Univ Access Inf Soc
DOI 10.1007/s10209-011-0265-5

LONG PAPER

Evaluating interface layout for visually impaired and mobility-impaired users through simulation

Pradipta Biswas · Peter Robinson

Abstract The authors have developed a simulator to help with the design and evaluation of assistive interfaces. The simulator can predict possible interaction patterns when undertaking a task using a variety of input devices and estimate the time to complete the task in the presence of different disabilities. This paper presents a study to evaluate the simulator by considering a representative application of searching icons, which was being used by able-bodied, visually impaired and mobility-impaired people. The simulator predicted task completion times for all three groups with statistically significant accuracy. The simulator also predicted the effects of different interface designs on task completion time accurately. The simulator is used to develop inclusive digital TV interfaces. A case study is presented to investigate accessibility requirements of a representative digital TV interface.

Keywords Human–computer interaction · Assistive technology · User model · Usability evaluation · Simulator · Digital TV

1 Introduction

The World Health Organisation (WHO) states that the number of people aged 60 and above will be 1.2 billion by 2025 and 2 billion by 2050 [17]. The very old (age 80+) is the fastest growing population group in the developed world. Many of these elderly people have disabilities, which make it difficult for them to use computers. The definition of the term ‘disability’ differs across countries and cultures, but the World Bank estimates a rate of 10–12% of population worldwide having a condition that inhibits their use of standard computer systems [18]. The Americans with Disabilities Act (ADA) in USA and The Disability Discrimination Act (DDA) in UK prohibit any discrimination between able-bodied and disabled people with respect to education, service and employment. There are also ethical and social reasons for designing products and services for this vast population. In particular, computers offer valuable assistance to people with physical disabilities and help to improve their quality of life. However, the diverse range of abilities complicates the designing of human–computer interfaces for these users.

We take a novel approach to designing and evaluating inclusive systems by modelling performance of users with a wide range of abilities. We have developed a simulator that can predict possible interaction patterns when undertaking a task using a variety of input devices, and estimate the time to complete the task in the presence of different disabilities and for different levels of skill [4–7]. In this paper, we demonstrate its use in evaluating interfaces for an application used by able-bodied, visually impaired and mobility-impaired people.

We are currently using the simulator to develop accessible digital TV interfaces and present a study for investigating accessibility issues of the programme selection menu interface of a digital television.

P. Biswas (✉) · P. Robinson
Computer Laboratory, University of Cambridge,
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
e-mail: pb400@cl.cam.ac.uk

P. Robinson
e-mail: pr@cl.cam.ac.uk

2 Background

Addressing a large variety of users is always a challenge to designers due to diverse range of abilities and differences in task, prior knowledge and situation. A user model may help to understand users and analyse interaction patterns. It is a representation of the knowledge and preferences of users [1]. Three main types of user model are in widespread use:

- The GOMS family of models, which were developed only for human–computer interaction (HCI).
- Models involving cognitive architectures, which take a detailed view of human cognition.
- Application-specific models.

The GOMS (Goal, Operator, Model, Selection) family of HCI models (e.g. KLM, CMN-GOMS, CPM-GOMS) is mainly suitable for modelling the optimal (skilled) behaviour of users [8, 10]. On the other hand, models developed using cognitive architectures consider the uncertainty of human behaviour in detail but have not been widely adopted for simulating HCI as their use demands a detailed knowledge of psychology. Application-specific models are developed by keeping only a single application in mind, and so they are hardly usable to model human performance in general.

There is not much reported work on systematic modelling of assistive interfaces. Most researches in assistive technology concentrate on a particular application or a set of users, which reduces the scalability of the overall approach. Furthermore, developing systems for a small segment of market often makes the system very costly [15]. A detailed literature survey on user models can be found in a separate paper [4].

In the present work, we have addressed some of the current problems of user modelling by developing a simulator inspired by model human processor [8]. The simulator embodies both the internal state of a computer application and also the perceptual, cognitive and motor processes of its user. Figure 1 shows the architecture of the simulator.

- The Environment model contains a representation of an application and context of use. It consists of the following:
    - The Application model containing a representation of interface layout and application states.
    - The Task model representing the current task undertaken by a user that will be simulated by breaking it up into a set of simple atomic tasks following the KLM model.
    - The Context model representing the context of use like background noise, illumination and so on.
- The Device model decides the type of input and output devices to be used by a particular user and sets parameters for an interface.
- The User model simulates the interaction patterns of users for undertaking a task analysed by the task model under the configuration set by the interface model. It consists of a Perception model, a cognitive model and a motor behaviour model.

