REPORT GENERATION USING MULTI-MODAL DATA ENTRY FOR OFFICE SYSTEMS

I.McKay [1] , M.A.Jack [1] , R.Thomson [2]

1. *Introduction :*

The purpose of this paper is to introduce work from the ongoing *Tentalus* project, which aims to explore multi-modal user interfaces (or MMUI's) in the context of report generation by a professional. This paper emphasises the user interface issues regarding such MMUI's, as well as the problems involved with the integration of multiple modes of data entry into one system such that the user can choose which mode to use to enter data. A range of user-interface components available to the system designer are assessed to characterise their relative merits, with a view to preparing their optimised combination. Data entry devices discussed here are: speech recognition, bar code reader technology, pen-based character recognition, the QWERTY keyboard, and mouse-activated pull-down lists.

Of primary concern to the research is the design of the user interface, whether in two dimensions, or in three, using virtual reality. The latter allows greater freedom of navigation through more intuitive menu structures, while presenting a more familiar (office-based for example) metaphor.

An experiment was carried out into the relative performance of a number of modes of data entry using Police Officers as subjects, the results of which are detailed also, with reference to the usability of that experiment and any consequences that might have on other MMUI's.

1.1 What is a MMUI ?

A Multi-Modal User Interface (MMUI) is an interface which allows the user to utilise more than one mode of data entry as required. There may be, for example a task that requires keyboard entry, followed by using a bar code, then a pull-down list, then speech recognition which could be said to be"multi-modal". There is no set definition of a MMUI. A three dimensional interface, in contrast to a two dimensional interface has the advantage of potentially looking more like the real world, and being able to use stronger visual metaphors. If an inexperienced user is presented with a likeness of the real world, then that user should be able to use his/her real world experience to interact with the interface. For example, an inexperienced user who is faced with a computer screen depicting a house has a good idea what objects that look like doors, desks and drawers should do. It is then a matter of finding out how to interact with them.

2. *Experiment Details*

This section details an experiment carried out at Fife Constabulary, examining multimodal data entry.

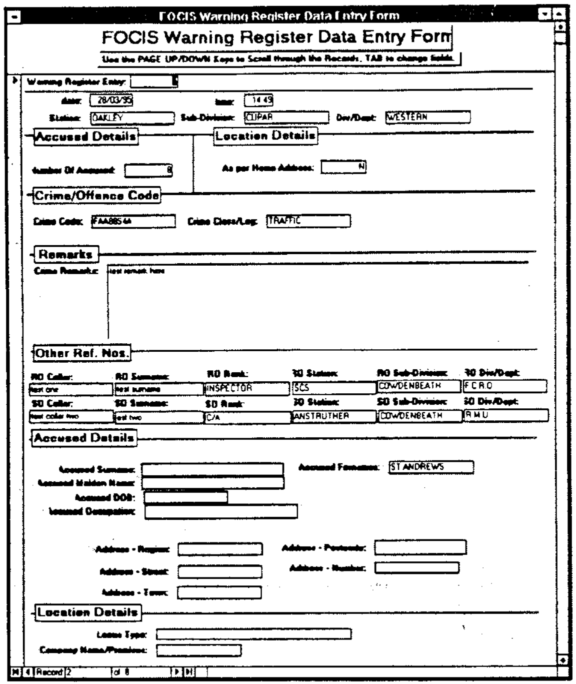
2.1 Scenario

The chosen scenario for the first experiment was that of a Police Officer making a report pertaining to a crime or warning, using a database such as the existing FOCIS system already in use. Since the FOCIS system uses

[1] 1 - Centre for Communication Interface Research, Department Of Electrical Engineering, University Of Edinburgh, 80 South Bridge, Edinburgh EH1 1HN

[2] Fife Constabulary Communications Department, Police Headquarters, Wemyss Road, Dysart, Fife. KY1 2YA

standard IBM-compatible personal computers to perform terminal emulation for the data entry task, the existing method of entry is solely the keyboard. The experiment uses a simulation of the FOCIS system, with different data entry schemes, one for each mode under investigation. The system runs under a Microsoft Windows database, as shown in 'Screenshot 1' below.



FOCIS Warning Register Data Entry Form

FOCIS Warning Register Data Entry Form

Use the PAGE UP/DOWN Keys to Scroll through the Records. TAB to change fields.

