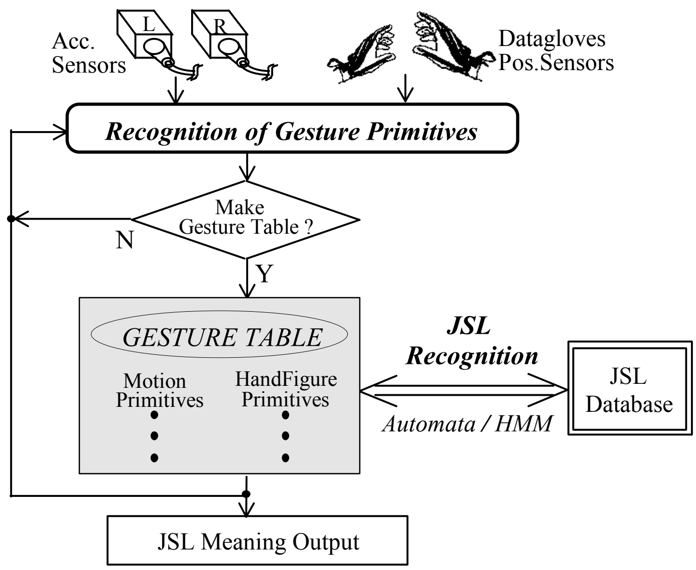
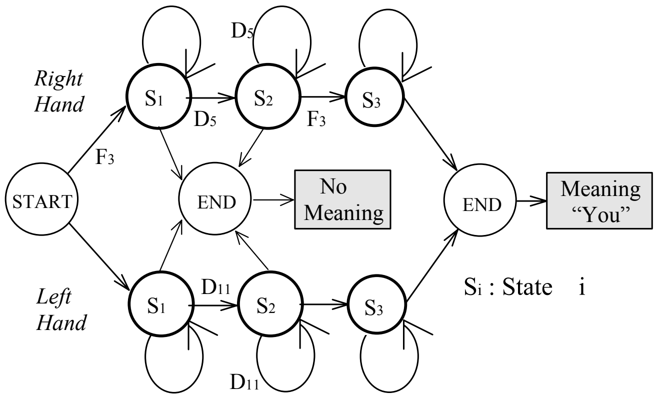


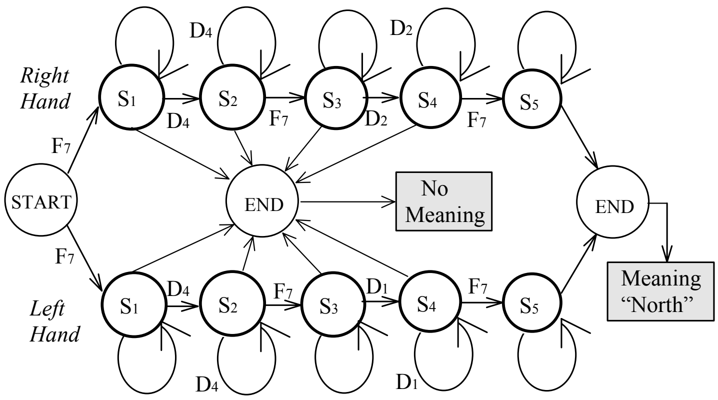
Figure 5. *Japanese sign-language recognition process*

3.3 JSL Recognition by Automata

Automaton is an information processing model consisting of finite number of states which transits deterministically according to an input[7]. Since the transition is determined by all the past input sequences, the final state of automaton is uniquely confirmed by the input sequence. By preparing necessary automata consisting of symbolic sequences of gesture primitives expressing JSL words, computerized recognition can be constructed. Two examples of automata used in the JSL recognition is shown in figure 6.



(a) Recognition of "You"

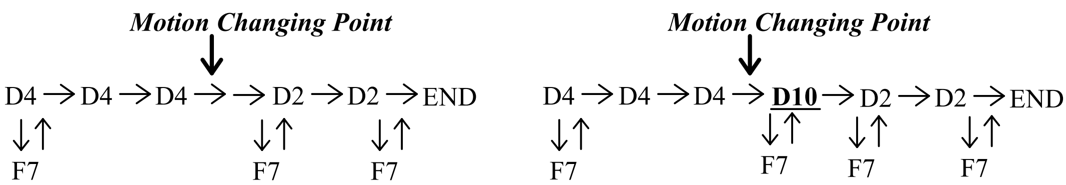


(b) Recognition of "North"

Figure 6. *Automata in sign-language recognition*

Table 5. *Experimental results of sign-language recognition by automata*

JSL words	Recognition Rate (%)	Misunderstood
You	100	-
Like	100	-
Good bye	93	No Meaning
Happy	87	No Meaning
North	83	No Meaning
Japanese 'a'	100	-
Japanese 'ta'	100	-



(a) Example of succeeded recognition *(b) Example of recognition failure at D10*

Figure 7. *Examples of success and failure in recognition of "North"*

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

Experimental results of recognition by the automata is listed in table 5. For the JSL words consisting of one shot motion such as "You" and "Like", 100 % recognition rate is obtained. On the other hand, in the recognition of the word "North", for example, whose motion pattern changes from D4 to D2, the resulting outputs are failed at the changing point in some recognition cases. In this case, the recognition results in "No meaning". This is caused by the property of automata whose transition is made deterministically. An example of the recognition failure process is schematically described in figure 7.

3.4 JSL Recognition by HMM

Another algorithm employing a HMM is also examined in this study. The HMM is a chain with finite states connected by the probabilistic transitions[8]. Each state is characterized by two sets of probabilities: one is a transition probability among states, and the other is output probability distribution which defines the probability of emitting each symbol from the state. These probabilities can be determined from the learning pattern sets by using learning algorithms such as the Baum-Welch algorithm, the Forward-Backward algorithm and the Viterbi algorithm. Figure 8 shows an example of the HMM having 3 states with the 2 output symbols. Although the learning procedure of the probabilities are necessary for the HMM recognition algorithm, and the recognition ability depends on the learning pattern sets and learning procedure, recognition failures found in the automata algorithm as shown in figure 7 are expected to be avoided.

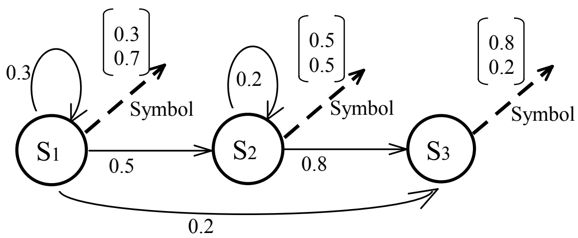


Figure 8. *Example of HMM*

HMM used for the JSL recognition is a left-to-right model having 3 states with a skip transition. And the Baum-Welch algorithm is adopted for the learning. Output symbols from each state are 24 symbols which consist of 10 motion primitives and 14 hand figure primitives. Table 6 shows several examples of experimental results. Comparing with the automata algorithm, JSL words with one shot motion tend to cause the misunderstood results in the HMM. The reason is that the mis-recognition of gesture primitives seems to be hard to be absorbed in the calculations of transition probabilities.

Average of 94 % recognition rates was obtained for the 30 JSL words stored in the database, on the other hand, automata algorithm resulted in the 89 % recognition rates. Although recognition ability was improved by adopting the HMM, the HMM outputs misunderstood results when it fails the recognition. On the other hand, the automata ends with the answer of "No meaning", to allow the system to interactively ask the user to input the gesticulation again.

Table 6. *Experimental results of sign-language recognition by HMM*

JSL words	Recognition Rate (%)	Misunderstood
You	100	-
Like	88	Interesting
Good bye	96	Happy
Happy	91	Good bye
North	100	-
Sad	100	-
Interesting	94	Like
Father	96	Interesting
Mother	100	-

4. CONCLUSIONS

The automata algorithm and the HMM are introduced for the recognition of JSL together with the construction of JSL database in which words are described as sequences of gesture primitives obtained from acceleration patterns and hand figures. A compact acceleration sensor which is considered to reduce the constraints of motion is introduced for the motion primitive extraction. In the current study of the JSL recognition, the HMM has better performance rather than the automata algorithm. The algorithms have to be further examined on their performance and reliability by increasing the JSL database scale. Although learning procedure by the repetition of pattern inputs in advance is necessary in the HMM method, it is considered to give robust recognition with the integration of gesture sensing devices.

5. REFERENCES

[1] T. Kurokawa, "Non-verbal Interfaces", IEICE (ed.), Ohm Co., 1994

[2] H. Sawada and S. Hashimoto, "Gesture Recognition Using an Acceleration Sensor and Its Application to Musical Performance Control", Trans. Electronics and Communications in Japan, Vol.80, No.5, pp. 9-17, 1997

[3] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models", Proc. of ISCV'95, 1995

[4] J. Kim, W. Jang, and Z. Bien: "A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)", IEEE Trans. on Systems, Man and Cybernetics -Part B:Cybernetics, Vol.26, No.2, April, 1996

[5] H. Sagawa, M. Takeuchi, M. Ohki: "Description and Recognition Methods for Sign Language Based on Gesture Components", 1997 Int'l Conference on Intelligent User Interface , pp.97-104, 1997

[6] H. Sawada, N. Onoe and S. Hashimoto, "Acceleration Sensor as an Input Device for Musical Environment", Proc. ICMC, pp 421-424, 1996

[7] Arto Salomaa, "Theory of Automata", Pergamon Press, 1969

[8] R.J. Elliott, L. Aggoun and J.B. Moore, "Hidden Markov Models, Estimation and Control", Springer-Verlag, New York, 1995

S-TEL: An avatar based sign language telecommunication system

Tomohiro Kuroda, Kosuke Sato and Kunihiro Chihara

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara, 630-0101, JAPAN

{tomo,sato,chihara}@is.aist-nara.ac.jp

ABSTRACT

Although modern telecommunication have changed our daily lives so drastically, the deaf cannot benefit from them based on phonetic media. This paper introduces a new telecommunication system for sign language utilizing VR technology, which enables natural sign conversation on analogue telephone line.

On this method, a person converses with his/her party's avatar instead of party's live video. As speaker's actions are transmitted as kinematic data, the transmitted data is ideally compressed without losing language and non-language information of spoken signs.

A prototype system, S-TEL, implementing this method on UDP/IP, proved the effectiveness of avatar-based communication for sign conversation via a real lossy channel.

1. INTRODUCTION

Although modern telecommunication has changed our social communication style so drastically, the audibly challenged cannot take benefits of them based on phonetic media. In order to associate their isolated community together and to increase the quality of their daily lives, a new communication system for signers is indispensable.

Today the deaf use TTY or facsimile instead of telephone. However, with these character based communication systems, they need to translate their sign conversation into descriptive language and to write down or type in. So, they eager a new telecommunication system which enables them to talk in signs.

Nowadays so many research works on computer aid for signers including telecommunication systems for signers are coming. Some of these research works have developed data compression techniques for video stream of sign language, others have developed script-based sign communication methods that translate signs into descriptive languages to reduce transmitted data. These methods succeeded to compress transmitted data, but the language and non-language information contained in signs is lost due to the above compression or the translation. Unfortunately, they cannot mediate natural sign conversation.

Therefore, Kuroda et al. (1995) introduced a concept of new telecommunication method for sign language integrating human motion sensing and virtual reality techniques. In this paper, a new realized telecommunication system based on this method is introduced.

In this system, a person converses with his/her party's avatar instead of the party's live video. Speaker's actions are obtained as geometric data in 3D space, the obtained motion parameters of the actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. Thus, it realizes optimal data compression without losing language or non-language information of given signs. Moreover, users can hide their private information without giving displeased feeling.

In this paper, the features of Japanese Sign Language are briefly explained in section 2 and foregoing studies on sign communication are mentioned in section 3. In section 4, the avatar-based communication method is introduced, and avatar-based communication and video-based communication is compared from bandwidth viewpoint. Finally, in section 5 a prototype avatar-based communication system, S-TEL, is experimented on UDP/IP by Deaf and sign experts.

2. JAPANESE SIGN LANGUAGE

Japanese Sign Language (JSL) is the mother tongue of Deaf for a hundred years in Japan (Komekawa, 1998). As JSL is a visual language, there are some features in comparison with phonetic languages.

- The meanings of signs are defined by hands' feature, position, movement, and direction (Fig. 1).
- Some signs cannot identify from frontal view because of occlusion (Fig. 2).
- Three-dimensional position of signs denotes the relation of the current communication. Persons and their ranks are shown by the position to indicate signs or the direction of upper body (Fig. 3).
- Non-Manual Signals (NMS) like nodding work as modifiers.

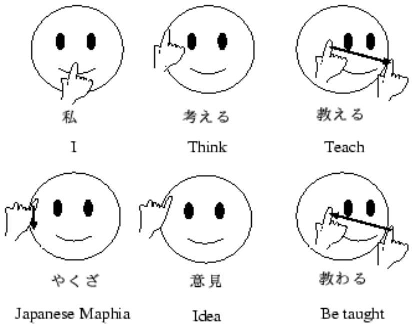


Figure 1. *The meanings are determined by hands figure etc.*

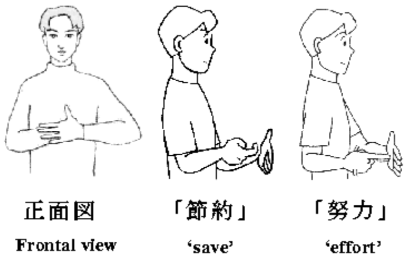


Figure 2. *Some signs cannot be identified from frontal view because of occlusions.*

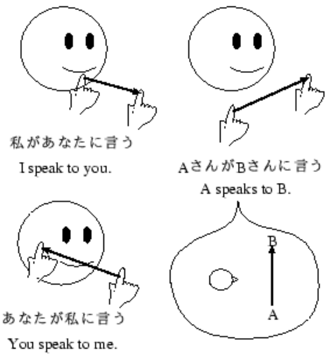


Figure 3. *The position of sign shows character and its rank*

3. FOREGOING RESEARCHES

There are many research efforts on computer aid for signers. However, most of them focus on sign translation or sign education. Only few attempts have so far been made at the communication among signer. These researches on communication among signer can be divided into two groups, which is script-based communication and video-based communication.

4.1 Script-based Communication

Jun et al. (1991) and Ohki (1995) proposed script-based communication system. These systems translate given signs into script language or phonetic language, transmit them, and produce sign animation on receivers terminal. Therefore, they can eliminate bandwidth of transmission so drastically. However, as these script-based systems have translation stage, they cannot forward given sign when their dictionaries have no entry for the given sign or they mistranslate it.

4.2 Video-based Communication

As sign language is visual language, videophone seems usable as telecommunication system for sign language. However, Yasuda (1989) pointed out following 'Eyes Torture' problem of videophones for daily life use.

- User's eyes may break into the other's private space through camera. The other's privacy may be trespassed.

Kamata (1993) experimented with some videophones and cleared that following problems make difficult of sign conversation on videophones

- 2D videophones images cannot show whole sign information, because sign language moved in 3D space.
- The received images do not have enough resolution, view, and frame rate for practical sign conversation.

As sign language includes fast hand motions and occluded postures, general methods for video compression are not suitable for sign communication. Sperling et al. (1985) and Gulska (1990) proposed video compression method for sign language, but their methods cannot transmit sign image sequences that has enough time/space resolution to read on analogue telephone line.

4.3 Bandwidth of Video-based Communication

Kamata (1993) argues that image sequence of QCIF mode ISDN videophones ($176 \times 144\text{(pixel)} \times 15\text{(fps)}$) doesn't always has enough time/space resolution for sign conversation. However, Sperling et al. (1985) says that three bits gray scale image sequence of $24 \times 16\text{(pixels)} \times 15\text{(fps)}$ can visualize 'enough intelligible' ASL. Assuming three bits gray scale image, Kamata's system requires at least 2.2Mbps and Sperling's requires 34Kbps for bi-directional communication. As these results vary so widely, we examined required bandwidth for video-based sign communication.

Firstly, to examine required frame rate, we selected two topics (79 seconds) from NHK sign language news, and measured how long frames each sign word continues. As Tab. 1 and Fig. 4 shows, some words continue less than 1/30 seconds. Moreover, Kanda et al. (1996) cleared that newscasters speak about 70% speed of normal sign conversation. Thus, the conclusion is that video rate (30 frames per second (fps)) is not sufficient for sign conversation. However as there are no faster display for home use, we assume video rate display in following discussion.

Table 1. *The Number of Frames Each Sign Continues. (1^- denotes less than 1 frame.)*

	News 1	News 2	Total
Average	5.72	9.36	7.84
Minimum	1^-	1^-	1^-
Maximum	21	30	30

Secondly, to examine required resolution, we selected 10 frames ($640 \times 480\text{(pixels)}$) from NHK sign language news, and measured the width of fingers. The narrowest finger's (pinkie of woman) width was five pixels. Thus, sign image requires 128×96 pixels to identify each finger.

From these discussions, assuming three bits gray-scale image, the required bandwidth for bi-directional video-based sign communication is 2.2Mbps. Therefore, video-based sign communication requires high bit-rate digital channel.

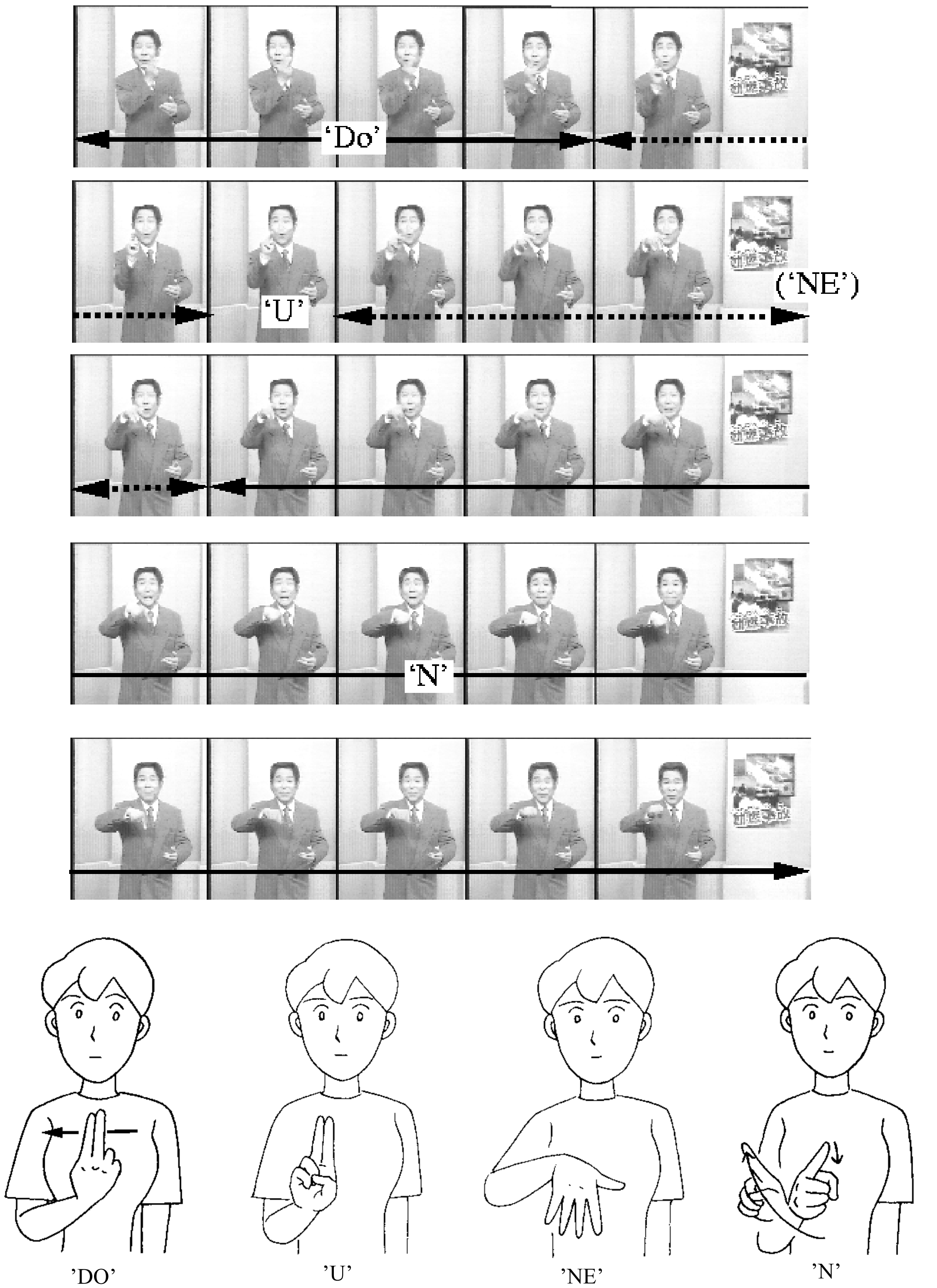


Figure 4. *Finger Spelling ‘DOUNEN’ (Power Reactor and Nuclear Fuel Development Corporation) from NHK Sign Language News. Finger Character ‘NE’ continues less than 1/30 seconds.*

4. SYSTEM DESIGN

4.1 Avatar-based Communication

To solve above problems of the previous telecommunication systems, we introduce new telecommunication system for sign language integrating human motion sensing and virtual reality techniques. This method solves natural sign conversation on conventional analogue telephone line.

In this method, a person converses with his/her party's avatar instead of the party's live video. Speakers actions are obtained as geometric data in 3D space, the obtained motion parameters of actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. This avatar-based communication has following advantages. Firstly, sending kinematic data without any translation process, this avatar-based communication realizes data compression without losing language or non-language information of given signs. Secondly, sending 3D motions, receivers terminal can produce signing avatar to increase readability of signs. Thirdly, this method can be applied to conferencing or party talking (Kuroda, 1997b). Finally, visualizing avatar instead of live video, users can hide their private information without giving displeased feelings for his/her party.

This system consists of following components as shown in Fig. 5.

- **Sender** obtains signs as geometric data and sends it. **Motion measuring part** measures hands, arms and upper body motions. **Encoder** encodes and compresses obtained data.
- **Receiver** receives kinematic data of signs and displays avatar. **Decoder** decodes given data into kinematic data. **Avatar producing part** produces avatar from given kinematic data and display it. This part makes use of reader's viewpoint information if needed. Produced avatar can be virtual CG avatar, robot, etc.

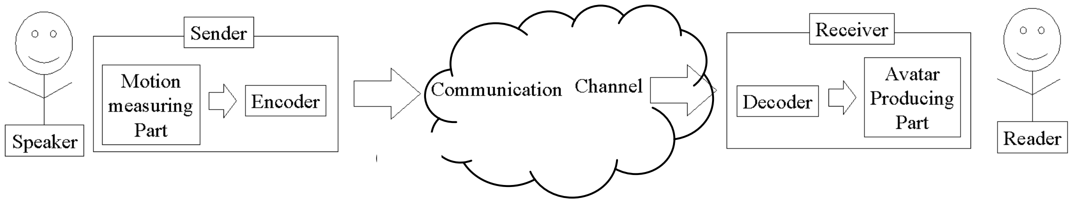
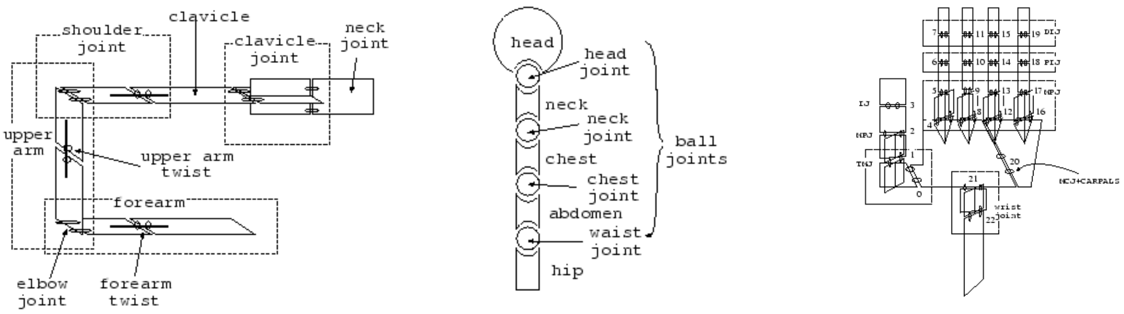


Figure 5. *Overview of Avatar-based Communication*

4.2 Bandwidth of Avatar-based Communication

Avatar must handle whole upper body as signer shows signs with hand, arm and upper body motions. Therefore, we assume skeleton of avatar as shown in Fig. 6. This model has following features.

- The spine consists of 33 or 34 vertebras, and small rotation between these vertebras produces backbone bend. Especially, 5 cervical vertebras and 7 lumber vertebras moves much more than the other vertebras. Thus, this model has joints at both sides of these two parts.
- Some sings like ASL 'why' visualized by shoulder motion. Therefore, this model has joints on neck side of clavicle.



Arm model | Body Model | Hand Model

Figure 6. *Skeleton Model of Avatar*

Suppose finger width is unit length as section 4.3. Finger width for woman is about 2 cm and the sitting woman's wrist move inside a sphere which diameter is 2m. Therefore, seven bits code can denote wrist position, and eight bits code can denote 360^o . This model has 72 degree of freedom and amount of rotation of each joint is as Tab. 2. Thus, 460 bits code can denote whole upper-body posture. Kuroda et al. (1996b) proposed a method to reconstruct upper body motion form position and rotation data of wrists and top of head. Applying this method, 415 bits code can denote whole upper-body posture. From these discussion, the required bandwidth of bi-directional avatar-based sign communication 16Kbps. Thus, avatar-based communication is available on conventional analogue telephone line.

Table 2. *Joints of Avatar*

Part	Degrees of freedom	Rotation	Bits
Hand	42	90	6
Wrist	4	180	7
Clavicle joint	4	90	6
Rotation around clavicle	2	270	8
Others	20	180	7

5. EXPERIMENTS

5.1 Prototype System S-TEL

A prototype, S-TEL, along the design discussed in section 4 is developed as Fig. 7. Kuroda et al. (1996a) cleared that 3D stereo scopic view has no effect on the readability of signs and that 2D CG reflecting readers motion parallax is sufficient to realize practical readability. Therefore, S-TEL uses normal 2D display as shown in Fig. 8.

S-TEL sender composed of Pentium 166MHz PC with Windows95, two CyberGloves and a Fastrak. S-TEL receiver composed of Intergraph TD-5Z workstation (Pentium 100MHz with OpenGL accelerator) with WindowsNT 3.51 and a Fastrak. All software components are built on World Tool Kit Ver. 2.1 for WindowsNT and Visual C++ 2.

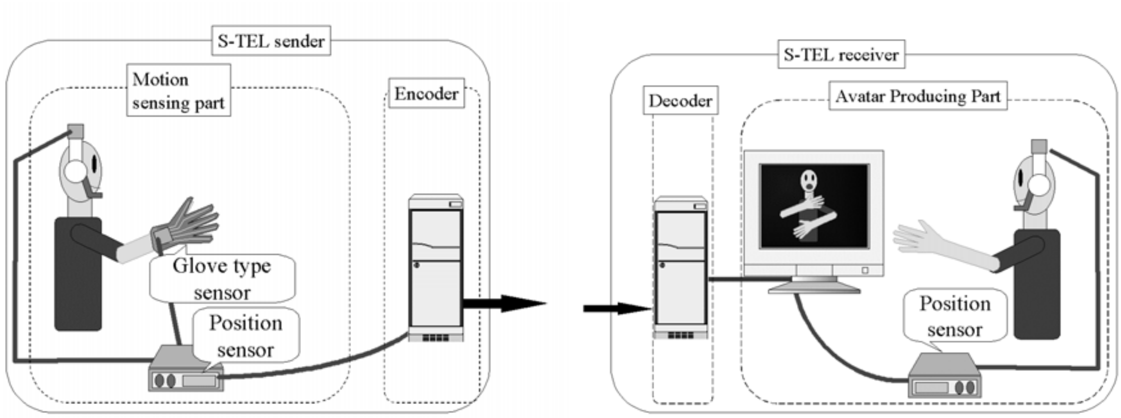


Figure 7. *Design Overview of S-TEL*

 Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

Figure 8. *Produced Avatar of S-TEL*

5.2 Bandwidth of S-TEL

S-TEL obtains signer's action with CyberGloves and Fastrak. CyberGlove obtains 18 finger joint bending as integer and Fastrak obtains position as 3 single float and orientation as single float quaternion (4 degrees of freedom). Therefore, the amount of data of one flame is 120 bytes. Assuming 30 fps, bandwidth of mono-directional S-TEL is 28.8Kbps.

5.3 Experiments

A developed prototype S-TEL was experimented with two types of UDP/IP communication channel.

Firstly, we connected S-TEL by Ethernet. Three deafs, three sign experts and four sign beginners tried to talk freely in sign language on S-TEL.

Secondly, we placed S-TEL sender at NAIST, Ikoma City (Nara) and S-TEL receiver at Kumamoto City (Kuroda et al., 1997a). The distance between two cities is about 700Km. Satellite JCSAT-1 connected two cities. The bandwidth of channel was 4.0Mbps bi-directional. By adding heavy traffic on the communication channel, there is 2% packet loss due to overload of the channel. Speakers were two sign experts and readers were one sign expert and two beginners. All testees are hearing. Testees tried to teach sign language on S-TEL. Testees also tried to talk in signs through NV/VAT Internet video chat, and they compared these two systems.

5.5 Results and Discussions

Experimental results are as follows.

- The realized frame rate of S-TEL is 26.1 fps. S-TEL can perform practical frame rate to read signs on consumer PC.
- 2% packet loss had no effect on the quality of visualized signs. Because, S-TEL detects erroneous flame and keep the continuity of the preceding flame image. This ensures S-TEL works properly on lossy channel.
- Readers could recognize 75% of spoken signs. Users can almost make themselves understood through S-TEL. Speakers signs which readers couldn't recognize are the following cases.
  - ◊ Readers couldn't recognize spoken signs which facial expressions modify, because S-TEL doesn't treat facial expressions.
  - ◊ Readers couldn't recognize finger spelling, because its finger size of the avatar is not sufficient to distinguish fingers.
- Sign experts completed a difficult task to teach 'self-introduction' for beginners over S-TEL. Usually, for beginner, one must teach at hand with utmost care and kindness. It is one proof that S-TEL can provide the environment where users can talk in signs as if they talk at hand.

- Testees said that they prefer S-TEL rather than videophones as sign communication media, because videophones could expose their privacy but S-TEL wouldn't. They also said S-TEL would enrich their daily lives.

These experimental results clear that avatar-based sign communication is effective as users can make themselves understood through S-TEL. Moreover, avatar-based communication is superior to video-based system in bandwidth viewpoint; avatar-base communication is available on lossy and narrow channel.

To increase readability of signs, avatar-based communication should treat facial expressions and visualize hand bigger. However, to make hand bigger causes another problems. Some researchers on sign animation including us did tried to make it bigger. Testees complained that unbalanced avatar makes signs unreadable and that testees sometimes feel hit by avatar when avatar stretches their arm to the front. Thus, to increase readability, avatars hand should expand when signer start to spell. It is needed to develop new method to identify whether spoken signs are finger spelling or not.

6. SUMMERY

In this paper, avatar-based sign communication system, which is an innovative system for sign language telecommunication, is presented. In this method, speakers actions are obtained as geometric data in 3D space, the obtained motions parameters of actions are transmitted to the receiver, and the speaker's virtual avatar appears on the receiver's display. As avatar-based communication treats speaker's actions as geometric data in 3D space, it allows to talk and read signs naturally through conventional analogue telephone line, increase readability of signs, and protects users' privacy.

The experiments to talk in signs through a prototype system, S-TEL, were performed. These clear the effectiveness of avatar-based communication and the superiority of avatar-based communication over video-based communication as a telecommunication media for sign language.

When S-TEL gets popular among Deaf widely, their isolated community would associate together. S-TEL would increase the quality of their daily lives. Authors are integrating the transmission of facial expressions of signers and the identification of finger spelling into S-TEL in progress.

Acknowledgement: This study is cooperated with Kamigyo Branch of Kyoto City Sign Language Circle "Mimizuku", and Kamigyo Branch of Kyoto City Federation of Deaf, Wide Project, Digital Research Inc. and Japan Satellite Systems Corp.

7. REFERENCES

S Gulska (1990), The Development of a Visual Telephone for the Deaf: Using Transputers for Real-time Image Processing, *In Transputer Research and Applications 3. Proceedings of the Third North American Transputer Users Group*, pp.7-16

X Jun, Y Aoki and Z Zheng (1991), Development of CG System for Intelligent Communication of Sign Language Images between Japan and China, IEICE transactions, **74**, 12, pp.3959-3961

K Kamata (1993), Sign Language Conversation through Video-phone – Experiment at School for the Deaf --, Technical Report of IEICE, ET92-104, pp.37-44, Japanese

K Kanda (1996), Sign Linguistic from Basics, Fukumura Press, Japanese

A Komekawa (1998), Japanese-Sign Language Dictionary, Japan Federation of Deaf, Japanese

T Kuroda, K Sato and K Chihara (1995), System Configuration of 3D Visual Telecommunication in Sign Language, *In Proceedings of the 39th Annual Conference of the Institute of Systems, Control and Information Engineers, ISCIE*, pp.309-310, Japanese

T Kuroda, K Sato and K Chihara (1996a), S-TEL: A Telecommunication System for Sign Language, *In Conference Companion of First Asia Pacific Computer Human Interaction*, pp.83--91

T Kuroda, K Sato and K Chihara (1996b), Reconstruction of Signer's Actions in a VR Telecommunication System for Sign Language, *In Proceedings of International Conference on Virtual Systems and Multimedia VSMM'96 in Gifu*, pp.429-432

T Kuroda, K Sato and K Chihara (1997a), S-TEL: A Sign Language Telephone using Virtual Reality Technologies, *In CSUN's 12th Annual Conference Technology and Persons with Disabilities*, Floppy Proceedings, KURODA_T.TXT

T Kuroda, K Sato and K Chihara (1997b), S-TEL: VR-based Sign Language Telecommunication System, *In Abridged Proceedings of 7th International Conference on Human-Computer Interaction*, pp.1-4

M Ohki (1995), The Sign Language Telephone, TELECOM '95, pp.391--395

G Sperling, M Landy, Y Cohen and M Pavel (1985), Intelligible Encoding of ASL Image Sequence at Extremely Low Information Rates, Computer Vision, Graphics, and Image Processing, **31**, 3, pp.335-391

H Yasuda (1988), TV-phone Now, Spectrum, **1**, 5, pp.88-102, Japanese

Improving the mobility of severely disabled

Pertti Peussa[1], Ari Virtanen[2] and Tiina Johansson[3]

[1,2]VTT Automation,
P.O.Box 13021, FIN-33101 Tampere, FINLAND

[3]Adaptation Training Centre for Disabled,
Launeenkatu 10, FIN-15100 Lahti, FINLAND

[1]*pertti.peussa@vtt.fi*, [2]*ari.virtanen@vtt.fi*, [3]*tiina.johansson@invalidiliitto.fi*

ABSTRACT

A modular mobility aid system is presented, which can be attached to most commercial electric wheelchairs to increase the independence and safety of the wheelchair user. The system consists of falling avoidance, obstacle avoidance, and beaconless navigation functions. Experiences from an evaluation period are presented. Also an idea is presented to combine map information to environmental perceptions in order to provide the wheelchair user a better awareness of his/her whereabouts. The navigation system can also be applied to different mobile robot applications.

1. INTRODUCTION

Safety and independence are very important for elderly people, and especially for the severely disabled. Many of them have to ask help any time they want to move outside their apartment; some need assistance even if they just want to go from one room to another. No one of us would like to be so dependent on other people.

The Adaptation Training Centre for Disabled and VTT Automation joined forces to create more safe and more economical moving aids. The project aimed at modular device prototypes, which could be attached to most commercial electric wheelchairs to enable persons who have not been able to steer a normal electric wheelchair before, to use it independently and safely. One important goal was to initiate industrial interest towards this kind of products, since without good and affordable assistive products no major improvements are possible in the elderly and disabled sector. Important things to consider were easy-to-use, low cost, and easy installation.

The project came up with three main functions: falling Avoidance, obstacle avoidance, and Navigation. To implement these functions, the development of three different but modular devices started. Three separate modules would have made it easy to choose a combination that would best fulfil the individual needs. The only exception to free combinations was, that the navigation module requires also obstacle avoidance module, since the navigation unit needs to know the environmental features, which only the obstacle unit can provide.

VTT delivered a prototype device to the Adaptation Training Centre for Disabled for evaluation about six months ago, and made an other one to further develop navigation methods and Falling&Obstacle avoidance technologies.

2. THE MOBILITY AID MODULES

2.1. Falling Avoidance

Falling avoidance is a safety function, that prevents the wheelchair from tumbling down stairs, decks and other similar places. For instance, backing a wheelchair out from an elevator in a confined staircase can be very dangerous. Also the obstacle avoidance system itself can cause problems since it won't "see" any descents; if one is passing an obstacle on the left, the wheelchair tends to turn to the right because it only sees a lot of free space there, and not e.g. stairs.

Falling avoidance is an unconditional on/off safety function, which the driver can not overcome it in any way. It is based on six optical sensors: two at the front, two at the back, and one on both side, see Figure 1. The sensors keep watch that there is a surface at a proper level around the wheelchair. If this is not the case,

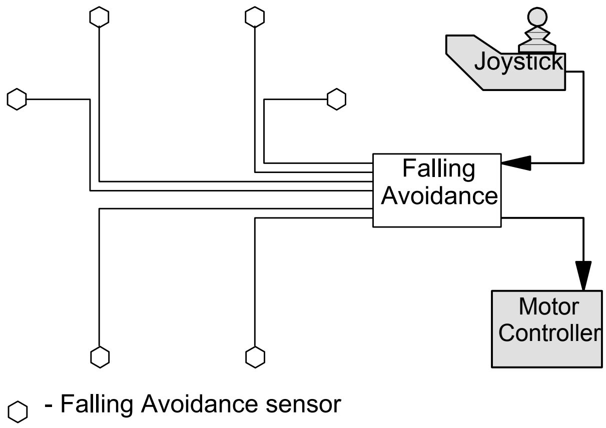


Figure 1. *The basic falling avoidance unit is connected between the joystick, and the motor controller.*

the unit does not allow to steer the wheelchair to that direction.

The Unit has been tested with a limited number of sensors. Some further work is still needed to find reliable low-cost sensors, and to find appropriate places for sensor attachments on different wheelchairs. The sensors must "see" far enough from the wheelchair so that there would be enough time to stop even from maximum speed. On the other hand they should also be very reliable despite the fact that they have to be able to detect the floor from an inclined angle, even if it may be very shiny (new, polished floor) or mat (dark carpet). During the project it was found sensible to let the falling avoidance share the same electronics as the obstacle avoidance in the pilot devices.

2.2. Obstacle Avoidance

Obstacle avoidance function is based on a modified potential field method. Thanks to the modifications, it enables driving through doorways, pushing doors open etc. These are simple, but important features, because it would be funny if you could not drive to a table or cupboard with your wheelchair! Potential method itself has many positive features, like that it is reliable and predictable. In many cases it is beneficial that it tends to steer the vehicle in the middle of the free space, which is makes the driving easier in e.g. narrow doorways. The performance of the obstacle avoidance unit can be seen on video clips; look for link at our website at http://www.kautre.aut.vtt.fi/.