The perception model [5, 6] simulates the phenomenon of visual perception (like focussing and shifting attention) and uses those to predict visual search time. We investigated eye gaze patterns (using a Tobii X120 eye tracker) of people with and without visual impairment. My model can reproduce the results of previous experiments on visual perception in the context of HCI and can also simulate the effects of different visual impairments (like maccular degeneration, colour blindness, diabetic retinopathy, etc.). Figure 2 shows the actual and predicted eye movement paths (green line for actual, black line for predicted) and points of eye gaze fixations (overlapping green circles) during a visual search task. The figure shows the prediction for a protanope (a type of colour blindness) participant, and so the right-hand figure is different from the left-hand one as the effect of protanopia [9] was simulated on the input image. It can be seen that the predicted points of eye gaze fixation and eye movement are almost same to the actual.

The cognitive model [4] simulates expert performance by using CPM-GOMS model [10]. It can also simulate performance of novices by using a dual-space model [14].

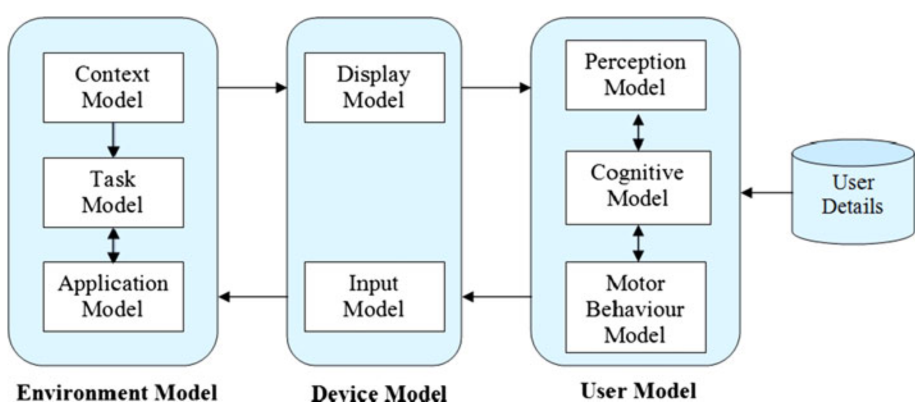


Fig. 1 Architecture of the simulator

The motor behaviour model [7] is developed by statistical analysis of cursor traces from motor-impaired users. We have evaluated hand strength (using a Baseline 7-pc hand evaluation kit) of able-bodied and motor-impaired people and investigated how hand strength affects HCI. Based on the analysis, we developed a regression model to predict pointing time. Figure 3 shows an example of the output from the model. The thin purple line shows a sample trajectory of mouse movement of a motor-impaired user. It can be seen that the trajectory contains random movements near the source and the target. The thick red and black lines encircle the contour of these random movements. The area under the contour has a high probability of missed clicks as the movement is random there and thus lacks control.

These models [4–7] do not need detailed knowledge of psychology or programming to operate. They have graphical user interfaces to provide input parameters and showing output of simulation. In the following section, we present a study to validate the simulator.

3 The Study

In graphical user interfaces, searching and pointing constitute a significant portion of HCI. Users search for many different artefacts like information in a web page, button with a particular caption in an application, email from a list of mails, etc. We can broadly classify searching into two categories:

- *Text searching* includes any search that only involves searching for text and not any other visual artefact. Examples include menu searching, keyword searching in a document, mailbox searching and so on.
- *Icon Searching* includes searching for a visual artefact (such as an icon or a button) along with text search for its caption. The search is mainly guided by the visual artefact, and the text is generally used to confirm the target.

We present a study involving an icon-searching task. We simulated the task using our simulator and evaluated the predictive power of the model by comparing actual with prediction.

3.1 Experimental design

We conducted trials with two families of icons. The first consisted of geometric shapes with colours spanning a wide range of hues and luminance (Fig. 4). The second consisted of images from the system folder in Microsoft Windows to increase the external validity (Fig. 5) of the experiment. Each icon bears a caption underneath. The first two letters and length of all the captions were kept same to avoid any pop-out effect of the captions during visual search.

The experiment was a mixed design with two measures and a between-subject factor. The within-subject measures were spacing between icons and font size of captions (Fig. 6). We used the following three levels for each measure:

- Spacing between icons
  - Sparse: 180 pixels horizontally, 230 pixels vertically. This was the maximum separation possible in the screen.
  - Medium: 150 pixels horizontally, 200 pixels vertically.
  - Dense: 120 pixels horizontally, 170 pixels vertically. This was the minimum possible separation without overlapping the icons.
- Font size
  - Small: 10 point.
  - Medium: 14 point as recommended by the RNIB [13].
  - Large: 20 point.