Warning Register Entry: 5

Date: 28/03/95 Time: 14:49

Station: OAKLEY Sub-Division: CUPAR Div/Dept: WESTERN

Accused Details

Number Of Accused: 8

Location Details

As per Home Address: N

Crime/Offence Code

Crime Code: FAA8854A Crime Class/Log: TRAFFIC

Remarks

Crime Remarks: test remark here

Other Ref. Nos.

RO Caller:	RO Surname:	RO Rank:	RO Station:	RO Sub-Division:	RO Div/Dept:
test one	test surname	INSPECTOR	SCS	COWDENBEATH	F C R O
SO Caller:	SO Surname:	SO Rank:	SO Station:	SO Sub-Division:	SO Div/Dept:
test caller two	test two	C/A	ANSTRUTHER	COWDENBEATH	R M U

Accused Details

Accused Surname:

Accused Forename: ST ANDREWS

Accused Maiden Name:

Accused DOB:

Accused Occupation:

Address - Region:

Address - Postcode:

Address - Street:

Address - Number:

Address - Town:

Location Details

Locus Type:

Company Name/Premises:

Record 2 of 8

Screenshot 1, the FOCIS-like experiment front end (keyboard mode shown)

The experimental procedure involved asking the subjects to enter the same data into the system, using one of each mode of entry (keyboard, speech recognition, bar-code, character recognition, pulldown lists) in a random order. The data for the scenario to be entered was chosen by selecting files from many different actual warnings in the existing FOCIS system. Long names were chosen for fields such as 'Station Name', in order that time differences between modes might be more apparent. Where some of the data to be entered did not belong to a discrete list, keyboard entry still had to be used[3]. For the scenario to be similar to a real one, these keyboard-only fields could not be avoided. Since the type of mode for each field is known, the performance of individual fields can be scrutinised. The experiment was completed by ten Police Officers.

3. Experimental Results

The FOCIS-like system was trialed at Fife Constabulary's headquarters for the convenience of the subjects (Traffic Division Officers), and to fit in with the scenario of entering data into a warning register at the station, during a hurried shift change.

3.1 Speech Recognition

Initially subjects tried to use the commercially available speech recogniser, to establish whether it would perform well enough to be used in the full experiment. Firstly, the recogniser was tried with an untrained, out-of-the-box user vocabulary. The recognition rates were unacceptably low over the first 25 minutes for the subject to enter and train enough fields for one warning record, even while missing out harder tasks like spelling the offender's name and address.

3.2 Handwriting Recognition

A similar trial of the handwriting recognition also proved to be unsuccessful, achieving about the same low level of recognition accuracy as the untrained speech recogniser. The recognition was seen to improve when

[3] A sheet of bar-codes or a pull-down of the alphabet could have been used to preserve the integrity of the experiment, but would have been tedious to the point of making the entry task prohibitively long.

hand-written letters had their differences exaggerated, a slow task. A second and third subject tried the handwriting recognition, and had similar accuracy rates. The method of selecting portions of bad text for deletion or replacement was found to be 'annoying' and 'frustrating'.

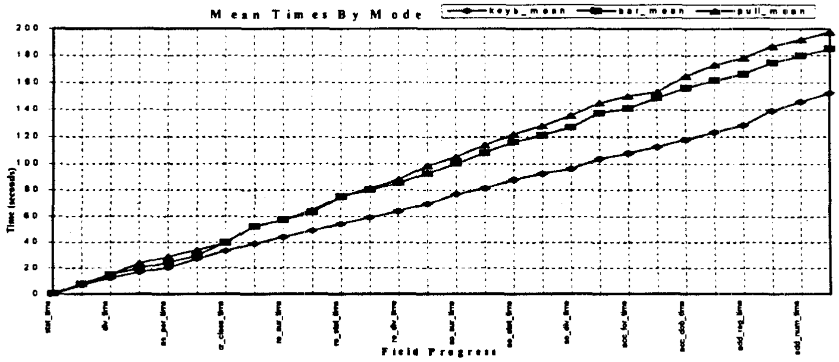
3.3 Other Modes

Because of the length of time taken to achieve any simple recognition with the speech or handwriting modes, it was decided that neither could be used in the main experiment unless an extra 30-40 minutes per subject could be found. With this figure being so impractical, these two modes were dropped from the main experiment. The main experiment then, consisted of asking ten subjects to enter the same scenario into the dummy FOCIS system three times, once each with a different randomly-ordered mode of entry.