The obstacle avoidance unit is attached between the wheelchair joystick, and motor controller, see Figure 2. Its basic function is to modify the analog joystick signals based on the ultrasonic sensor readings. There is also a version which is compatible with a digital wheelchair control bus called the M3S bus. This version can be changed to an other bus called the DX bus. At the moment most of the wheelchair control signals are analog, but the relative amount of wheelchairs utilising serial buses is growing.

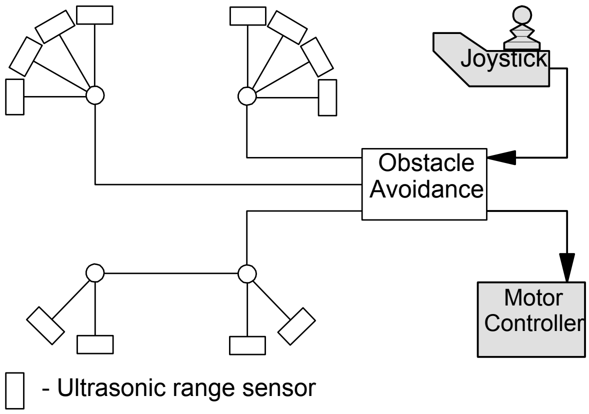


Figure 2. *The obstacle avoidance unit is based on a low-cost 8-bit microcontroller, and 12 low-cost ultrasonic sensors. The analog version is attached between the wheelchair joystick, and motor controller as above. M3S, and DX compatible units are connected to the bus, and then configured to modify the joystick messages accordingly.*

The obstacle avoidance unit does not take over the control from the driver. The unit merely assists the driver by adding a corrective component to the joystick signal. The driver remains the head of the steering all the time. When approaching an obstacle right ahead of the wheelchair, the unit reduces the speed, but continues to approach the obstacle as long as the driver consistently steers towards it. This way one can push doors open, or drive to a table.

The obstacle avoidance software is not limited to ultrasonic sensors. Therefore it is easy to apply also optical sensors, radars, or other technologies. In serial production the size of the unit can be reduced to one fifth of the present size. The component costs are about 330ECU. So far the unit has been successfully tested with three different wheelchairs made by Permobil, Invagear, and Ortoped, see Figure 3. A Finish company is interested in developing this unit into a product.

Figure 3. *Here the obstacle avoidance unit is attached to a standard electric wheelchair. Notice the ultrasonic sensors in the front, and the electronic obstacle avoidance unit at the back. The rear sensors can not be seen from this angle.*

2.2.1. Operating principle. The obstacle avoidance unit is developed to assist persons, who have enough control over their decision making. This means that a wheelchair is not allowed to manoeuvre by itself, and the pilot must have the possibility to collide with anything she/he wants to e.g. to push them aside. obstacle avoidance cannot be bullet proof, because identifying different situations (the pilot wants to collide, or the pilot didn't perceive the obstacle) is impossible. Nevertheless the system must act so clear, that the driver can feel the unit interference in operation. This helps the driver to rely on obstacle avoidance, and also to judge whether the unit is beginning to fail.

Obstacle avoidance algorithm used is based on virtual force field (VFF) -method described by Borenstein and Koren (1991a). Normally in VFF method an obstacle creates virtual pushing force and a target point creates virtual pulling force. These forces are combined and the resultant is considered as a new control of the mobile robot. As the position and heading of the mobile robot are known, the used method needs quite a lot of computing power and one of its disadvantages is that doorways are difficult to pass because of opposite forces. A more developed method is Vector Field Histogram, VFH (Borenstein and Koren 1991b), which is much faster but it also uses position information.

The obstacle avoidance unit does not use wheelchair position, heading, or target information. An obstacle generates a virtual force as in VFF-method. These virtual forces are combined to single force vector, see Figure 4. Sum of this force vector F and the original control vector O is the resultant vector R. In VFF-method R is used as a new control vector. In our method, original control vector O (given by user with a joystick) is turned to the same direction as the resultant vector R. This new control vector O' is then sent to the motor controller. When O is a zero vector, wheelchair does not move. The user alone controls the speed of the wheelchair. This method allows driving against wall at full speed, if user drives straight toward it. For this reason, one must check directions straight ahead and back separately. If any obstacles are inside the safety limit, wheelchair speed is automatically restricted to a crawling speed. Thus it is still possible to push a doors open or drive to a table, but safely. The algorithm is fast. It takes only about 15 ms to run it with an 8 bit processor.

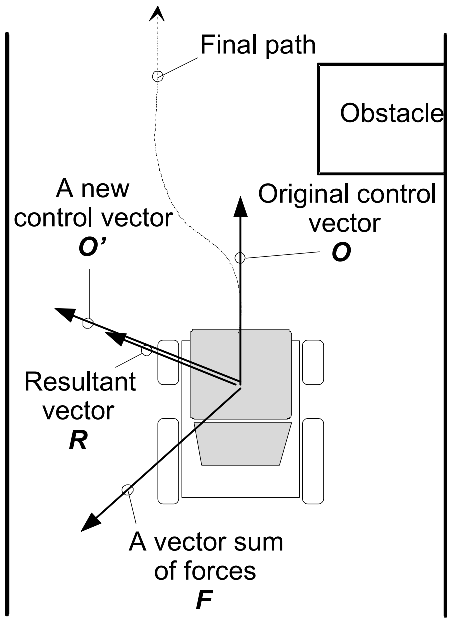


Figure 4. *The vectors of the obstacle avoidance algorithm.*

2.2.2. Encountered problems. The device can not avoid humans, if they happen to move very quickly towards the wheelchair across its route. Usually this is not the case, since those persons usually have their inborn "obstacle avoidance system", and they can dodge the wheelchair far better than the wheelchair could ever bypass them. Some difficulties may arise with standing persons, if they have very soft and thick clothes, which reflect only a very weak ultrasonic echo. Fortunately this is found to be rare during the performance tests that have been carried out. Also obstacles that are above the driver's chest, or less than 10 cm above the floor are not "seen" too well by the sensors, and are therefore not bypassed. This shortcoming will be studied in the future, especially if the pilot devices get good acceptance. One possibility is to use sensors, whose operating principle is other than ultrasonic.

Sometimes the potential field method used by the obstacle avoidance unit may steer the wheelchair in an

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

undesired way, if there is an obstacle on the other side of the path, and a dangerous area like stairs or a street on the other side. Since the obstacle avoidance unit does not 'see' any obstacles on the stairs/street side, it tends to drive the wheelchair towards that direction in order to stay away from the obstacle. This is why the augmented joystick signal from the obstacle unit may be passed to falling avoidance unit (if it is attached), which ultimately decides whether it is safe to steer to the current direction.

The obstacle avoidance unit does not get any information whether the wheelchair has really turned to the direction given. This may be problematic, if the driving surface is uneven, or if the castor wheels of the wheelchair have abrupt driving geometry: the augmented signal is not necessarily strong enough to turn the wheelchair even if it has perceived obstacles. Also dead zones in the motor controller add up to the problem. This problem could be overcome by installing wheel encoders, but then the device installation procedure would become more complex. However, this is not a problem, if also the navigation unit is installed. Wheel encoders namely come along with the navigation system, because of the required odometry function.

2.3. Navigation System

The navigation unit is currently under development. It will be able to guide the wheelchair to a given location within a building, or restricted outdoor area. The pilot will not participate to steering, the navigation unit takes care of it, but she/he may interrupt it at any time. The steering signals are sent to the obstacle avoidance unit for further processing just like normal joystick signals. The navigation unit requires three additional sensors: two encoders for odometry, and a low-cost electric compass. Physically the navigation unit means these sensors, and a laptop PC, which contains route planning, and route following software together with maps and landmark information. The laptop is connected to the obstacle avoidance unit by a cable. Figure 5 presents a system which contains all the functions described in this article. At the moment all the maps and landmark information must be imported manually, but it is believed that this could be made easier later.

The starting point for the navigation system design was, that no artificial landmarks or beacons should be used. This dates back to the requirement of maximum independence: the system should operate in any building - not only in user's home. This requirement can also be analysed from the opposite perspective: if the navigation would be based on e.g. radio beacons, it could only be used in areas which are equipped with the specific beacon type. This would limit the user's "independence" only to his or hers own apartment or day centre, since there are no general beacon standards emerging in the foreseen future.

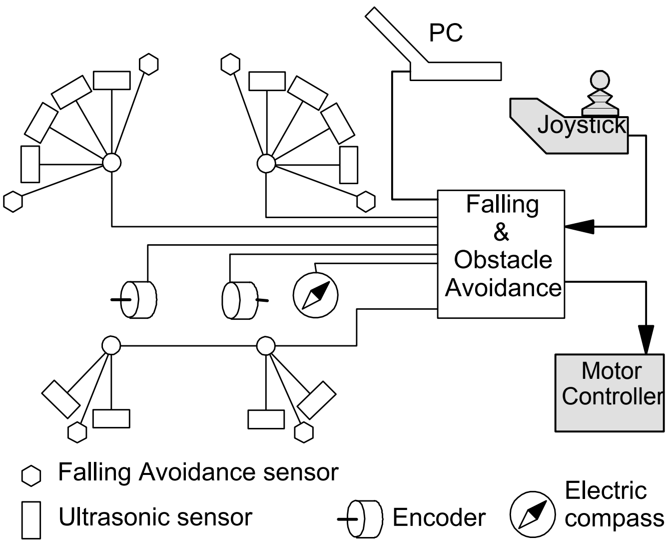


Figure 5. *All the falling avoidance, obstacle avoidance, and navigation hardware installed on a wheelchair. The navigation PC is connected via serial, bidirectional CAN bus.*

The navigation system is not based on accurate positioning or trajectories. It is based on detecting landmarks, and going from one landmark to another, just like we advise someone to e.g. a railway station: "Follow this street about a kilometer, then turn right at the lights. Follow that street about 200 meters, and turn left right after the City Hall. The station is 100 m ahead." These instructions can be elaborated to a more formal state machine representation, where one state could contain e.g. the following statement:

```
if travelled_distance = 1000m AND at_traffic_lights = true,
```

```
then Follow_current_street=off; Turn_right_90 = on
```

Since the wheelchair is going to move in a rather structured environment, all the landmarks and actions can be described beforehand, like in Table 1.

Table 1. A *structured environment usually contains only a limited number of landmarks, and actions that are appropriate in that context.*

Landmark	Action
a wall or large obstacle in front, side distance increased X cm, side distance decreased X cm, a door (when bypassing), a corridor opening (when bypassing), a pillar (when bypassing), a doorway (when entering)	go ahead, turn right, turn left, enter through doorway, follow wall, stop

The obstacle avoidance unit passes all the sensor signals to the navigation system. The navigation PC carries out pattern recognition routines to detect, and classify the landmarks. Since nothing is very precise in the real world, all the information above is handled using fuzzy logic. The pattern recognition is not always an easy task, since the wheelchair may not be moving along straight lines due to the possible obstacles on its route.

Even with fuzzy logic and other precautions, the actions made by the navigation unit would probably cause the wheelchair to hit the walls, to fail to enter doorways, etc. This is another area where the navigation systems relies on the obstacle avoidance unit. As said before, the potential field method tends to keep the wheelchair in the middle of free space, and makes entering through the doorways easier. So when the navigation unit is sending steering actions, they are sent to the obstacle avoidance unit (and these are sent to the falling avoidance unit, if it is installed), which ensures that the wheelchair adapts its actions to the real world, see Figure 6.

3. SOME EXPERIENCES

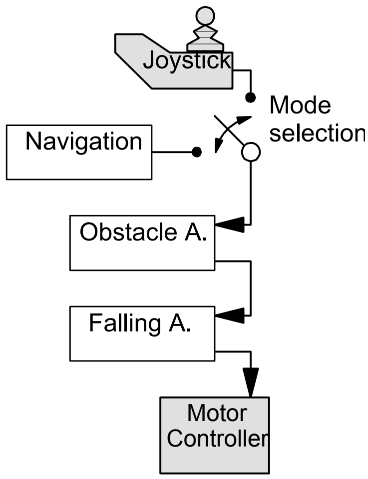


Figure 6. *The flow of the steering signals, when all the modules are installed.*

A prototype device was delivered to the Adaptation Training Centre for Disabled for evaluation about six months ago. It did not have any navigation system, and due to assembly reasons the falling avoidance system was ‘watching’ only backwards. The experiences have been the following:

As with any prototype also this one has not been working properly all the time. Since the evaluation takes place in another city, even very simple malfunctions have caused problems, since repairing them is not possible for others than the ones that have developed the prototype. Most of the problems may be due to the contact problems in the D-connectors used for sensors.

Ultrasonic sensors are not perfect sensors. They may fail in four different ways:

– The obstacle is too small to create any echo strong enough to be detected; e.g. thin wires.

– The shape of the obstacle does not reflect the echo back to the sensors; e.g. a glass door from inclined angle.
– The obstacle absorbs the ultrasonic pulses and only a weak echo is generated; e.g. thick fur coats.
– The obstacle is not within the ultrasonic beam; e.g. an obstacle at the level of the wheelchair pilots face, when the sensors are placed near the floor.

Severely disabled are the main user group. They need very individual seat arrangements, and therefore the sensors should be easily repositioned. This is especially important if the wheelchair is going to be used by an institution, where the user changes frequently.

Many of the disabled who have been involved in the project would like to get this kind of devices to their wheelchair. Naturally they also have certain prejudice towards technical devices. This is why this kind of devices should operate with 101% reliability all the time in all situations. Also the technical support and service must be of high quality. Further development should emphasise the applications for children.

Technology is not the only thing to be developed; also the assisting personnel needs to be trained how to utilise new technology. However, part of the assisting personnel in nurse homes etc. are not technically inclined, and they may consider new devices as a nuisance despite of what the patients think of them. This could even introduce confrontation, if the patient uses his/her improved independence to disagree with personnel.

In all cases this kind of moving aids would provide their users considerably more independence, and experiences that they have succeeded in something they previously felt difficult. Taking care of their own things using their own schedules is an essential part of independence.

4. VR APPLICATIONS

Gundersen et al. (1996) described a wheelchair, where a remote ‘helper’ operator could assist the wheelchair to enter difficult places, like narrow doorways or crowded elevators. Here the suggested approach is different, and is based on providing simple environment information to the wheelchair user to assist her/him to move around more independently.

If we assume that a the described navigation system is widely used, then there are also corresponding maps available. This kind of simplified but structured information itself could be utilised with different VR -interfaces for e.g. training etc. But a wheelchair with prescribed attachment units can combine the map, and real environment information. This could be very useful for persons with degraded cognitive capabilities, since both information formats are very simple. The map could tell that the user is in a lobby, and that the elevators are on the left while the environment information could tell things like “a doorway on the left”, and “obstacle on the right”. This kind of information requires very low bandwidth, and is therefore easier to communicate than e.g. video frames. This approach could provide a simple augmented reality description of the environment to the wheelchair user, and thanks to the simplicity it would be an affordable device. The interface has to be selected according to the functional senses the user has available.

This kind of information could benefit the user in three ways. First it enables her/him to gain a better understanding of hers/his whereabouts, secondly it will activate hers/his mental activity, and last but not least it will make her/him feel more safe. All these are important elements of independence.

The key requirement is that this kind of maps are easily available. The only way this could happen is, that they become part of the information infrastructure of public buildings, which means that they have to be beneficial to all of us - this kind of a system is not going to be established widely only for special groups. The strong investment into telematics research in EU raises hopes that something similar may come up for mass market. There are already experimental telematics services, where a cellular phone user can get his trip planned and guided (text format) automatically from his present location to the destination using all the public multimodal transport services available. Similar services would be welcome to large airports and railway stations as well, and this is where this kind of map/navigation service could gain its commercial feasibility.

5. CONCLUSIONS

This paper presents a set of auxiliary mobility aid device prototypes for electric wheelchairs, and the experiences gained during an evaluation period. The devices may increase the independence and safety of elderly people, and especially those with severe disabilities. The devices combine map, and environmental information which - besides being used for navigation, etc - could also be provided to the wheelchair user via an appropriate interface, if he/she has cognitive difficulties. However, the best way to achieve true progress in the field of assistive technology is that the solutions can be utilised by all of us in a commercially feasible way. Here the hopes lie on the telematic applications for travellers, and on the infrastructure of wireless communications.

Acknowledgements: The authors would like to acknowledge the support of the TIDE programme of EU, and the Adaptive and Intelligent Systems Applications programme of the Technology Development Centre (TEKES).

6. REFERENCES

J Borenstein and Y. Koren (1991a), Potential Field Methods and Their Inherent Limitations for Mobile Robot Navigation, *Proc. 1991 IEEE Conf. on Robotics and Automation*, Sacramento, pp. 1398-1404.

J Borenstein and Y. Koren (1991b), The Vector Field Histogram - Fast Obstacle Avoidance for Mobile Robots. *IEEE Trans. on Robotics and Automation*, **7**, 11, pp. 278 - 288.

R. Gundersen, S. Smith and B. Abbott (1996), Applications of virtual reality technology to wheelchair remote steering systems, *Proc. 1st ECDVRAT*, Maidenhead, pp. 47-56.

Development and evaluation of a virtual reality based training system for disabled children

Manuel Desbonnet[1], Sara L Cox[2] and Abdur Rahman[3]

[1,2]Mechanical & Aeronautical Engineering Dept.,
[3]Electronic & Computer Engineering Dept.,
University of Limerick, Limerick, IRELAND.

[1]*manuel.desbonnet@ul.ie*, [2]*sara.cox@ul.ie*, [3]*abdur.rahman@ul.ie*

ABSTRACT

Children need mobility for normal development. An electric wheelchair can provide mobility for severely disabled children, but this requires training, which may be difficult. This virtual reality based training system is aimed at solving these difficulties in a practical, cost-effective way. The project involved the construction of two virtual environments for training these children. This was achieved by developing a software solution using WorldToolKit and AutoCAD.

1. INTRODUCTION

Children need mobility for normal development, and the absence of this mobility can cause a cycle of deprivation. Bertenthal et al. (1984) found that self-locomotion is very important in the development of a child's perception of spatial orientation, social communication and concept formation. For severely disabled children, an electric wheelchair can provide a substitute for "normal" mobility, however, there are some difficulties associated with training children to use them.

Before they are one year old, children begin to interact with their environment, and influence it. This was called the "contingency experience" by Brinker and Lewis (1982) who stated that a lack of this experience can lead to a "learned helplessness" whereby the child has lost motivation and is no longer interested in controlling or interacting with their environment. Brinker and Lewis stated that one of the reasons a child may miss out on the contingency experience is a lack of locomotive ability, and hence an inability to interact with their environment. Some disabled children are denied these stimuli and this can lead to a cycle of deprivation (Nisbet 1996, Figure 1). Verburg et al. (1984) have shown that for disabled children, mobility using powered wheelchairs can replace some of this lost developmental experience.

Traditionally, there has not been a structured system for training a child to use a powered wheelchair. The Occupational Therapist at St. Gabriel's School for disabled children, stated that the children were placed in their powered chairs without any guidance and simply encouraged to explore the environment (Personal Communication). However, Furumasu et al. (1995) have been working on the development of training programmes to standardise the training procedure.

Ultimately, using a 'real' electric wheelchair is undeniably the most effective way of training children to use their wheelchairs. There are, however, several problems associated with this:

- Potential damage. Electric wheelchairs are heavy and can move at speeds of up to 3mph. A poorly controlled chair can cause damage to the furniture in a home or injury to the child or other persons.
- Frustration. Initially, children learning to use electric wheelchairs are often frustrated, as they do not have adequate control over the chair, and bump into walls and other obstacles.
- Motivation. It can sometimes be difficult to motivate a child to use an electric wheelchair, particularly if they find the task difficult or frustrating.
- Space. A large area is required to enable the child to learn how the wheelchair responds to their control inputs with minimum risk of injury through collision. While a child is learning, she/he must be closely supervised.

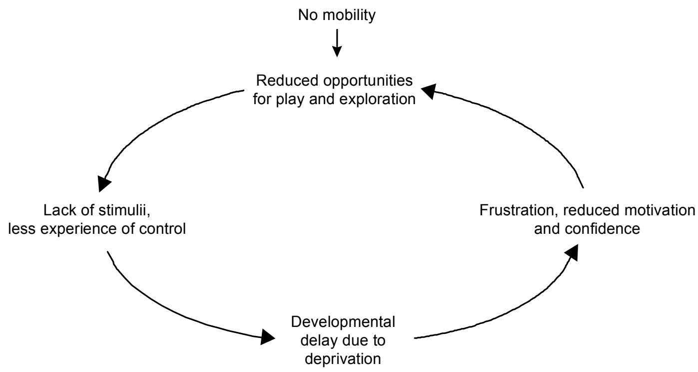


Figure 1 *Cycle of Deprivation (Nisbet 96)*

The use of a virtual reality simulator can address these problems in a number of ways.

- There is no danger to the child or the surroundings while using the simulator since they are not actually moving.
- When starting on the simulator the child may initially be placed in a large specially designed 'practice room' which will provide motion cues, but without obstacles, so that they can become accustomed to the response of the wheelchair to their hand/arm movements.
- While the simulator will mimic the motion of an electric wheelchair, the environment will have a game-like feel, making the simulator more enjoyable to use. The child can be given a task or a goal to instil a sense of purpose in their movement through the environment.

The VR simulator approach has also been developed by a group in the U.S. under Dean Inman at the Oregon Research Institute (Inman et al. 1995). However, there are differences between their Virtual Mobility Trainer (VMT) and our Virtual Reality Training System (VRTS):

The VMT wheelchair platform incorporated shaft encoders to physically measure the motions of the wheelchair. The CPU used these measurements to update the wheelchair's location in the virtual environment. The VRTS uses an interface system to directly read the joystick speed direction variables from the wheelchair control system. The VMT platform did not incorporate any motion feedback, whereas, in the VRTS simulator, it is ultimately intended that the wheelchair will be mounted on a platform that can be tilted, providing valuable motion cues to the user.

The VMT system used a Head Mounted Display to provide the visual feedback. The VRTS uses the more simple 'Window on World' display where the user views the virtual environment through a standard PC monitor. Although using this approach results in the loss of a certain degree of realism, this decision was taken to avoid the potential negative effects of using immersive VR systems (Wilson, 1996). It also has the positive effect of making the system less complex and more cost effective.

2. SIMULATOR DEVELOPMENT

At the start of this project two very important areas were researched.

1. Research into the childrens' disabilities, how children are currently trained, and how this might be improved using the simulator.
2. Development of the simulator software and wheelchair interface.

These two areas are closely linked. The limits of the children's' abilities will affect the development of the software, the introduction of new concepts or removal of non-useful ones. Identifying criteria that will give the best training experience will involve several iterations over the design loop.

Initial research has identified and assessed the problems of both disabled children learning to use electric wheelchairs and their therapists while training them. St. Gabriel's School, Limerick, a school and centre for disabled children, was visited where training methods were observed. Children trying to control an electric wheelchair for the first time, were recorded photographically. This allowed personal experience and observation of some of the problems described above.

To give the children a broad range of training scenarios, it was aimed to develop three different virtual environments:

- A domestic setting. Mastering this environment would help the child cope with the necessary task of getting about in their own homes. In this environment, the child can learn all the basic tasks required to control a powered wheelchair: starting/stopping, negotiating doors and corridors.
- A road crossing situation. This would simulate a potentially dangerous scenario involving moving objects.
- A shopping centre. This environment would introduce the child to more difficult situations - those which involve places/buildings which may not cater as effectively for their disabilities. The introduction of other mobile obstacles (people and mobile equipment) make this a more challenging setting.

Due to the limited processing power of the computer system, only the first two of the above environments have been created, and most of the emphasis was put on the home environment since this can be used to learn the basics of wheelchair control.

2.1 Software

The software used to execute the virtual environment is the WorldToolKit (WTK) from Sense8 Corporation, which is a C language library of functions with a built-in simulation manager. The virtual environment is run on a computer system based on a 200 MHz Pentium Pro processor.

The WorldToolKit library implements functions for the creation and manipulation of a virtual environment in which the real-time simulation manager presents the virtual environment to the user. The toolkit manages the tasks of rendering, reading the input devices, importing the model geometry and performed simulation functions (such as moving cars) which simplifies the development of a virtual environment.

Creating the virtual environment involved writing program code for the WorldToolKit and the construction of a three dimensional model which the WorldToolKit displayed. This required a CAD(Computer Aided Design) program with three dimensional capabilities. AutoCAD from Autodesk was used for this operation.

AutoCAD provides a powerful and flexible development environment. It has a wide variety of commands for creating 3D meshes, and is ideal for creating 'square' objects such as buildings and rooms. However (at release 13) it is not possible to completely specify the models with AutoCAD unless some customisation software is written. For example textures and arbitrary RGB colours can not be applied in a manner that WTK can understand (WTK can import DXF files and allows for texture application through the AutoCAD layer mechanism, but this is fairly crude and not feasible for large complicated meshes).

This difficulty was overcome by using AutoCAD's built in programming language, AutoLisp, to write a custom software extension which allowed for storing extra data with the objects in AutoCAD. In this way colour and texture information could be applied. A further extension allowed for the application of basic tasks such as translation, rotation, and scaling. For example the rotation task was used extensively to allow doors to open when the user approached them. These features facilitate the development of future environments since they can be nearly completely described inside AutoCAD, and with a minimum of extra WTK code development.

All this extra data was stored as extended entity data and written out as a DXF file. A C++ program was written to read in a DXF file including extended entity data, process it, and write an output file in the WorldToolKit's own model file format. Figure 2 shows the VRTS with it's various components and how they interact with each other.

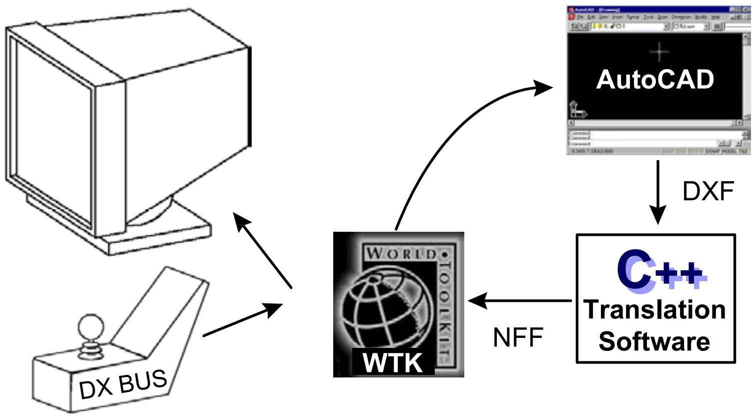


Figure 2 *VRTS setup*

In creating the virtual environments, there is a balance to be achieved between having realistic looking sophisticated environments, and the speed with which the simulator can run these environments. A minimum frame rate of about 10-12 frames per second(fps) is recommended by Bryson (1992), and Wilson (1996). If the frame rate falls below 6 fps, the simulation becomes jerky and difficult to use, especially for some disabled children who are visually impaired. During the development of the home environment, the emphasis was initially placed on maximising the realism, making much use of texture application. However, it was found that this reduced the frame rate to unacceptable levels (< 5 fps). Removing many of the textures, especially large textures that were used to give a view outside the house, dramatically increased the simulation speed with very little detraction from the visual appeal of the environment.

2.2 Wheelchair Interface

The interface between the virtual environment and the user is an important component of this simulator. To optimise the usefulness of the simulator as a training tool, it was designed to interface directly with the controller on the child's wheelchair. Ideally it should be possible to interface to a wide variety of different makes of controllers on different wheelchairs, however, this requires an open standard interface specification, and although such a standard does exist (M3S 1998), it is not yet widely used in wheelchairs in Ireland.

St. Gabriel's School, whose pupils are collaborating in the evaluation of the simulator, have decided that all new electric wheelchairs that they acquire will be fitted with the DX system produced by Control Dynamics Ltd. Therefore, it was decided to tailor the VRTS to this wheelchair control system. The DX system is also suitable because it is a sophisticated system consisting of various modules communicating with each other at a high level over the DX bus which is based on the CAN communications protocol (Bosch 1991). A DX bus module, the DX key was purchased from Control Dynamics Ltd., and this provides a direct interface between a personal computer and the DX bus system.

This DX key interface enables the user of the simulator to navigate within the virtual environment using the controller fitted to their own chair. This controller is usually a joystick, but may be one of a range of other devices such as push button switches, position sensing devices or sip-puff switches. This is transparent to the simulator since the DX key monitors the speed and direction requirements sent to the power module, and these are independent of the modules which generate these values. Other features of the DX bus controllers can also be integrated into the simulator, for example, lights and environmental controls allowing the user a greater degree of interaction with the environment.

3. DISCUSSION

A preliminary evaluation has been performed, and the simulator system running the home environment was demonstrated at St. Gabriel's School to the occupational therapist, and four disabled children. Three of the children had previous experience using powered wheelchairs, and successfully maneuvered their wheelchair around the virtual environment. The fourth child had no previous experience with powered wheelchairs, and could not navigate independently in the environment. She demonstrated poor grasp of the concepts required to control the wheelchair, i.e. she would drive into a wall and not stop. She did not explore of her own volition, and only went where she was told.

The feedback from this evaluation was positive, and all four children enjoyed using the simulator, one boy was very enthusiastic. The therapists suggested improvements to the simulator that would make it more useful and accessible for the children. These were as follows:

1 More sound. Sound is a very important aspect of simulation since some of the children have visual impairments. Sound also makes the simulator much more exciting and novel.

2 More interaction. Introduce more sophisticated dynamic objects such as falling objects: if the chair hits a table, then have a cup or a plate falls off and breaks. Hide and seek - hide an object in the house and have the child go and search for it.

3 Develop the outside of the house, have paths leading from front door around to the back. The occupational therapist reported that in her experience children had greater difficulty manoeuvring outside than inside a building possibly due to the lack of reference data points.

The simulator was improved and some of the suggestions above were implemented. A second more thorough evaluation was then undertaken with five disabled children who had not taken part in the previous evaluation. Figure 3 shows the simulator being used by one of the children. The results of this evaluation showed that four of the children were able to use the simulator with degrees of ability varying from very good control to only gross directional ability. Only one child required constant help in using the simulator.

A significant point that arose in both trial runs was that children's ability to use their powered wheelchairs was reflected in their competence in the use of the simulator. For example one girl who had mediocre control in the simulator, and found especially difficult to go through doors, also had the same problem in reality with her own chair (Occupation Therapist, personal communication).

Figure 3 *Wheelchair simulator in use*

4. CONCLUSIONS

The evaluations of the simulator have shown that the system can play a useful role in training disabled children to use powered wheelchairs. The fact that children who could use their powered chairs could also manoeuvre successfully in the simulator shows that the virtual wheelchair bears a reasonable resemblance in look and feel to a real chair. However, the current system is of limited benefit as a training tool since due to the limited processing power of the simulator computer, there is limited visual realism, and relatively crude modelling of the wheelchair (i.e. the virtual wheelchair does not behave in the simulation exactly like a real wheelchair).

Future development of the system would involve the up-grading of the computer hardware and the construction of a platform on which the child's wheelchair could be mounted. This will enhance the system by providing motion feedback, which will greatly reinforce the visual cues received from the monitor.

Acknowledgements: We are grateful to the University of Limerick Foundation for funding this project and to Maura O'Leary and the children of St. Gabriel's school for their co-operation in the evaluation of the VRTS.

5. REFERENCES

B I Bertenthal, J J Campos, K C Barrett (1984), Self-produced locomotion: An organizer of emotional, cognitive, and social development in infancy. *Continuities and discontinuities in development*, pp. 175-209.

R P Brinker, M Lewis (1982), Making the world work with microcomputers, *Exceptional Children*, **49**, 2.

J Furumasu, P Guerette, D Tefft (1996), The development of a powered mobility program for young children. *Technology and Disability*, **5**, pp. 41-48.

D P Inman, K Loge (1995), Demonstrating and Evaluating a Virtual Reality Training Program for Teaching Mobility Skills to Orthopedically Impaired Children, *3rd Annual International Conference on Virtual Reality and Persons With Disabilities*, San Francisco.

G Verburg, E Snell, M Pilkington, M Milner (1984), Effects of powered mobility on young handicapped children and their families. *Proceedings of 2nd International Conference on Rehabilitation Engineering*, Ottawa.

J R Wilson (1996), Effects of participating in virtual environments: A review of current knowledge. *Safety science*, **23**, 1, pp. 39-51.

S. Bryson (1992), Virtual Environments in Scientific Visualization, *Proceedings of 2nd International Conference on Artificial Reality and Tele-existence*, Tokyo.

P. Nisbet, J. Craig, P. Odor, S. Aitken (1986), 'Smart' wheelchairs for mobility training, *Technology and Disability*, **5**, pp 49-62.

M3S(1998), M3S Dissemination office, http://www.tno.nl/m3s/

Bosch (1991), Bosch CAN Specification, http://mot-sps.com/csic/techdata/refman/can2spec.pdf

Neural-network virtual reality mobility aid for the severely visually impaired

M R Everingham[1], B T Thomas[1], T Troscianko[2] and D Easty[3]

[1]Advanced Computing Research Centre,
[2]Perceptual Systems Research Centre,
[3]Department of Ophthalmology,
University of Bristol, Bristol BS8 1UB, UK

[1]*everingm@acrc.bris.ac.uk*

ABSTRACT

This paper describes a new approach to image enhancement for people with severe visual impairments to enable mobility in an urban environment. A neural-network classifier is used to identify objects in a scene so that image content specifically important for mobility may be made more visible. Enhanced images are displayed to the user using a high saturation colour scheme where each type of object has a different colour, resulting in images which are highly visible and easy to interpret. The object classifier achieves a level of accuracy over 90%. Results from a pilot study conducted using people with a range of visual impairments are presented in which performance on a difficult mobility-related task was improved by over 100% using the system.

1. INTRODUCTION

Many people who are registered as blind nevertheless retain some residual vision, and are said to have "low vision". Examples of conditions resulting in low vision are cataracts, diabetic retinopathy, age-related maculopathy, and retinal detachment. The precise form of visual impairment varies according to the particular medical condition it results from, but often a person with low vision experiences an extreme loss of perception of high frequencies resulting in gross blurring of the visual scene over a significant area of their field of view. The visual impairment is typically such that the person is unable to be mobile without some form of assistance such as a guide dog. In recent years an increasing amount of research has attempted to apply techniques from image processing and computer vision to the needs of people with low vision, with most of this work concentrating on improving the visibility of higher spatial frequencies in a scene. A common factor in these techniques is that since they work with no knowledge of the semantic content of an image, they tend to enhance noise and irrelevant detail in a scene equally to important visual information, resulting in cluttered images which are difficult to interpret. The Computer Science Department at the University of Bristol, in conjunction with the Psychology Department and Bristol Eye Hospital, is developing a mobility aid for low vision which uses artificial intelligence techniques to recognise objects in a scene so that visual enhancement may be carried out taking into account which components of a scene are important and which may be considered irrelevant detail or noise. This paper describes how such content-driven enhancement may be carried out, and presents results of a pilot study in which the concept was evaluated by a number of people with a variety of low vision conditions.

1.1. Structure of Paper

The outline of this paper is as follows. Section 2 reviews previous work in applying computer vision techniques to the needs of people with low vision. Section 3 describes in detail our content-driven approach and demonstrates how results may be expected to be superior to methods using only image processing. Section 4 presents results of the performance of our system in technical terms and the results of a pilot study evaluating the concept of the system. Finally, in section 5 some conclusions are made and scope for further work discussed.

2. BACKGROUND

Over the last fifteen years an increasing amount of research has attempted to apply techniques from computer vision to the needs of people with low vision. Peli and co-workers have been working on improving vision in cases of loss of contrast sensitivity by contrast enhancement in specific spatial frequency bands important to tasks such as face recognition (Peli et al, 1994). The development of head-mounted display units has led to some research in application of these for use by people with low vision. Massof and Rickman (Massof and Rickman, 1992) have produced a head-mounted system which provides magnification and real-time contrast enhancement of selected spatial frequencies using a single camera and dual cathode-ray-tube (CRT) displays. Development of this device was continued in conjunction with NASA (NASA, 1993). Peli (Peli, 1995) has implemented a device which uses a single display unit to display binary images with adjustable brightness and contrast enhancement of high 1-dimensional spatial frequencies, and a 2x zoom facility. Goodrich and Zwern (Goodrich and Zwern, 1995) have used a commercially available colour head-mounted display to provide variable contrast and magnification for people with low vision.

All of these devices work by amplifying the contrast of an image in specific frequency bands in order to make important image features, typically edges, more easily visible. Such operations are easy to achieve using video electronics which can currently be implemented in real-time, but this approach has fundamental limitations. As an example, Figure 1 shows an image of a typical urban scene, and the same image enhanced by the adaptive filtering technique described in Peli (1991), which amplifies the contrast of high spatial frequencies using an adaptive method to adjust to varying low frequency luminance across an image. As one can see, the technique does something to improve the visibility of important edges in the scene such as the outlines of pavement and cars, but this is at the expense of introducing much noise by the amplification of textural properties of the image which are not necessary for interpreting the scene, for example the cloud texture in the sky, and the noisy road appearance. This occurs because the technique, in common with all others using basic image processing, is unable to differentiate between image content which is semantically important or merely noise or textural detail. Such techniques are in general unable to perform specific enhancement of important high-level scene properties, for example dangerous objects such as vehicles or important visual properties such as the boundary between road and pavement.

Figure 1. *(a) Original image, (b) Enhanced by Peli method*

Other recent techniques have not attempted to retain all information in the original scene, but concentrated on detecting only objects which may be hazards or of particular importance for mobility. For example, Molton et al (Molton et al, 1998) are working on a system using stereo vision to find obstacles in the path of a mobile user. Such systems might be seen as being an accessory to people with low vision rather than a complete mobility solution, and are not discussed further here.

3. CONTENT-DRIVEN IMAGE ENHANCEMENT

3.1. Approach

The approach being developed at Bristol differs from other techniques in that it attempts to identify the semantic content of an image such that images can be enhanced in a content-driven manner, taking into account which features of an image are important, and which may be disregarded as noise or unnecessary detail. In our approach, image classification forms the basis for image enhancement. In our enhanced images, each object in the scene is coloured using a solid high saturation colour corresponding to the type of object, for example the road is black, pavement white, vehicles bright yellow and so on. In current work we use a scheme whereby all objects are classified as one of a set of eight object classes, so for example the “obstacle” class contains such things as bounding walls, lamp-posts, pillar-boxes etc. We judge that it is easy to distinguish eight separate colours when they are high saturation and high contrast, and that the information present in the scene using this set of object classes is sufficient for mobility in a typical urban environment. Figure 2 shows as an example the image from Figure 1 and its appearance when enhanced by our method.

Due to lack of colour printing facilities, figures have been reproduced in monochrome here and it should be noted therefore that the contrast between object types in these images is lower than in the original colour images.