The between-subjects factors are the following:

- Group
  - Able-bodied
  - Visually impaired
  - Motor-impaired

Each participant undertook 8 trials for each combination of the within-subject measures. The sequence of the trials was randomized using a Latin square.

3.2 Material

We used a 1,280 × 800 LCD colour display driven by a 1.7-GHz Pentium 4 PC running the Microsoft Windows XP operating system. We used a standard computer Mouse (Microsoft IntelliMouse® Optical Mouse) for clicking on the target.

3.3 Process

The experimental task consisted of shape-searching and icon-searching tasks. The task was as follows:

1. A particular target (shape or icon with a caption) was shown.
2. A set of 18 candidates for matching was shown.
3. Participants were asked to click on the candidate which was same as the target in terms of both icon and caption.

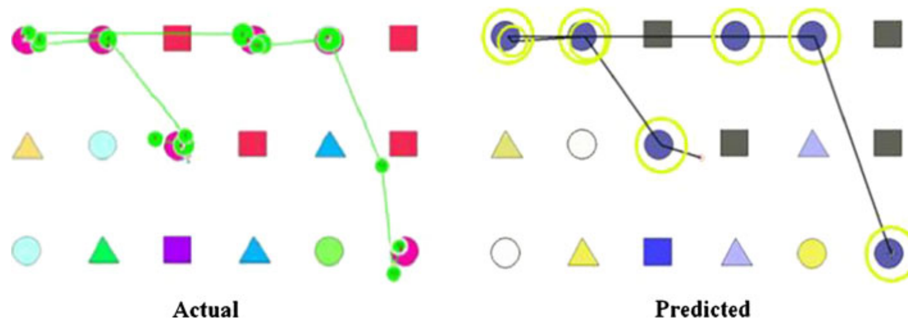


Fig. 2 Eye movement trajectory for a user with colour blindness

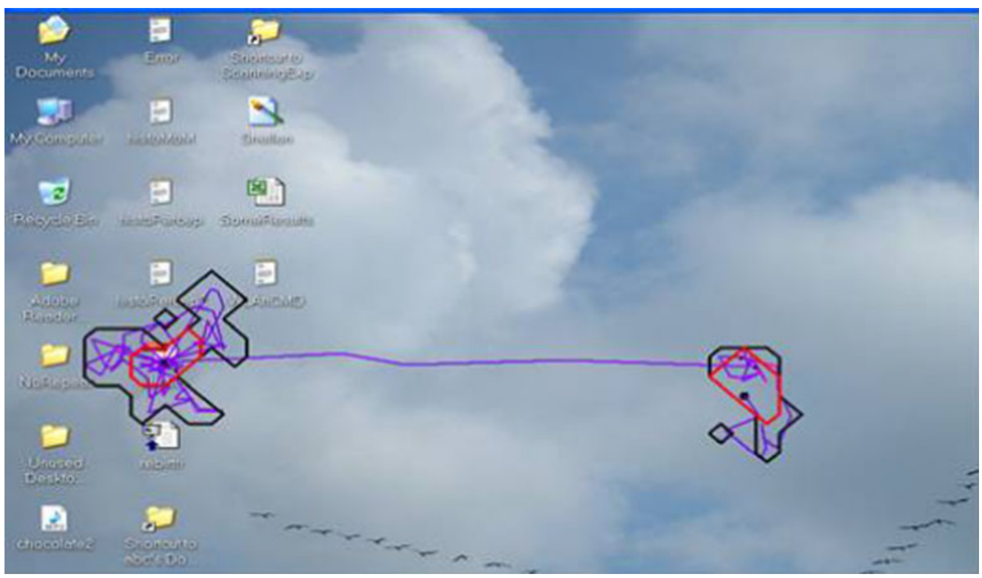


Fig. 3 Mouse movement trajectory for a user with cerebral palsy

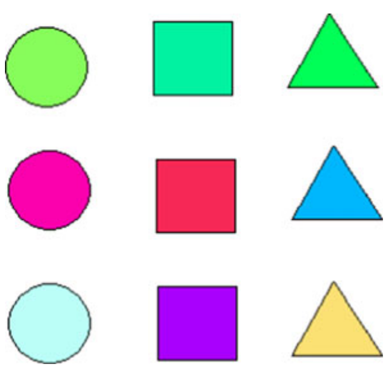


Fig. 4 Corpus of shapes

Fig. 5 Corpus of icons

Each participant did 72 searching and pointing tasks in total. They were trained for the task before start of the actual trial. However, one of the participants (P4) retired after undertaking 40 trials.