Each subject was given the scenario to enter by means of a printed sheet of instructions. The scenario was deliberately chosen to be none to complex, in order that it was a subject's entry rate under investigation and not their mental ability to recall information. Each subject took approximately 15-20 minutes to complete the three modes of entry, after which they were given two questionnaires to fill in. One asked background information on the subject's age and experience (both Police and computer usage), while the second asked specific questions about the experiment.

At the end of the experiment, ten subjects had been succesfully processed. In general, all had been very positive about taking part, especially upon their realisation that the goal was to find them an easier and less time-consuming method of entering data.

The mean times for each mode upon exit from every field were calculated, and plotted on Graph 1, shown below:



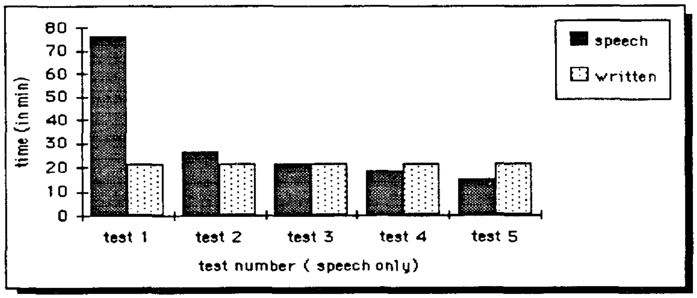
Graph 1 : Mean entry time upon exit from each field

The mode with the largest deviation from the mean was the pulldown mode and the keyboard was the fastest mode of entry of the three under investigation. The keyboard mode introduced spelling errors into those fields that the pulldown or bar-code could not though , as many as four incorrect fields in one record in one case. Even though the keyboard was the quickest mode of entry for each subject, only half said that it was their preferred method of entry. The reason for the other half not choosing it could be attributed to comments made, such as "you have to keep checking what you type, and are forced to think of the spelling etc.".

Bearing in mind that the bar-code and pulldown (in this context) was new to all subjects, then there is no measure of the learning effect. Ideally, the experiment would have involved entering a series of scenario to see if the entry rates improved (and if so, by how much) with repetition.

4. Measuring Speech Recognition Accuracy

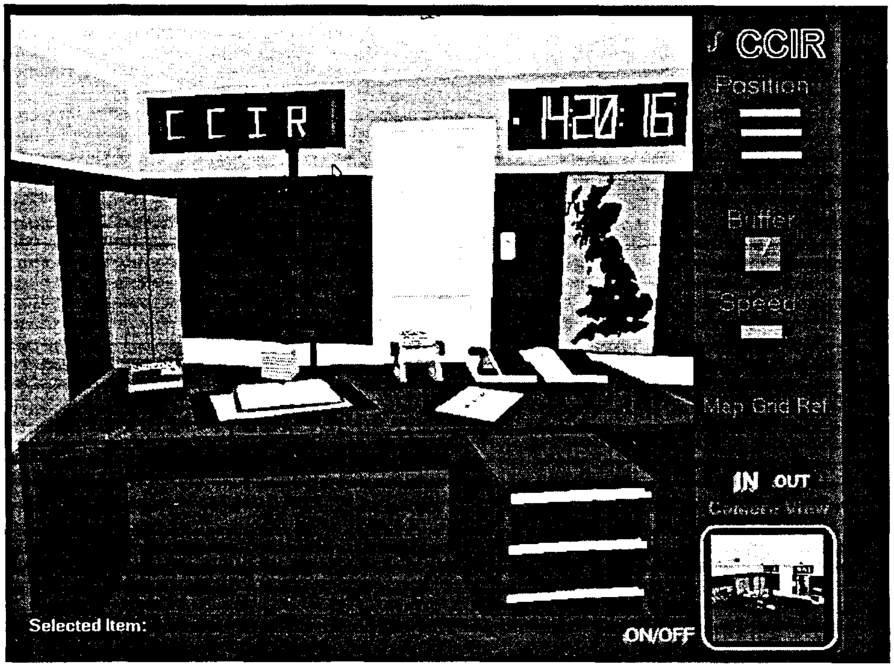
As a side issue, an experiment was carried out into measuring the recognition accuracy of the same speech recogniton package with a single (previously unheard) user. The time taken for the user to complete a scenario was measured against the time taken to hand-write it, and the process repeated over multiple iterations. Because an incorrectly recognised word means that the user has to retry that word and train the recogniser, the total time taken on that scenario is a good guide of the overall accuracy with different levels of training. The results of this are shown in *Graph 2* below, showing that using Speech Recognition in this particular case became faster than handwriting after three iterations, the biggest improvement coming between the first and second.