Figure 2. *(a) Original image, (b) Enhanced by Bristol method*

There are many advantages to using a scheme such as this which uses object classification as the basis for image enhancement. In our enhanced images, the colour of an object is independent of it's original visual properties which may make the object less visible, so that here for example cars are coloured bright yellow regardless of their actual colour which may be hard to see. Irrelevant detail such as textural properties of an object is also removed by classification, which gives the images a very simple and uncluttered appearance which is easy to interpret. The form of output is easy to customise to a particular user's requirements; in this case we have used a predefined set of high saturation colours, but these colours may be customised by a user to improve visibility according to his or her particular visual impairment. Other modifications to the visual presentation scheme are equally easy to make, for example overlaying of edges on the image, or causing particular object types such as obstacles to flash when they are large in the scene to attract attention to them. Additionally, because our system gathers semantic information about the content of the scene, it is not even necessary to use a visual output medium – the information could instead or additionally be presented using existing touch, sound or speech output devices.

We have primarily been investigating the use of the saturated colour output scheme, and Figure 3 demonstrates the effectiveness of this scheme. Here we see the image from Figure 1 in original form, enhanced by a Peli method (Peli, 1991), and by our method. In this case, all three images have been blurred by an equal amount in order to simulate a visual impairment causing reduced perception of high spatial frequencies.

Figure 3. *(a) Blurred image, (b) Peli method, (c) Bristol method*

Even without colour reproduction, Figure 3 clearly demonstrates the effectiveness of our method. Arguably the Peli method has reduced the visibility of the scene by reduction in the overall contrast of the image, whereas in the image enhanced by our method one can clearly see the boundary between pavement and road, which is invisible in the other two images, and the three vehicles which appear as bright yellow blobs. In the original colour image, it is also possible to see by colour the two lamp-posts in the scene, which are again invisible in the other images due to their poor visibility in the original scene.

3.2. Architecture

Figure 4 shows a block diagram of our image enhancement system, which uses a neural network to perform classification, building on work in outdoor scene classification at Bristol (Campbell et al, 1997). Typically input to the system would come from a head-mounted camera, and output be displayed on a head-mounted display. The *image segmentation* stage segments an image into a number of regions which are deemed to correspond to a single object or object part. The *feature extraction* stage takes the pixels of a region and calculates a set of numeric features or "feature vector", which describes the visual properties of the region and its context in the image. The *neural net classifier* takes the feature vector of a region as input and its output corresponds to an object class to which the region belongs, for example road, vehicle etc. The final *image labeller* stage produces the coloured image in which the pixels of each region are coloured with the high saturation colour corresponding to the determined object class for the region. During a separate training phase, a *training algorithm* uses previously collected *ground truth* data in the form of hand-classified images to train the neural network. Each of these stages of processing is now discussed.

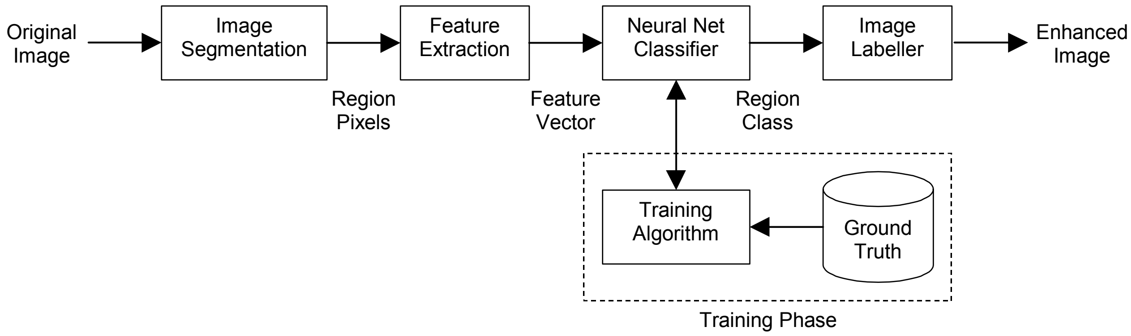


Figure 4. Architecture of Image Enhancement System

3.3. Segmentation

The aim of image segmentation is in general ill-defined, but in this context we will characterise it as being to divide an image into a set of connected regions such that each region corresponds to a single object or object part. Figure 5a shows an edge-map of an approximately ideal segmentation of the original image from Figure 1 performed by a skilled human operator. Automatic segmentation of general outdoor scenes remains an unsolved problem with much research being carried out in this field, see (Pal and Pal, 1993) for a review. Most techniques regard a region as being homogenous with respect to some visual properties such as colour or texture, and we have tried many and diverse techniques using a variety of homogeneity measures. We find that none give near-perfect results, but that good results can be achieved using a surprisingly simple technique, based only on homogeneity of grey-level luminance across a region.

Figure 5. *(a) Ideal Segmentation, (b) Automatic Segmentation*

Figure 5b shows the quality of segmentation obtainable using this technique. Segmentation is performed using the well-known K-means clustering algorithm which is an iterative clustering technique which converges to a local minimum (Selim and Ismail, 1984). We define that regions are of homogenous grey-level, and cluster all pixels in the image using grey-level alone. Individual regions are then extracted using a connected-components algorithm such that all connected pixels assigned to the same cluster belong to the same region. The K-means algorithm requires only a single parameter, the number of clusters, and this was set to 5 for all images in our work, on an empirical basis. Clustering noise is removed by merging regions with area less than 16 pixels (determined empirically) to a neighbouring region. The result is typically a moderate over-segmentation of the scene into 100-400 regions.

3.4. Feature Extraction

In order to classify regions into an object class using a neural network classifier, the raw image information in the form of pixels of a region must be reduced to a small set of numeric *features* which describe the region sufficiently well for it to be classified. This is achieved by the feature extraction stage. We have used features which are intuitively appealing, corresponding to visual concepts which we as human beings typically use to describe objects, and where possible we have used features which have some psychophysical basis. The main features we extract in this manner are Colour, Shape, and Texture, and these are described here.

3.4.1. Colour

Colour is intuitively useful for recognising objects in a scene, for example grass is green, and roads are grey. Colour can also be used in cases where it might be expected to be less useful, for example cars are produced in many colours, but typically they have in common that the colour is of high saturation. Many colour spaces exist (Foley et al, 1990) which are candidates for generating colour features. We have experimented with a number of these including RGB, YUV, HSV and L*u*v*, and found that colour spaces which separate the luminance and chromaticity components of a colour are more successful. We found the CIE L*u*v* colour space (Wyszecki, 1982) to be most successful overall, and this is an appealing result as it is a space which has a good perceptual basis, in the sense that colours which are equidistant in the space are approximately equidistant in terms of perceived difference. In order to calculate three colour features for a region, the mean RGB value of the pixels in the region is taken from the original image, and converted to L*u*v*, with the L, u and v values being used as features.

3.4.2. Colour Constancy

When using colour to recognise objects, we should be careful to note that the colour of pixels in an image *do not* correspond to the actual colour of an object, but it's perceived colour under a general unknown illuminant, for example daylight. In addition, different image capture systems produce images with different colour properties. If we want our classifier to be robust to such changes as different weather, lighting or cameras, we must attempt to recover the *actual* colour of an object, rather than its colour under a particular illuminant and using a particular image capture system. The human visual system exhibits this invariance (Wyszecki and Stiles, 1982) known as colour constancy. We attempt to achieve this property by using a colour correction technique to normalise images. The key is to approximate the colour skew caused by illumination/capture properties and remove this. Our technique is based on the "grey world" model (Buchsbaum, 1980) which determines colour skew by assuming that the mean reflectance in a scene is achromatic i.e. has equal RGB components. Given mean RGB and grey-level components of an image, each of the RGB components is scaled to make the individual means equal to the overall grey-level mean using a

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = \begin{pmatrix} I_\mu / R_\mu & 0 & 0 \\ 0 & I_\mu / G_\mu & 0 \\ 0 & 0 & I_\mu / B_\mu \end{pmatrix} \begin{pmatrix} R_u \\ G_u \\ B_u \end{pmatrix}$$

tri-diagonal matrix:- where R_c, G_c, B_c are the corrected RGB components of a pixel, R_u, G_u, B_u are the uncorrected RGB components, R_μ, G_μ, B_μ are the mean RGB components across the image, and I_μ is the mean grey-level across the image, equal to $(R_\mu + G_\mu + B_\mu)/3$. In addition to correcting colour skew in this manner we also normalise mean brightness and contrast by scaling grey-levels to full range and moving the mean to half range. By disregarding the top and bottom 5% of the grey-level histogram in this process, an approximate invariance to varying daylight conditions is obtained.

3.4.3. Shape

Shape should intuitively be a very powerful feature for discriminating object classes. We use here a shape representation which is invariant to translation, rotation, and scaling. We consider such invariance important for obtaining robustness of classification. The representation we use is a Fourier shape descriptor, which considers the outer boundary of a region as a periodic sequence of two dimensional points, where transformation to the frequency domain allows the significant properties of the shape to be characterised in a small number of values. The particular formulation of Fourier descriptor used is that of Shridhar and Badreldin (Shridhar and Badreldin, 1984), who achieved very good results for hand-written character recognition using the descriptor. Formulation of the shape features for a region proceeds in three steps:-

First, the outer boundary of a region is traced using a conventional boundary following algorithm to give an ordered, closed list of 2-D co-ordinates. Secondly, the two co-ordinate sequences corresponding to the x and y co-ordinates of the boundary points are separately transformed to the frequency domain using two 1-D Fast Fourier Transforms (FFT). Finally, the frequency domain components are transformed to give the required invariance:-

$$r(n) = \sqrt{|a(n)|^2 + |b(n)|^2}$$

where a(n) is the FFT of the x co-ordinates and b(n) is the FFT of the y co-ordinates. Discarding r(0), r(n) can be shown (Badreldin et al, 1980) to be invariant to translation and rotation. Dividing by r(1) gives scale invariance:-

$$s(n) = r(n) / r(1)$$

Using a sub-set s(2..k+2) of the k low-frequency coefficients of s(n) gives a good approximation of the shape even when k << L where L is the length of the boundary. In our work we have used 10 coefficients as shape features for a region. Figure 6 shows a region from an outdoor image and the approximation of it's outer boundary using 10 and 30 shape coefficients, showing how an increasing number of coefficients gives higher accuracy of shape approximation, particularly for objects with sharp corners.

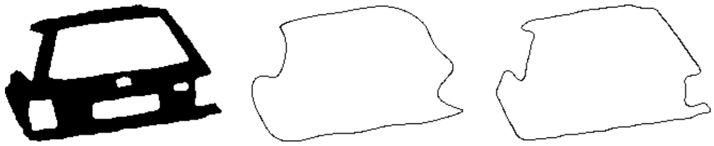


Figure 6. *(a) Region silhouette, Boundary reconstruction from (b) 10 coefficients, (c) 30 coefficients*

We can also observe from Figure 6 that, in common with most shape descriptors in use, because only the outer boundary is traced, any information about holes in the region is lost, and we judge that exactly this information may be most important for identifying objects by shape. We therefore present a novel modification to the shape descriptor which incorporates information about holes in a region. Figure 7 shows the shape approximations obtained using the new descriptor.

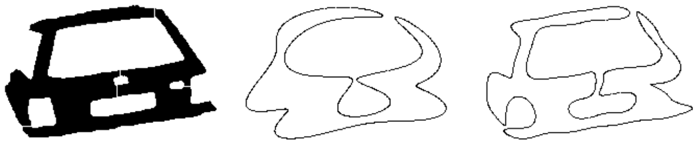


Figure 7. *(a) Cut Region Silhouette, Boundary reconstruction from (b) 10 coefficients, (c) 30 coefficients*

This novel shape descriptor retains information about holes in the region by first cutting the region silhouette to link internal holes to another hole or to the area outside the region so as to form a *single* boundary which traverses both the original outer boundary and the boundaries of the holes in the region. The cuts made are chosen to be minimal in length to obtain stability of the scheme.

3.4.4. Texture

Texture is also intuitively useful for discriminating between objects, for example brick walls have a highly periodic and directional texture where grass has a higher frequency more random texture. To obtain texture features for a region we use Gabor filters, which are a particular type of filter tuned to respond to a specified range of spatial frequencies and orientations. Gabor filters have been to shown to have similar characteristics to simple visual cortical cells (Marcelja, 1980) and therefore sound psychophysical basis. Figure 8a shows an example of the real part of a complex-valued Gabor filter in the spatial domain; the imaginary part is equal but subject to a 90° phase shift. In the spatial domain, the filter has the form of a complex sinusoid modulated with a Gaussian oriented along the direction of the filter, while in the spatial frequency domain it is an orientated Gaussian. Using a complex-valued Gabor filter, the magnitude response is constant given an input sinusoid image with frequency and orientation matching those of the filter. Jain and Farrokhnia (1991), and Dunn and Higgins (1995) provide good overviews of the mathematics of Gabor filters.

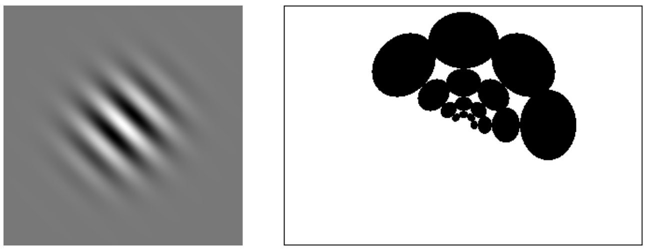


Figure 8. *(a) Gabor filter in Spatial Domain (real part), (b) Gabor array in spatial-frequency domain (half peak)*

Using a bank of n filters and taking magnitude output gives a vector of n real values characterising the texture content of an image at a particular pixel. We use the scheme of Jain and Farrokhnia (Jain and Farrokhnia, 1991) to determine the properties of our filter bank, in which the centre frequencies of the filters are in a power of two series, with 45° orientation step, and the frequency and orientation bandwidths are set to 1 octave and 45° respectively. Figure 9b shows this partitioning of the spatial-frequency domain, in which the half-peak magnitude responses of the filters are shown, with a total of 16 filters being used; the asymmetry in the spatial frequency domain is caused by the use of complex-valued filters. The magnitude response of a Gabor filter across a region of constant texture can be well characterised as a Rician distribution (Dunn and Higgins, 1995), but as estimation of the parameters of this distribution requires the costly solution of non-linear equations, we instead use a Gaussian distribution, which has been shown to be adequate in most cases (Dunn and Higgins, 1995). Thus to form our region texture features, we take the mean and standard deviation of each Gabor filter output across a region, giving 32 features in all. In practice however, we find that using means alone gives almost as good results, so we use a reduced set of 16 texture features formed by the mean magnitude responses.

3.4.5. Combined Feature Set

Combining colour, shape and texture features with other simple features describing a region's position, size and orientation, we arrive at the complete set of 35 features shown in Table 1.

Table 1. *Feature Set*

Feature	Description
1	Size (proportion of image)
2-3	Position (scaled x, y co-ordinates)
4-5	Orientation (sin & cosine of angle)
6-8	Colour (mean L*u*v*)
9-18	Shape (invariant Fourier coefficients)
19-35	Texture (mean Gabor magnitude)

3.5. Neural Network Classifier

A multi-layer perceptron (see Bishop, 1995) with one hidden layer is used as a classifier, with a 35-n-8 architecture, and with hidden and output units having logistic sigmoid activation functions. The 35 inputs correspond to the feature vector, and each of the 8 outputs indicates membership of a particular object class. A one-of-n coding scheme is used to encode membership of each class, and a winner-takes-all scheme is used to determine the object class chosen by the network. Inputs were normalised to have zero mean and unit variance across the entire training and test data sets. In our work we experimented with a number of hidden units between 5 and 40, finding 23 hidden units to be optimal.

3.5.1. Training Procedure

The neural network classifier is trained using a back-propagation with momentum algorithm (Bishop, 1995). More elaborate techniques such as Scaled Conjugate Gradient (SCG) were tried but were found to give no better results on this problem at the expense of increased training time. Training data was taken from two sources, primarily the Bristol Image Database (Mackeown, 1994). This is a database of 200 high-quality outdoor suburban and rural scenes which have been hand-classified by a skilled human operator. A second database, the Bristol Blind Mobility Database, containing 10 scenes captured in more challenging urban situations was used in addition. The data was split randomly into 70% training regions and 30% testing regions.

4. RESULTS

4.1. Classifier Performance

Table 2 shows the classification performances achieved by our neural network classifier. Performance is quoted for the two image databases as percentage correct using two metrics – percentage correct by regions, and more meaningfully, percentage correct by area. In addition, the maximum *a priori* (MAP) probabilities of correct classification are shown with the corresponding most likely class. This information represents the maximum performance possible by a Bayesian classifier with no *a posteriori* information i.e. no feature data.

Table 2. *Neural Network Classifier Performance*

Database	% by Region	% by Area	MAP % by Region
Bristol Image	79.1	92.5	31.8 vegetation
Blind Mobility	70.1	80.4	36.7 vehicle

We see that overall the performance of the classifier is very high, with up to 92.5% of image area correctly classified using images from the Bristol Image Database, and 80.4% using the Bristol Blind Mobility Database. Figure 10 shows this level of performance qualitatively, which can be seen to be very good. The lower performance on the Bristol Blind Mobility database may be explained by the higher complexity of the scenes in this database, and the low overlap with the Bristol Image Database images which dominate the neural network training. This performance should be improved by collecting more training data.

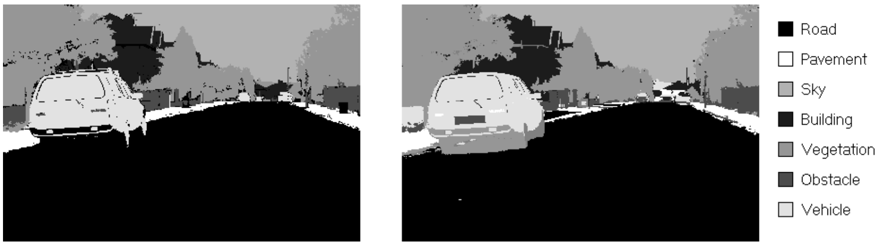


Figure 10. *(a)Ideal Classification, (b) Neural Network Classification*

4.1.1. Colour Constancy

Our technique for achieving colour constancy proves to be very effective. Figure 11 demonstrates qualitatively the results of using uncorrected or corrected colour features. Figure 11a shows the neural network classification of an image from the Bristol Image Database. The colour content of the image was then modified to give a very slight yellow tint, approximating the type of images obtained from our camcorder. Figure 11b shows the classification of the modified image by a network trained using uncorrected colour features, and Figure 11c shows the classification using corrected colour features. As can clearly be seen, the network trained using uncorrected colour features fails completely on most object classes due to the shift in the colour space which was not present during training, whereas the network trained using corrected colour features obtains a classification approximately as accurate as using the unmodified image.

Figure 11. *(a) Classification, (b) Uncorrected Classification, (c) Corrected Classification*

4.1.2. Processing Speed

As we are aiming eventually at a real-time implementation of our system, processing speed is important. At present, because of the use of Gabor filters for texture description, which are computationally expensive, performance does not approach real-time, with each image requiring 9.5s total processing time on a 300MHz Intel Pentium-II processor. However, discarding texture features we are able to achieve a classification accuracy of up to 86.5% by area in a processing time of just 300ms. We anticipate that with optimisation and current trends in increased processing speed, we may achieve true real-time performance in the near future.

4.2. Pilot Study

In order to test the effectiveness of our technique for enhancing images by classification, a pilot study was conducted in the Bristol Eye Hospital. 16 registered-blind subjects with a variety of visual impairments including age related macular degeneration, retinitis pigmentosa and optic atrophy participated in the experiment, ranging in age from 38 to 87 years with mean age 69 years. A repeated measures design was used with the subjects being shown a total of 45 images in each of three conditions – original images, image enhanced by a Peli technique (Peli, 1991) and images enhanced by our classification technique. The set of 45 images was divided into three groups of 15, and for each group a different task was set – pointing to at least two obstacles such as lamp-posts, pointing to at least two vehicles, and tracing the line between both pavements and the road. The order of the tasks was counterbalanced using a Latin Square technique. Images were displayed on a computer monitor with 50cm diagonal screen, at a distance subtending 54.6° of the subject's visual field. To avoid effects due to difference in luminance between the image types, all images were normalised to have a mean luminance of 19.29cdm^{-2}. Percentage correct scores were recorded for each patient under each condition of image type and task.

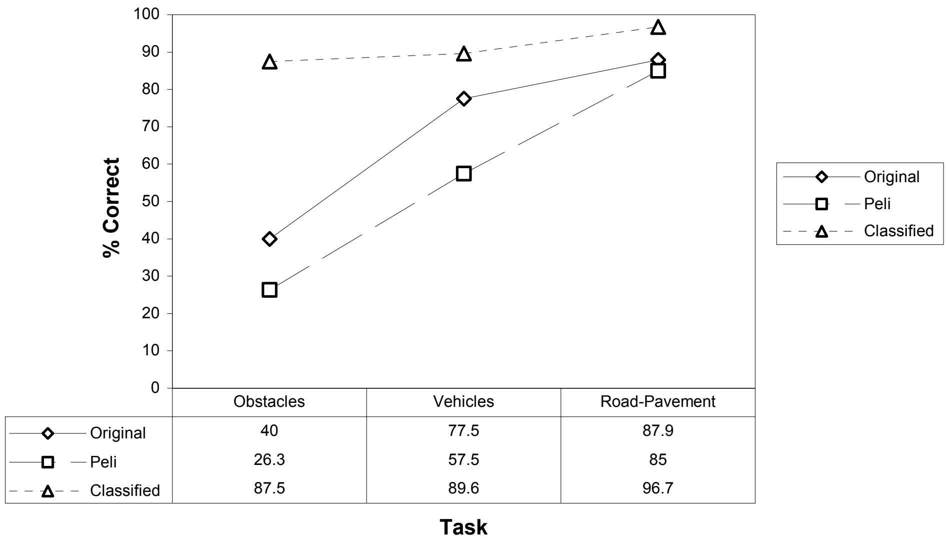


Figure 12. *Mean % correct by Image Type and Task*

Figure 12 summarises the main results of the experiment. A 2-way repeated measures ANOVA was performed on the individual percentage correct scores, using Image Type and Task as two factors, each with three levels. Two significant ($p < 0.001$) main effects were found. The main effect on Image Type indicates that the type of images did change performance on the tasks. The main effect on Task indicates that the tasks varied in difficulty. A significant interaction was found between Image Type and Task indicating that the type of image used improves performance on the different tasks to varying degrees. Post-hoc analysis was performed using the Newman-Keuls technique and reveals the following significant ($p < 0.001$) results.

Mean performance of subjects is significantly better using our classification technique than using either original images or images enhanced by the Peli technique. On the most difficult task, recognition of obstacles, mean performance across subjects increased from 40% to 87.5% using our technique. Peli-enhanced images gave significantly worse performance on the first two tasks than the original images. There is also no statistically significant difference between performance over the three different tasks using our technique, indicating an improvement in performance on the tasks proportional to their difficulty.

In summary, our technique improved overall performance significantly and resulted in consistent high performance across tasks of varying difficulty. On the most difficult task, obstacle recognition, performance increased by over 100%.

5. CONCLUSIONS

This paper has described a novel content-driven approach to image enhancement for people with low vision in the context of mobility in an urban environment. The classifier used as the basic model for image enhancement achieves a very high level of accuracy of up to 92.5% by image area. A pilot study has demonstrated the effectiveness of the technique, which improved object recognition performance by people with a range of visual impairments on a difficult mobility-related task by over 100%, and resulted in consistent very high performance over three tasks important for mobility. In future work, we aim to improve the classifier accuracy and it's robustness to varying environmental conditions and situations. We also aim to improve the speed of the system so that it might be implemented to run in true real-time. A further study is planned to assess tolerance to region misclassifications in a mobility task.

Acknowledgements: This work is supported by the National Eye Research Centre. We thank British Aerospace PLC, Sowerby Research Centre, Bristol for their support with the development of the Bristol Image Database. We also thank Nital Karia and Manoj Kulshreshta for collection of data for the pilot study.

©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

6. REFERENCES

A Badreldin, A K C Wong, T Prasad, M Ismail (1980), “Shape descriptors for N-Dimensional curves and trajectories”, *IEEE Proceedings on Cybernetics and Society*, pp. 713-717.

C M Bishop (1995), *Neural Networks for Pattern Recognition*, Oxford University Press.

G Buchsbaum (1980), “A spatial processor model for object colour perception”, *Journal of the Franklin Institute*, **310**, pp. 1-26.

N W Campbell, W P J Mackeown, B T Thomas, T Troscianko (1997), “Interpreting Image Databases by Region Classification”, *Pattern Recognition*, **30**, 4, pp. 555-563.

D Dunn, W E Higgins (1995), “Optimal Gabor Filters for Texture Segmentation”, *IEEE Transactions on Image Processing*, **4**, 7, pp. 947-964.

J D Foley, A van Dam, S K Feiner, J F Hughes, *Computer Graphics: principles and practice*, Addison-Wesley.

G L Goodrich and A Zwern (1995), “From virtual reality to large print access: Development of a head-mounted virtual display”, *Low Vision Shared Interest Group Meeting, American Academy of Ophthalmology,* Atlanta, Georgia 30th October.

A K Jain and F Farrokhnia (1991), “Unsupervised Texture Segmentation using Gabor Filters”, *Pattern Recognition*, **24**, 12, pp. 1157-1186.

W P J Mackeown (1994), “A Labelled Image Database and its Application to Outdoor Scene Analysis”, *PhD Thesis*, University of Bristol.

S Marcelja (1980), “Mathematical description of the responses of simple cortical cells”, *Journal of the Optical Society of America*, **70**, 11, pp. 1297-1300.

R W Massof and D L Rickman (1992), “Obstacles encountered in the development of the low vision enhancement system”, *Optometry and Vision Science*, **69**, pp. 32-41.

N Molton, S Se, J M Brady, D Lee, P Probert (1998), “A stereo vision-based aid for the visually impaired”, *Image and Vision Computing*, **16**, 4, pp. 251-263.

NASA (1993), “High-tech help for low vision”, *NASA Tech Briefs*, **17**, 2, pp. 20-22.

N R Pal, S K Pal (1993), “A Review on Image Segmentation Techniques”, *Pattern Recognition*, **26**, 9, pp. 1277-1294.

E Peli (1991), “Image enhancement for the Visually Impaired – Simulations and Experimental Results”, *Investigative Ophthalmology and Visual Science*, **32**, 8, pp. 2337-2350.

E Peli (1995), “Head-mounted display as a low vision aid”, *Unpublished manuscript*.

E Peli, E Lee, C Trempe and S Buzney (1994), “Image enhancement for the visually impaired: the effects of enhancement of face recognition”, *Journal of the Optical Society of America*, **11**, pp. 1929-1939.

S Z Selim and M A Ismail (1984), “K-means-type algorithms: A generalized convergence theorem and characterisation of local optimality”, *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, **6**, 1, pp. 81-87.

M Shridhar and A Badreldin (1984), “High accuracy character recognition algorithm using fourier and topological descriptors”, *Pattern Recognition*, **17**, 5, pp. 515-524.

G Wyszecki, W S Stiles (1982), Color Science: Concepts and Methods, Quantitative Data and Formulae, Wiley, New York.

Robotic travel aid for the blind: HARUNOBU-6

Hideo Mori and Shinji Kotani

Department of Electrical Engineering, Yamanashi University,
Takeda-4, Kofu 400-8511, JAPAN

forest@es.yamanashi.ac.jp, kotani@es.yamanashi.ac.jp

ABSTRACT

We have been developing Robotic Travel Aid (RoTA) "HARUNOBU" to guide the visually impaired in the sidewalk or campus. RoTA is a motor wheel chair equipped with vision system, sonar, differential GPS system, dead reckoning system and a portable GIS. We estimate the performance of RoTA in two viewpoints, the viewpoint of guidance and the viewpoint of safety. RoTA is superior to the guide dog in the navigation function, and is inferior to the guide dog in the mobility. It can show the route from the current location to the destination but cannot walk up and down stairs. RoTA is superior to the portable navigation system in the orientation, obstacle avoidance and physical support to keep balance of walking, but is inferior in portability.

1. INTRODUCTION

Among 307,000 visually impaired in Japan 65,000 are the complete blinds. Most of them lost their sight in the elderly age. It is very difficult for the aged to learn to walk with the long cane or the guide dog, because they are not so rich in the auditory and haptic sensing and have not good memory for the cognitive map.

We have been developing Robotic Travel Aid (RoTA) "HARUNOBU" since 1990 to guide the visually impaired in the sidewalk or campus (Kotani S., Mori H. & Kiyohiro N.,1996). RoTA is a motor wheel chair equipped with vision system, sonar, differential GPS system, dead reckoning system and a portable GIS(Geographic Information System). In designing the RoTA, we add a guidance function and a safety function to the conventional mobile robot functions.

MoBIC Project (the mobility of blind and elderly people interacting with computers) was carried out from 1994 to 1996 with support of the TIDE program of the Commission of the European Union. It developed MoBIC travel Aid (MoTA) which consists of MoBIC Pre-Journey System (MoPS) and MoBIC Outdoor System(MoODS) (Pertie H., et al.,1996). MoPS is a simulator that helps the exploration of a previously unknown area and the selection and preparation of a route before an actual walk. MoODS is a portable system that gives assistance during the walk. It consists of a small wearable PC kernel of 16 x 11 x 7 cm in the size, a GPS, an electronic campus and a pair of special earphones that prevent masking the ambient sound essential for echo location. The system provides on-route information about the current position. The system informs the traveler automatically when they are leaving the chosen route or if the accuracy of the system has degraded. A prototype of MoTA was developed and estimated through a field test and found the design philosophy was useful in the human navigation.

RoTA is superior to the guide dog in the navigation function, and is inferior to the guide dog in the mobility. It can show the route from the current location to the destination but cannot walk up and down stairs. RoTA is superior to MoTA in the orientation, obstacle avoidance and physical support to keep balance of walking, but is inferior in portability. The functional comparison of RoTA, the guide dog and MoTA is shown in Table 1.

In the road environment the most important objects may be the car and the pedestrian. Conventional methods for the car and pedestrian detection seem to simulate the perception of the human beings. We get the idea of objects discrimination from the study of ethologist (Tinbergen N.,1969). He shows that the animal behavior is represented by a chain of fixed action patterns even if the behavior is an advanced and complex one. To explain the mechanism of the behavior Tinbergen proposes three concepts: sign stimulus, Central Excitatory Mechanism (CEM) and Innate Releasing Mechanism (IRM). The animal dose not recognize objects as human being does, it makes a response not to the whole of the object but to the part inherent in the object. The part of the object that activates the fixed action pattern is called sign stimulus. CEM is similar to the modern multi-tasking system in the modern computer. All the fixed action patterns are in the dormant state, and when a sign stimulus appears the IRM activates one of the fixed action patterns corresponding to the sign stimulus. We think that Tinbergen's concepts are useful to configurate the vision based mobile robot. We use sign pattern(SP) as the technical term instead of the sign stimulus. Sign pattern is different from the landmark in three factors as shown in Table 2. The purpose of the landmark is to verify the current location, on the other the sign pattern is used to activate and guide the fixed action of the robot. We think the basic fixed action patterns are Moving-along SP, Moving-toward SP, Following-a-person, Turning-corner, Avoiding-obstacle, Moving-for-sighting SP.

Table 1. *RoTA, MoTA and guide dog robot*

	Obstacle avoidance & orientation	Mobility	Portability	Navigation
RoTA	O	O	×	O
MoTA	×	×	O	O
Guide dog	O	O	O	×

Table 2. *Comparison of landmark & sign pattern*

	Sign pattern	Landmark
Purpose	To guide the fixed action pattern	To verify the cuuent location
Object	Permanent and temporal objects	Permanent objects
Representation	Simple feature; edge, rhythm, shadow	2-D & 3-D model

2. GUIDANCE

A Geographic Information System (GIS) is required as the base of the navigation system of RoTA. The GIS of RoTA has to include the robot guide information and the human guide information. The robot guide information should give the sensor system of the robot the information about the environment.

2.1 Sign pattern

The robot does not recognize the total environment but it recognizes only two kinds of signals required to guide the robot in the environment. One is a sign pattern and the other is a landmark. For instance SP in Moving-along means a signal used to correct the location and heading errors of the dead reckoning system. RoTA uses as the SP of Moving-along an elongated feature on the road such as road boundary, lane mark, fence, tactile block and so on. We define the rhythm of walking as the SP of Following-a-person. As the SP of Avoiding-car we define the shadow underneath the car.

2.2 Robot guide information

To keep safe and to follow the Japanese traffic regulation RoTA should move on the sidewalk and zebra crossing marks. For this reason we define the path on which RoTA and the blind can move safely. When the road has the sidewalk of enough width for RoTA, the path is specified on it. When the road has not a sidewalk, the path is specified right or left roadside which is free from falling into creek, downstairs and depression. The digital map of the GIS includes a road network, a path network, sign patterns and land marks. The road network includes road information such as the type, the distance, the direction and absolute location of the street and the junction. After route searching the GIS feeds to the locomotion control system the robot guide information along the route. Fig.1 shows a snapshot of the display of RoTA in Moving-along SP in Yamanashi University campus. The upper middle part shows a video image in which a SP searching window is described by a large square and SP tracing windows are described by small squares. The right upper part shows the robot coordinate system in which a detected sign pattern is described by a line segment. The left upper part shows the heading of RoTA. The lower right part shows the digital map of the campus, the center of which shows the current location of RoTA. The left lower part shows a differential GPS sky map in which solid circles show received satellites and open circles show non-received satellites.

2.3 Map learning by practice

To make the digital map to guide the robot, one should select landmarks sign patterns in the course and measure the distance and orientation between intersections. This measurement requires much efforts, moreover visual sign patterns change with the time (morning/daytime/evening), weather (sunny/cloudy) and the season. For these reasons RoTA has a function of the map learning by the practice.

Before the practice the operator gives RoTA a rough map represented by a list sections. The section is defined as the part between intersections and is specified by an approximate distance and a direction. In the first practice RoTA moves along the course based on the rough map and detects a SP and corrects the lateral location error based on the SP. In the SP detection RoTA works in two modes; searching mode and tracing mode. In searching mode the vision module detects SP candidates with a wide view angle and selects one which matches with the section of the rough map in the direction. Then the vision mode begins, it traces the SP obtained in searching mode with a narrow visual angle in the predicted position. It takes less processing time in tracing mode than in searching mode. The sign pattern information obtained by the first practice can improve the performance of traveling not only in the traveling time but also in the safety of locomotion.

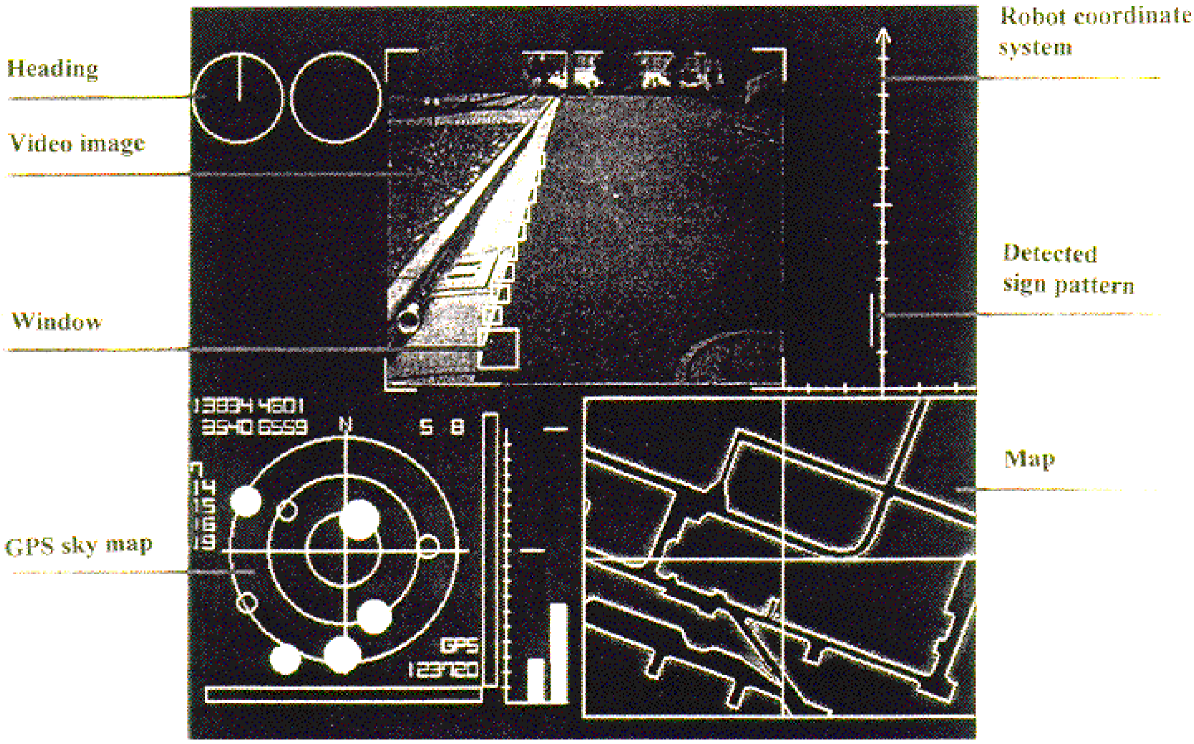


Fig.1. *A snapshot of the display of RoTA in Moving-along SP in the campus*

During the first practice the vision module stores the trajectory and SPs with their location and direction. After the first practice the learning process omits the noisy SPs and then fills gaps between neighboring SPs. The new SPs are used to update the rough map. The new map includes SP information about its location and direction. The second practice can improve its performance by using the new map.

Fig.2 (a) and (b) show the first and the second practice of HARUNOBU-4 in our campus. . A broken line shows the trajectory, and a line segment shows a SP candidate. A small closed circle indicates a searching point. In the first practice, four closed circles at corner N_2 show that the vision module repeat searching until it gets the real SP of the direction N_2N_4. At T-shape intersection N_3 the vision module misses SP and after three searching processes it detects SPs at the opposite(right) side of the passage, and tracing one of them. HARUNOBU-4 reaches at point B and finds the traveled distance is over the specified approximate distance and makes a U-turn immediately. After two searchings it finds the real SP of direction N_3N_4. In the second practice as shown in Fig.2(b) the searching point drastically decrease from fifteen to six.

2.4 Human guide information

We are developing a human guide information system. Its basic concept is almost the same as MoPS] When the blind is unsure the current location, he/she push the button, then the system tell the current location through a synthesized voice maker. When the blind wants to know future path to the goal, the system answers the time, distance and the number of turning to the goal.

3. PEDESTRIAN DETECTION BY RHYTHM

Conventional human motion tracking method includes the modeling of human body and the matching the model with the real data. The stick figure model is a well-known model of human body, but it should be modified by the distance, the clothes that will be changed by man or woman, summer or winter.

When one walks in the sidewalk, the rhythm of the walk is almost constant. The rhythm can be seen in the swing motion of feet and hands and in the up down motion of the head and shoulder. Among these motion the feet motion is the most detectable by the computer vision, because their rhythm is clearer than those of the head and the shoulder, the background of the feet is simpler than those of head and shoulder, and the clothes and another part of the body do not cover the feet in the image. The rhythm of the feet is a good sign pattern as it is easy to detect by the computer vision.

The difficult process of scaling to fit the object image to the model is not needed. It is free from the distance, the clothes and the weather. The implemented method is as follows(Yasutomi S, Mori H. & Kotani S.,1996).