3.4 Participants

We collected data from 2 able-bodied, 2 visually impaired and 3 motor-impaired participants (Table 1). The motor-impaired participants were recruited from a local centre, which works on treatment and rehabilitation of disabled people, and they volunteered for the study. Other participants are students and staff members of our university. All were expert computer users and used computers more than once a week.

3.5 Simulation

Initially, we analysed the task in the light of our cognitive model [4]. Since the users undertook preliminary training, we considered them as expert users. We followed the GOMS analysis technique and identified two subtasks:

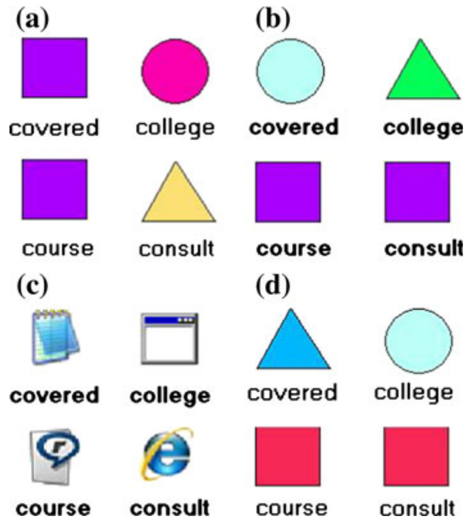


Fig. 6 Sample screenshot of the study. **a** Dense *Spacing* Big *Font*. **b** Medium *Spacing* Medium *Font*. **c** Sparse *Spacing* Medium *Font*. **d** Dense *Spacing* Small *Font*

Table 1 List of participants

	Age	Sex	Impairment
C1	27	M	Able-bodied
C2	30	M	
P1	27	M	Myopia (−4.5 Dioptre)
P2	26	M	Myopia (−5.5 Dioptre)
P3	30	M	Hypokinetic motor impairment resulted from cerebral palsy, restricted hand movement, wheelchair user
P4	42	M	Cerebral palsy, restricted hand movement, also suffering from tremor in hand, wheelchair user
P5	45	M	Hyperkinetic motor impairment resulted from stroke, significant tremor in fingers, wheelchair user

- Searching for the target.
- Pointing and clicking on the target.

So, the predicted task completion time is obtained by sequentially running the perception model [5] and the motor behaviour model [7]. The predicted task completion time is the summation of the visual search time (output by the perception model) and the pointing time (output by the motor behaviour model).

3.6 Results

Figure 7 shows the correlation between actual and predicted task completion times. We also calculated the relative error $\frac{\text{Predicted}-\text{Actual}}{\text{Actual}}$ and show its distribution in Fig. 8. The superimposed curve shows a normal distribution with

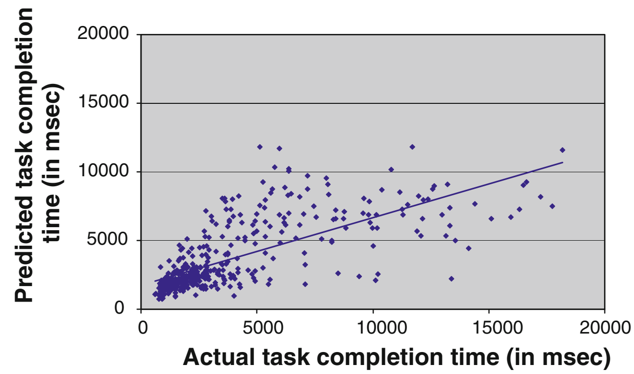


Fig. 7 Scatter plot between actual and prediction

same mean and standard deviation as the relative error. We found that the correlation is $\rho = 0.7$ ($p < 0.001$), and 56% of the trials have a relative error within ±40%. The average relative error is +16% with a standard deviation of 54%. The model did not work for 10% of the trials, and the relative error is more than 100% in those cases. For the remaining 90% of the trials, the average relative error is +6% with a standard deviation of 42%.

We also analysed the effects of font size and icon spacing on the task completion time and investigated whether the prediction reflects these effects as well. So, we conducted two 3 × 3 ANOVA (*Spacing* × *Font* × *Group*) on the actual and predicted task completion times, respectively. We investigated both the within-subject effects and results of a multivariate test. In the ANOVAs, we did not consider the trials for which the relative error was more than 100% as the model did not work for those trials. Participant P4 did not also complete the trial, leaving us with 40 rows of data ($N = 40$).