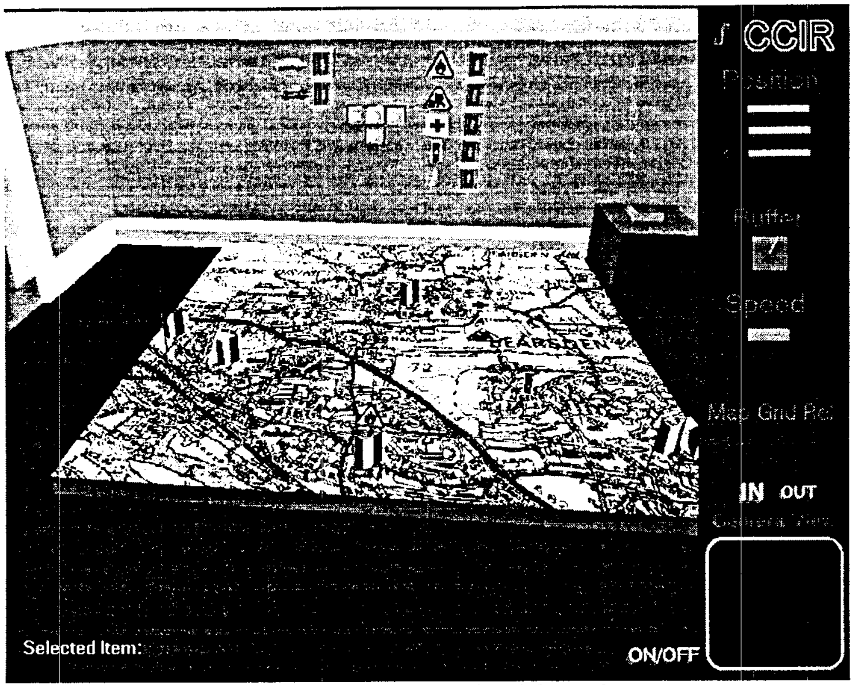
Graph 2 : Comparing dragon dictate with handwriting

5. The VR Office - a concept demonstrator

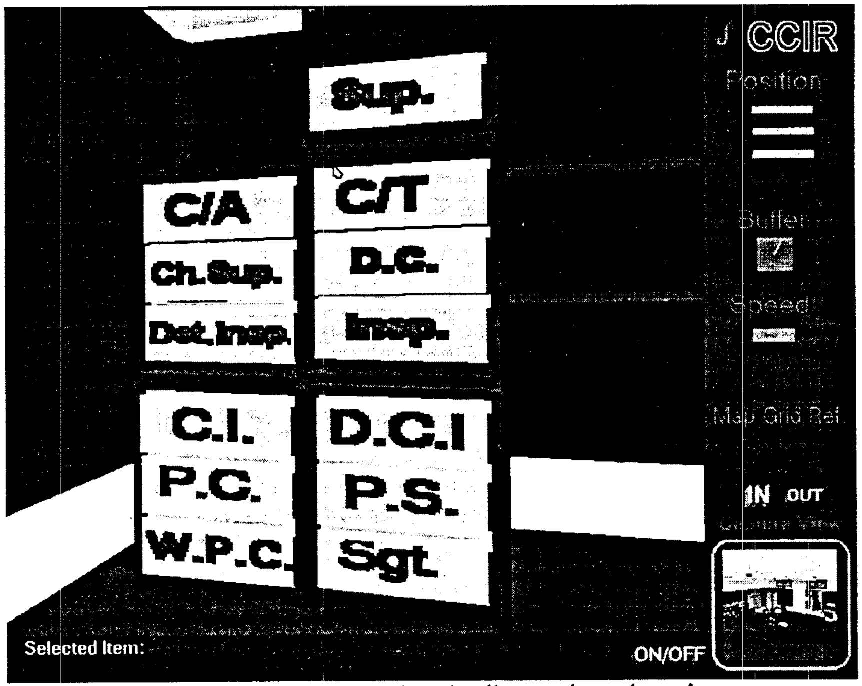
As an extension of the FOCIS experience, a virtual office has been created on a desktop PC, using a commercial 3D virtual reality modelling package, investigating such a crime reporting scenario. The term 'Virtual Reality' (or 'VR') was first used in 1965 [1], to mean an environment in which the user feels 'immersed'. The VR Office features scrolling maps, voice-controlled movement, desktop-planning tools, virtual communication links, speech-recognition of crimes and Acts of Parliament. Some objects in the world had to be created specifically for the desired scenario, such as the those needed to perform command and control functions, while others are placed in the world merely to help it adhere to the office metaphor, such as the coffee table with a coffee machine on it [2]. Pictures of the virtual office follow. Screenshot 2 shows the main data entry room, with a door into the command and control room (Screenshot 3), and the 'rank' selector (Screenshot 4). The latter is the three-dimensional analogue of a two-dimensional pulldown list. As in any software design that involves input from the user, sound Human Factors principles [3,4,5,6,7] should and have been applied in order that the user can recognise any metaphor, have the concept of localisation in any dialog, and not be bombarded with information and possible actions to be taken.