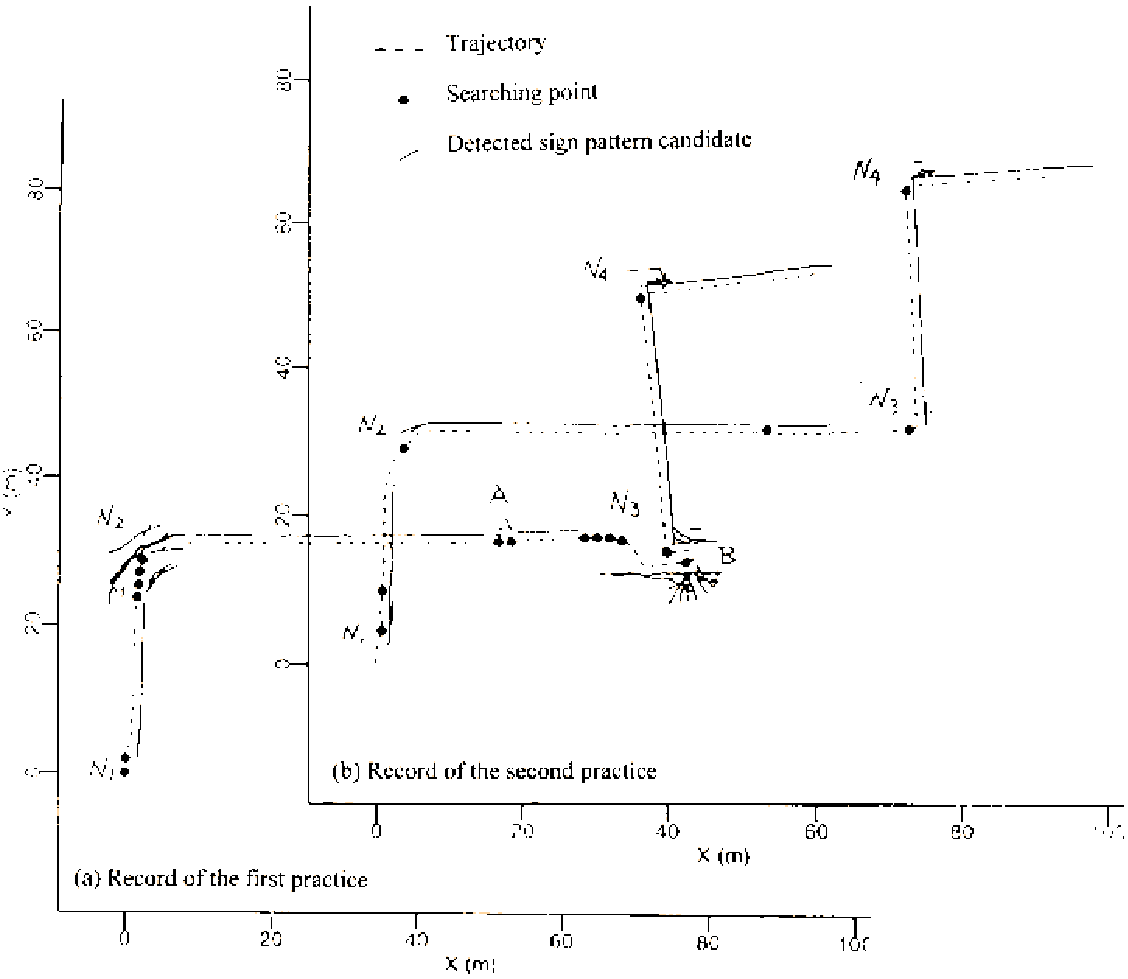


Fig. 2. *An example of SP learning*

3.1 Motion segmentation

The frame subtraction is applied to detect moving objects as shown in Fig.3. So this method is effective when the video camera is in the stationary state. A horizontal projection is operated after binarizing the subtracted image. The horizontal projection is sliced by a threshold to obtain *H* segment that may represent the height of a person. A vertical projection is operated and sliced by another threshold to obtain *V* segment that may represent the width of the person. If *V* segment satisfies the threshold of width, window W of *HV* in size is assumed to be the head to feet window of the person. Then the right and the left foot window, W_R and W_L, each of which is 1/5 of the *H* segment and 1/2 of *V* segment are set up in the lowest part of W. This process is called finding process, and is followed by tracking process as follows. Window W_R and W_L of the last frame are a little enlarged in length and width to trace the feet in the next frame, and the horizontal and vertical projections are operated on the new binarized subtracted image. New W_R and W_L are obtained by the same slicing operation as the finding process.

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

3.2 Rhythm matching

The most significant features of W_R and W_L are (1)the ordinate of the bottom of the windows and (2)the area of the binarized subtracted image in the windows, as the ordinate shows the distance of the person, and the periodic change of the area of the sliced image depends on the rhythm of walking.

Auto-correlation function is operated on time series of the area of W_R and W_L. When the primary components of the power spectrum of the two time series are satisfied with 2σ of the mean rhythm of walking, W_R and W_L are judged as the feet of a person. An example of time series of the area in W_L and its power spectral are in Fig 4.

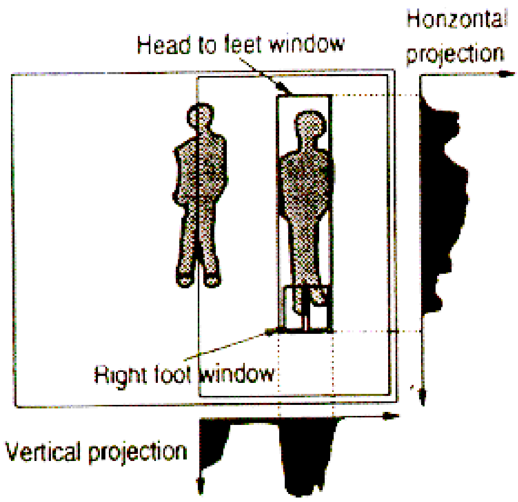


Fig. 3. *Setting of three windows*

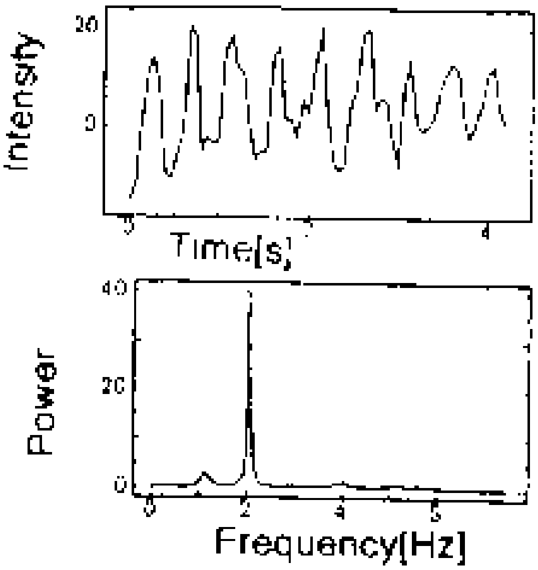


Fig. 4. *An Example of the time series of the area of WL and its power spectral*

	Samples	Correct	False
Pedestrian	407 (334:pants,45:short pants & short skirt, 28:long skirt)	94.9%	5.1%
Non-pedestrian	106	96.2%	3.8%
Total	513	95.0%	5.0%

Table 3. *Results of pedestrian detection*

3.3 Results of pedestrian detection

Pedestrian detection algorithm for stationary camera is implemented on a monochromatic image processing system (HITACHI Co.Ltd.,IP-2000). It samples a moving object every 67ms and takes 64 samples (4.3sec) to judge the object by the rhythm. We fixed a video camera in our campus 1m in the height and 15in the depression angle, and recorded 407 pedestrians and 106 non-pedestrian including bicycles and dogs on a videotape in a cloudy day. Among the pedestrians 82% of them wear pants, 11% of them wear short pants or short skirt and 7% of them wear long skirt. The experimental results on the videotape are shown in Table 3. Five % of errors are caused by (1) noise of video signal that makes jitters on the image, (2) swaying of trees and grass that make the same rhythm as that of pedestrian, (3) the same color of shoes as that of the asphalt paved road.

4. DANGER ESTIMATION AT AN INTERSECTION

When the driver and the blind keep the traffic regulation perfectly, they will not meet with any accident. However as they often pay less attention to the right and the left sides of an intersection, they will occasionally have an accident. According to the statistics of the traffic accidents in Japan, about 50% of them occur at or near intersections. To avoid collision we should estimate danger level of vehicles at or near the intersection (Kotani S.,Mori H. & Charkari N.M.,1996).

4.1 Car detection by shadow

The sunlight and the diffused sky light do not reach the underneath a car. The image intensity of the underneath part is almost noise level in the video image. Its intensity is lower than any other part such as a wet part or a patched part repaired by new asphalt even in the cloudy day as shown in Fig.5.

These phenomena are applied to the SP definition of the car. A window is set up in the lane, and .the vertical projection of the window is obtained. When the projection curve shows a flat bottom of a certain width with cliffs at the right and left side as shown in Fig.5, we define the bottom as the sign pattern of a car. Three levels of danger are defined in this work, 0: *Safe* 1: *Warning* 2: *risky*. The robot detects the location s_i and moving direction r_j of the car by its sign pattern and predicts its future path based on the Japanese traffic regulation. The danger coefficient d_{ij} for a vehicle at (s_i,r_j) is defined as follows. When the future paths of the vehicle and the robot do not cross ($d_{ij}=0$), possibly cross ($d_{ij}=1$), surely cross ($d_{ij}=2$).

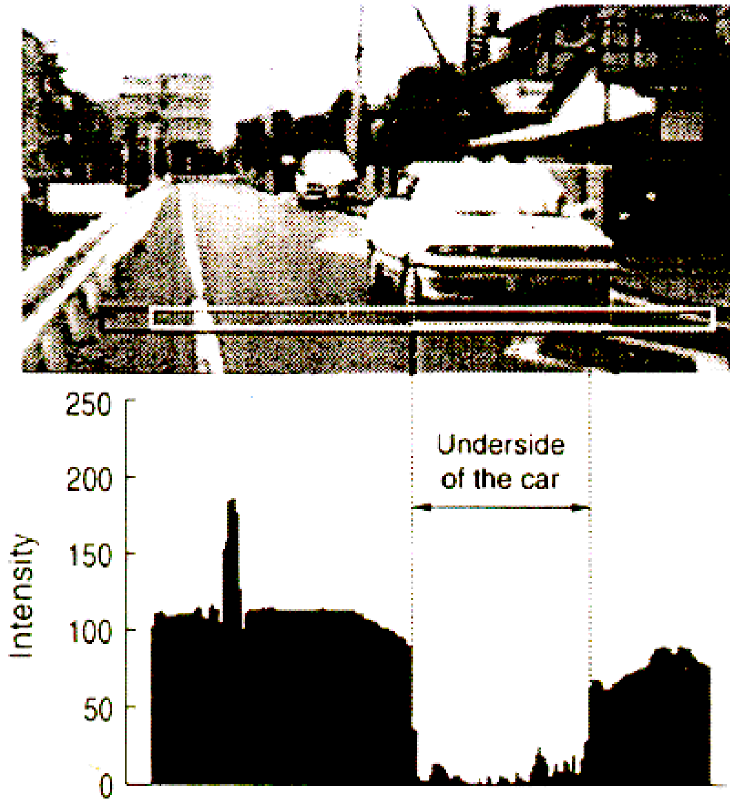


Fig. 5. *An intensity curve in a window which is set up underneath the car*

4.2 Japanese traffic regulation

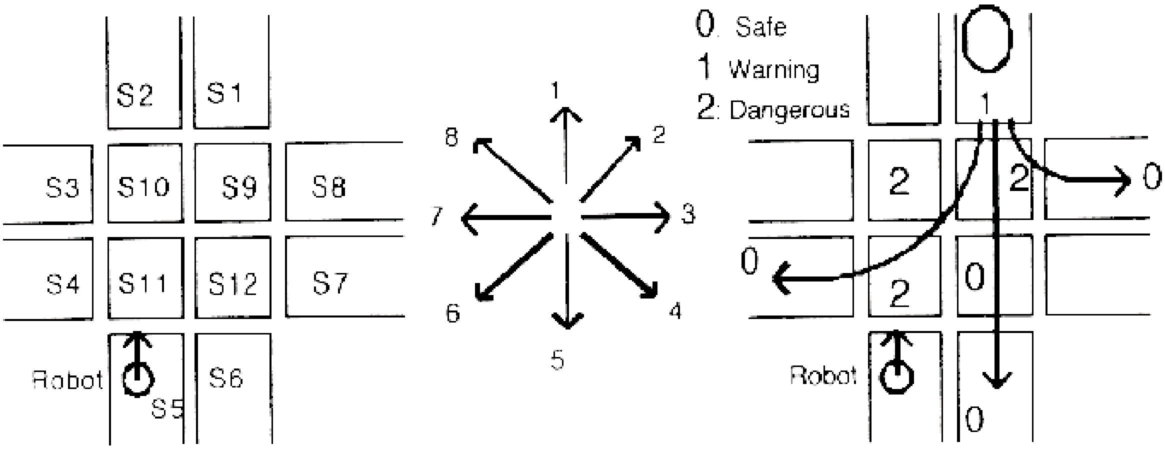
We formulate the traffic regulation including the behavior of the careless driver as follows. (J1) Vehicles move along the left lane mark. (J2) Vehicles follow typical path. (J3) When the driver moves straight, he will only pay attention to the front. When he turns left, he will pay attention to the front and the left. When he turns right, he will wait until all the straight moving and right turning cars pass by. (J4) When the blind starts moving across the intersection, the car must not obstruct his/her way.

4.3 Robot's traffic regulation

We consider that the robot follows the same traffic regulation as the guide dog. (R1) The robot moves along the left side of the road. (R2) When the danger estimate value is safe, the robot sends the blind the permission message to start crossing. (R3) After the blind receives the permission message, he gives the robot a start command. (R4) The robot has an intelligent disobedience function. The robot does not follow the blind's command before the danger estimate value becomes safe. Based on the traffic regulation of the car and the robot, danger matrices d_{ij} are given as shown in Fig.6.

4.4 Results of danger estimation

The car detection algorithm is implemented on a personal computer of CPU486 (100Mhz) with an image processing board (HITACHI Co.Ltd., IP-2000). The car tracking is performed every five frames. The danger level is estimated while a car passes an intersection within 3 - 4 sec. A video camera was fixed 1m in the height and 2.5m apart from a T shape intersection as shown in Fig.7. It recorded 105 cars that pass the intersection on a videotape. We looked B at the display and judged the estimation made by the computer. The performance of the computer for 105 cars shows 90% of success as shown in Table 4. Among 10 misjudgments, eight misjudgments were caused by two or three cars in successive running less than 20m apart. The vision system is successful in tracking the first car, but often fails in detecting the second car. Remaining one was caused by the trajectory of an ill-mannered driver. The last one was caused by the mis-tracking of a car that was too large to process at video rate. Fig.8 shows examples of three cases. In the right side of the display six parameters are shown; "TIME" indicates the quantized time, "Moving car" shows the result of the process, "DIST" shows the obtained distance of the car from the video camera in meters, "Speed" shows the estimated velocity in KM/h, "AP-TIME" indicates the predicted arrival time at the intersection. The Trajectory of the vehicle is shown at the bottom of the right side part. Finally the danger coefficient of the car is represented *safe*, *warning* and *risky*.



(a) Sections at the cross *(b) Quantized directions* *(c) Danger estimation for a front vehicle*

Fig.6. *Danger estimates of a vehicle at an intersection*

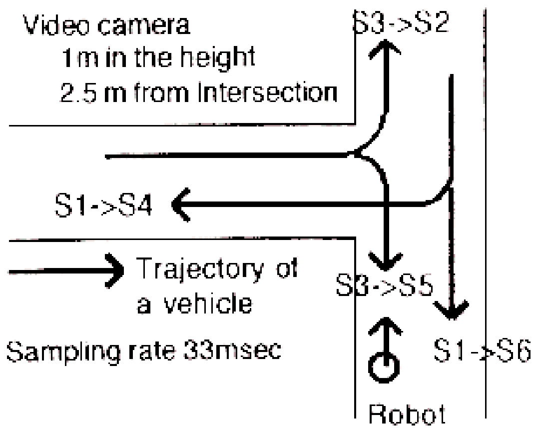


Fig.7. *Experimental set up at a T shape intersection*

Course	No. of cars	Correct	False
S1 → S6	36 (100%)	30 (83%)	6 (17%)
S1 → S4	16 (100%)	15 (93%)	1 (7%)
S3 → S2	23 (100%)	22 (95%)	1 (5%)
S3 → S5	30 (100%)	28 (93%)	2 (7%)
Total	105 (100%)	95 (90%)	10 (10%)

Table 4. *Results danger estimation at the T shape intersection*

5. EXPERIMENTAL RESULTS

We have implement the concept of RoTA on a color vision-based mobile robot HARUNOBU-6 as shown in Fig.9. It has a motor wheel chair (SUZUKI Co.Ltd., MC14) as the undercarriage part, a color video camera with pan/tilt control(Sony EVI-G20) and a real time image processing board (HITACHI I Co.Ltd., IP-2000) as the vision module, two sonar range sensors(IZUMI I Co.Ltd., SA6A-L2K4S, 130kHz), an optical obstacle sensor(SUNX Co.Ltd., PX24ES), a dead reckoning system with an optical gyroscope (HITACHI WIRE I Co.Ltd., OFG-3) and a differential GPS system (MATSUSHITA DENKO I Co.Ltd., GS-5). The performance of these sensors are shown in Table 5. The vision module is used to get the information of orientation and navigation. The sonar range sensor is used to get mobility information. The optical obstacle sensor is used for reflective obstacle avoidance. A horizontal bar is attached the rear of HARUNOBU-6. By touching the bar the blind can keep his balance in walking and can feel the surface of the ground through its vibration. He/she can get the mobility and orientation information through the motion of HARAUNOBU-6.

The performance of RoTA "HARUNOBU-6" was tested by three test courses.

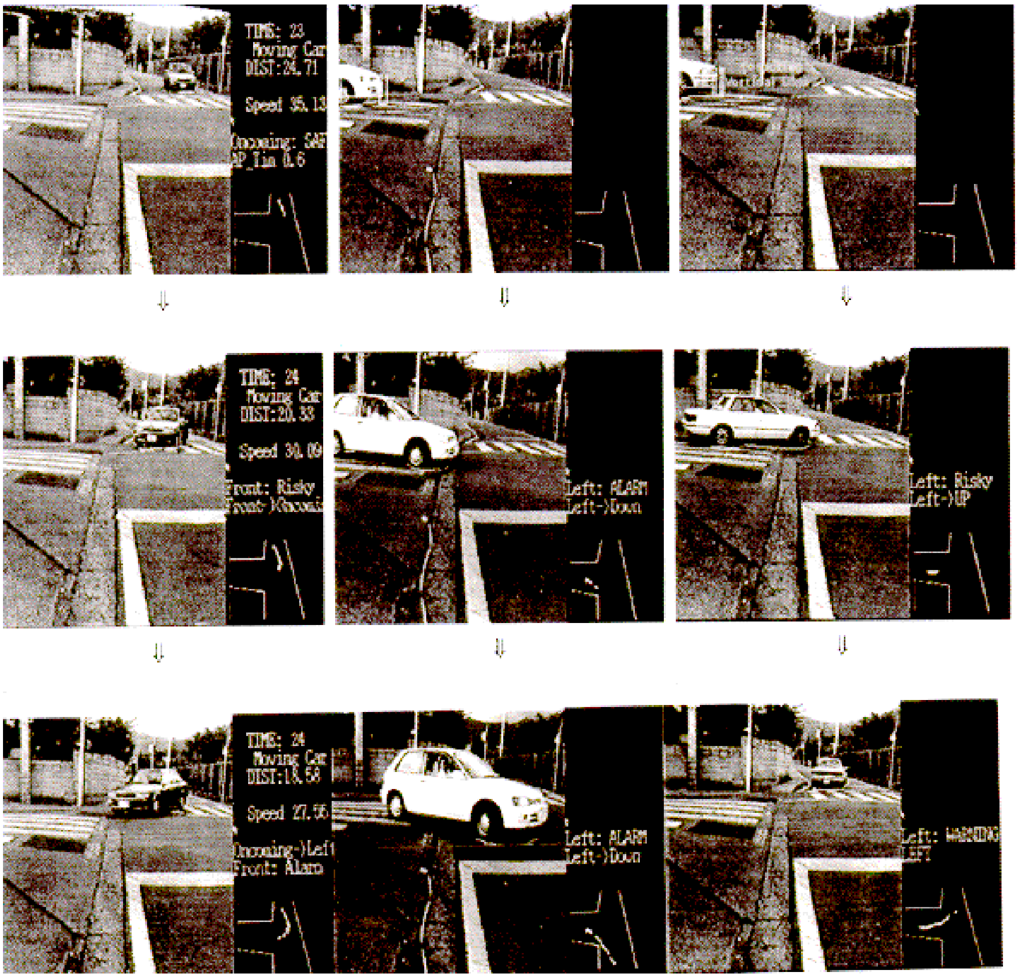
The first test course is set up in a small zone of our university campus of 50m by 50m. In this course HARUNOBU-6 changes 360 degrees in its heading. The illumination of sunlight changes from back light to counter light. From the technical point of view this experiment give us the problem of iris control. A blind who lost his sight by *retinosis pigmentosa* tested HARUNOBU-6. He said the robot was useful for him to move from building to building. He suggested us that a step attached the rear of the robot would be useful to rest himself during the locomotion. He can escape from the accidents by getting off the step.

The second test course is set up in an open space of Kofu stadium, In such open field the bind feels difficulty in orientation because he cannot use the echo location. Although the position error(3σ) of the differential GPS is 2 meters, it is useful in only open space. The open space is a good place to guide RoTA by the differential GPS.

The third test course is set up in the hospital of YAMANASHI MEDICAL UNIVERSITY. To guide a patient of ophthalmology from the doctor's office to his/her ward a nurse is required. Instead of the nurse our RoTA is expected. The illumination of the corridor is not homogeneous, therefore it is difficult to detect SP and obstacles by the vision. The sonar range sensor and the optical sensor are used in the hospital.

Table 5. *Performance of sensors of HARUNOBU-6*

Sensor	Detected objects	Range
Vision module	Road edge, car, Pedestrian	2 – 30 [m]
sonar range sensor	Right and left side wall	0.2 – 2 [m]
Optical obstacle sensor	Suddenly appearing obstacle	0.1 – 1.5 [m]



(a) Enter from ahead, turn left *(b) Enter from left, come here* *(c) Enter from left, go ahead*

Fig.8. *Some results of the danger estimation system*

Fig. 9. *HARUNOBU-6*

6. CONCLUDING REMARKS

We have a plan to develop several RoTAs in corporation with Japanese companies to make field tests by two kinds of the blind. The first is the blind who can walk with the guide dog and the second is the diabetic who loses his sight recently and cannot walk without a helper.

The guide dog user will want to walk in the crowded streets for visiting and shopping. We think he can use PC with voice interface to communicate with RoTA. The difficult problem in this case is how to make the map data base.

The blind of diabetes will want to learn walking in a safe place such as the campus of a hospital or a park. The difficult problem is the human interface because the diabetic loses not only the vision but also auditory sense and the haptic sense. He will not be able to use PC.

This work is supported by Grant-in-aid for Scientific Research on Priority Areas No.07245105, Grant-in-Aid for Scientific Research(B)(2) No.07555428 by the Ministry of Education, Science, Sports and Culture, The Mechanical Industry Development & Assistance Foundation (MIDAF).

7. REFERENCES

Kotani S., Mori H. & Kiyohiro N.(1996), Development of the robotic travel aid "HITOMI, *Robotics and Autonomous Systems*, *17*, pp.119-128

Kotani S., Mori H. & Charkari N.M.(1996), Danger estimation of the Robotic Travel Aid(RoTA) at intersection, *Robotics and Autonomous Systems*, *18*, pp.235-242

Mori H., Charkari N.M. and Matushita T.(1994), On-line Vehicle and Pedestrian Detection Based on Sign Pattern, *IEEE Trans. On Industrial Electronics*, **41**, *4*, pp384-391

Pertrie H., Johnson V., Strothotte T., Rabb A., Fritz S. And Michel R.(1996), MOBIC: Designing a Travel Aid for Blind and Elderly People, *Journal of Navigation*, **49**, *1*, pp.45-49

Tinbergen N.(1969), *The study of instinct*, The Claredon Press Oxford

Yasutomi S., Mori H. & Kotani S.(1996), Finding pedestrians by estimating temporal-frequency and Spatial-period of the moving objects, *Robotics and Autonomous Systems*, *17*, pp.25-34

Generating virtual environments to allow increased access to the built environment

G T Foster[1], D E N Wenn[1] and F O'Hart[2]

[1]Distributed Systems Research Group, Department of Cybernetics, University of Reading, Whiteknights, Reading. UK.

[2]University of Dublin, Trinity College, Dublin, IRELAND

grantfoster@cyber.rdg.ac.uk, cybdenw@cyber.rdg.ac.uk, Fergal.Ohart@cs.tcd.ie

ABSTRACT

This paper describes the generation of virtual models of the built environment based on control network infrastructures currently utilised in intelligent building applications for such things as lighting, heating and access control. The use of control network architectures facilitates the creation of distributed models that closely mirror both the physical and control properties of the environment. The model of the environment is kept local to the installation which allows the virtual representation of a large building to be decomposed into an interconnecting series of smaller models. This paper describes two methods of interacting with the virtual model, firstly a two dimensional representation that can be used as the basis of a portable navigational device. Secondly an augmented reality called DAMOCLES is described that overlays additional information over a users normal field of view. The provision of virtual environments offers new possibilities in the man-machine interface allows intuitive access to network based services and control functions to a user.

1. INTRODUCTION

The problems of ergonomic access and control are particularly relevant to people with special needs when faced with poor building design and information support. This is highlighted in studies of disabled or elderly people in the community for instance in the 1991 RNIB survey by Bruce et al, 1991 a comparison is made between the level of independent mobility of young registered blind people in 1991 and the level found by Gray and Todd in 1965. It was found that the percentage level of independent mobility had not increased over time in spite of the increase in available mobility training. The difficulties in accessing the built environment are a major factor for this lack of mobility and are common across a broad range of people with special needs:

These issues are being addressed in the EU TIDE (Telematics Initiative for Disabled and Elderly) project ARIADNE "Access, Information and Navigation support in the *Labyrinth* of Large Buildings". ARIADNE is exploiting a new generation of networked intelligent buildings to provide an infrastructure that will enhance accessibility to the built environment. This is achieved through the use of new microwave smart card sensor technology that will allow the building to sense a user is in a particular place in a building and respond in an appropriate way to that users requirements. The smart card readers are networked around the building and can trigger local environment changes in the building or transmit information directly to the user through hand held devices and sophisticated talking signs. The ARIADNE network offers a powerful information and control resource that can be used by user interfaces with different levels of sophistication. More details of the ARIADNE project can be seen in Foster 1998a.

Exploiting networked control technologies as promoted by products such as ARIADNE, it is possible to create a virtual environment that is a mirror of the real environment, a concept promoted by Gelernter 1992. This is possible because the network infrastructure allows the real time passing of information and control commands between the virtual model and the actual environment. This control and information infrastructure ensures that the Virtual Environment remains concurrent with the real world, producing a mirror of the actual environment as can be seen in Figure 1.

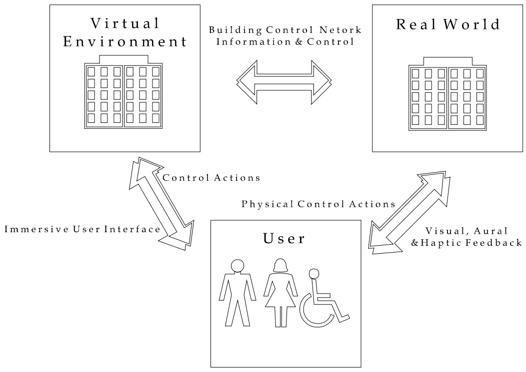


Figure 1. *Interaction either with the Real World or a Virtual Environment.*

The use of a Virtual Environment offers the user an alternative method of interacting with the building. When interacting with the Real World the user has to physically deal with the environment and receives physical feedback from the building itself. Having a virtual model of the environment offers an alternative control route as information contained in the model can be presented in a variety of ways which are tailored to the requirements of the particular user. Through the network infrastructure, control and information can flow seamlessly into the real environment and vice versa. This approach benefits the human user but can also be used to structure the environment of mobile robots within the built environment as presented by O'Hart 1998.

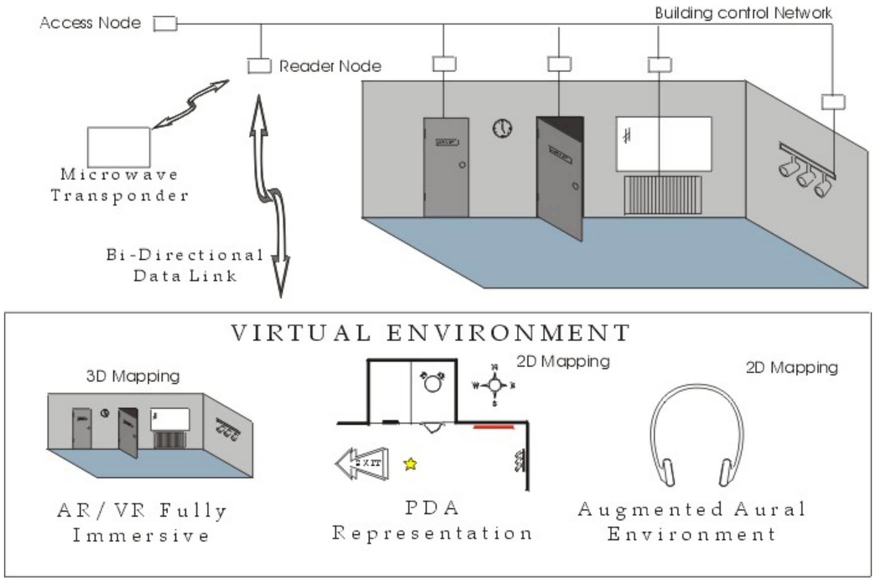


Figure 2. *Obtaining a virtual model of the local environment.*

The underlying concept of this paper is that the model of the environment can be provided by the building itself. The entire building is split down into manageable sections, often rooms and corridors, and the model of the local environment is held locally on the network nodes that are physically installed in the building as can be seen in Figure 2. Moving around the site allows a User with a suitable transponder to extract the local information via a microwave link managed by a Reader Node. Associated with the Reader Node is an Access Node which is a device that holds the virtual representation in a Local Object Model. Once a virtual model is available it can be presented accordingly, two main methods being presented in this paper:

1. **Three Dimensional Mapping using AR/VR Immersive Environments.** A 3D augmented Reality suite called DAMOCLES is presented, the Head Mounted Display (HMD) of which is based on the 'I-Glasses' headset from Virtual I/O Ltd. These have onboard tracking, and are cheap and lightweight but allow the generation of both fully immersive Virtual Environments and semi immersive Augmented Environments that superimpose additional information upon a users normal field of view. This device and associated computer could be made portable by mounting on an electric wheelchair. Original background to this research being presented by the author and Hammond (1996) of an augmented reality designed to interoperate with the control system of a smart house.

2. **Two Dimensional Augmented Aural Environments.** A prototype of a novel navigational tool is described based on an audio based device translates the environments into useful speech facets. The

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

device contains a compass and pre-recorded sampled speech generation. This device could be the basis of a navigation device for visually impaired people.

2. GENERATING LOCAL MODELS

One of the major problems with the generation of virtual environments is that models quickly become large and unwieldy, this is a particular problem if the VR device is incorporated into a mobile unit. To combat this and to fit the models within the memory constraints of the network infrastructure the entire building model is decomposed into a series of smaller models called Local Object Models (LOM). A more detailed account of Local Modelling techniques can be seen in Foster 1998b. In a typical installation, Reader and Access nodes are placed strategically at 'decision points' within the building. These decision points are the places where routes diverge forcing a user to make a navigational decision.

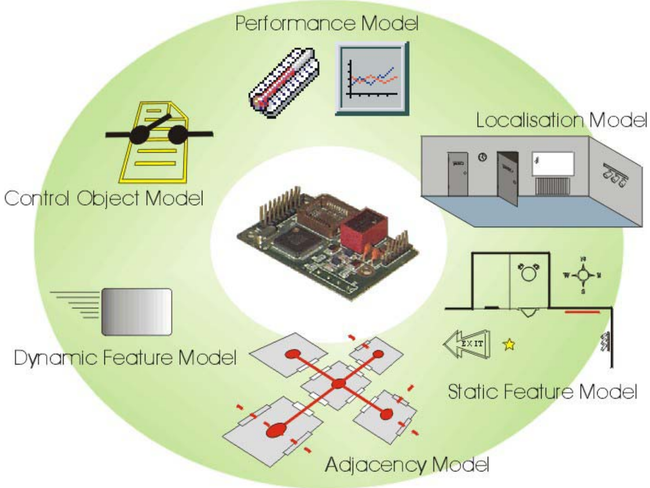


Figure 3. *Model classes within the Local Object Model stored on a network node.*

The Local Object Model can be seen in Figure 3, and is implemented on each Access Node in the ARIADNE network. Contained within the Local Object Model there are several model classes which form a distributed data resource across the network. These can be discussed in more detail:

- **Localisation Model.** The localisation model contains a physical description of the environment local to the node. Generally this is represented using VRML or a Cartesian / Polar 2D mapping.
- **Static Feature Model.** The Static Feature Model contains a list of objects that are not accessible directly through the network but are of interest to a user, for instance telephones, seating areas, toilets etc.

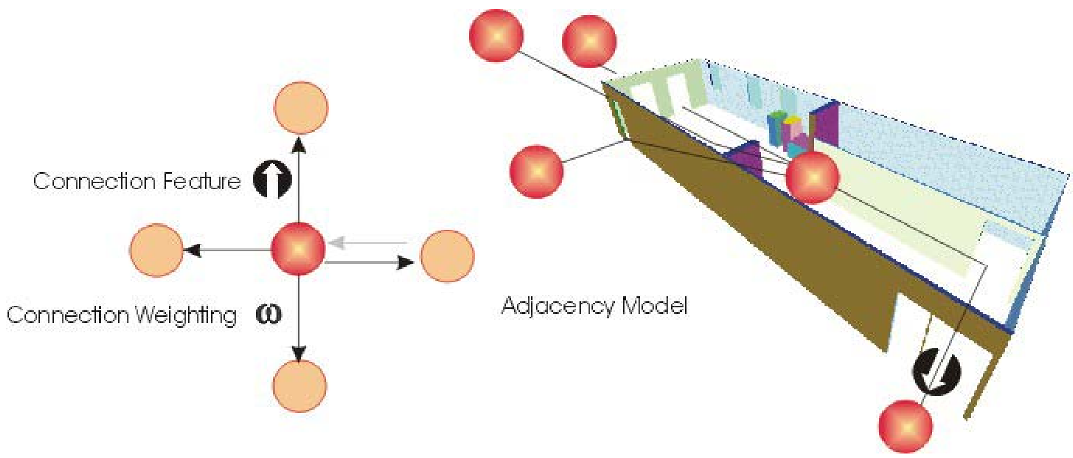


Figure 4. *The Adjacency Model.*

- **Adjacency Model.** The adjacency model represents the navigation links between neighbouring Localisation Models as can be seen in Figure 4. Connections between areas are given a weighting depending on distance and difficulty and the route must be achieved by passing through a connection feature which is typically a door or archway. A holistic view of the entire network produces an interconnecting lattice of adjacency models that provide a navigational map of the building an example being shown in Figure 5. This lattice can be used as a framework for network wide services such as the parallel searching agents as presented by the author in Foster 1997.

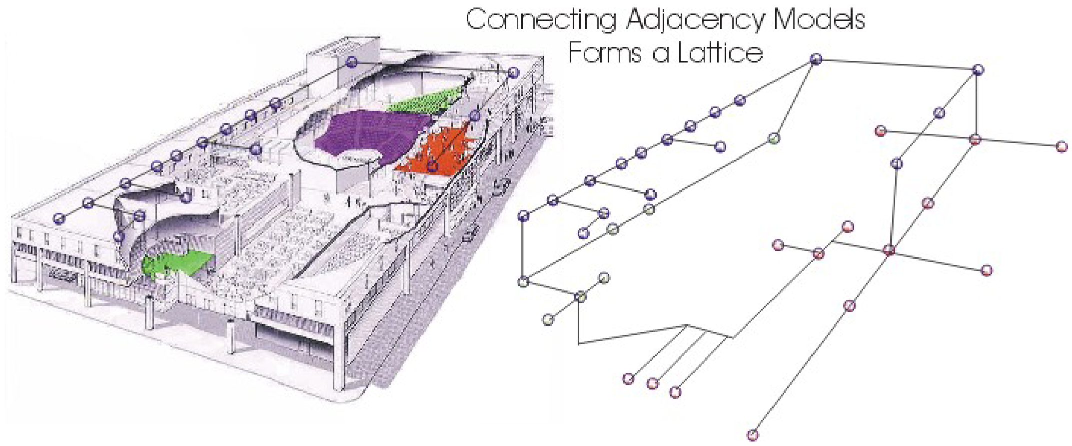


Figure 5. *A navigational topology lattice generated from connected adjacency models.*

- **Control Object Model.** The Control Object Model contains a list of devices that are able to be controlled through the network such as window openers, lights and heating.
- **Performance Model.** The performance model is related to network maintenance and keeps a model of the performance of the control objects in case of sensor or actuator failure.
- **Dynamic Feature Model.** The Dynamic Feature Model keeps a record of the tagged devices that are currently in the local area. This model can be used to perform a network search to locate a particular person or tagged object.

3. INTERPRETING THE VIRTUAL MODEL

3.1 A Navigation Aid Using A 2 Dimensional Environment.

To interpret the virtual model from the building a prototype of a hand held navigational device has been developed. It consists of a small microprocessor, currently a Neuron 3150 device which provides audible output through a ISD 33180 speech chip. This contains up to 180 seconds of analogue sampled audio which can be split up into an arbitrary number of sample segments. The device also has a basic compass facility through the use of a Vector 2X magnetometer from Precision Navigation, which provides bearings with a resolution of 1°.

The virtual model is downloaded to the device via the microwave link. In this case a two dimensional environment can be compactly represented as a set of *n* features **F** local to the decision point in the building. (1).

$$F = \{ f_0, f_1, f_2, ... f_n \} \tag{1}$$

$$f_n = \{ i, \omega, \delta, \phi \} \tag{2}$$

Where a feature f_n contains the following information i an id number that is directly associated to a audible message on the hand held device, ω is the target size of the feature, δ is the distance and ϕ the bearing direction to the feature. (2).

A typical usage scenario can be considered, the user moves around the building until a decision point is reached, at this juncture, being in range of a Reader Node, the hand held device receives the two dimensional local model from the building and alerts the user via a simple buzzer. The user can then place the device into a 'locate' mode via a simple button press. In this mode points the user scans the device around the room, when the bearing from the compass device matches the feature bearing given in the model, that particular feature is announced to the user who then is able to orient themselves at the feature they are pointing towards.