Screenshot 2, one room of the Virtual Office



Screenshot 3, The Command and Control room



Screenshot 4, The 'rank and collar number selector'

5.1 Integrating Speech Recognition and Synthesis into the System

The two chosen 'off-the-shelf' systems to implement the speech recognition and synthesis were 'Dragon Dictate for DOS' and Creative Lab's 'SBTalker'. Two DOS-based systems were chosen since the VRT system runs under DOS, and it was hoped that these TSR-based programs might co-exist with VRT. The Dragon Dictate systems functions by replacing standard keyboard input to DOS applications with recognised spoken words. Once run, all the functions are available through a hot key on the keyboard, and a small 'pop-up' dialog box appears under DOS. Unfortunately, the VRT system uses a different method of performing the video output, which becomes corrupt if Dragon Dictate's dialog box pops up. It was thought that C-code to run the SBTalker external DOS executable could be written with the VRT SDK. This was tried, but with little success - the SBTalker executable refused to recognise that the TSR's had already been loaded. Thus, neither the VRT or SBTalker applications would run on the same PC as the VRT. Another approach was needed, which would involve less interference between the component parts of the system.

5.2 The Solution

It was decided that due to the above limitations the speech recognition and synthesis systems would be better placed on separate PC's, communicating with the main VRT PC by means of serial cables. This has the advantage that each function is logically abstracted to a separate PC, and allows each to run faster and more efficiently. The disadvantage of this approach of course is that it is more costly in terms of computer hardware, needing two more PC's. Also, it is not a trivial task to redirect input to the off-the-shelf systems from the keyboard to a serial port. A layer of C-code is needed under the VRT, adding serial port access functions to the API, and some internal VRT functions to manage serial port buffers, and talk to the recognition and synthesis PC's. This system works, and is shown in Diagram 1 below :

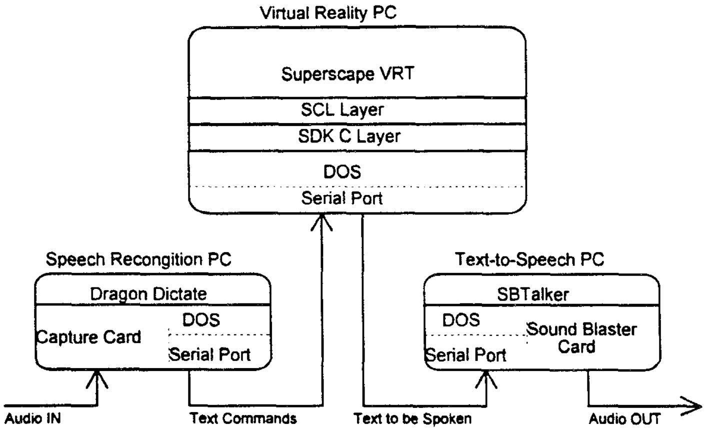


Diagram 1 : Integrating Speech Recognition and Synthesis into the VR System

The speech recognition system was trained with movement commands and words found in a warning register. A book object was included in the office, which, when selected, brought up a 2D representation of a data entry form, allowing fields to be entered from the keyboard and received from the second PC. This form followed the format of the FOCIS simulation described earlier, but did not have as many fields. The user inhabiting the virtual office can now navigate through the office without having to use the 3D Mouse, by simply saying the words 'left', 'right', 'forward', 'fast' and so on, allowing the user to navigate in three dimensions. Those commands recognised by the isolated-word Speech Recognition system are given in *Table 1* below. Relevant issues concerned with user interaction can be found in [8].