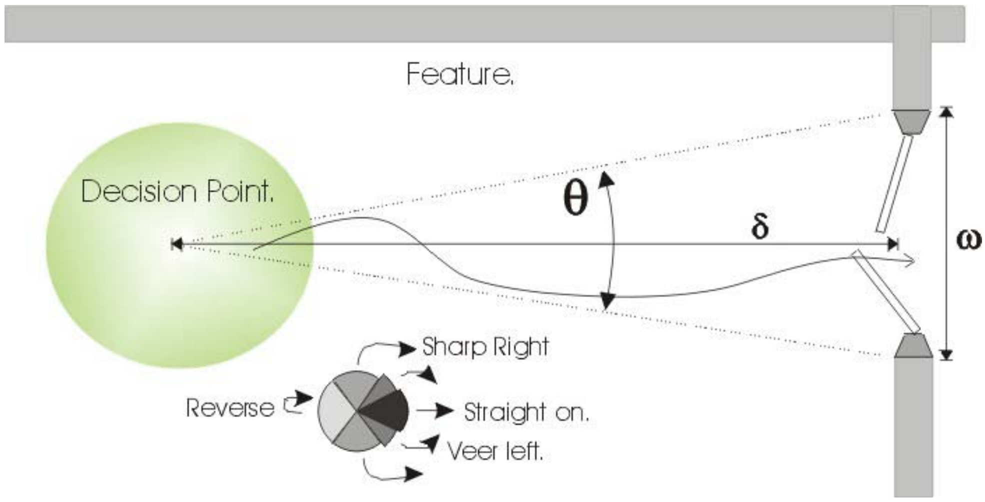


Figure 6. *Taking a bearing from a decision point to a target feature.*

Having located the feature the user can then track towards the object by placing the device into a 'follow' mode. With reference to Figure 6, the bearing to the object is taken via the compass device and the user can set off towards the object. To stay on course the user must stay within an error angle θ in order to reach the goal successfully.

$$\theta = \tan^{-1}\left(\frac{\omega}{\delta}\right) \tag{3}$$

The desired error angle is a straightforward calculation (3). If the user continues within that given error angle the destination should be found, however if the user deviates from the bearing simple audible messages are given to nudge the user back on track. A nominal template for this facility can also be seen in Figure 6, where the user is prompted to veer, make a sharp turn or reverse to get back on course. The effects of any errors in following the desired bearing are summative so successive errors can place the user in an unknown position until another decision point is reached. However assuming the user is able to keep more or less on track, it is possible to navigate through the building by travelling from decision point to decision point.

There are some drawbacks to this approach, the major problem is shared with many VR applications and that is tracking the user. The broad area covered by the microwave readers cannot locate the user very accurately, roughly in a 10 metre radius which adds error to the initial starting point for the following mode. This problem can be addressed by using other communication media such as a directed IR link from the ceiling which will locate the user to within roughly a 2 metre radius before initiating the following mode.

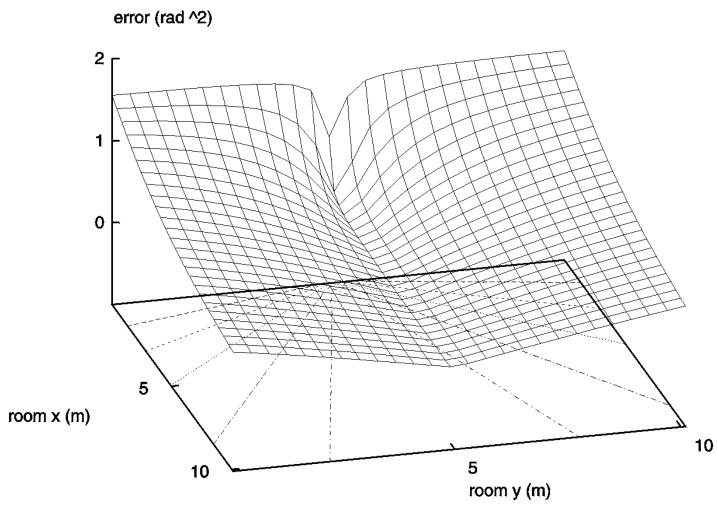


Figure 7. *A surface depicting the bearing error across a 10m×10m room space.*

Using bearings to navigate means that the system is sensitive to positional errors, considering a 10m×10m space with an origin at (10,5) and a target at (0,5) the surface in Figure 7 shows the extent of bearing errors related to position. There is a zero error directly along the bearing between the origin and the location of the target feature. Imagining that the target bearing is to go directly north, when the user is far away from the destination feature the bearing error is tolerable. However this error quickly increases the closer the user is to the destination.

The other drawback with the device as it stands is that the compass is affected by magnetic fields and soft iron within the building, this can mean that magnetometer readings are distorted by metalwork within the building structure, furniture and also magnetic fields from VDUs. The Vector 2X has a built-in hard-iron calibration algorithm. This compensates for magnetic fields generated by a host system such as the housing and electronics. When mounting the compass in the host system, care should be taken to minimize the possible sources of magnetic interference. For instance, mounting screws should be non-ferrous. Unfortunately it is extremely difficult to compensate for dynamic changes in the magnetic field strength. However in practice the device is intended to be used in relatively open areas away from such interference which quickly drops off with distance due to the inverse square law. Another problem with the magnetometer is that it must be kept perfectly level for good results. In order to accurately measure the X and Y components of the magnetic field, the compass needs to be aligned parallel to the surface of the Earth. Typically this is achieved by using a gimbal mount or placing the compass in a bubble of oil allowing gravity to level the compass. Tilting the device gives increasing bearing error per degree of tilt and, due to the Earth's magnetic field, this is an effect that is more pronounced the nearer the magnetic pole the device is used.

3.2 An Immersive 3D Augmented Reality Environment - DAMOCLES.

Another method of representing the virtual models contained on the network is through an immersive 3D environment where the user wears a Head Mounted Display (HMD) and stereoscopic information is produced. An augmented reality overlays image information on the normal field of view of the user much like a head up display. The augmented reality suite presented in this paper is DAMOCLES, so called because the original image used to test the HMD was a large sword which appeared to float above the head of the user.

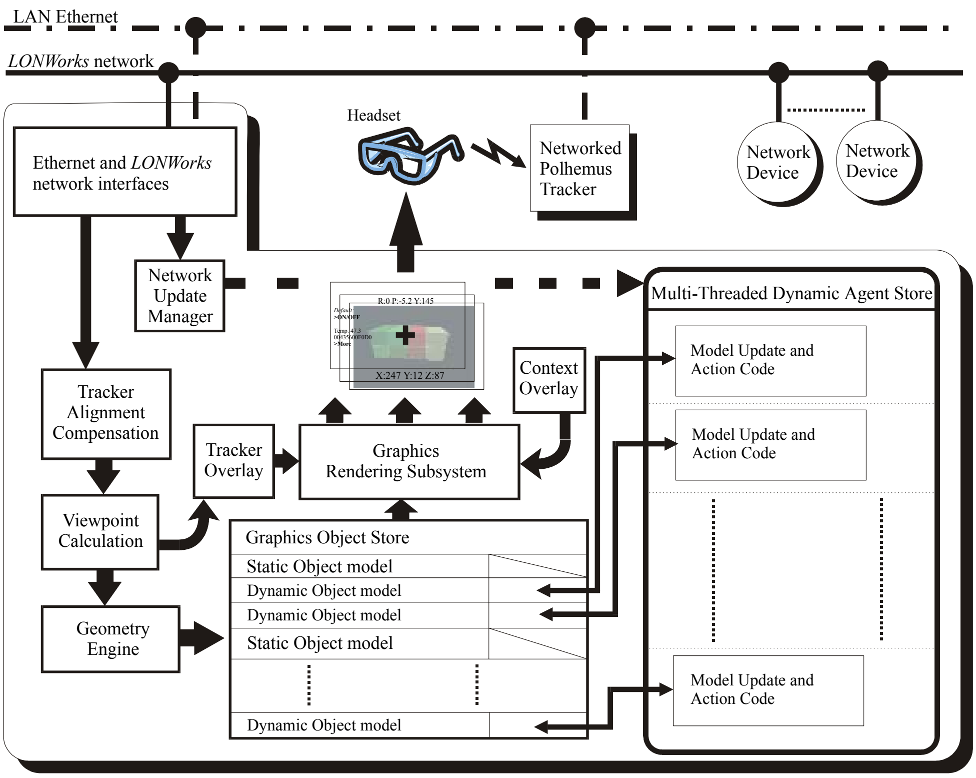


Figure 8. *A block diagram of the DAMOCLES system.*

Augmented reality falls between the two extremes of complete VR and full reality and depending on the system in question there exists a continuum of possible degrees of augmentation that can be offered to the user as discussed by Milgram 1994. This ranges from simple wireframe overlays to full colour enhanced images to present the abstracted display of measured values. A human user is certainly more comfortable with data that only slightly modifies his view of the environment, this approach eliminates the nausea effects often associated with fully immersive environments.

The aim of the DAMOCLES system is to permit a user to be physically present within a real world environment and to obtain enhanced interfaces to a variety of different objects within that world. This is achieved by the use of a stereo-graphic HMD with half silvered optics allowing the user to view overlaid computer generated imagery on the real surroundings. The ultimate goal is to allow a user to freely wander around a built environment in a similar way to the *Touring Machine* proposed by Feiner 1997. However unlike the unidirectional touring machine system DAMOCLES has the requirement for detailed information about devices in the network and the ability to interact with them.

A block diagram of the system is shown in Figure 8. At any point in the network a client computer may be connected, and at the same location a magnetic tracking system is sited, the Polhemus tracker in this case. Given a knowledge of the location of the base point for the tracker the location and orientation of the HMD can be determined. With the current tracking device this is restricted to a radius of 1-2m from the transmitter. There are considerable problems associated with using a magnetic tracker in most locations, not least of which is its spatial calibration to ensure good static accuracy as discussed by Azuma 1995.

Once connected to the network the remote client obtains information from the network about the devices in its location. Unlike the 2D mapping system DAMOCLES uses a more centralised client server architecture rather than a truly distributed approach so the initial transactions are with the central database server. It would be more efficient if all the of the location information could be obtained without referring to the database server. However this would require that the users computer know where it was within the network, and since one of the features of the network is the location transparency of a device, it is simpler to refer to a main database for the initial datum at present. Once the virtual environment is generated it is then displayed upon the HMD with due regard to position and orientation.

Given localisation information from a server DAMOCLES then proceeds to build a graphical model description. The system first takes graphical data from a device constructing an object within the virtual part of the AR. Attached to each device is a description of the dynamic behaviour of that object. This takes the form of code modules or agents that act upon the model description within the AR. If the construction of the object within the virtual world is complicated because of regular changes to the device, then instead the code module contains all the information required to construct a simulacrum of the device, these two scenarios are shown in Figure 9.

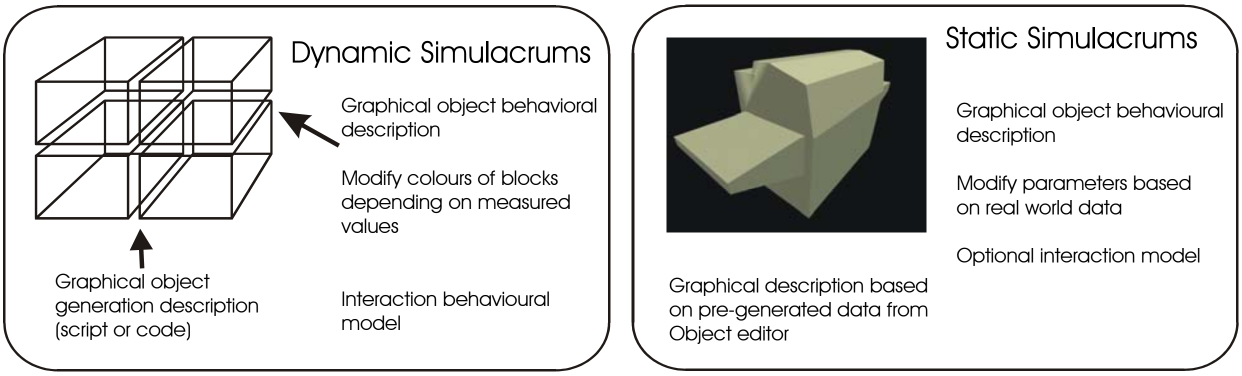


Figure 9. *Graphical object model derivation*

Both OpenGL and DirectX rendering libraries have been used, currently the latter has been selected due to performance considerations on the PCs and the option to easily implement high speed stereoscopic rendering. The choice of rendering system has an influence upon the type of graphical models that can be stored. If fast and efficient model transfer is to take place, size and complexity is dependant on the graphics library being used. Since the models are to be stored on the devices themselves then they must be as compact as possible whilst still conveying the required detail.

After the localised world has been generated from the Local Object Models in each of the nodes in the area interaction can then occur between the user and any devices that support external stimulation via the Local Object Model. Two forms of object update relevant for display purposes may occur. Firstly, changes in the physical world sensed and supplied by a device will entail changes in the graphical object within the virtual part of the augmented reality, for instance doors opening will require angular transformations of their simulacrums. Secondly, multimodal data sets require abstracted information to be generated and displayed. Depending on the type of the measured parameter this can be quite complex. For example displaying a colour representation of black body temperature requires translation into a suitable range and then mapping to some colour. In both cases the modification to the graphical object is handled by a software agent running in a separate thread within the virtual world. The source of this code is either a system database or from the device itself. By breaking the overall system down into compact discrete objects the size of model update agents is small, and the virtual environment does not have to deal with a flood of information trying to update thousands of objects.

When using a mirror world any device in the network that has an actuator attached to it can be controlled, this introduces a number of issues. Safety must be considered, particularly if the users interactions could harm other people or cause the controlled object to go out of limits. Concurrency problems occur when the users requests conflict with the control algorithms already operating within the device. Concurrency is also an issue when multiple users may be interacting with the model and the actual device leading to issues of prioritising control. At the present time restrictions have not been imposed upon what a user of DAMOCLES can do to the controlled environment but there is provision within the existing agent control framework to add priority management and process limiting features.

DAMOCLES was designed to operate with a diverse range of networked devices. Consequently an interaction method that is consistent for all objects has been implemented by using a context sensitive linguistic paradigm. A method whereby each dynamic code agent implements a default verb specific to that device. For instance an electric window opener has a default reflective action that is to open or to close. Highlighting a selected window object within the users field of view selects that device, a button press then invokes the default verb for the object, and in this case the window opens. More complex verb sets are accommodated in the software by permitting a selection to linger on the object, after a delay a list of available actions is presented to the user via the HMD. A consistent approach has been presented to the user that is easily understood whilst allowing flexibility for more complex functions to be implemented.

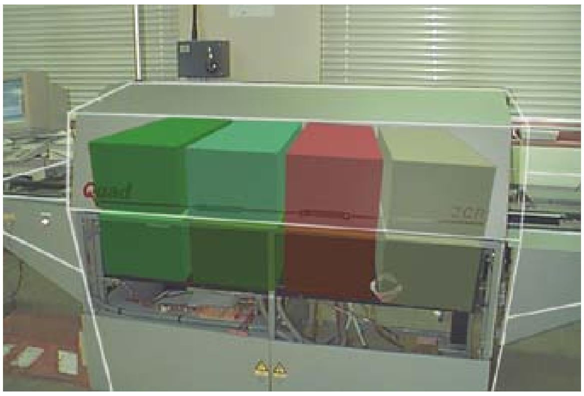


Figure 10. *An AR snapshot of heated zones in an oven.*

Figure 11. *Wearing the DAMOCLES HMD.*

A snapshot of the augmented reality system can be seen in Figure 10, here the system is installed in the Distributed Systems Research Lab. Looking at the network controlled machinery in the lab displays the state of that equipment. Figure 10 showing an oven which has various heated zones with temperatures displayed in a suitable colour. In this case the third zone from the left is actually coloured red, indicating a high temperature being detected by the networked node controlling the device. Note some static registration errors on the wireframe overlay, a result of tracker calibration.

Figure 11 shows a picture of the DAMOCLES HMD in action, the Polhemus tracker is visible mounted above the head of the user. This picture has actually been enhanced with DAMOCLES highlighting of control features, in this case the wooden framed glass door in the background is actually highlighted as a controllable network object.

4. CONCLUSIONS

This paper has presented the use of networked building control systems as an infrastructure to support distributed virtual environments. Two methods of interacting with these distributed models have been discussed, the first being a prototype of a hand held navigational device which could be used by a person with visual impairments as an orientation and navigation device. This device shows considerable potential although there are some basic problems associated with the use of a compass within a building.

The second system presented is the DAMOCLES augmented reality system that offers a novel method of controlling devices within the built environment. The system augments the normal field of view of a user by superimposing additional information derived from the network, and controllable objects can be selected and activated simply by the act of looking at them. Presently the system is a laboratory prototype and considerable development is required before it could be a truly portable device.

5. REFERENCES

R.T. Azuma (1995) "Predictive Tracking for Augmented Reality", PhD dissertation UNC Chapel Hill Dept. of Computer Science technical report TR95-007.

I Bruce, A McKennell and E Walker (1991) *Blind and partially sighted adults in Britain: the RNIB survey*, Volume 1, Published by HMSO

S. Feiner, B. MacIntyre & T. Hollerer, (1997) "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment", *Proc. ISWC '97,* Cambridge, MA, pp. 74-81.

G.T.Foster & J.P.N. Glover. (1997). Supporting Navigation Services Within Intelligent Buildings. *Proceedings of the First Mobinet Symposium.* Athens.

G.T. Foster, D.E.N. Wenn & J.P.N. Glover. (1998a) ARIADNE - Exploiting a new generation of Intelligent Buildings. In *Improving the Quality of Life For The European Citizen. Technology For Inclusive Design and Equality.* pp 388-391. Porrero & Ballabio (eds.) pub. IOS Press 1998.

G.T. Foster. (1998b) "*The use of local modelling techniques to facilitate information and resource management schemes in distributed control systems*." PhD thesis. University of Reading.

D. Gelernter (1992) "*Mirror Worlds*" pub. Oxford University Press. ISBN 0-19-507906.

Gray & Todd.(1967) *Mobility and Reading Habits of the Blind*, Published by HMSO

J.C.Hammond, P.M. Sharkey & G.T. Foster (1996) "Integrating Augmented Reality With Home Systems" *Proc. 1st European Conference on Disability, Virtual Reality and Associated Technologies.* ECDVRAT, Maidenhead, UK, 1996.

P. Milgram & F. Kishino (1994) "A taxonomy of mixed reality visual displays", *IEICE (Institute of Electronics, Information and Communication Engineers) Transactions on Information and Systems*, Special issue on Networked Reality, December.

F.J. O'Hart & G.T. Foster.(1998) "Creating dynamic mobile robot environments from an intelligent building" Proc. *Third European Robotics and Intelligent Systems and Control Conference EURISCON.* Athens, June 1998.

Preliminary findings on a virtual environment targeting human mental rotation/spatial abilities

A Rizzo[1], J G Buckwalter[1], P Larson[2], A van Rooyen[2], K Kratz[2], U Neumann[3], C Kesselman[4] and M Thiebaux[4]

[1]Alzheimer's Disease Research Center, University of Southern California,
3715 McClintock Ave., Los Angeles, CA 90089-0191

[2]Fuller Graduate School of Psychology, 180 North Oakland Ave., Pasadena, CA. 91101

[3]Computer Science Dept., University of Southern California, Los Angeles, CA 90089

[4]Information Sciences Institute, University of Southern California,
4676 Admiralty Way, Suite 1001, Marina del Ray, CA 90292

[1]*arizzo@mizar.usc.edu*

ABSTRACT

Virtual Reality technology offers the potential to create sophisticated new tools which could be applied in the areas of neuropsychological assessment and cognitive rehabilitation. If empirical studies demonstrate effectiveness, virtual environments (VE's) could be of considerable benefit to persons with cognitive and functional impairments due to acquired brain injury, neurological disorders, and learning disabilities. Testing and training scenarios that would be difficult, if not impossible, to deliver using conventional neuropsychological methods are being developed which take advantage of the attributes of virtual environments. VE technology allows for the precise presentation and control of dynamic 3D stimulus environments, in which all behavioral responding can be recorded. A cognitive domain where the specific advantages found in a virtual environment are particularly well-suited, is with human visuospatial ability. Our paper outlines the application of a virtual environment for the study, assessment, and possible rehabilitation of a visuospatial ability referred to as mental rotation. The rationale for the Virtual Reality Spatial Rotation (VRSR) system is discussed, and the experimental design that is being used to collect data from a normal, aged 18 to 40 population is presented. Our research questions are then outlined and we discuss some preliminary observations on the data that has been collected thus far with the system.

1. INTRODUCTION

Virtual reality technology is increasingly being recognized as a potential tool for the assessment and rehabilitation of human cognitive and functional processes (Foreman et al, 1997; Pugnetti et al, 1995; Rizzo and Buckwalter, 1997; Rose, 1996). Virtual environments (VE) allow for the creation of dynamic stimulus environments, in which all behavioral responding can be recorded. This technology could potentially offer testing and training options that are unavailable with the use of conventional neuropsychological methods. It is our belief that computer-generated interactive simulated environments can be used to assess and rehabilitate cognitive abilities, much like an aircraft simulator tests and trains piloting abilities. Flight simulators have been used for over fifty years to train both military and commercial pilots, and the benefits of this technology have been demonstrated (Johnston, 1995). In this regard medical applications that use VE's are now showing promise as a way to train the visualization and procedural skills needed to perform surgery (Satava, 1996). Persons with cognitive and functional impairments due to traumatic brain injury, neurological disorders, and learning disabilities, could also benefit from the advantages of VE-based assessment and rehabilitation. VE's are now being developed and tested which focus on component cognitive processes including: memory (Rose et al, 1997), executive functions (Pugnetti et al, 1998, Mendozzi et al, 1998), spatial skills (Foreman et al, 1997; McComas et al, 1998; Rizzo et al, 1998a), and attentional processes (Brown et al 1997; Wann et al, 1997). VE functional training scenarios have also been designed to test and teach basic activities of daily living such as: street-crossing (Strickland, 1997; Inman et

ISBN 0 7049 1141 8

al, 1997), common object recognition (Strickland, 1997), supermarket shopping (Cromby et al, 1996), use of public transportation (Mowafy and Pollack, 1995), and wheelchair navigation (Inman,1997). VE projects such as these give hope that the 21st century will be ushered in with new and exciting tools to advance a field that has long been mired in the methods of the past.

Our work has focused on the development of a VE for the study, assessment, and rehabilitation of a visuospatial ability referred to as Mental Rotation (MR). Everyday life situations which rely on this ability to use imagery to turn over or manipulate objects mentally are quite common. These include automobile driving judgments, organizing items in limited storage space, sports activities, and many other situations where one needs to visualize the movement and ultimate location of physical objects in 3-D space. High level mathematics performance has also been linked, in large part, to MR ability (Casey et al, 1995). Indeed, in a recent Los Angeles Times interview, it was noted that world renown physicist, Stephen Hawking, "…translates mathematics into geometry, and turns around geometrical shapes in his head." (Cole, 1998). The initial MR investigations began almost 30 years ago with the work of Shepard and Metzler (1971) who tachistoscopically presented pairs of two-dimensional perspective drawings to subjects and required them to make judgments as to whether the 3-D objects they portrayed, were the same or different (see figure 1). A near perfect linear relationship was found between the amount of angle rotation difference between the pairs of objects, and the reaction time to decide whether or not the objects were the same or different. Since precise mathematical relationships between hypothesized mental representations and behavioral performance are relatively rare, MR has been the focus of much research over the years.

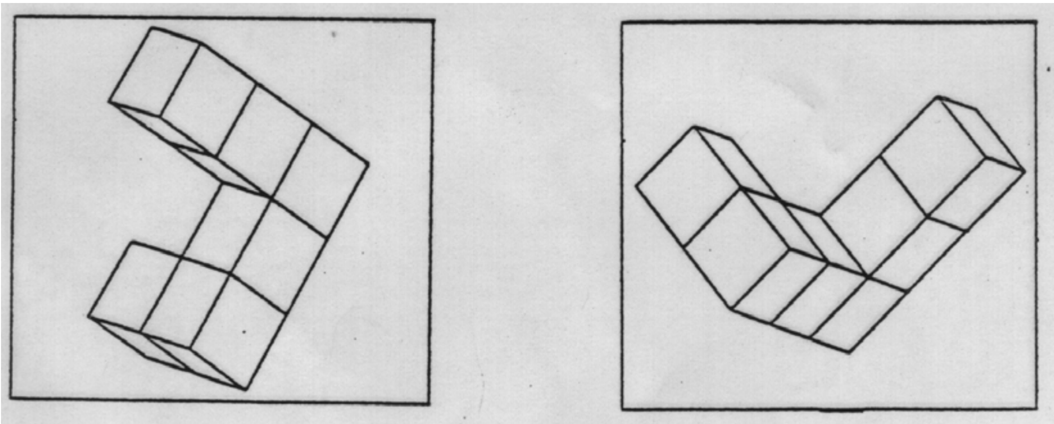


Figure 1. *Mental Rotation Stimuli*

Tests of spatial ability, including the MR variable, have commonly been used for the study of brain/behavior relationships particularly regarding sex differences in cognition. Mental rotation ability has been shown to produce the most consistent and sizable sex differences, in favor of males, in the cognitive literature (Voyer et al, 1995). Consequently, a lively literature has emerged examining MR, in addition to cognitive variables where female advantages have been shown (i.e., verbal fluency and fine motor skills, etc.). Studies have revealed differential cognitive performance due to such hormonal factors as onset of menopause, estrogen and testosterone administration, and stage of the menstrual cycle (Gouchie and Kimura, 1991; Kampen and Sherwin, 1994; Silverman and Phillips, 1993) However, these findings remain controversial. Several studies have attempted to explain cognitive sex differences as the product of sociocultural influences, and on non-specific testing performance factors related to the use of timed tests and "reluctance to guess" factors (Richardson,1994; Qubeck, 1997; Delgado and Prieto, 1996). Also, while it has also been suggested that the effect size in gender differences is decreasing with time, meta-analytic research argues against such conclusions (Masters and Sanders, 1993; Voyer et al, 1995).

Spatial ability is an important domain to consider in the assessment of neurological disorders, traumatic brain injury, and neuropathological conditions of aging. Spatial orientation abilities have been shown to be an important variable in the differential diagnosis of dementia. For example, research indicates that victims of Alzheimer's disease have an 84% incidence of spatial orientation impairments compared to only a 4% incidence in frontotemporal dementia (Miller et al, 1997). Impairments in spatial orientation were also shown to be more common in Alzheimer's disease compared to both normal elderly and those with vascular dementia (Gianotti et al, 1992; Signorino et al. 1996). Similar impairments have been observed following the occurrence of traumatic brain injury and stroke (Lezak, 1995). In light of these issues, (which are on-going research interests at our lab at the USC Alzheimer's Disease Research Center), and our interest in the potential usefulness of virtual technologies, we began development of the Virtual Reality Mental Rotation/Spatial Skills Project.

Traditional measures used for the assessment of mental rotation have produced intriguing findings, yet lack the precision needed to better understand this spatial ability. The most common test uses two-dimensional stimuli that portray three-dimensional objects and requires complete mental processing of the stimuli without any motoric involvement (Vandenberg and Kuse, 1978; Peters et al, 1995). We have developed and are collecting data on a measure of spatial rotation ability that is administered in a VE to more precisely evaluate and possibly rehabilitate this cognitive process. The use of a VE for the assessment of cognitive abilities allows for better standardization of stimulus presentation as well as quantification of multiple characteristics of the stimuli.Further, responses of the subjects can be quantified on a range of characteristics that cannot be evaluated using traditional psychometric instruments. The combination of greater control and description of the stimuli along with more precise measurement of responses should allow for characterization of the cognitive processes involved in spatial skills in a more discrete fashion than is possible with standard measures. Comparison of performance in the VE with performance on standard measures offers the potential to better understand this crucial cognitive ability. Also, by examining changes in spatial performance following VE exposure, useful rehabilitation options may emerge and be developed. This is based on our view that immersive VE-delivered physical rotation training with the MR stimuli could help improve imaginal mental rotation abilities. This assertion is bolstered by a recent study which concluded that rotary object manipulation and mental object rotation share a mutual process that is believed to direct the dynamics of both imagined and actual physical object reorientation (Wohlschlager and Wohlschlager 1998). By conducting future studies on this VE system with the elderly, and persons with brain injury or neurological disorders, the feasibility and effectiveness of this novel technology for assessment and rehabilitation purposes with these groups can be addressed.

The useful application of VE's in the areas of assessment and rehabilitation of cognitive/functional abilities, while intuitively appealing, cannot progress until basic cost/benefit, feasibility, and clinical effectiveness issues are examined. These include factors relating to the selection of appropriate training and target variables, system costs, clinical population characteristics, optimal levels of presence/immersion, interface and navigational demands, side effects, learning and generalization, and data analytic strategies. These issues are more fully explored in other papers, along with detailed descriptions and rationales for VE's addressing psychological and cognitive variables (Rizzo et al, 1998a,b). Our research program has been designed so that many of these issues can be economically addressed, while at the same time, data can be collected regarding our cognitive variable of interest -- visuospatial mental rotation. This approach allows for the investigation of VE specific concerns (side effects, generalization), factored with both clinical applications (assessment and rehabilitation of clinical groups) and general experimental studies (sex difference investigations). This multi-purpose approach was a definate "selling point" in getting acceptance and resources for the development of this system.

The following describes our Virtual Reality Spatial Rotation (VRSR) system and details the experimental design that is being used to collect data from a normal, aged 18 to 40 population. We will outline our research questions and present some preliminary observations on the subjects that have been evaluated with the system. Also, at the time of the conference, it is expected that we will have results available from the full data set.

2. METHOD

2.1 Subjects

Fifty-four subjects (23 males and 31 females) between the ages of 18-40 were tested. Subjects included employees recruited at the Information Sciences Institute of the University of Southern California, graduate students from the Fuller Graduate School of Psychology, and undergraduate students from the University of Southern California and California State University at Los Angeles.

2.2 Virtual Reality System

The Virtual Reality Spatial Rotation (VRSR) system uses an ImmersaDesk drafting table format virtual prototyping device. The Pyramid Systems ImmersaDesk employs stereo glasses and magnetic head and hand tracking. This projection-based system offers a type of VR that is semi-immersive. It features a 4 X 5-foot rear-projected screen positioned at a 45 degree angle. The size and position of the screen give a wide-angle view and the ability to look down as well as forward.

The VRSR assessment and training system was designed to present a target stimulus that consists of a specific configuration of 3D blocks within a virtual environment (similar to Figure 1). The stimuli appear as "hologram-like" three dimensional objects floating above the projection screen. After presentation of a target stimuli, the participant is presented with the same set of blocks (control object) that needs to be rotated to the orientation of the target and then superimposed within it. The participant manipulates the control object by grasping and moving a sphere shaped "cyberprop" which contains a tracking device and provides tactile feedback. Upon successful superimposition of the control and target objects a "correct" feedback tone is presented and the next trial begins.

2.3 Procedures

The experimental sessions take place over a two hour period. After informed consent is obtained, basic demographic information, computer experience and usage, and spatial activities history (Newcombe et al, 1983) are recorded. Female subjects complete a brief survey of reproductive history. Next, a baseline measure of mental rotation ability is assessed using a redrawn version (Peters et al, 1995) of the Mental Rotation Test (MRT-A) of Vandenberg & Kuse (1978), a twenty item, 2-dimensional paper and pencil task. Subjects then complete a comprehensive neuropsychological battery administered under standard conditions. Following the completion of the neuropsychological battery, subjects complete the Motion History Questionnaire (Kennedy and McCauley, 1984) and Simulator Sickness Questionnaire (Kennedy et al, 1993), which includes a pre-VE exposure symptom checklist. Experimental subjects then participate in the fifteen minute VE task that both assesses and trains mental rotation abilities. After 5 non-rotational practice trials, each subject's VE spatial rotation baseline performance is assessed over 20 trials using a VE version of the items from the pencil and paper MRT. Next, 100 training trials of increasing stimulus complexity are administered. After a one minute break, the original 20 VE MRT trials are administered again to measure changes in VE spatial rotation ability. Control subjects are given a filler task (crossword puzzle) of matching duration instead of the VE exposure. The Simulator Sickness Questionnaire, which contains a post-VE exposure symptom checklist is then given to each subject. Finally, an alternate form of the paper and pencil MRT is administered to assess changes in mental rotation performance.

2.4 Testing Instruments

The neuropsychological battery included a diverse collection of instruments. Mental rotation ability is assessed using the Mental Rotation Test. This test uses line drawings of block stimuli and consists of two 10-item sections in which the subject is required to match two of the four choices to a target figure. Incorrect choices are mirror images of the target or alternative block configurations. Standard administration provides for a five minute time limit. The alternate form of the MRT uses the same drawings but reorders their presentation and switches position of the target stimuli.Verbal attention and mental control is assessed with the Digit Span Forward and Backward test from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981). Visuoconstruction abilities are measured by the Block Design subtest of the WAIS-R. The Trail-Making Tests A and B are used to evaluate executive control processes and attention (Army Individual Test Battery, 1944). The Judgment of Line Orientation test is used to evaluate visuoperceptual skills (Benton et al, 1978). The California Verbal Learning Test is employed to assess verbal learning and memory (Delis et al, 1983). Nonverbal memory is evaluated by the Visual Reproduction subtest of the Wechsler Memory Scale-Revised (Wechsler, 1987). These tests are all commonly used for neuropsychological assessment of these cognitive processes and as such have widely used normative information available. Finally, surveys of simulator sickness and motion sickness history are administered.

2.5 Data Analysis and Research Questions

We have collected data from a variety of domains. These included: 1. Neuropsychological performance on tests of cognitive functioning (attention, verbal and visual memory, visuospatial abilities, etc.); 2. Demographic factors (education, gender, reproductive history, etc.); 3. Spatial Activity History (a self-report scale of participation in everyday activities that contain spatial components); 4. Computer Usage History Questionnaire (a self-report measure that we have developed which assesses computer use, programming activities, use of computer games, etc.); 5. Side effects assessment; and, 6. VE data: all movement is digitized in real time, allowing for playback of each response. While we anticipate developing more sophisticated analytic techniques, we are currently analyzing time to completion per trial, ratio of actual movement path to optimal movement path as a measure of efficiency, and various compiled measures, such as the total time for first 20 VRSR items vs. last 20.

From this data, we will attempt to answer the following research questions:

1. What is the level of side-effects that occur with use of the VRSR system and are there sex differences?
2. Is the occurrence of side effects low enough to justify a future VRSR trial with elderly subjects, persons with dementia, persons with traumatic brain injury, and individuals with other neurological impairments?
3. What is the relationship between various performance measures on the VRSR system and performance on standard neuropsychological tests of attention, memory, and other visuospatial variables?
4. How well does the paper and pencil MRT predict performance on the VRSR system and how does this vary contingent on how VRSR performance is quantified (i.e., total time vs. efficiency ratio)?
5. Do the same sex differences that are seen on the pencil and paper MRT appear on VRSR performance?
6. In women, do these performances vary contingent upon hormonal differences due to day of the menstrual cycle?
7. Will advanced data collection methods of the VRSR system enable us to delineate common gender specific strategies for spatial rotation?
8. Does VRSR performance improve with practice (100 training trials) as seen by comparing 20 pre-training VR MRT items with 20 identical post-training VR MRT items (intra-method generalization)?
9. Does VRSR training improve post-training pencil and paper MRT performance in participants who score low on the pretest MRT compared to practice effects in the control group. (inter-method generalization)?
10. If training and transfer effects are found (questions 8 and 9), are there sex differences?
11. Does history of self-reported computer usage play a role in the above?
12. Does history of self-reported spatial activities influence the above?

3. RESULTS AND CONCLUSIONS

While we are still collecting the data, a few anecdotal observations can be made. There were minimal negative side effects reported by participants in the VRSR condition. Of those reported, they have been mainly related to fatigue. Our control group has made similar reports and this may be primarily due to the "mentally taxing" nature of the battery of neuropsychological tests administered. This observation provides encouragement for the future use of this system with the elderly and with neurologically impaired groups.

Regarding generalization issues, it appears that for low scorers on the MRT pretest (scores less than 20 out of a possible forty), the VRSR condition produces higher gains on the post MRT compared to practice effects in the control group on this second administration of the equivalent form MRT (mean =+ 9.3 for VE vs. + 2.3 for controls). This result was significant ($p < .05$) with the 19 participants in this subset of our sample. However, this observation needs to be interpreted with caution as it is derived from an early exploratory analysis on a small number of participants. Statistical tests of significance on the other questions outlined above are now being conducted on our data set. Following data analysis of these samples, and contingent upon the continued minimal occurrence of side effects, we will begin running a normal elderly (age 65+) group through the system. Again, following side effect and data analysis of the aged sample, persons with Alzheimer's disease, and a brain injury group will be tested. It is hoped that this measured approach to applying VE technology to these groups will lead to the development of safe, new, and useful assessment, diagnostic, and rehabilitation strategies. The complete results of our first study will be available at the time of presentation to the conference.

4. REFERENCES

Army Individual Test Battery (1944). *Manual of Directions and Scoring*. War Department, Adjutant General's Office, Washington D.C.

A L Benton, NR Varney and K S Hamsher (1978), Visuospatial judgment: a clinical test, *Archives of Neurology*, **35**, pp. 364-367.

D J Brown, S Kerr and J R Wilson (1997),Virtual environments in special-needs education, *Communications of the ACM,* **40**, *8*, pp. 72-75.

M B Casey, R Nuttall, E Pezaris and C.P Benbow (1995), The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples, *Developmental Psychology,* **31,** pp. 697-705.

K C Cole (1998), Hawking's universe is open and shut, *Los Angeles Times*, March 14th, 1998, pp. A1, A19.

J J Cromby, P J Standen, J Newman and H Tasker (1996), Successful transfer to the real world of skills practised in a Virtual environment by students with severe learning difficulties*, Proceedings of the European Conference on Disability, Virtual Reality and Associated Technology,* Maidenhead, UK, pp.103-107.

A Delgado and G Prieto (1996), Sex differences in visuospatial ability: Do performance factors play such an important role?, *Memory and Cognition,* **24,** *4*, pp. 504-510.

D C Delis, J H Kramer, E Kaplan and B Ober (1983), *California Verbal Learning Test*. Psychological Corporation, New York.

N Foreman, P Wilson and D Stanton (1997), VR and spatial awareness in disabled children. *Communications of the ACM,* **40**, *8* , pp. 76-77.

G Gainotti, V Parlato, D Monteleone and S Carlomagno (1992), Neuropsychological markers of dementia on visual-spatial tasks: a comparison between Alzheimer's type and vascular forms of dementia, *Journal of Clinical and Experimental Neuropsychology,* **14,** pp. 239-252.

C Gouchie and D Kimura (1991), The relationship between testosterone levels and cognitive ability patterns, Psychoneuroendocrinology, **16**, pp. 323-324.

D P Inman, K Loge and J Leavens (1997), VR education and rehabilitation*, Communications of the ACM,* **40,** *8*, pp. 53-60.

R Johnston (1995), Is it live or is it memorized?, *Virtual Reality Special Report,* **2**, *3*, pp. 53-56.

D L Kampen and B B Sherwin (1994) Estrogen use and verbal memory in healthy postmenopausal women. *Obstetrics and Gynecology,* **83**, pp. 979-983.

R S Kennedy and M E McCauley (1984), *The Motion History Questionnaire,* Essex Corporation, Orlando, FL.

R S Kennedy, N E Lande, K S Berbaum and M G Lilienthal (1993), Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology,* **3**, pp. 203-220.

M D Lezak (1995), *Neuropsychological Assessment,* Oxford University Press, New York.

M S Masters and B Sanders (1993) Is the gender difference in mental rotation disappearing?, *Behavior Genetics,* **23**, *4*, pp. 337-341.

J McComas, J Pivik and M Laflamme (1998) Children's transfer of spatial learning from virtual reality to real environments, *CyberPsychology and Behavior*, **1**, *2*.

L Mendozzi, A Motta, E Barbieri, D Alpini and L Pugnetti (1998), The application of virtual reality to document coping deficits after a stroke: Report of a case, *CyberPsychology and Behavior*, **1**, *1,* pp. 79-91.

B L Miller, C Ikonte, M Ponton, M Levy, K Boone, A Darby, N Berman, I Mena and J L Cummings (1997), A study of the Lund Manchester research criteria for frontotemporal dementia: clinical and single-photon emission CT correlations, *Neurology,* **48**, pp. 937-942.

L Mowafy and J Pollack (1995), Train to travel, *Ability,* **15**, pp. 18-20.

N Newcombe, MM Bandura and DG Taylor (1983) Sex differences in spatial ability and spatial activities, *Sex Roles,* **9**, *3*, pp. 377-386.

M Peters, B Laeng, K Latham, M Jackson, R Zaiyouna and C Richardson (1995), A redrawn Vandenberg and Kuse Mental Rotations Test: different versions and factors that affect performance, *Brain and Cognition,* **28**, pp. 39-58.

L Pugnetti, L Mendozzi, A Motta, A Cattaneo, E Barbieri and S Brancotti (1995), Evaluation and retraining of adults' cognitive impairments: Which role for virtual reality technology?, *Computers in Biology and Medicine,* **25**, *2*, pp.213- 227.

L Pugnetti, L Mendozzi, E A Attree, E Barbieri, B M Brooks, C L Cazzullo, A Motta and F D Rose (1998), Probing memory and executive functions with virtual reality. Past and present studies, *CyberPsychology and Behavior*,**1**,*2.*

W J Qubeck, (1997), Mean differences among subcomponents of Vandenberg's Mental Rotation Test. *Perceptual and Motor Skills,* **85**, *1*, pp. 323-332.

J T Richardson (1994), Gender differences in mental rotation. *Perceptual and Motor Skills,* **78**, *2*, pp. 435-448.

A A Rizzo and JG Buckwalter (1997), Virtual reality and cognitive assessment and rehabilitation: The state of the art, In *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical, and Methodological Issues in Assessment and Rehabilitation* (G Riva, Ed.), IOS Press, Amsterdam, pp.123-146.

A A Rizzo, J G Buckwalter, U Neumann, C Kesselman and M Thiebaux (1998), Basic issues in the application of virtual reality for the assessment and rehabilitation of cognitive impairments and functional disabilities, *CyberPsychology and Behavior,* **1**, *1*, pp. 59-78.

A A Rizzo, M Wiederhold and J G Buckwalter (1998), Basic issues in the use of virtual environments for mental health applications, In *Virtual Environments in Clinical Psychology: Scientific and Technological Challenges in Advanced Patient-Therapist Interaction* (G Riva and B Wiederhold Eds.), IOS Press, Amsterdam, in press.

F D Rose (1996), Virtual reality in rehabilitation following traumatic brain injury*, Proceedings of the European Conference on Disability, Virtual Reality and Associated Technology,* pp. 5-12.

F D Rose, E A Attree and B M Brooks (1997), Virtual environments in neuropsychological assessment and rehabilitation. In *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical, and Methodological Issues in Assessment and Rehabilitation*, (G Riva, Ed.), IOS Press, Amsterdam, 147-156.

R M Satava (1996), Medical Virtual Reality: The current status of the future*. Proceedings of the Medicine Meets Virtual Reality 4 Conference*, San Diego, CA., pp. 100-106.

R N Shepard and J Metzler (1971) Mental rotation of three-dimensional objects, *Science,* **171**, pp. 701-703.

M Signorino, E Pucci, E Brizioli, G Cacchio, G Nolfe and F Angeleri (1996), EEG power spectrum typical of vascular dementia in a subgroup of Alzheimer patients, *Archives of Gerontology and Geriatrics,* **23**, pp. 139-151.

I Silverman and K Phillips. Effects of estrogen changes during the menstrual cycle on spatial performance, *Ethology and Sociobiology,* **14**, pp.257-270.

D Strickland (1997), Virtual reality for the treatment of autism In *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical, and Methodological Issues in Assessment and Rehabilitation* (G Riva, Ed.), IOS Press, Amsterdam, pp. 81-86.

S G Vandenberg and A R Kuse (1978), Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills* **47**, pp. 599-604.

D Voyer, S Voyer, and MP Bryden (1995), Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables, *Psychological Bulletin*, **117**, pp. 250-270.

J P Wann, S K Rushton, M Smyth and D Jones. Virtual environments for the rehabilitation of disorders of attention and movement, In *Virtual Reality in Neuro-Psycho-Physiology: Cognitive, Clinical, and Methodological Issues in Assessment and Rehabilitation*, (G Riva, Ed.), IOS Press, Amsterdam, pp.157-164.

D Wechsler (1981), *WAIS-R Manual,* The Psychological Corporation, New York.

D Wechsler (1987), *Wechsler Memory Scale-Revised*, The Psychological Corporation, New York.

A Wohlschlager and A Wohlschlager (1998), Mental and manual rotation, *Journal of Experimental Psychology,* **24,** *2*, pp. 397-412.

Virtual reality in vestibular diagnosis and rehabilitation

D Alpini[1], L Pugnetti[1], L Mendozzi[1], E Barbieri[1], B Monti[1] and A Cesarani[2]

[1]Scientific Institute S. Maria Nascente, Fondazione don Gnocchi, via Capecelatro 66, 20148 Milan, ITALY
[2]University of Sassari, Institute of Otolaryngology, via Mancini 5, Sassari, ITALY

[1] *melau@tin.it*

ABSTRACT

While vestibulo-oculomotor and vestibulo-spinal functions are usually investigated by means of electronystagmography and stabilometry, environmental exploration and navigation cannot be easily studied in the laboratory. We propose that virtual reality (VR) can provide a solution, especially for those interested in the assessment of vestibular influence over spatial cognitive activities. Subjects exposed to immersive VR show rotatory behaviors during exploration that are the result of both a lateralized vestibular dominance and of the interplay with ongoing cognitive activity. The effect of vestibular dominance over exploratory behavior disappears in non-immersive VR conditions, but certain patterns of exploratory movements still seem to be associated to cognitive performance. On these grounds, we propose the use of VR to improve current techniques of vestibular rehabilitation based on visual feedback. We describe a new equipment that combines the functions of a digital stepping analyzer with those of a PC VR workstation. The patient controls the navigation of virtual environments by means of appropriate displacements of the center of gravity. The combination of a closed-loop feedback to control displacements of the center of gravity and active exploration of the environments makes an otherwise static exercise contingent on a viridical representation of spatial navigation.

1. INTRODUCTION

Human equilibrium is a complex psycho-sensory-motor function which serves a number of purposes, such as exploration of the environment (using visual informations via oculomotor control), maintenance of a desired body position against the force of gravity (posture), and movement within the environment against the force of gravity (walking, running, jumping,..).

The role of the vestibular system in equilibrium is to integrate multisensorial cues from the eyes (Kohen-Ratz, 1977), the labyrinths and proprioreceptors and to provide a continuous neural representation of both segmental body orientation, i.e. the position of body segments relative to one another, and orientation of the body within the environment (spatial orientation) (Andersen et al., 1993). Orientation is, therefore, the basis of any correct motor behaviour, including the exploration of the environment. In fact, the latter can be conceived as essentially an ocular-motor affair in which the head and the body help overcoming the constraint of a narrow field of view. Our abilities to localize objects in a space, orient relative to them, move toward or away from them, reach and manipulate them, all depend critically on receiving, processing and integrating spatial information from gravity, the visual field, and our own body. The vestibular system provides the multisensorial integration that serves to maintain a stable map between spatial localization, spatial orientation, and physical space (McNaughton, 1987).

While vestibulo-oculomotor and vestibulo-spinal functions are usually investigated by means of electronystagmography and stabilometry, environmental exploration cannot be easily studied in the laboratory. Virtual reality (VR) can provide a solution for the assessment of vestibular influence over spatial cognitive activities such as orientation and navigation.

This paper introduces ongoing studies dealing with vestibular-mediated behavior during the exploration of virtual environments and outlines the rationale for the use of VR as an assistive tool to retrain vestibular function.

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

2 .VESTIBULAR ACTIVITY AND THE EXPLORATION OF VIRTUAL ENVIRONMENTS

Although the interaction that can take place in a virtual environment (VE) is still limited to a small variety of relatively simple motor activities, an important advantage is that the latter can be planned and executed as in a real space. The activity of the vestibular system is generally discussed in the framework of immersive VR experiments (Stanney et al., 1998). However, some features of exploratory behavior in non-immersive conditions could also be interpreted as the expression of vestibular mechanisms. This hypothesis is based on the non-randomness of turning behaviors of healthy and impaired subjects in different experimental settings.

A simple manouvre is used clinically to show that vestibular activity is lateralized. Subjects are asked to perform a simple 180° rotation from a standing upright position. This movement has no explicit purpose and is performed in the real environment and repeated a few times. The direction (rightward or leftward) of each rotation is noticed. As this test was performed by a group of 50 healthy individuals and by two matched groups (n. 25 subj. each) of adults suffering from right or left vestibular dysfunction, subjects with normal vestibular function showed a clearcut preference for rightward rotations, which we interpreted as a sign of lateralized vestibular dominance. This conclusion was supported by the finding that in subjects with left vestibular hypofunction the preference for rightward rotations was less but still visible, while in subjects with right vestibular hypofunction a prevalence of leftward rotations was observed .

The exploration of a VE wearing a headset connected to a tracking device (immersive VR setup) provides a unique opportunity to study spontaneous head and body rotations performed during the course of a cognitive activity. In a recent experiment (Alpini et al., 1996) 39 healthy righthanded subjects, aged 22 to 43 were asked to carry out a cognitive task in an immersive VE. Subjects sat on a revolving chair in order to avoid postural stance preferences and muscle fatigue during the VR session. Sideward rotation of the head and/or the whole body triggered a counterclockwise displacement of the visual image to simulate the natural change in viewing angle and to allow a 360° exploration of the environment. The latter was made of decagonal rooms connected by corridors. Each room had 5 doors of different shape and color located on alternate walls. We asked subjects to select an exit door only after they had carefully observed the entrance door. Hence, as they entered a new room, they had necessarily to turn 180° to observe the door which they had just passed through and then turn again to make the appropriate selection. To do this, subjects were free to rotate clockwise (rightward) or counterclockwise (leftward). Each subject explored a total of 32 rooms, which took an average of about 30 minutes. The first two body rotations after entering each room were classified as rightward or leftward turns. The analysis showed a prevalence of clockwise movements (rightward) on both the first and the following (back) movement. The majority of events occurred as a combination of back and forth turning movements (Right-Left and Left-Right). Only a few instances of right-right (R-R) rotations and even less of left-left (L-L) rotations were seen. We have then looked at turning behavior. Subjects were then divided into three groups according to their cognitive performance: good (9 subjects), average (15 subjects), and bad performance (15 subjects). Subjects in the first group were both fast and accurate while subjects in the third group were the slowest and made many selection errors. Subjects in the first group showed a clearcut preference for leftward rotation on the first movement followed by a rightward return. Subjects in group 2 did not attain a "preferred" turning behavior until room 15, whereas subjects in group 1 did it much earlier. Group 3 subjects did not show a clearcut preference for any of the two heterologous turning behaviors. As for the clinical diagnostic manouvre, rotations in the VE were not guided by the necessity to explore, but rather to follow an instruction. In other words, subjects got important visual information for the task they were carrying out only after rotation was over, not <u>during</u> rotation. This experiment suggested that immersive VR should be further investigated as a tool to assess in the laboratory vestibular function during exploratory activity. It also confirmed the interplay between ongoing cognitive activity and seemingly pure automatic behaviors (preselected sensory-motor schemata) such as turning backwards.

The use of an immersive VR setup, however, may not be safe or practical for some subjects (Stanney et al., 1998). Cybersickness is an obvious contraindication. Therefore, in a further experiment we asked 26 healthy subjects to explore four rooms of a virtual house by handling a joystick while watching a 17” color PC monitor (non-immersive VR) (Pugnetti et al., 1998); this setup is known to reduce the risk of VR side-effects (Stanney et al., 1998). The task was to search for an object which, in fact, was not there. In this condition, the exploration is based on pure ocular-motor strategies that can be inferred from the direction of rotation of the virtual environment caused by lateral bends of the handle of the joystick. This, of course, excludes any direct stimulation of the labyriths and neck proprioceptors and has some analogy with the exploration of space which occurs during smooth driving. Nine of the subjects explored first the right hemispace after entering each of the four rooms, 10 subjects preferred a leftward exploration, while 7 subjects made rightward and leftward explorations in various combinations. The latter showed a significantly better recall of objects present in the virtual rooms than the former two groups.

Though still incomplete, these findings suggest that the lack of a direct involvement of the afferent systems involved in head turning reduces the expression of a lateralized vestibular dominance; another hypothesis worth of further study is that smooth shifts of the field of view performed to get continuous visual information (true exploratory rotations) also reduce the expression of vestibular dominance, i.e. visual processing takes precedence over orientation. However, in both immersive and non-immersive VR conditions there seems to be an interaction between certain rotation patterns (i.e. vestibular activity) and cognitive performance (Alpini et al.,1996).

3.VIRTUAL REALITY IN VESTIBULAR REHABILITATION

Vestibular rehabilitation is a well known technique that combines physical and instrumental exercises to reduce vertigo and dizziness caused by a vestibular dysfunction. Instrumental rehabilitation is usually based on stabilometry: a platform records the shifts of the center of gravity (CoG) of a standing subject who is made aware of them and improves his static equilibrium through visual feedback on a computer screen (Kohen-Ratz, 1977). The assumption is that when the patient regains control over his own CoG, then he will be able to transfer this benefit to the largely automatic control of CoG necessary during dynamic activities of daily life. This is not always true. Navigation, for example, is a complex function that requires the continuous matching of internal and external landmarks in an automatic and unconscious way. Usually we have no conscious representation of the CoG as a "landmark", but we know that we are along the correct way because we compare the internal references (with special regards to the direction of gravity) with external visual references (Wapner and Witkin, 1950): a door, a wall, the horizon,... It follows that navigation strategies are different in closed and in open spaces and that the qualitative aspects of an environment such as lighting conditions, colours and obstacles influence spatial orientation and navigation.

Virtual reality represents a more specific tool to approach vestibular rehabilitation from an instrumental (feedback) perspective because it mimicks the experience of navigation and spatial orientation. Furthermore, VR allows motor performance (body control of the displacements in the virtual space) and cognitive performance (i.e. active exploration of the environments) to be combined in a variety of meaningful ways.

Here we outline the characteristics of the equipment we have used to get preliminary insights into this potential new VR application.

3.1 .Hardware

We combined a low-cost immersive VR workstation with a digital stepping analyzer (D.S.A.) to detect of head and trunk movements (Fig. 1). A modified joystick is attached to a platform on the floor and subjects control their movements inside the VE by voluntary displacements of their CoG while standing close to the vertical of the joystick. The sensitivity of the joystick to sideward bending of its modified handle can be calibrated and adjusted by an external command box. This allows to personalize the patient/VR interaction. The D.S.A has been developed by the Bioengineering Department of the don Gnocchi Foundation and is composed by a solid state camera placed 120 cm above the standing subject and providing a compound video signal at 50 Hz with a resolution of 625 lines. The camera is synchronized with the other components through a Video Sync Generator. The subject wears four refrangent 1 cm diameter hemispheres on the rignt and left shoulders, and on the front and the back of the head. The camera is equipped with a solid state sensor (CCD) with high black/white resolution and a 3.5 mm lens. Between the sensor and the lens an InfraRed filter (870 nm) cuts the visible components of the images. An InfraRed Flash with 120 LD IR and a 880nm emission makes the CCD sensitive only to IR images reflected by the markers. The video-images of the subjects and of the markers are then filtered and resynchronized. The resulting signals are elaborated by three counters (X,Y,D) connected and synchronized by a common clock. The counters provide X/Y coordinates for the first over-threshold pixel referred to each marker and the duration (D) of its state. X,Y and D informations are sequentially ordered and divided by a Multiplexer. In this way they are stored in a FIFO (First In First Out) buffer as paked 8 bit units. A Universal Asynchronous Receiver Transmitter (UART) connected with a standard serial RS232 port of a Pentium 100 PC, allows the transfer of data from the buffer to the PC at a maximum rate of 230.400 bit/sec. A specific C++ program running under Windows 95 controls the raw data flux from the DSA, the dynamic display during acquisition, and computes parameters and stores results in a database. Each trial requires at least 2Mb of HD memory to store the raw data.

The system proved to reliably recognize displacements of the head and shoulder markers in different conditions: standing, stepping, or performing any complex head and trunk movements, such as those required to navigate a virtual environment. Therefore, the system should make it possible to compare times and travelled distances in different subjects or in the same subject before and after vestibular rehabilitation.

Comparisons are based on the following main parameters:

- the lateral and longitudinal components of the velocity of each marker
- the absolute mean velocity of each marker (cm/sec)
- the lateral and longitudinal components of the displacements of each marker
- the total distance travelled by each marker; it is an estimate of the displacement of the subject into the virtual world
- the global displacement, which measures the absolute displacements of the subject in the real world

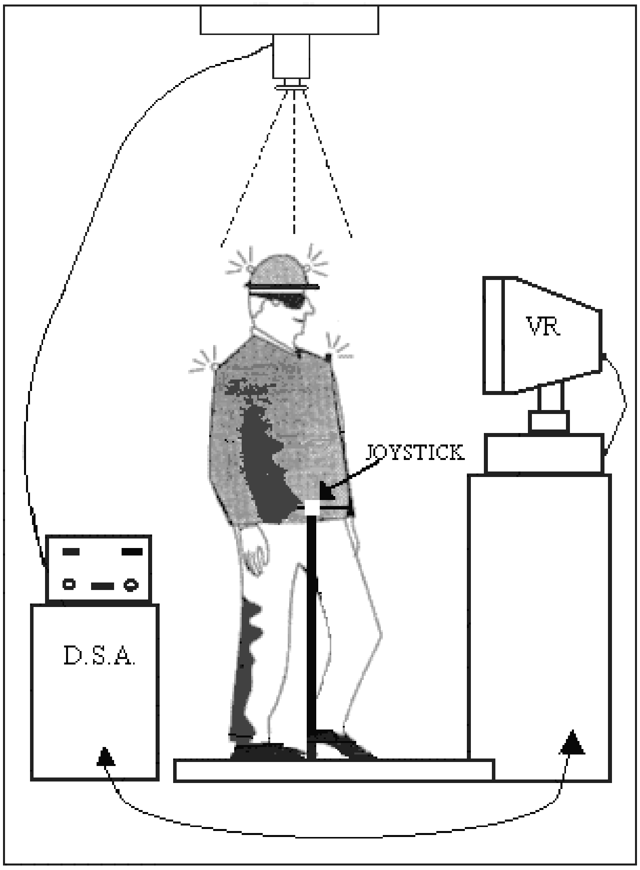


Figure 1. *A schematic drawing of the equipment for vestibular rehabilitation*

3.2 . VR Software

The two VEs currently used for vestibular rehabilitation were both developed under Superscape VRT. The first program features a green landscape with a few landmarks and scattered arrows that indicate the direction to travel. By displacing his/her CoG along antero-posterior and lateral directions by appropriate utilization of ankle-hip strategies, the trainee must follow an invisible path that connects an arrow with the next as quickly and accurately as possible. This environment stimulates CoG control while the subject is "walking" distances within the VE. Thus, CoG controlled displacement is made dynamically contingent on purposeful movement and vice-versa, as it is in real walking. This is expected to make an individual more confident on his/her ability to control shifts of the CoG and to be better able to generalize the benefit to real conditions.

The second program features an indoor path with rooms of variable dimension, obstacles to negotiate and corridors of variable lenghth. Some of the environments are painted with high contrast vertical stripes or have squared patterns on the floors to induce a "natural" optokinetic stimulation during movement. In order to navigate this VE the trainee needs to exert a more accurate control over CoG and learn to ignore the interference of optokinetic stimulation. The data of two healthy subjects F. (female 27 yrs) and D. (male, 40 yrs) performing on the outdoor virtual track are reported in Table 1.

Table 1. *Velocities and displacements of each marker in two healthy controls performing on the outdoor VR path*

F., female, 27 yrs.

Marker	Mean velocity (cm/sec)	lateral velocity (cm/sec)	longitudinal velocity (cm/sec)	longitudinal path (m)	lateral path (m)	total distance (m)	global displacement (m)
Front	19.36	15.00	7.82	0.86	1.65	2.13	0.09
Rear	2.75	1.66	1.25	1.06	1.41	2.34	0.06
Left	3.59	1.73	2.13	1.51	1.23	2.55	0.08
Right	4.91	2.81	2.79	1.20	1.21	2.11	0.05

D., male, 40 yrs.

Front	14.00	11.00	5.73	1.72	3.30	4.20	0.02
Rear	8.40	5.24	4.61	2.86	3.25	5.21	0.08
Left	5.07	2.78	2.85	3.51	3.42	6.24	0.06
Right	4.62	2.67	2.50	3.25	3.47	6.00	0.07

Table 2. *Velocities and displacements of each marker in an MS patient before and after vestibular rehabilitation*

FIRST TRIAL

Marker	Mean velocity (cm/sec)	lateral velocity (cm/sec)	longitudinal velocity (cm/sec)	longitudinal path (m)	lateral path (m)	Total distance (m)	global displacement (m)
Front	8.75	6.03	3.94	11.47	17.54	25.45	0.12
Rear	7.84	4.69	4.32	4.67	5.07	8.47	0.14
Left	6.59	4.05	3.36	10.74	12.96	21.09	0.13
Right	6.89	3.90	3.91	12.09	12.04	21.29	0.14

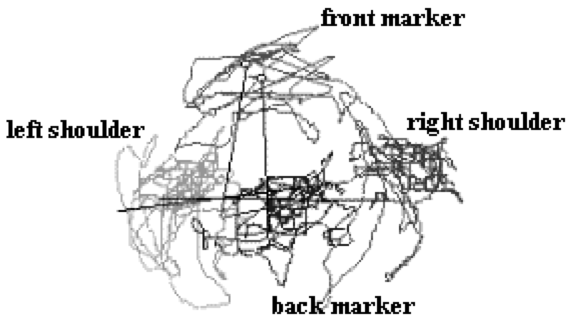
II TRIAL

Front	12.22	9.11	5.30	4.35	7.47	10.02	0.11
Rear	4.76	2.69	2.52	3.71	3.95	7.00	0.07
Left	8.63	3.85	6.09	5.54	3.50	7.85	0.11
Right	5.76	2.84	3.63	4.43	3.47	7.03	0.17

In both subjects the front marker moves faster as if the movement of the head was performed with a posterior fulcrum while the velocities of the right and left shoulders were comparable. The velocities in both the longitudinal and lateral directions are also similar: the subjects remained on the spot (low global displacement values) while bending in both directions. However, D. travelled a longer distance in the VE than F.; i.e. he had more difficulties in the control of his shifts of CoG.

In Table 2 the data of a patient with multiple sclerosis (M., female 51 yrs. old) who was trained with the same VR program are reported. Figure 2 compares the raw traces of a healthy subject with those of the patient during her VR retraining sessions. The first trial was recorded at the beginning of treatment, the second after a period of two weeks of rehabilitation, including physical and traditional instrumental techniques. In the first trial velocities were generally lower than those of healthy subjects (compare to Table 1). The movements of the head and the trunk were similar: i.e. the patient was not able to dissociate the head from the trunk, and she moved her head more laterally than longitudinally. The global displacement was also higher than in controls (she could not avoid stepping away on the platform) and the distance travelled in the VE (total distance) was higher than expected.

In the second trial, the velocities of the shoulder markers were not modified, but the movements of the head and trunk became more independent, as shown by an increase of front velocities values and a contemporary decrease of the back values. Furthermore, the total distance travelled in the VE was smaller than in the first trial. We suggest that vestibular rehabilitation allowed the MS patient to optimize her head/trunk and ankle\hip strategies to control her travel along the same virtual path.



Healthy control (40 yrs.)

MS patient (51 yrs.), 1st trial

MS patient (51 yrs.), 2nd trial

Figure 2. *Trajectories of the 4 markers during vestibular VR training sessions with the program simulating an outdoor path.*

4. CONCLUSIVE REMARKS

Consistent with the definition of equilibrium as a psycho-motor performance in which the vestibular system provides a continuous integration of multisensorial cues, modern assessment of vestibular dysfunction and vestibular rehabilitation should benefit from methodologies based on psycho-motor multisensorial stimulation.

The human equilibrium system derives much of its strength and plasticity from the influence of permanent stimulation from the world we live in. Failures of the sensory inputs as well as of the central equiliubrium regulation may lead to vertigo, nausea and vomiting, blurred vision, nystagmus, head and body instability, changes in cardiac rythm, and metabolic alterations. The system has many inborn possibilities for internal stabilisation and compensation (Smith and Darlington, 1991). This term describes a type of central nervous counterregulation which occurs as a result of functional damage. It utilises supplementary functions which "mask" the underlying dysfunction which, however, continues to exist, and can, in special conditions, manifest itself. Specifically, vestibular compensation means the disappearance of all asymmetries (static and dynamic) in the ocular and spinal vestibular responses along with a more central plastic reorganization of the reflexes. This process may be defined as an error-controlled goal-directed learning. The goal of vestibular rehabilitation is to train the patient to learn substitutive and compensatory behavioural strategies. The main problem of current visual-feedback procedures is the poor transfer of motor skills learned during rehabilitative sessions into daily life activities.

VR is a technology that is already known to be capable of involving patients in a wide range of learning experiences (Rizzo et al., 1997). Also in the case of vestibular rehabilitation, VR may provide the right "dynamic cognitive and spatial milieu" to link CoG control to activities of daily living such as exploration and navigation of continuously varying environments. This would stimulate in a very complex way the compensatory mechanisms and would prevent the establishment of too specific associations between the control of equilibrium and the environmental conditions in which that control is achieved. The use of a reliable methodology to quantify learning of CoG control is suggested to be as important as the combination of the appropriate virtual environments with the appropriate cognitive tasks to optimize the outcome of the retraining.

Studies are only at the beginning, however. More systematic investigations need to be carried out in at least two main directions: to understand more concerning the interplay between vestibular function and cognitive activities related to orientation and the exploration of the surrounding space, and to assess the efficacy and the generalization of VR-based vestibular rehabilitation.

5. REFERENCES

D Alpini, L Pugnetti, A Cesarani, L Mendozzi and E Barbieri (1996), Evaluation of vestibular strategies in human exploration of the environment by means of immersive virtual reality, In *Giddiness and Vestibulo-Spinal Investigations* (CF Claussen, MV Kirtane, L Constantinescu & D Schneider, Eds), Elsevier Science, Amsterdam, pp. 315-323

RA Andersen, LH Snyder, CS Li and B Stricanne (1993), Coordinate transformations in the rapresentation of spatial information, *Curr. Opin. Neurobiol.*, **3**, pp. 171-176.

R Kohen-Ratz (1977), *Psychobiological aspects of cognitive growth*. New York: Academic Press.

BL Mc Naughton (1987), Neural association of movement and space: preliminary steps toward a non-cartographic theory of spatial rapresentation and learning, *Neurosci. Lett.* ,**Suppl. 29**, pp. 143-144.

L Pugnetti, L Mendozzi, EA Attree, E Barbieri, BA Brooks, CL Cazzullo, A Motta and DF Rose (1998), Probing memory and executive functions with virtual reality: past and present studies, *Cyberpsychol. Behav.,* **1**, 2, pp. 151-161.

A Rizzo, JG Buckwalter and U Neumann (1997), Virtual reality and cognitive rehabilitation: a brief review of the future, *J. Head Trauma Rehab.,* **12,** pp. 1-15.

PF Smith and CL Darlington, (1991), Neurochemical mechanisms of recovery from peripheral vestibular lesions (vestibular compensation), *Brain Res. Rev.*, **16**, pp. 117-133.

K Stanney , G Salvendy, J Deisinger et al., (1998) Aftereffects and sense of presence in virtual environments: development of a research and development agenda, *Int. J. Human-Comp. Interact.,* **10**, 2, pp. 135-187

S Wapner and HA Witkin (1950), The role of visual factors in the maintenance of body balance, *Am. J. Psychol.*, **66**, pp. 385-408

Virtual reality as a communication aid for persons with aphasia

Elisabeth Ahlsén and Vladimir Geroimenko

Department of Linguistics and SSKKII Center for Cognitive Science
Göteborg University, Box 200, SE-40530 Göteborg, SWEDEN

eliza@ling.gu.se, vladimir@ling.gu.se

ABSTRACT

The paper presents a prototype using virtual reality as the basis for a picture communication aid for persons with aphasia (acquired language disorder after a focal brain injury). Many persons with aphasia have severe word finding problems and can produce no speech or very little speech. This problem is often connected to problems of abstraction, which makes it problematic to use picture communication based on classifying pictures by semantically superordinate categories and searching for them via these superordinate categories. The use of virtual reality makes it possible to use purely visuo-spatial orientation as a search strategy in a picture database. In the Virtual Communicator for Aphasics, we are exploring these possibilities. By "moving around" in a virtual environment, based on video-filmed panoramas, the user can search for pictures essentially in the same way as he/she would search for objects in the real world and communicate by pointing to the pictures. Pointing for communication can be used directly in the panoramas, in overview pictures accessed by hot-spots, and in arrays of pictures of objects also accessed by pointing to hotspots in panoramas or overview pictures. Speech output is possible. Some of the potential advantages and limitations of using virtual reality based on photorealistic panoramas and pictures in this type of application are discussed.

1. AIM

The main aim is to develop a prototype for a communication aid which uses virtual reality and a spatial orientation strategy for picture communication. Persons with aphasia will try the prototype for evaluation and further development. The study is part of the project Intermodal translation (supported by KFB), which investigates the conditions for using information in different modalities for communication and for the transfer of communication between modalities in relation to disabilities which prevent the use of one modality.

2. GENERAL BACKGROUND

Within the project Intermodal translation, different sub-projects are exploring the conditions for transferring communication between different modalities or different manners of communication. One of the challenging areas are persons who have disabilities preventing them from communicating by using spoken or written language. How much communication can be achieved by using pictures and how can this type of communication be made reasonably fast, efficient and easy to use? How can the different needs of different disabled groups be met by the use of emerging technology? One of the target groups are persons with aphasia and severe language output problems.

Persons with aphasia generally suffer from anomia, i.e., they have more or less severe word finding problems. This pervasive feature of aphasia is believed to be related to a difficulty in using abstraction or decontextualization, affecting the ability to categorize. For example, many persons with aphasia have considerable problems filling in the last word of sentences like "Dachshund and poodle are both types of ________." or "Volvo and Fiat are both types of _______." etc.

A number of aphasic persons have severe problems of linguistic expression, i.e., they might not be able to speak at all or only be able to say a few stereotypic utterances and perhaps yes and no. These persons are badly in need of a means of expression and many efforts are being made to provide aids using communication technology. Some of these aphasics can have relatively well preserved language comprehension, whereas others have more global problems. All of the, however, need to be able to communicate.

Since written communication is most often not available to the severely aphasic persons, most attempts to develop communication aids for them are based mainly on using pictures that are linked to synthetic or digitized speech. (Cf., however, Todman et al, 1995, Todman and Grant 1995, Dennis et al 1995, for approaches using written text and speech.)

It is well known that picture communication involves some problems. It is not always readily used by aphasic persons for expressing needs and wishes, even when it would be possible. Perhaps this depends on the somewhat indirect and unusual or artificial nature of picture communication, as opposed to speech and gesture. The area of picture communication in general and the specific conditions for picture communication in relation to aphasia is still in need of a great deal of basic as well as applied research, involving a number of disciplines, such as psychology, linguistics, cognitive science, semiotics, sociology and neuroscience.

3. SEARCH STRATEGIES AND PICTURE COMMUNICATION

Since low-tech communication boards, picture books and card stacks have been replaced by computers allowing the use of dynamic displays and the storage of large picture databases, the problems of organizing search strategies in such databases have come into focus. Given that the user has a large number of pictures stored in his/her computer but only the screen size for interface, how does he/she find the pictures needed for communication in a specific situation in a reaonably fast and efficient way?

There are two main search strategies used in this context:

1) Search based on semantic categories
2) Search based on visuo-spatial orientation

(A third possibility is the use of associations of mixed types, i.e. visual, semantic., phonological etc., used in Minspeak applications, cf. Baker 1987, which will not be treated here, but can perhaps be seen as an expansion of the first search alternative above.)

Let us look a little closer at the conditions of the two search strategies.

Search by semantic categories is, by far, the most common way of organizing picture communication aids. This means having an overview screen with pictures representing superordinate categories, e.g. persons, clothes, actions, furniture, vehicles, food etc. Sometimes the subordinate-superordinate relation of pictures to categories is supplemented with part-whole or location relations, such as a kitchen for kitchen related objects or a town for shops.

The challenging tasks in this enterprise are to find a intuitive, transparent and consistent taxonomy of categories and, not least, to find pictures to represent these superordinate categories. The latter can be done by, for example:

1) choosing a typical (or prototypical) object from the category to represent the whole category, (e.g. a hammer for tools, an apple for fruits)
2) having a more abstract representation that is supposed to give associations to the category (e.g. a pictogram, a simplified line drawing or a sign like lightning for actions, i.e. signs resembling traffic signs)
3) trying to give a more complex concrete picture, including several objects from the category (e.g. a small drawing of a kitchen for kitchen related objects or a set of clothes (sock, trousers, shirt) for clothes. (These pictures usually become quite hard to decipher, since they include many details on a minute surface.)

Some examples of communication aids using this type of search strategy are C-VIC/Lingraphica (Steele et al 1989, 1992) and PicBox (Johnsen and Linell, 1995).

The search strategy based on semantic categories presumably works fine for a person (aphasic or not) who can figure out, learn and remember the system for categorizing and the pictures for the superordinate categories, thus knowing in what category to look for what more specific pictures. It is, however, not possible to use for persons with severe problems in using abstraction.

Already in the systems based on semantic categories, we could see that locations or situations, like a kitchen or a garden, could sometimes be used as a complement.

The alternative way of orienting in a picture database is to build as much as possible on visou-spatial orientation, which is assumed to be preserved in most of the persons with problems of categorization/abstraction.

The idea behind visuo-spatial navigation for communication is to avoid the categorization/abstraction problem as much as possible, by using a map-like orientation based on locations and giving the user the feeling or illusion of walking around in the world, communication by looking for things in places where they are usually found and pointing to them.

This idea has been exploited by the Danish communication aid Genlyd (Rygaard, 1990). In Genlyd, the user finds himself in a miniature town with a main street including many shops and offices etc. There is also a house which can then be used as the home interface and in the different rooms different object of everyday life can be found. Genlyd uses simple line drawings. The main overview pictures are the town, the house and the rooms of the house.

4. VIRTUAL REALITY AND VISUO-SPATIAL ORIENTATION

The example programs mentioned above were designed a few years ago. Today, more general picture databases, where pictures can be selected and included in individually adapted, multi-layer communication windows using dynamic displays, e.g. for activity based charts or semantic category charts, provide a great potential for experimenting with picture communication.

The basic principle of spatial orientation has, however, not been fully explored and exploited. Virtual reality provides a new and exciting means for making the illusion of truly spatial search in a real-world like environment more realistic and potentially more useful to a group of persons with no other possibility for expression. An interface where the user moves around in a photorealistic r artificial virtual environment is quite different from the so far used map-like overviews which give more of a bird's eye perspective.

5. THE VIRTUAL COMMUNICATOR FOR APHASICS (VCA) PROTOTYPE

5.1 Content and function

The VCA prototype utilizes digitized videorecorded panoramas as the basic overview interfaces on the screen. The user can move around in the panorama using the mouse and move further into new panoramas using relevantly located hot spots.

The presently available panoramas are:

- The town with two streets surrounding the house
- The shopping street
- The office street
- The interior of the house with the different rooms seen from the entrance hall
- The interior of each room

From the panoramas, it is possible to move to more detailed overview pictures, such as the interior of a cupboard and from those pictures to individual pictures, e.g. of objects usually found in such a cupboard.

An example of how the present structure can be used is presented in the example shown in Fig. 1.

1. GENERAL INTERFACE PANORAMA:

HOUSE WITH TWO SURROUNDING STREETS

Click HOUSE

2. HOUSE INDOOR PANORAMA:

VIEW FROM THE ENTRANCE WITH OPEN DOOR INTO BEDROOM, BATHROOM, KITCHEN AND LIVING ROOM

Click KITCHEN

3. KITCHEN PANORAMA

Click REFRIGERATOR DOOR

4. REFRIGERATOR INTERIOR:

OVERVIEW PICTURE WITH DIFFERENT TYPES OF FOOD

Point to CHEESE

Fig. 1. *Example of the spatial search structure in the prototype for virtual communication: search for the word cheese*

An alternative addition to the search route in Figure 1 would be to click the *overview picture* in order to get a set of *pictures of different types of foods* to and then click *cheese* in the separate picture of cheese, but this would probably not be necessary if a cheese is clearly visible in the overview picture. An addition of this type would, however, be a way to find more clearly visible pictures of different objects from an overview picture or even objects that can not be seen in the picture but could be expected in it, since an overview picture of, for example, a refrigerator, can not contain all the items that could possibly be kept in a refrigerator.

The panoramas can be used in full-screen format or leaving the leftmost part of the screen open for arrays of more specific pictures, such as pictures of different kinds of food in response to clicking on the overview interior of the refrigerator, in its turn reached by clicking on the door of the refrigerator, in our example.

By clicking at objects in the pictures the user can get recorded speech output from the computer. It is, thus, possible to, for example, create a shopping list for a relative over the phone . (This is a function also included in some of the existing communication aids.)

The panoramas can, of course, be extended to interiors of shops and offices, to outdoor scenarios etc.

The VCA prototype, thus, exploits the same kind of content as Genlyd, the crucial differences being

1) the use of photorealistic pictures and especially panoramas, intended to give the feeling of moving in a virtual scenery
2) the speech output
3) the potential future use of actual photos and panoramas from the user’s own home etc.