5.3 General

At present, virtual drawers can be opened to select the officer's rank, a globe and map can be scrolled to pinpoint a location, model cars can be moved around a desktop map, and flags on a map can be selected to indicate police stations.

<Map>	<up>	<forwards>	<select><calculator>
	<down>	<backwards>	<pen>
	<left>	<fast>	<clock>
	<right>	<slow>	<zoom><in>
<door>	<open>	<stop>	<out>
	<close>	<reset>	also : 8 police stations and 12 ranks

Table 1 : Verbal Commands Recognised by the Virtual Office

The basic function of the virtual office is to allow the user to enter data into an underlying database by interacting with a 3D world as they would in the real world. No real underlying database exists as yet, all data gets held within the virtual reality software. The novelty in using a 3D interface [9,10] such as the virtual office lies in the fact that it can embed other modes of data entry within it, as well as showing 2D interfaces as well if need be. At the moment, the speech and bar-code entry modes are the only ones integrated alongside the standard keyboard and 2D mouse interaction.

Police Officers have shown great enthusiasm for the 'command-and-control' capabilities of the 'virtual office', especially if tied into the general data-entry tool. A computer has great advantages when it comes to modeling, differing viewpoints, networked training, and recording and playback of simulated scenarios. Other 3D interfaces have been suggested for information input and retieval [11], and some of the more imaginitive and ground-breaking suggestions have come from novellists[12].

6. Conclusions

Keyboard entry was the fastest mode for every subject, though if time had allowed for the learning effect to be measured, the results might have been different. All of the subjects came from a keyboard data-entry background, and took some 'getting used' to the pulldown list especially. A second experiment is proposed to investigate such a learning effect. Despite being the quickest method, only half of the subjects preferred it, the main objection being that it forced them to think of spelling. Some subjects wanted bar-code and pulldown entry to work since they forced correct spelling, and marked them as being their preferred choice.

Both the speech recognition and pen-based character recognition systems chosen were not accurate enough to be used in a trial without user enrollment. The two systems used did however act as good indicators of how the technology might work in the future, when out-of-the-box recognition rates become practical and users are allowed to become familiar with the technology. For any real FOCIS system using bar-code entry to be practical, the sheets of bar-codes would have to be bound into a large book due to the large number of crime codes, making the task slower and less practical.

A virtual reality system is a good framework for building a MMUI since it allows both 2D and 3D scenes to be displayed to the user, combining the benefits of both. One major caveat is that any 2D interface must fit logically into or alongside the 3D interface and metaphor, a 2D dialog box being represented on a book page for instance.

References

[1] "The Ultimate Display", Ivan E. Sutherland, Proceedings of the IFIP conference 1965, pages 506-508

[2] "The Psychology of Everyday Things", Norman. D. A., Basic Books, NY, 1988

[3] "Handbook of Human-Computer Interaction", Helander. M (ed.), North-Holland, 1988

[4] "Interfaces", the British HCI Group Newsletter.

[5] "Human Computer Interaction", P.Johnson , McGraw-Hill , 1992

[6] "The Magical Number Seven Plus or Minus Two : some limits on our capacity for processing information", Miller. G. A., *Psychology Review 63*, 1956.

[7] "The Ten Commandments of Color", A.Marcus, *Computer Graphics Today*, Nov 1986

[8] "Interaction in a Virtual Environment", Mark R. Mine, SIGGRAPH'94 Course Notes.

[9] "Why is 3D interaction so hard and what can we really do about it", J.E.Gomez, SIGGRAPH'94

[10] "Is visualisation REALLY necessary ? The role of visualisation in Science, Engineering, and Medicine", Nahum D. Gersun, SIGGRAPH'94

[11] "The Information Visualiser: An Information Workspace", Card. S. K., Robertson. G. G., Mackinlay. J. D., *Proceedings of ACM.SIGCHI '91*, ACM/SIGCHI, New York., pp 181-188

[12] "Cyberpunk Novelists Predict Future User Interfaces", Marcus. A., Norman. D. A., Rucker. R., Sterling. B., Vinge. V, *Proceedings of ACM.SIGCHI '92*, ACM/SIGCHI, New York., pp 435-437

Printed and published by the IEE, Savoy Place, London WC2R 0BL, UK.