5.2 Technology

Our prototype is based on one of the most interesting and powerful VR technologies available at this time - Real VR from Live Picture Corp. The basic elements of this technology are 360 degree panoramic images of real places in JPEG file format. The high level of compression enables us to create full-screen high-resolution panoramic views with the file sizes under 70K. This means that our prototype is fully Internet-based and can be downloaded easily in real time even via a 28.8 kbps modem.

The panoramic virtual environments based on the Real VR technology are VRML 2.0 compatible and can include all types of multimedia objects with complex interactivity and behavior. First at all, users are able to navigate through VR panoramas using the mouse or keyboard: look around, up and down, zoom in and out. Then they can interact with any active parts of a panorama (called hotspots] or interactive objects inside a panorama. The Real VR technology has unlimited possibilities of integrating VR with multimedia elements. In principle, all existing types of multimedia can be integrated into this kind of VR, for example: 3D VRML objects and environments, 2D image-based on-screen and world objects, 3D image-based on-screen and world objects, 2D (GIFs) and 3D animation, audio and streaming audio (incl. 3D spatial sound), video and streaming video, multimedia databases, hotspots, links, FlashPix images, etc.

Although we are using only some of the above multimedia elements in our current prototype, the Real VR technology gives us great opportunities to use any multimedia that may be necessary for achieving our present aim as well as for the further development of the prototype.

At this time, we are developing the following two version of the VCA prototype:

1) The plug-in version that includes all available elements and functions. The only drawback of the version is that users have to download a plug-in (Live Picture Viewer) in order to experience the prototype. Cncerning patforms and browsers, the plug-in is available for both PC and Macintosh and for Netscape Navigator as well as Internet Explorer.
2) The Java version that does not require any plug-in and works on the principle “just click and see” on any computer. Unfortunately, at this time the Java version is not able to provide us with all the objects and functions that can be implemented in the plug-in version.

6. VCA AND APHASIA – POSSIBILITIES AND PROBLEMS

6.1 Possibilities

Some of the possibilities of using a VR tool, like the VCA prototype, are:

The spatial search strategy: Preserved ways of orientation can be used for communication, thus avoiding dependence on categorization and abstraction skills.

The photorealistic panoramas and pictures: The user does not have to deal with visual abstraction to line drawings or simplified/stylized picture representations.

The natural/intuitive search way: The aid should be easy to use, by just pretending to walk around looking for things in the ‘world’ on the screen.

The vocal output: Turntaking/attention getting is facilitated, and phone messages are possible.

The possibility of individual adaptations: It is not impossible for an individual user to have more or less of his/her own environment and own objects inserted.

The possibility to add any information that can be spatially represented: What is needed can, in principle, be put in.

The virtual access to different locations: Not all users can move around in their real life environment and use pointing. Many potential users are hemiplegic and some of them use wheelchairs. They can, by using VCA, at least refer to the places in the interface and objects in those places.

6.2 Limitations

The VCA is a first attempt to make use of at least some of the possibilities listed above. There are, however, also many potential limitations and problems motivating further research.

The spatial strategy – the search vs. size problem revisited: How much picture information is it possible to place in natural locations that are easily found by search in a virtual environment? As picture based communication systems grow larger, search problems will appear, regardless of which main search strategy is employed.

Non-picturable and non-localizable information: What about information that is not easily picturable or located in a specific type of place? This problem is partly shared with category based strategies, when it comes to picturability and natural place in the search procedure. A major problem is that only part of the information that a person needs in order to communicate is easily picturable at all, mainly objects, persons, places and to some extent actions, events and properties and mainly concrete instances of these categories.

Photorealism and abstraction: Can photorealistic pictures be a problem? This has to be investigated by letting potential users try the aid. Although search in a photorealistic VR world seems very intuitive and straightforward, it might be a problem to even recognize a panorama representation of one's own home. Maybe this also requires some kind of abstraction, And what about photographic picture representations of objects as communication tools? It may be problematic to use a photo, which is a highly concrete and specified representation, to represent an object which with respect to some of the specific features is not exactly like the one in the photo. Persons with aphasia have, for example, been known to reject colored drawing of a yellow sock when searching in a picture database in order to say that they want their own gray socks. This is one of the reasons for sometimes choosing more simple rudimentary line drawings or pictograms that are not so highly specified for details in picture communication systems. Far too little is known, at this point in time, to make any hasty predictions about how individual persons with aphasia will react to the photos and panoramas. One question here, is, of course, whether an artificial VR environment would be more suitable for some aphasics and whether a mixture of video-panoramas, photos and simple drawings should in some cases be used. The photorealistic VR environment (like most category based picture communication systems) is also, by necessity, very culture specific.

Individual adaptation: A related question to the above is how much individual adaptation would be needed and how much would be realistic for most users.

Other limitations: Like other communication systems, there could also be problems for the user related to visual half field defects, apraxia and more general problems in managing the computer as well as problems related to the portability of the communication aid (Ahlsén, 1996).

7. WHAT'S NEXT FOR VCA?

The VCA prototype will be a rudimentary picture communication aid which can demonstrate the possibilities of a VR communication aid and to some extent be evaluated. Persons with aphasia will be able to try the prototype. Interviews with aphasic persons and with experienced speech-language pathologists will also be used for evaluation and further planning.

A related case study of the ordinary daily communication of a couple of non-aphasic elderly persons (many aphasics are elderly persons) will be carried out. The purpose of this study is to obtain more information about what types of more specific information could and would need to be included in VCA to make it a useful tool for communication in every day communication.

8. FURTHER POTENTIAL DEVELOPMENTS

The next step would be to make a highly individualized VR application for a single person in cooperation with that person and to evaluate it in real-life use. At the same time, a standard version which is more rich in information and, thus, has a more advanced system for spatial search should be developed.

There is also a need for experimental studies evaluating the features and usefulness of different types of picture representations for persons with aphasia (as well as other persons).

9. OTHER POSSIBLE APPLICATIONS OF VCA

The VCA prototype is designed specifically for persons with aphasia, since their need for spatial navigation seems to be great. This group might, however, not be the only one that could profit from the use of spatially based picture communication. Other potential users could be persons with Alzheimer's disease or other forms of dementia and persons with mental retardation. If and how these persons could use VCA is an open question, having to do with how easy and intuitive the search strategy will turn out to be.

Extending the scope of VCA would also make it a possible tool for language learning that could be used both for first and second language education in training basic vocabulary for objects etc. in various environments.

10. REFERENCES

E Ahlsén (1996) Aphasia as a challenge to AAC. In *Aphasia as a Challenge to AAC*. The Stroke and Aphasia Federation of Finland, Helsinki, pp. 17-24.

B Baker (1987) Minspeak. *Byte*, September, pp. 186-202-

F Dennis, A Waller, N Whitehead, Y Cairns, J K Brodie, A F Newell and K Morrison (1995) Evaluating TalksBac and AAC system for dysphasic adults. In *Widening the Perspective*. Augmentative Communication in Practice: Scotland. Collected papers 1995 (S Millar and A Wilson, Eds) CALL Centre, Edinburgh, pp. 62-64.

B Johnsen and S Linell (1995) Kan man samtala via bilder? NUTEK R 1995, 3, Stockholm.

K. Rygaard (1990) *Hvofor tro ny teknologi kan udvide afatikerens kommunikative kompetence*. (Psykologisk kandidatstudie Speciale), University of Copenhagen, Department of Psychology.

R Steele, M K Kleczewska, G S Carlson and M Weinrich (1992) Computers in the rehabilitation of chronic, severe aphasia: C- VIC 2.0 cross modal studie. *Aphasiology,* **6:2**, pp. 185-194.

R Steele, M Weinrich, M K Kleczewska and G S Carlson (1989) Computer-based Visual Communication in Aphasia, *Neuropsychologia*, **14**, pp. 409-426.

J Todman, L Elder and N Alm (1995) Evaluation of the content of computer-aided conversations. *AAC Augmentative and Alternative Communication,* **11**, pp. 229-234.

J Todman and S Grant (1995) Using the TALK system for informal conversation. In *Widening the Perspective*. Augmentative Communication in Practice: Scotland. Collected papers 1995 (S Millar and A Wilson, Eds) CALL Centre, Edinburgh, pp. 43-44.

Implementation and capabilities of a virtual interaction system

Arthur W. Joyce, III and Anna C. Phalangas

Alfred I. duPont Hospital for Children / University of Delaware
1600 Rockland Road, Wilmington, DE 19803, USA

joyce@asel.udel.edu, phalanga@asel.udel.edu

ABSTRACT

Virtual Interaction refers to a technique for interacting with computer generated graphics. Graphical objects are overlaid on live video of the user. A chromakey separates the user (foreground) from the background resulting in a silhouette of user. The computer causes the graphical objects to move in relation to the silhouette so that they appear to interact with the user. This paper presents the implementations of the system, some techniques for interaction and discusses using the system as a tool for physical therapy.

1. INTRODUCTION

Virtual Interaction is the name of a low immersion, virtual reality system developed at the Alfred I duPont Hospital for Children and the University of Delaware. The users view themselves participating in virtual worlds, consisting of computer generated graphics, with real-time interactions. The purpose of this project is to create interactive environments, which can enhance therapy by providing motivating activity tailored to each user's interests and physical abilities. Virtual interaction provides physical activity for individuals with disabilities without requiring direct interaction with real physical objects. It also offers the potential for tracking clinically relevant performance information.

The system is moving into its third implementation. This paper presents these implementations and discusses the basic requirements of any implementation. It also discusses several interaction modes and their computational requirements.

2. BACKGROUND

In the early 1970s, Myron Krueger developed a system called VideoPlace, which provided a number of interactions between users and computer generated graphics (Krueger 1990). The users were back-lit to provide a strong contrast in the video signal. This almost binary signal was easily separated into the foreground/background regions needed for the interactions. Users saw their silhouettes interacting with the graphics.

Researchers at the Massachusetts Institute of Technology (MIT) have developed the IVE (Interactive Video Environment) and the ALIVE (Artificial Life IVE) systems (Maes 1995). These systems use color segmentation to separate the user's silhouette from the background without any special backdrops. The MIT research has focused primarily on intelligent, artificial creatures that react to the user's gestures. The users see themselves in their physical surroundings along with the computer-generated characters. The graphical characters are carefully generated so that they may move both in front of or behind the user. The gesture recognition and character responses occur over several seconds.

The Vivid Group (Toronto, Canada) currently markets the Mandala VR System that uses a blue backdrop and a chromakey to separate the user's image from the background. The blue background is replaced with a computer generated scene and interactive graphics. The interactions are immediate, video game like responses. The Vivid Group uses proportional control based on areas of interest to control the action in many of their games (MacDougall 1996).

Research being performed at the Alfred I. duPont Hospital for Children and the University of Delaware explores the use of this technology as a means of providing therapeutic and exercise activities for individuals with disabilities (Joyce 1995). The project, called Virtual Interaction, has the goal of further exploring the benefits noted by Krueger and developing an inexpensive implementation. The focus of the development is to provide interactivity with low-latency feedback. Although the interactions do not have to take the form of a game, this paper often refers to the motivational activity as playing within the context of a game. Real therapy still requires real work, but the goal of this research is to add fun, game-like elements to it.

An example of how Virtual Interaction might be used is the implementation of the video game, *Breakout*. (The images in Figure 3 are based on this game.) Atari introduced Breakout as an arcade game in 1976. A ball bounces between a paddle and a wall of bricks. Each time the ball hits a brick, that brick is removed. (In order to create a more playable game, the bricks have been moved from their original vertical orientation to the brick ceiling shown in the pictures.) With Virtual Interaction, the user becomes the paddle. The player steers the ball toward the remaining bricks while trying to keep the ball from hitting the bottom of the screen.

3. VIRTUAL INTERACTION IN PHYSICAL THERAPY

In most instances, the ability to understand the physical world is learned from direct experience. The ability to move, understanding relationships with one's environment and other physical concepts are learned through practice and refinement. For individuals with disabilities, the opportunities to interact and learn may be limited. Virtual reality may be able to offer individuals with disabilities the opportunity to manipulate objects and experience situations not otherwise available to them.

Krueger found that people are motivated to respond when they see their images. Their sense of curiosity led them to explore when the graphical objects in the image responded to their actions. Krueger hypothesized that the engrossing nature of the activity could be used in physical therapy to provide distraction from repetitious and tedious exercises. He saw from first hand evidence that individuals with disabilities could be motivated to push their capabilities in order to interact with a virtual environment. (Krueger, 1990, p198)

Although no formal studies have been done, anecdotal evidence agrees with Krueger's hypothesis. A Virtual Interaction system was demonstrated at RESNA 97, held in Pittsburg, Pennsylvania, and attendees were invited to try the system. A young man with Cerebral Palsy played a remarkable game of Breakout using his head to control the ball. He played the game with great intent and, based on his respiration rate, got an exceptional upper trunk workout. Others have tried the system in the laboratory and have responded favorably to its potential as a therapy tool.

Virtual reality has potential in cognitive or educational applications. (Helsel, 1992) Although the Virtual Interaction project has focus primarily on physical therapy, an example of a simple cognitive exercise is its implementation of the Milton Bradley game *Simon*. This memory game presents a sequence that must be repeated back by the player. The sequence increases in length each round. A number of possible adaptations are envisioned, such as adjusting the number of choices to the cognitive level of the user. A physical challenge might be added by programming the buttons to move a little further away each round.

As part of the Virtual Interaction project, time has been invested observing therapy sessions in a variety of settings and discussing the system with therapists. This provided ideas for potential applications and has helped to begin to establish a design methodology. In general, therapists have been enthusiastic over the possibilities of Virtual Interaction. A common theme among the discussions was that therapists are always looking for ways to keep therapy and exercise interesting. The discussions confirmed the hypothesis that motivation is an important element in physical therapy.

Virtual Interaction is a low-immersion virtual reality environment that does not burden the user with goggles or sensors. This avoids problems associated with high-immersion virtual reality systems, such as motion sickness and cumbersome equipment. Like all virtual reality, the Virtual Interaction system allows the environment to be defined in any manner. Gravity, mass and inertia can all be ignored or redefined to tailor the interaction to the abilities of a user. While the system is intended to require the users to physically participate in the environment, there is no real physical interaction. The lack of physical contact, for example, makes it an appropriate tool for providing exercise for burn patients with painfully sensitive skin. The behavior of the virtual objects can be customized for each particular activity. For instance a virtual breeze may be added in order to gently push objects in a direction and emphasize activity on that side of the user.

A diverse population with different levels of physical and cognitive abilities must be considered when creating activities for therapy. Programs should be designed so that the therapist has a user-friendly interface to adapt the activity quickly and easily. There may be other factors that require consideration when designing activities. It has been noted that some young children (and cognitively young individuals) passively listen to audio stimulation and do not resume physical activity until the sound has stopped. (Fell 1994, Contaldi 1993) What was intended to be motivational becomes their sole focus and inhibits their participatory activity. It is likely that a similar effect will be seen with video interactions. As long as there is movement on the screen, some individuals may simply sit and watch. The graphical objects may have to cease activity completely in order to solicit participation from these users.

The system can log the user's activity and potentially track their performance. For instance, Breakout logs each point of contact between the ball and the silhouette. The aggregate provides some insight into the ability of an individual to reach different areas of the playing field. When distracted and motivated by the game, individuals will perform at their maximum capability. Data-logging can be used to show growth in accomplishments over time. It might also demonstrate differences between initial expectations and actual achievements. One of the most powerful lessons that individuals can learn is that there might be a difference between their self-perceived capabilities and their actual abilities.

4. THE VIRTUAL INTERACTION SYSTEM

The basic Virtual Interaction configuration is shown in Figure 1. The user sitting in front of a blue backdrop, facing a camera and video display (TV). The camera and the TV are connected to the Virtual Interaction system's electronics, which reside in a personal computer. (Note that there are two video displays, the TV and the computer monitor.) An alternative configuration uses an overhead camera and a blue table to provide a desktop interaction field. Other modes of use might include multiple users, or focus on isolated body parts, for instance focusing on the leg for kicking exercises.

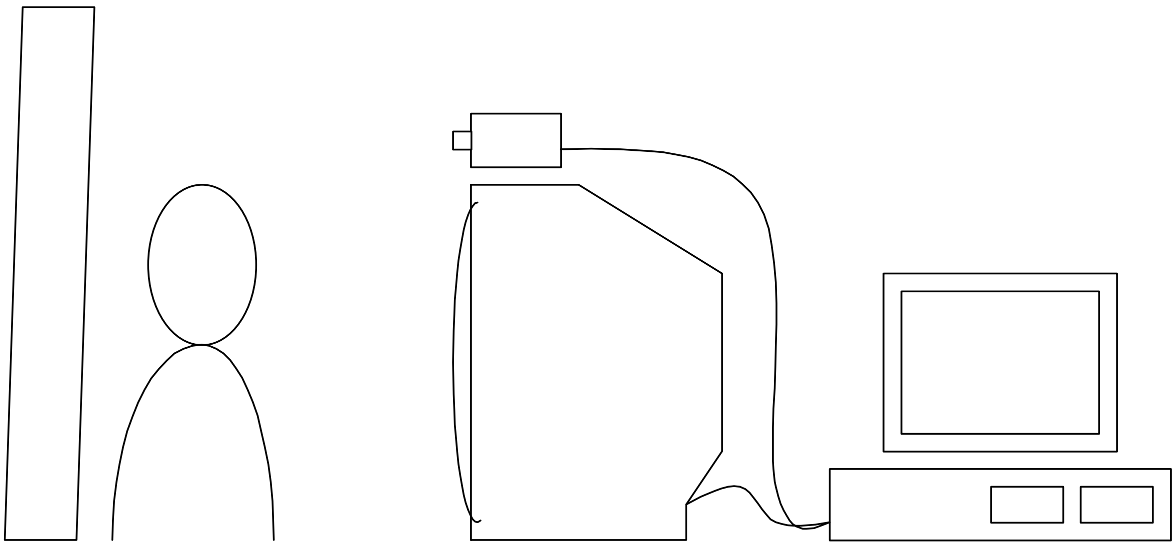


Figure 1. *The user sits in front of a blue backdrop. The camera on top of the TV sends the video image to the computer. The computer mirrors the video, adds graphical objects to it and sends it out to the TV, where the user sees him or herself interacting with the virtual objects.*

There are several considerations that must be reviewed when setting up a system. The camera must be far enough away from the user so that the camera's field of view is wide enough to provide a reasonably large playing field. Video conferencing cameras generally have wide-angle lenses with 45-degree fields of view. The full span of the arms can be seen if the user is five feet or so from the camera. This is also a good distance for users to be from the TV. They can comfortably see the entire screen but are still close enough to see detailed interactions. The backdrop needs to be as large as the camera's field of view. It should be as flat as possible and lit evenly from the front. Typical office lighting is generally good enough as long as the backdrop does not create a shadow over itself or the user, and the user doesn't create a shadow on the backdrop.

5. DISCUSSION OF SYSTEM REQUIREMENTS

The basic requirements of the system are an overlay system and foreground/background separation. The overlay system merges the live video with computer generated graphics to produce the video that the users see. The foreground/background separation produces the silhouette used for controlling the interactions. In addition, a large display is desired for output. Incorporating audio output can enhance the illusion of interaction as well as providing other game playing elements.

Figure 2 shows how the video flows through the electronics. The video from the camera is fed into the live video frame buffer and a chromakey. The chromakey output is also stored in a frame buffer, which the computer queries for its foreground/background information. The computer-generated graphics are drawn in the graphics frame buffer. These two video streams are merged and converted back to an analog video signal, which is sent to the TV display. It is possible to eliminate the chromakey and its frame buffer by adding computer access to the live video frame buffer. The computer makes the foreground/background decisions based on what pixel values are read from the live video frame buffer. This also makes alternative tracking schemes, such as flesh detection, possible.

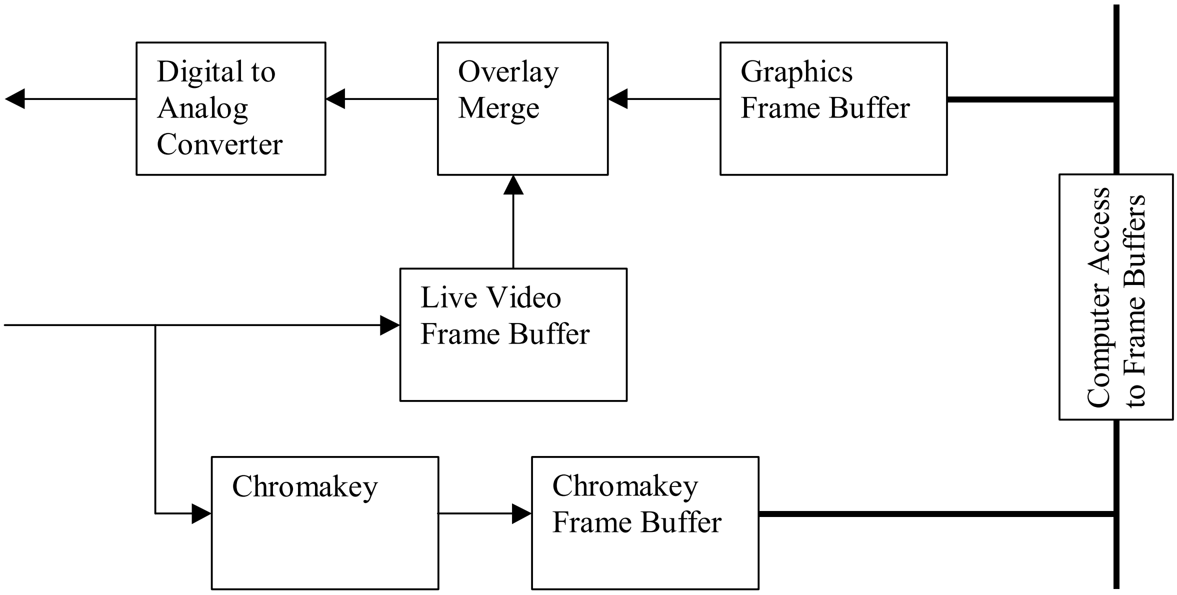


Figure 2. *The video flows from the camera into the Live Video Buffer and the Chromakey. The computer reads the Chromakey buffer for its foreground/background information. The computer generates the graphical objects in the Graphics Buffer. This is merged with the live video and sent to the TV display.*

Figure 3 shows examples of images at various points in the system. The camera produces an image of the user, which is digitized and sent to the Live Video Frame Buffer and the Chromakey. The Chromakey produces a binary foreground/background image. The computer samples this image and causes the graphical objects to react to the user's silhouette. The graphical objects are drawn in the Graphics Buffer. (The image in the graphics buffer is not shown in Figure 3.) The live video is mirrored by reversing the horizontal scan and is merged with the graphics video stream. The resulting video is shown to the user.

Figure 3. *Example images at various points in the Virtual Interaction system. The left image is typical of the camera output. The center image is the user's silhouette, as produced by the chromakey. This is the image that the system reacts to. The image on the right is the resulting image shown to the user.*

Most users are familiar with their images in a mirror and are able to use the visual feedback to position themselves. A camera and a monitor provide the view of an outside observer, which is not a model of interaction that most users are comfortable with. Horizontally flipping the video image produces the mirror-like feedback that users are accustomed to. This is such a natural feedback mechanism that most users do not recognize the difference until it has been pointed out to them.A life sized, or near life sized video image provides the user with some sense of immersion. A large screen output helps to create a one-to-one correspondence that allows the user to treat the system as a magic mirror. Large screens also help users to see detailed graphics and intricate interactions. Television sets are currently the most cost-effective means of providing this large format. It is therefore important that the output video be available in NTSC or PAL format.

Sound can enhance the illusion of interaction. The sound seems to define the moment of impact as well as indicate that some amount of energy has been expended. The system requirements are an ordinary sound card and a few wave files. Synthesized musical instruments can also be used to produce other auditory elements or provide the auditory output for a virtual musical keyboard.

6. IMPLEMENTATIONS

The project is in the early stages of its third implementation. The goal has been to develop a system that would be widely available to therapists and clinicians. If the cost can be held down, it might be economical enough to be used in private homes to continue physical therapy outside of clinical settings. Off-the-shelf technologies have been used wherever possible and all of the implementations have been based on IBM compatible PCs.

The first implementation integrated commercially available products with some custom electronics. A Creative Labs Video Blaster was used to merge live video with graphics generated under Microsoft Windows 3.1. The resulting video was seen on the computer's VGA display. A Video Blaster version FS 200 was used to vertically flip the video in real time. Turning the camera over on its back and vertically flipping the video produced the requisite mirrored video. Since accessing the frame buffer on a Video Blaster caused the video to freeze, an alternative method for sampling the color content of the image was required. A Micro Search ChromaKey+ (an analog chromakey) and a custom 1-bit deep digitizer were used for the foreground/background separation.

An AverMedia Averkey converted the VGA signal into NTSC format. The computer's VGA display mode had to be 640 by 480 pixels with the entire screen occupied by the output video. Keyboard input was used to activate pop-up dialogs to adjust the game parameters. Since the NTSC output simply echoed the computer's display, these dialogs were also seen on the user's display.

All of the electronics have been integrated onto a single, fully custom printed circuit card for the second implementation. This has been done to overcome some limitations of the first system, push the technology and add some features. This card takes NTSC video in and puts the resulting video out in NTSC format. The Live Video and Graphics buffers are double buffered for real-time, transparent access to the digitized video and smooth animation. Because the video and graphics buffers are completely separated from the computer's VGA display system, the VGA display no longer shows the output video. Therefore, the VGA display can be used to present other data, for instance to show the therapist the user's performance or allow adjustments to the game while it is being played. It might also be used to allow the therapist to have participatory input.

In addition to the basic game requirements, the custom card contains extra resources that were included to facilitate research toward more sophisticated systems. There is an FPGA-based processor connected to the capture bus and an additional graphics buffer referred to as the Background Buffer. The FPGA-based processor is intended to process the video stream on a per-pixel basis. The envisioned uses are to implement a hardware chromakey, hardware based region finding, or simple image processing algorithms such as edge detection. The Background buffer's output is multiplexed to either go to the FPGA-based processor or go to the output video multiplexer. The foreground/background separation can be accomplished without the blue backdrop if the Background buffer is filled with a reference image, which is then compared to the live video by the FPGA-based processor. Alternatively, the FPGA-based processor could be used to implement a chromakey, which could control the video multiplexer to allow a background image to replace the blue backdrop.

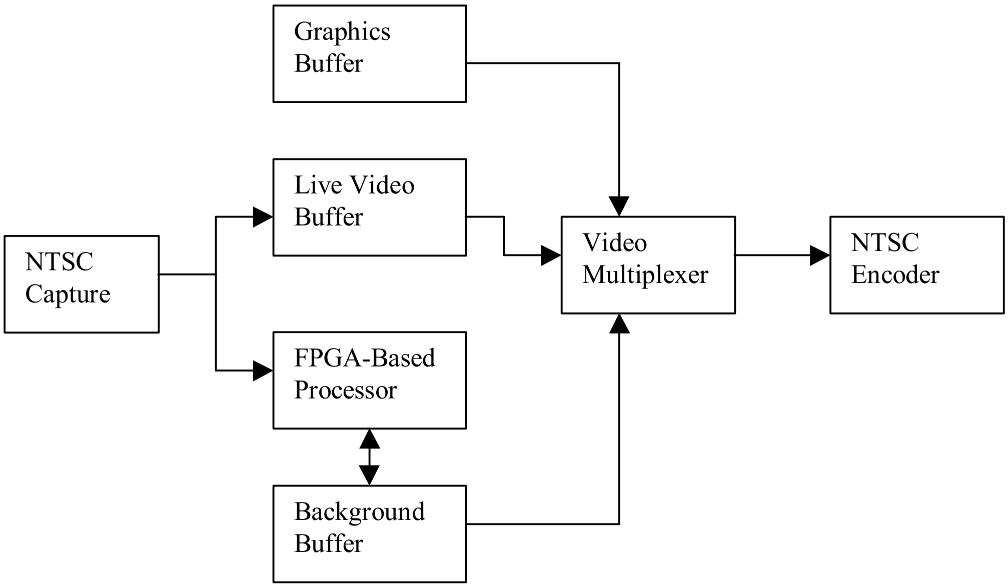


Figure 4. *The architecture of the custom Virtual Interaction card contains additional resources. An FPGA-based processor and a background frame buffer have been included. These are intended to support research into more sophisticated scene generation and more sophisticated foreground/background separation with feature tracking.*

The custom card contains four Motorola XC56166 digital signal processor chips, one for each of the three frame buffers and one for the FPGA-based processor. Each frame buffer consists of four Texas Instruments TMS55161 VRAM chips, thereby giving each frame buffer two pages of 1024 by 512 memory. The memory controllers are implemented in Xilinx XC3142 FPGAs, which allows the memory models and addressing modes to be changed to suit an application. The NTSC decode is performed by a Brooktree Bt819 and the NTSC encoder is a Bt856.

The card has a 16-bit ISA bus interface and uses a single 16-address slot in the IO address space. The PC does not have direct access to the frame buffer memory but rather issues commands to the DSP to read from or write to the memory. Although this provides mediocre performance when operations are handled on a pixel by pixel basis, the DSPs are a significant resource and can be programmed for higher level functions. For instance, if run-length encoding is used for the graphical objects, the fill rate improves from about 60 thousand pixels per second to roughly 2 million pixels per second.

The third implementation uses the newly available VigraVision card by VisiCom (San Diego, California, USA). This card provides video capture, NTSC/PAL output, fast access to the video buffer, and can be set to mirror the video. This card, introduced in the spring of 1998, is the first commercially available product that meets all of the system requirements at reasonable cost. Although this system uses an overlay capture system similar to the Video Blaster, the display adapter does not have to use the 640x480 mode. Therefore, over half of the area on the computer monitor is still available for control functions and performance monitoring.

In addition to the video capture and overlay requirements, any implementation must meet certain rendering requirements, which are application dependent. Basically a system must be able to create smooth animation, which requires both processing power and double buffering. None of the demonstration programs are particularly demanding, so the rendering power has not been an issue. The VideoBlaster system uses Windows 3.1, which does not support double buffering. To prevent the erase and redraw operations from creating visible artifacts, both operations are combined into a single BitBLT operation. This operation copies a bitmap that has been expanded, so that when it is copied it restores the area previously occupied by the graphical object to the background color of the overlay. The custom implementation has a dedicated frame buffer for the graphical overlays. The buffer is driven by a DSP, which performs many of the pixel level operations. The VigraVision system uses Microsoft Windows 95 with DirectX. It incorporates the features of a modern video card with double buffering and a graphics accelerator, as well as the requisite live video overlay.

All three implementations use chromakeying with blue backgrounds to determine the user's silhouette. Any chromakey is subject to the fundamental bandwidth limitations of the incoming NTSC signal. Color changes are smeared over four to eight pixels, which makes the vertical edges of the silhouette ragged. Interactions must filter out this noise and recover gracefully from the inevitable errors. Controlled lighting can produce superior definition of the silhouette but the goal of this project has always been to provide a system that can be placed in a clinical setting with as little physical disruption to the environment as possible. Instead of controlling the lighting, the focus has been on working around the problems created by a less than perfect chromakey separation.

7. INTERACTIONS

The basic implementation of *Breakout* uses a bounce interaction, which tests for the silhouette in the path of the ball (or any other object using the bounce interaction). The program tests at several points around the periphery of the ball in order to detect a glancing blow as well as frontal contact. For a medium speed game, the ball moves about 6 pixels during a 33ms frame period. (That equates to moving across the screen in about 3-4 seconds.) Therefore the bounce interaction tests about 30 pixels per frame period. When a contact is made, further checks are performed to eliminate spurious noise and then the normal of the silhouette is calculated. The dot product of the normal and the motion vector produces the magnitude of the impact. Adding twice that amount of motion in the direction of the normal gives a mirror-like deflection.

Interestingly, the addition of a collision sound as the ball bounces off of the user's silhouette enhances the illusion that the purely graphical ball actually made physical contact with the user. Sound is associated with the expenditure of energy and hearing a sound seems to provide the user with the sense that a physical, energetic collision occurred.

Simple switches can be used to activate anything for which the application calls. The Simon game mentioned above uses simple switches. An application that demonstrates the use of switches as well as audio interactions is the virtual musical keyboard. Each switch is tied to a different note, which is played though the MIDI interface by a Sound Blaster FM Synthesizer. In order to debounce a switch, the program tests a small region for activation and tests a larger, overlapping region once activated. The silhouette must therefore be well outside of the activation region before the switch is reset and a new switch contact can be initiated.

An application that uses a drag and drop interaction is an anagram unscrambling game. The computer presents a set of letters and the user drags the letters into their proper sequence. The drag and drop interaction is activated with touch and the graphical object then moves with the user's silhouette. The interaction continues until the program tells it to drop, either because the user held still, or the object reached a goal. The drag expects to track an appendage which is longer than it is wide, for instance, an arm, or on a smaller scale, a finger. The tracking continually calculates the major axis of the appendage, using the normals along the side to estimate the direction of the axis. The most extreme point along the major axis is used as the tip and is the location of the drag point.

An interaction mode that allows users to swat objects is being developed. The swat interface must not only determine that the ball has been hit, but where the silhouette was before it hit the ball. The program being developed uses an anticipatory proximity measure, which measures how close, and in what direction the silhouette is to where the ball is going to be. When contact is made, changes in the proximity measurement determine the magnitude and direction of the deflection. Rapid hand motions by the user may be on the order of 30 pixels in a single frame period (assuming a typical upper body and arm span camera view). This is about the same as the diameter of the ball or the thickness of the hand viewed edgewise. Because the silhouette can come from any direction, looking ahead by two frame periods requires the program to examine regions that are about 150 pixels on a side. A sparse search is used to find the silhouette and then the estimate of the closest point is progressively refined. Because of the high-speed access to the video buffer, the VigraVision system is able to store the raw video and delay the search until contact has been made. The custom card system must perform the search during each frame period.

8. ERRORS AND OTHER CONSIDERATIONS

The system is not perfect and it makes mistakes. Given the particular application and diversity of abilities expected among the users, graceful error recovery is important. For instance, in Breakout, the ball can become imbedded in the silhouette. Once inside, it is surrounded by the silhouette and bounces continuously. The net effect is a ball that does little more than wiggle until the user moves enough to free it. Requiring the users to make significant movements is not a viable option for error recovery. Instead heuristics are employed. For this error, the heuristic prevents the ball from bouncing continuously by insuring that the ball moves some minimum specified distance without striking the silhouette. If the silhouette continues to be found, the ball is considered trapped and it simply continues in its present direction until it emerges out of the silhouette. It then reverts to its normal bounce mode.

The swat interface is being developed because experience with the Breakout demo has shown that users want to swat and push the ball, which is how they would interact with a real one. The bounce interface tests for contact along the leading edge of the ball under the assumption that the motion of the ball causes all contacts. Therefore, it assumes that the contact point is along the edge of the silhouette. However, when the silhouette has moved since the last test, that contact point may be inside of the silhouette and there is no edge on which to base the normal. Another error that occurs is push through. The user hits the bottom of the ball intending to push it up. The program correctly sees the silhouette and bounces up. But the hand continues to move and pushes through the ball. The ball then sees the bottom of the hand in front of it and bounces downward. These experience points out that a user's expectations and perception of the interface must be considered when designing the interactions.

There are other sources of errors, not the least of which is the misclassification of the foreground and background. Interactions must be designed to work with less than perfect silhouettes. The fundamental NTSC bandwidth issue and motion blur both help to create ragged edges. Controlled lighting and a camera with a high-speed shutter can provide superior separation, but the resulting system would result in a significantly higher impact on the clinical environment.

9. FUTURE RESEARCH

Future system development can explore more sophisticated interactions. It is not possible to track a feature inside of the silhouette. For instance, the hand cannot be used to drag an object across the body. Once the silhouette of the hand is inside of the silhouette of the body, it disappears as a trackable feature. An application may be able to compensate for this limitation with heuristics. The ALIVE System at MIT estimates the locations of the hands based on the body outline. If the arms are away from the body, they are easily tracked. If the arms do not protrude from the basic silhouette of the body, then the contour of the body is used to estimate the orientation of the arms. A change in width of the body at the hips indicates that the hand is at the user's side. If there is a change in the mid-torso region, the arm is bent at the elbow. Ideally, the system might provide more sophisticated feature tracking perhaps by using flesh detection.

The system will be studied as a tool to aid physical therapy. One of the reasons behind the migration to the VigraVision card is the need for stable, commercially available hardware for use in clinical settings. The technology demonstration programs need to be modified and extended in order to create activities that can be used with a variety of individuals.

10. CONCLUSION

Virtual Interaction can provide interactions that are engaging and believable. Experience with the demonstration applications has shown that the interactions are somehow tangible but the environment is a harmless cartoon world. An indication of the tangibility of the objects can be seen in the amusement users get when balls bounce off of their heads. They eagerly explore what effect their actions elicit. The mirrored self-image provides easily understood feedback.

Virtual Interaction should provide a unique, customizable, motivational tool for therapy. The applications should include good exercise and learning environments for individuals with disabilities. By making these activities more fun, the system has the potential to motivate a person to do more in therapy, thereby improving their well being.

Virtual Interaction can be implemented with reasonably inexpensive, off-the-shelf technology. Current modes of interaction should be extended and further modes, such as the simple gesture recognition demonstrated at MIT, developed. Improved technologies, such as flesh detection and more sophisticated feature tracking, should allow more capabilities and open new modes of interactivity.

Acknowledgements: This research has been funded by the U.S. Department of Education, Grant #H129E20006 from the Rehabilitation Services Administration and Grant #H133P30003 from the National Institute on Disability and Rehabilitation Research, The National Science Foundation, Grant #HRD-9450019, and the Nemours Research Programs.

11. REFERENCES

E. Contaldi, P. Taney, (1993) Introduction to Computer Based Music for the Consumer, *Presentation at DATI Annual Conference*, Oct 4-5, 1993, Newark, Delaware

H. J. Fell, H. Delta, R. Peterson, L. J. Ferrier, Z. Mooraj, M. Valleau, (1994) Using the Baby-Babble-Blanket for Infants with Motor Problems: An Empirical Study, *ASSETS '94*, p77-84, Marina del Rey, California (Author's note: The statement concerning the cessation of activity was part of the presentation but was not noted in the paper.)

S. Helsel (1992) Virtual Reality and Education, *Educational Technology*, May 1992, p 38-42

A. W. Joyce, III, A. Phalangas, (1995) Virtual Interaction; An Interface for Individuals with Disabilities, *Proceedings of RESNA '95*, Vancouver, Canada, p 425-427

Myron Krueger, (1990) *Artificial Reality II*, Addison-Wesley Publishing Co., Reading, Massachusetts

P. Maes, T. Darrell, B. Blumberg, A. Pentland, (1995) The ALIVE System: Wireless, Full-body Interaction with Autonomous Agents, *M.I.T. Media Laboratory Perceptual Computing Technical Report No. 257*

F. MacDougall, (1996) Video Image Based Control System, *U.S. Patent Number 5534917*, July 1996

Anthropometric models of children

Matthew T Beitler[1] and Richard A Foulds[2]

[1]Center for Human Modeling and Simulation, Department of Computer & Information Science, University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104-6389, USA

[2]Research Laboratories, duPont Hospital for Children, 1600 Rockland Road, Wilmington, DE 19803-3616, USA

[1]*beitler@graphics.cis.upenn.edu*, [2]*foulds@asel.udel.edu*

ABSTRACT

The objective of the work presented in this paper is to create a complete database of Anthropometric Models of Children, which can not only describe the anthropometric attributes of the individuals, but the functional abilities and growth behaviors as well. This work also includes a prototype system, which is being used during an assessment session to automatically gather the anthropometric and functional information of the subject and incorporated it into the population data of the model database.

1. INTRODUCTION

Three-dimensional simulation of anthropometrically correct human figures has been researched extensively (Badler et al, 1979; Bapu et al, 1981; McDaniel et al, 1988; Badler et al, 1993), but this research has all focused on full grown male and female populations. In this paper we present the Anthropometric Models of Children database which is based on the anthropometric data gathered for the Consumer Product Safety Commission in 1977 (Snyder et al, 1977). This model database currently contains accurately scaled body segment dimensions for children, ages 2 through 18, at the 5th, 50th, and 95th percentiles of the sampled population. Additionally, the models in the Anthropometric Models of Children database are compliant with the standardization efforts of the VRML Humanoid Animation (HANIM) working group of the VRML Consortium (HANIM 1997). The goal of HANIM is to create a standard structure for humanoid figures, to allow animation data sets and kinematic control systems to manipulate a variety of humanoid models. Additionally these models are able to be controlled by the body object bitstream of the MPEG4, which was defined by the Synthetic Natural Hybrid Coding Ad-Hoc Group on Face and Body Animation (SNHCFBA 1998). Compliance and integration with these standardization efforts is especially necessary for research efforts relating to the rehabilitation and disability fields, because there is a smaller population that directly benefits from developments in these areas. It therefore becomes important that researchers try to "piggyback" their efforts onto research developments and standardization efforts, which have a broader population benefit and adapt those developments to serve the needs of the smaller population of individuals with disabilities.

2. BACKGROUND

2.1 Consumer Focused Design

This research began as a small part of a consumer focused design program investigating methods of product design that are adaptable to consumer needs, are cost effective and exploit new methods of rapid prototyping. This design program included exploring ways in which virtual models and computer visualization software could be used to reduce the cost of designing rehabilitation devices. Part of this work focused on creating

* The work described in this paper was performed while M. T. Beitler was at the Research Laboratories of the duPont Hospital for Children.

 ISBN 0 7049 1141 8

software tools and resources for designers, which would assist them in the creation of new rehabilitation devices.

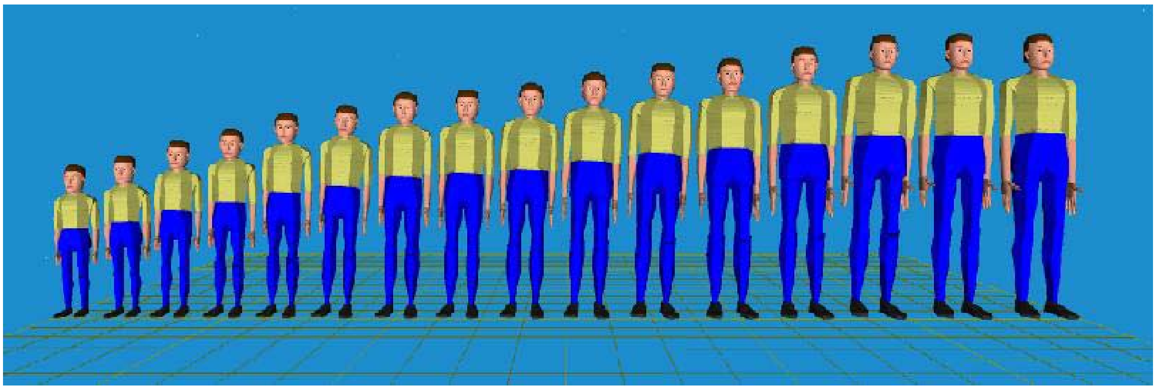


Figure 1. *The Anthropometric Models of Children, Ages 2-18.*

Since there are a diverse range of physical abilities from one user to the, there exists a need for designing devices which are user specific. This design cycle is complicated by the complexity of the device and the small market size as well as numerous biomechanical factors that need to be considered. The techniques of virtual design and rapid prototyping are ideally suited for assessing these factors as well as rapidly designing and manufacturing one-of-a-kind rehabilitation devices.

2.2 Device Generation from D-H Parameters

The initial work of this consumer focused project concentrated on a system that would allow the rapid creation of virtual devices based on kinematic specifications of the device. The system allows the designer to accomplishes this by having him input the Denavit-Hartenberg parameters (Denavit and Hartenberg, 1955) of the device they are prototyping. The Denavit-Hartenberg notation is a common kinematic protocol used by designers for defining a device's movements. After the designer has inputted the Denavit-Hartenberg parameters the system automatically generates a 3-D graphical model of the device (Figure 2) that can be manipulated by the user.

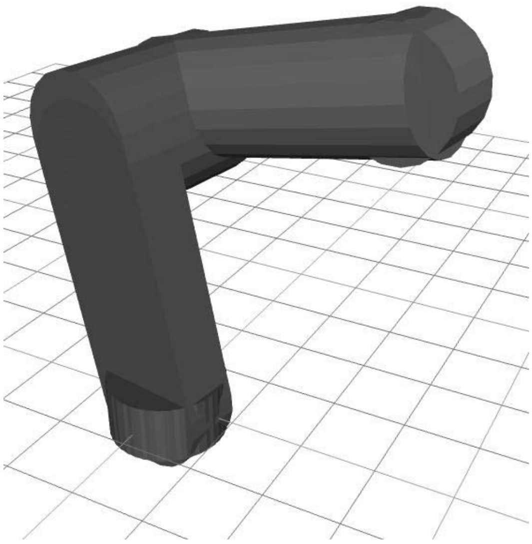


Figure 2. *A virtual device automatically generated from D-H parameters.*

An additional feature of the system is that it also allows the designer to define a set of external inputs to control the movements of the multiple joints of the device or to interact with the device using a Cartesian based control method. A Spaceball 6DOF isometric joystick, which has proven effective in controlling rehabilitation devices in the past (Wisaksana et al, 1995), allows the designer and user to experiment with controlling the virtual prototype device. The different degrees of freedom of the virtual prototype device can also be controlled with keyboard inputs. The model of the device can also be altered interactively by the designer, so that he can customize it according to verbal feedback from the user and the evaluation process. The initial software system successfully facilitated the design process of rapid prototyping by:

- allowing experimentation with different scenarios of controlling a mechanism
- allowing quick implementation of control scenarios designed by others
- enabling the designer to discover possible design flaws before actual fabrication of a mechanism
- making the design of "one-of-a-kind" control scenarios an affordable possibility, because of the ability to change the control mapping in seconds
- making the design of "one-of-a-kind" mechanisms an affordable possibility since unnecessary time and resources must not be spent on fabricating physical prototype devices

2.3 Anthropometric Customization of D-H Parameters

The final stage of the virtual device creation system was to include an “anthropometric adaptation” component, which could alter the D-H parameters based on the anthropometric information of a particular user. For this stage of the work we had proposed to use the ™JACK software application, which allows ergonomic and anthropometric simulation of human figures. The JACK human simulation application has been underdevelopment for almost 20 years (Badler et al 1979) at an estimated cost of over 15 million dollars. It is widely considered to be the most comprehensive software for graphical human modeling and simulation. However, since most of the funding for the development of JACK was provided by NASA and the United States Army, the anthropometric data which was incorporated into the JACK human simulation system was entirely based on studies which were gathered from adult male and female populations.

3. METHOD

3.1 Anthropometric Data of Children

Due to the deficiencies of the available human simulation software, the focus for the final stage of the virtual device creation system need to be changed. After a review of previous work we discovered that there was an obscure SAE study (Snyder et al, 1977) done in the mid-seventies which gathered anthropometric data for children for the purposes of design safety in automobiles.

However the information from the study was only available in journal form, so we were required to manually scan in each page of the 200 plus page study and evaluate those image files with optical character recognition software to extract the anthropometric information of the studied populations. Since optical character recognition is by no means an exact science, the database was sequentially verified with a custom designed piece of software, which examined the information in a number of ways to insure the integrity of the information.

The data of the oldest population set was first compared with the anthropometric data of the JACK system, to insure that none of the information deviated from a set of preset norms. In the event the data did vary, the original document of the anthropometry study was manually examined to determine the proper value of the anthropometric table cell. Once the oldest population set was verified, the second oldest population was put through the same procedure, however this time the oldest children population data set was used as the baseline for evaluating the preset norms, instead of the JACK data set. Each population age range was recursively verified using this method.

3.2 Extrapolation of Missing Data

In some cases, we were not able to directly transfer the anthropometric data from the original anthropometry study. This was due to a lack of standards in anthropometric assessment. For example, in the case of the hand, anthropometric information on the thickness of the hand was not gathered, so this information was estimated using the thickness of the index finger, thumb and wrist as control variables. Only four anthropometric values needed to be estimated in this manner.

™ JACK is a trademark of the University of Pennsylvania

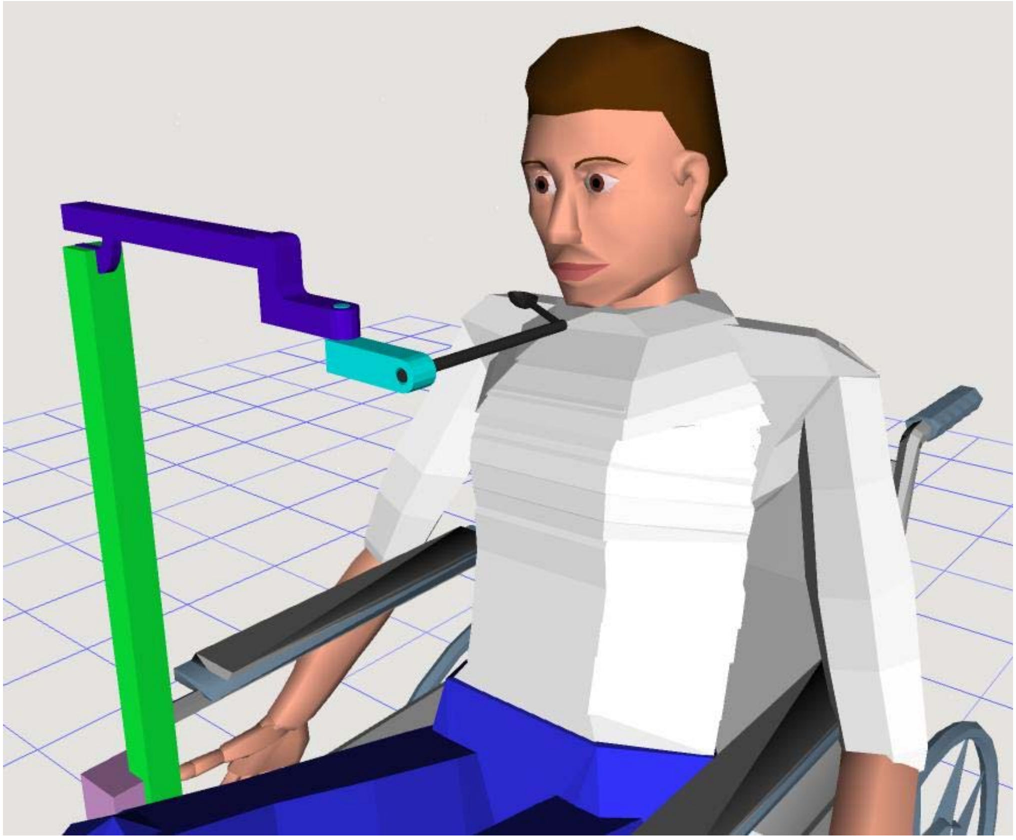


Figure 3. *An anthropometric model of a 14 year old child with the Magpie feeding device.*

4. MOTIVATION

4.1 Rehabilitation Aids

The first implication of this research is in the design of rehabilitation aids for children. A major consideration in the recommendation of rehabilitation aids for children is the amount of usability a child can get out of a device. The major disadvantages of devices, which are currently on the market, are their inability to be readily adapted, their long manufacturing time and their high cost. To avoid these problems in the future, a method of designing and fabricating rehabilitation devices has been proposed that attempts to separate the user interface design from the supporting features design (Orpwood 1990). The fundamental assumption of this method is that most of the supporting features can be acquired from conventional design techniques, whereas the user interfaces must be customized for each individual. Our research project has adopted this methodology and attempts to help the designer in identifying the configurable and extensible features of a device by enabling them to interact with a 3D graphical model of both the device and the child. By having the designer experiment with virtual computer models, as opposed to the costly and time-consuming fabrication of physical prototype models, this approach reduces the cost of device design and production. Additionally, an anthropometrically correct model, that can be interactively aged, allows a designer to better visualize the consumer of the rehabilitation device they are designing. The Anthropometric Models of Children allow the designer to predict how many years a device will be ergonomically usable by a consumer and enables the designer to alter the design of the device so that it will be more adaptable to the consumer's aging. The child models also help the designer realize what aspects of the device design can be fixed and which aspects need to be variable from user to user. Anthropometric alteration also helps the designer to visualize which components need to be variable as the consumer naturally changes over time, whether it be by natural aging or due to the effects of their disability.

4.2 Telerehabilitation

A second area this research has implications in is the field of telerehabilitation. The field of telerehabilitation has emerged over the past few years, as telecommunications and bioinstrumentation technologies have advanced and cost of those technologies has fallen. Bioinstrumentation can allow qualitative and quantitative information to be gathered about the child's physiological activity during a therapy session. This information can then be transmitted to a doctor, physical therapist or clinician at a remote site for diagnosis and evaluation. In this venue, the Anthropometric Models of Children database allows the proper body model to be chosen for a child and it provides an abstraction by which body animation parameters can be extracted. Those body animation parameters can then be encoded in a standard MPEG4 bitstream, transmitted to the remote site, decoded and reanimated for the individual who is monitoring the therapy session or making a diagnosis of the child's condition. By creating an abstraction between the definition of the body and the data stream which controls that body, the Anthropometric Models of Children help to reduce the telecommunications bandwidth which will be one of the more limiting factors in developing techniques for the effective practice of telerehabilitation.

4.3 Weaknesses

Admittedly, a weakness of the database is the fact that the populations, from whom the anthropometric data was gathered, were not disabled. The populations of children who participated in the original anthropometric study exhibited natural patterns of development and growth, whereas the physical development of children with disabilities is often effected by their particular disability as well as the treatment and therapy which they receive. This is an area that needs future improvement, but it does not prevent the database from being used effectively.

5. RESULTS

The developments of this project has produced 16 sets of male children models, each with a 5th, 50th and 95th population percentile sizing. The data for the additional 16 sets of female children models has been gathered, but have not yet been added to the database. Additionally, a prototype system is still underdevelopment to automatically incorporate assessment data into the population data of the model database.

An enormous amount of work has gone into making the Anthropometric Models of Children database a reality. For a research project that was never formally proposed, the Anthropometric Models of Children have been quite successful. The models are currently being using in at least three projects outside of the duPont Hospital for Children and the VRML model database of the children takes an average of 400+ unique IP hits per week.

Additionally, the developments of this research project have been of great interested to the industry leaders in human modeling and simulation. Transom Technologies, the company that is commercially developing JACK, was so impressed by the Anthropometric Models of Children, they contacted us requesting permission to incorporate them into their Transom JACK 2.0 product release.

Further information and HANIM 1.0 compliant VRML 97 versions of the Anthropometric Models of Children can be found at: http://www.asel.udel.edu/~beitler/children

Acknowledgements: This research is supported by the National Science Foundation (Grant #MIP94-20397) as well as the Rehabilitation Engineering Research Center on Rehabilitation Robotics of the National Institute on Disability and Rehabilitation Research (NIDRR) of the Department of Education (Grant #H133E30013). Additional support has been provided by the Nemours Research Programs of the duPont Hospital for Children.

6. REFERENCES

N Badler, C Phillips, and B Webber (1993), *Simulating Humans: Computer Graphics Animation and Control*, Oxford University Press.

N Badler, J O'Rourke, and H Toltzis (1979), A Spherical Representation of a Human Body for Visualizing Movement, *IEEE Proceedings, 67(10),* pp. 1397-1403.

P Bapu, S Evans, P Kitka, M Korna, and J McDaniel. (1981), *User's guide for COMBIMAN programs*, Technical Report AFAMRL-TR-80-91, Univ. of Dayton Research Institute, U.S.A.F. Report.

J Denavit, and H Hartenberg (1955), A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices, *Journal of Applied Mechanics*, pp. 215-221.

Humanoid Animation Working Group of the VRML Consortium (1997), *HANIM 1.0 Specification*, http://ece.uwaterloo.ca./~h-anim/spec.html.

J McDaniel, M Korna, P Krauskopf, D Haddowx, S Hardyal, M Jones, and J Polzinetti (1988*), User's Guide for CREW CHIEF: A computer graphics simulation of an aircraft maintenance technician*. Technical report, Armstrong Aerospace Medical Research Laboratory, Human Systems Division, Air Force System Command, Wright-Patterson Air Force Base, OH.

MPEG4 Synthetic Natural Hybrid Coding Ad-Hoc Group on Face and Body Animation (1998), http://www.es.com/mpeg4-snhc.

R Orpwood (1990), A Design Methodology for Aids for the Disabled. *Journal of Medical Engineering and Technology*, pp. 2–10.

R Snyder, L Schneider, C Owings, H Reynolds, H Golomb, and M Schork (1977), *Anthropometry of Infants Children and Youths to Age 18 for Product Safety Design*, Society of Automotive Engineers.

A Wisaksana, G Verburg, and S Naumann (1995), The Development of An Interface Between the Spaceball, *RESNA 95 - Proceedings*, page pp. 478-480.

The invisible keyboard in the air: An overview of the educational, therapeutic and creative applications of the EMS *Soundbeam*™

Tim Swingler

463 Earlham Road, Norwich, NR4 7HL, UK

tim@soundbeam.prestel.co.uk

ABSTRACT

Soundbeam is a 'virtual' musical instrument, an invisible, elastic keyboard in space which allows sound and music to be created without the need for physical contact with any equipment. The system utilises ultrasonic ranging technology coupled to a processor which converts distance and movement information into MIDI. Originally developed for dancers - giving them a redefined relationship with music - Soundbeam has proved to have dramatic significance in the field of disability and special education, because even those with profound levels of impairment are, even with the most minimal movements, able to compose and to instigate and shape interesting, exciting and beautiful sounds. Individuals who may be especially difficult to stimulate can benefit from what may for them be a first experience of initiation and control. A continuum of applications - ranging from the fundamentals of 'sound therapy' (posture, balance, cause-and-effect) through to more creative and experimental explorations in which disabled children and adults become the composers and performers of enthralling musical collaborations, and beyond to interactive installations and multimedia performance - can be described.

Soundbeam first appeared in prototype form in 1984 and was taken up in a serious way in special schools in the UK and subsequently in Scandinavia, the Netherlands and elsewhere following its launch in 1990. A totally redesigned version of the machine will be ready in the Autumn of 1998.

1. INTRODUCTION

Recent advances in movement sensor and other technology have utilised various electronic media to create human - machine interfaces with apparently significant but largely unexplored educational potential. Such systems have often been conceived and developed for purposes outside the sphere of formal school education or therapy (e.g. performance art, dance, electroacoustic composition, VR environments, video animation), yet preliminary research demonstrates that it is the educational and therapeutic environment in which such technology can contribute most dramatically; specifically to the educational experience of young children with severe disabilities who may have missed the foundation stones of early learning, and for whom a first experience of enjoyable control and initiation can be a crucial educational motivator.

Soundbeam™ is one example of such technology. It emits an ultrasonic beam of variable range. Movements within the beam generate data interpreted by any MIDI instrument. Described as 'an invisible, elastic keyboard in space', Soundbeam was first conceived by its originator - composer Edward Williams - as a tool for dancers. Soundbeam has proved to have dramatic significance in the disability field, because even with profound levels of impairment the most minimal movements can instigate and shape interesting sound effects, trigger rich and exotic aural textures, or perform soaring improvisations. Individuals especially difficult to stimulate can benefit from what may for them be a first experience of initiation and control.

The idea of a musical instrument which could be played without any physical contact was first developed by Leon Theremin, the Russian composer whose 'Thereminvox' astounded audiences in the 1920's, and which can be heard on the soundtracks of countless low-budget sci-fi movies. and famously on the Beach Boys' *Good Vibrations*. Most people have never heard of the Thereminvox, but nearly everyone has heard one played. In spite of its 'scary monster' associations, the Thereminvox (which is currently enjoying something of a comeback) is intended as a serious musical instrument. Repertoire exists for it, and the performances of Theremin virtuoso Clara Rockwell attest to the instrument's expressive power.

Soundbeam's development was inspired by the Thereminvox. There are, however, two essential differences between the two machines. Firstly, whereas the Theremin creates a fixed playing zone close to the device itself, Soundbeam (designed for large-scale performance) incorporates a variable ranging control which allows the invisible beam to be compressed into a few centimetres or stretched out to cover an entire stage area. In practice this means that the invisible instrument can be varied in size to accommodate the movements that the player wishes to perform, or is capable of performing. The second key difference is that whereas with the Theremin variations of timbre were not available (you were more or less stuck with the scary monster sound), Soundbeam - on its own - produces no sound at all. Instead, the machine functions as an information processor which translates distance and movement data into signals which are understood by electronic instruments, so that any sound which a given synthesiser or sampler generates can be 'played' with the beam.

Soundbeam works by emitting an invisible beam of high frequency sound (ultrasound) inaudible to human ears. The ultrasonic pulses are reflected back into the device's sensor by interruptions of and movements within the beam. Information about the speed and direction of this movement is translated into a digital code (MIDI) which is understood by a growing proliferation of electronic musical instruments, from relatively simple and inexpensive home keyboards to professional studio-quality synthesisers, samplers and expanders.

The system incorporates a number of control parameters. Variations in the send/receive switching speed of the transducer allows variation of the beam’s range from a minimum of 25 centimetres to a maximum of 9 metres. Shorter beams concentrate note information into a relatively small space, a set-up which has proved to be of significant value in special education where the player's ability to move expansively may be limited by disability. Longer beams allow a complete performance space to be 'live' with sounds. *Transpose* settings allow semitonal modulation of the scales contained within the beam, useful where the beam is played alongside other electronic or acoustic instruments. *Mode* settings govern the realisation of note information contained in the beam. Each of the Mode presets comprises four variables: the number of notes potentiated (between one and sixty-four), the relationship of those notes to one another (scales, chords and arpeggios), the articulation required to activate the note (the dynamics of the movement in relation to the sensor), and secondary information such as velocity, pitchbend and modulation depth. In addition to the resident Soundbeam 'memories' it is also possible to load several user-defined sequences of notes and chords into the beam. This has the benefit of enabling a considerable degree of compositional exploration without the necessity for a commensurate level of keyboard skill: an idea can be programmed in at the user's pace and then performed in real time with dance or simple movement. Theme-based note strings and chord progressions can also be entered into the machine’s memory for subsequent real-time improvisations in the beam. Up to four sensors can be attached to the controller, making possible the creation of a three-dimensional playing space, with beams each of different length, each angled in a different plane, and each with its own sound or 'voice'. Alternatively one or more of the sensors can be used to operate MIDI controllers such as volume, panning, modulation wheel, pitchbend etc.

2. SOUND THERAPY

In traditional music therapy, the less the client is able to say something with sound because of a physical or cognitive disability, the heavier becomes the therapist's responsibility for empathy and interpretation . The main focus and engine for the mood and meaning of the music which is happening is on the therapist, and this creative and interpretative role is increasingly shifted away from the child with more profound levels of disability. Consequently, as the liberating potential of musical expression increases, it becomes correspondingly less achievable. This allocation of creative 'power' may have no clinical or therapeutic rationale, it may simply result from what is physically possible. New evidence suggests that technology like Soundbeam can provide answers to this problem.

The experience of *initiation* is central to the success of this approach, especially for individuals with profound disabilities. If one’s overall experience of life is essentially passive, it may be difficult to develop any concept of 'selfhood', of oneself as a separate individual. What devices such as Soundbeam offer, perhaps for the first time and regardless of the individual's degree of immobility, is the power to *make something happen*. This is the vital experience of "that was me!", which can function as the foundation stone for further learning and interaction. This use of sound as the source of motivation is an extremely simple but crucially important application of the technology; it is impossible to overstate its value.

‘Aesthetic resonation’ is a term coined by Dr Phil Ellis (‘96, ‘97) to describe special moments experienced by individuals described as having profound and multiple learning difficulties, in which they achieve total control and expression in sound after a period of intense exploration, discovery and creation. Enjoyment and self-motivation are key aspects of this work, which Ellis describes as *Sound Therapy.*

Ellis’ work has provided the first systematic long-term evaluation of Soundbeam's potential for children and adults with disabilities. The beam is positioned so that as soon as the child begins to move an interesting sound is triggered, motivating further movement and, eventually, radically enhanced posture, balance and trunk control. All of this is accomplished in parallel with a strong sense of fun and achievement. For the child, the therapeutic dimension of what is happening is irrelevant. Ellis also discusses some of the broader aesthetic issues connected with his approach:

> “There are differences between sound and music, but the term ‘music’ may now encompass a broader sound spectrum due to the possibilities which have emerged during this century through the increasing use of electricity in music.... Furthermore, through sound synthesis electricity has made it possible to discover and create sounds which have never before been heard, and which could not be created any other way. In addition, we can simulate... a range of acoustic environments - concert halls, rooms, cathedrals or other large spaces for example, or may create acoustic environments which are impossible to encounter in the external physical world. This aural richness and variety provides the internal motivation which lies at the heart of this approach. In addition the technology also provides physical access for the disabled.
>
> ....Sound itself is the medium of interchange... This approach contrasts with traditional models of music therapy, with its emphasis on ‘treatment’, direct intervention and imposition of external stimuli determined by an outside agent. Even where a music therapist may claim to be ‘responding’ to a patient’s music, this is a personal response on the part of the therapist. Often the therapist uses, or moves towards, a traditionally based musical language comprised of melody, harmony and rhythm, so limiting the soundscape and genre of ‘musical’ discourse. The ‘patient’ or ‘client’ is viewed in a clinical way, with a condition which needs to be treated or ameliorated. There are clearly defined goals with these treatments, with success measured according to how effective the treatment has been in terms of the clinical or medical condition. The *modus operandi* of these approaches is essentially from the outside -in, with an emphasis on clinical intervention rather than independent learning.
>
> In Sound Therapy a different, contrasting approach is taken. Whilst progression and development remain a key focus, the essence lies in the internal motivation of the child, in working from the inside - out. This internal motivation is produced through the use of sound within a carefully controlled environment. At all times the child is given the opportunity to independently take control of the situation as far as possible. Certain aspects are controlled externally - notably the sonic environment - but the essence lies in allowing the child freedom to act as she or he chooses within available parameters which remain as open as possible. In this way, learning occurs incidentally. As a result we can see progression and development in a variety of ways across a range of disabilities. Such progression is not prescribed in advance, but happens as a natural and additional part of activity, all stemming from the internal motivation of the child - a phenomenon referred to as *aesthetic resonation*. This is made possible through a particular use of sound as the primary medium of interaction, and through giving access through the use of technology, so enabling even profoundly handicapped children the opportunity for expression and control - in other words the encounter with and development of communication skills - through sound.”

From systematic analysis of videotape session records, Dr. Ellis has identified nine criteria of progression and development:

1. from involuntary to voluntary
2. from accidental to intended
3. from indifference to interest
4. from confined to expressive
5. from random to purposeful
6. from gross to fine
7. from exploratory to preconceived
8. from isolated to integrated
9. from solitary to individual

He notes that even profoundly disabled children respond to Sound Therapy by:

- performing, listening, verbalising, 'composing' with sound;
- often showing 'aesthetic resonance' through most telling facial expressions;
- being actively involved for extended periods of time;
- revealing an ability for concentration not apparent elsewhere;
- beginning to discover, explore, give expression to and communicate their own feelings;
- making significant physical responses - movements and gestures which hitherto have not been seen, or have not previously been made independently.

> "...in addition, a change has been seen in behaviour patterns beyond the immediate environment of Sound Therapy. Some children are now more self-aware and are interacting...Other children show more tolerance and a growing awareness of other people, moving towards interpersonal skills."

Kathryn Russell (Russell '96), working in Australia has also identified a number of areas of development with the special students she involves with Soundbeam, including:

- AESTHETIC AWARENESS (includes the capacity to make choices and judgements as to what sounds or movements to select and manipulate)
- IMAGINATION (anticipating sounds and movements well ahead of time - perhaps in the week between classes)
- LISTENING SKILLS (listening to the effects of moving or standing still)
- CHOICE-MAKING SKILLS (will I choose Soundbeam? which sound will I choose? which part of my body will I move? where will I move?)
- CONCEPTUAL SKILLS (especially 'beginning, middle and end' - how will I begin? what will I do then? how will I make an ending? specific musical concepts such as high, low, fast, slow, variation; the concepts of linkage - words with movement, feeling with tone colour...)
- MOTOR PLANNING SKILLS (which movement will I make now to produce...?)
- REFLECTIVE COGNITION (how did I feel about the piece I just invented? what could I have done differently? What did it remind me of?)
- MEMORY SKILLS (can I remember which sound I liked last time? do I want to use it again?)
- SPATIAL ORIENTATION (where in space is that dog barking sound?)
- LANGUAGE SKILLS (describing what I did and how I felt, giving a title to my work).
- EXPLORING A HYPOTHESIS (I remember that if I move this way, that sound happens. If I move the same way, will I get the same sound?)
- SOCIAL SKILLS (waiting for a solo turn, sharing the beam to produce joint improvisations).
- CONFIDENCE (this is something that I *can* do).

> "Bearing in mind the extremely short attention span of many children with special needs, students have demonstrated a remarkable capacity to focus on their improvisations for long periods of time, thought previously to be beyond their abilities.....Those using Soundbeam for music education have discovered that children who are able to take control of their music making develop not only expressive and practical movement capabilities, but also create improvisatory music which has relevance and validity"

Aural stimulation - the use of sound as a motivator - lies at the heart of this kind of application of the technology. However it is possible to extend the experience by including visual stimuli, and by using *vibration*. Originally developed to allow individuals with hearing-impairment to use Soundbeam, *Soundbox*, *Soundchair* and the larger *Soundbed*, are vibro-acoustic resonators upon which the user may stand, sit or lie, thereby *feeling* the physical vibrations of the music generated by the beam. The addition of a graphic equaliser allows specific frequency ranges to be enhanced or diminished. Although first envisaged as a recreational resource (its use by no means exclusively restricted to deaf users) the various clinical and therapeutic benefits of this kind of vibration are increasingly well-documented (Wigram, 199).

The concept of *Snoezelen* pioneered by Ad Verheul at the Hartenberg Institute in the Netherlands (Verheul, 1987) has been highly influential in the provision of resources in schools and other centres for people with disabilities. Typically installed as a multi-sensory room (MSR), such environments incorporate a range of equipment designed to stimulate the primary senses in a pleasing and relaxing way. Projectors, bubble columns, fans, mirror balls, soft play, ball pools, coloured lights and sound-light floors are all characteristic. Traditionally, these environments have been used with a strong emphasis on relaxation, though there is now a growing awareness of the potential of MSRs as a tool for learning and control. Bradford-based company SpaceKraft have been in the vanguard of the design and manufacture of systems which allow users of this equipment to effect direct control over the various stimuli in the environment, rather than being passively subjected to experiences chosen by carers, by using pressure mats, tracker balls, squeeze and paddle switches, etc. The Soundbeam *Switcher* applies this principle by allowing the sensitivity of the beam to be used for switching. The beam can be divided into eight, four, two or a single portion, each corresponding to a particular effect in the room.

The dilemma facing many carers, especially where the exigencies of timetables and staffing make long periods of ‘wait and see’ activity impractical, is that for the more ‘unreachable’ students, the ‘that was me!’ moment of realisation - the revelatory experience of making something happen (perhaps for the first time) - may take hours, weeks or even months to happen. It is extremely difficult if not impossible to make an accurate assessment as to how long it will be worth persisting with activity which involves such opportunities for control. Yet these are the very students for whom the benefit of such an experience will be most profound.

One solution which this kind of configuration makes possible is to create a sensory world which can be enjoyed passively, but which is also permanently ready for interaction and control as soon as the student begins to investigate it by moving. Two sensors are used, angled in such a way as to intersect. Used in this way, the machine independently creates complex rhythmic and melodic structures without the need for any human intervention. These intersecting beams are positioned horizontally about 0.5m above the surface of the Soundbox resonator. A bank of eight coloured pinspots connected to the Switcher are positioned nearby. Lying on the resonator, the user feels and hears the sound structures generated by the beams and at the same time sees projected above patterns of coloured light which change in parallel with the sound. Hand or arm movement effects changes in these stimuli. For example, movement of the right hand will alter the pattern of notes and rhythms, whereas movement of the left hand changes the switching of the lights. This configuration of equipment is of value in three main ways:

- It allows for an optimal level of sensory input, providing auditory, visual, and tactile stimulation, combined in a dynamic and aesthetically pleasing manner.
- The experience is available to users who are unable or disinclined to exploit it in an interactive way, because the intersecting beams create the changing patterns of sound, light and vibration autonomously, without the need for human intervention. However the option of interaction and control is ever present. The idea is to create a stimulating environment which can be used passively, but in which interaction and control are totally accessible at a point when the user wants to or is able to respond. Relaxation/interaction ceases to be an either/or dichotomy.
- As learning develops, users are able to exercise increasingly sophisticated levels of choice, moving to trigger favoured colours, sounds and frequency ranges.

3. FROM THERAPY TO PERFORMANCE

Russel’s findings, like those of Ellis, provide a systematic assessment of physical and affective responses, giving quantifiable data about Soundbeam's clinical possibilities. In parallel with this approach, many projects in the disability field are exploring the creative performance-based paths opened up by this new technology. At 'The Ark' in Bracknell, a multi-arts project for people with learning difficulties, dancer Penny Sanderson and musician David Jackson's workshops always involve a narrative theme, and include live music, dance and drama, providing an excellent example of the way in which technology can be used to complement and enhance a successful established activity (rather than a sterile 'technology-led' approach) in a way which allows participants a fuller involvement. The sounds and tunes triggered by the students' movements in the beam have a place in the story and are tailored to the personalities and moods of the individual students. Dinosaurs, waterfalls, butterflies, monsters, princesses, lions, explosions - this aural dimension is all controlled and modulated by the students’, and this in turn reinforces and remotivates their involvement.

Fransisco Borges da Souza, a music therapist working in Portugal, has formed a rock band with Soundbeam as one of the key instruments. The astounding expressiveness of the disabled Soundbeam player clearly reveals a talent that has been unlocked.

Special schools in England are now starting to collaborate with so-called 'mainstream' schools on music projects involving Soundbeam. As the children are able to learn and perform on an equal basis, the disabled/non-disabled barriers can be broken down. English special schools are often isolated and the arts can clearly give a strong focus for integration where this can be enabled by appropriate technology. Meldreth Manor School in Hertfordshire and the Ormerod School in Oxford are outstanding examples.

We are taught that 'serious' musicianship demands years of dedication. So what are we to make of devices which allow musical expression to happen almost immediately, and how can the musical 'validity' of what we hear be assessed? With conventional instruments, designed for those with average or above-average physical, mental and sensory functioning, the time gap between musical imagination and musical realisation takes years to develop. Good technology radically shortens this gap. It extends the limits of selected-scale or percussion based work, and it asks the player to learn not the technical skills of the traditional instrumentalist but the freedoms and disciplines of improvisation. This kind of music is difficult to evaluate because there are no right or wrong ways of playing it - no performance of a piece of music played with Soundbeam will ever sound the same twice; but it is possible to assess the extent to which the student enjoys it and gets a feeling of achievement from it, and some of the work reviewed here indicates strongly that the attainment of significant milestones in the physical, cognitive and social development of individuals with a range of disabilities can be radically assisted by the use of such technology.

4. REFERENCES

Ellis, P, 'Sound Therapy' in *Special Children*, Nov/Dec 1995, pp 36-39.

Ellis, P, 'Incidental Music' video with booklet 'Sound Therapy, The Music of Sound', The Soundbeam Project, 1996.

Ellis, P, 'Special Sounds for Special Needs: Towards the Development of a Sound Therapy'. *International Society for Music Education* 1994 Conference proceedings.

Ellis, P, 'Developing Abilities in Children with Special Needs - A New Approach' in *Children and Society*, 1995, 9:4, pp. 64-79.

Ellis, P, 'Sound Therapy' in *Primary Music Today*, issue 3, September 1995.

Ellis, P, 'Incidental Music: a case study in the development of sound therapy' *British Journal of Music Education*. (1995), 12, pp. 59-70

Ellis, P, 'The Music of Sound: a new approach for children with severe and profound and multiple learning difficulties'. *British Journal of Music Education*, 1997, 14:2, pp. 173-186.

Ellis, P, 'Soundbeam !' in *Music in the Curriculum*, issue 31, 1996, p.8.

Jackson, D, 'Ark in Performance' (video interview transcript). Soundbeam Project 1992.

Jackson, D, 'Soundbeam Artistry' creative user manual, Soundbeam Project 1995.

Jackson, D, 'More than the Music' in *Special Children*, Nov/Dec 1994, pp 31-32.

Rickson, D, 'Soundbeam'. New Zealand Conference on Technology for People with Special Needs. <http://nz.com/webnz/ability/csn_papers:html>

Russell, K, 'Imagining the Music, Exploring the Movement: Soundbeam in the Sunshine State'. *Queensland Journal of Music Education*, 4,1, 1996. Pp 41-48.

Russell, K, and Russell, N, ' "See that? That's magic": New Sounds and Sights in Music Movement Improvisation - the Soundbeam Experience. ASME, Brisbane, July 1997.

Swingler, T, 'Unlocking Musicality: using Soundbeam as a new key to eloquence' in Krout, R. : *Integrating Technology*, Music Therapy Perspectives, NAMT, 1994, 12, 1, pp. 4-5.

Swingler, T, 'Dinosaurs and Butterflies' in *Extending Horizons*, National Council for Educational Technology, 1995.

Swingler, T, 'Liberation or Limitation?' in *Extending Horizons* op. cit.

Swingler, T, 'The drum is not the only weapon' in *Nordic Journal of Music Therapy*, 1995, 4 (2), pp. 103 - 107.

Swingler, T, ‘Choreography as Composition - Movement into Music: using Soundbeam as a new key to creative eloquence’. Proceedings of the Biennial World Conference, ISME, 1994,

Swingler, T, ‘La Coreografía como composicíon el movimiento en la música: la utilización del Soundbeam como un nuevo acceso a la elocuencia creativa’ in Violeta Hemsy de Gainza (Editor) *La Transformación de la Educación Musical a las puertas del Siglo XXI.* ISME /Editorial Guadalupe 1997. Pp. 75-82.

Williams, E, ‘Introduction to vibroacoustic therapy’. Soundbeam Project 1997

Soundbeam Project web site: <http://www.soundbeam.co.uk>

Author Index

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998
©1998 ECDVRAT and University of Reading, UK; ISBN 0 7049 1141 8

Keyword Index