For calculating the within-subject effects, the Greenhouse–Geisser correction was used if the Mauchy's test detected violation from sphericity assumption [9] giving fractional values for the degrees of freedom. In this study, the main effect of *Spacing* did not violate sphericity assumption ($W = 0.854$, $\chi^2 = 5.69$ in actual, $W = 0.99$, $\chi^2 = 0.374$ in prediction, $p > 0.05$), while the main effect of *Font* ($W = 0.825$, $\chi^2 = 6.935$ in actual, $W = 0.836$, $\chi^2 = 6.429$ in prediction, $p < 0.05$) and the interaction effect of *Spacing* and *Font* ($W = 0.244$, $\chi^2 = 49.939$ in actual, $W = 0.539$, $\chi^2 = 21.913$ in prediction, $p < 0.05$) violated sphericity assumption.

We have found the following significant effects on both actual and predicted task completion times (highlighted in bold in Table 2):

- A main effect of *Spacing* (F (2, 74) = 5.44, $p < 0.05$) on actual task completion time.
- A main effect of *Spacing* (F (2, 74) = 6.95, $p < 0.05$) in predicted task completion time.

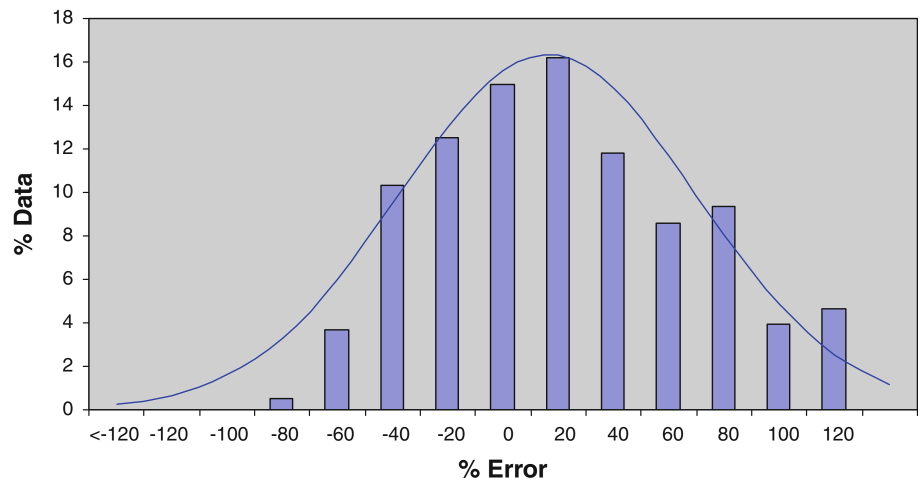


Fig. 8 Relative error in prediction

Table 2 Test of within-subjects effects on task completion time

Source	Actual			Predicted		
	df	*F*	Sig.	*df*	*F*	Sig.
Spacing	**2.0**	**5.44**	**0.006**	**2.0**	**6.95**	**0.002**
Spacing* × *Group	**4.0**	**3.15**	**0.019**	**4.0**	**4.64**	**0.002**
Error (*Spacing*)	74.0			74.0		
Font	1.7	0.22	0.770	1.7	2.89	0.071
Font* × *Group	**3.4**	**5.02**	**0.002**	**3.4**	**3.75**	**0.012**
Error (*Font*)	63.0			63.6		
Spacing × *Font*	2.3	1.03	0.370	3.3	1.54	0.204
Spacing × *Font* × *Group*	4.7	0.83	0.528	6.5	1.32	0.250
Error (*Spacing* × *Font*)	86.3			121.0		

Table 3 Multivariate test on completion time

Effect	Actual			Predicted		
	df	*F*	Sig.	*df*	*F*	Sig.
Spacing	**2**	**5.62**	**0.008**	**2**	**6.28**	**0.005**
Spacing* × *Group	**4**	**2.78**	**0.033**	**4**	**3.97**	**0.006**
Font	2	0.31	0.739	**2**	**4.05**	**0.026**
Font* × *Group	**4**	**6.39**	**0**	**4**	**5.05**	**0.001**
Spacing × *Font*	4	1.41	0.253	4	2.18	0.093
Spacing × *Font* × *Group*	8	2.15	0.043	8	1.74	0.106

- An interaction effect of *Spacing* and *Group* (F (4, 74) = 3.15, $p < 0.05$) on actual task completion time.
- An interaction effect of *Spacing* and *Group* (F (4, 74) = 4.64, $p < 0.05$) on predicted task completion time.
- An interaction effect of *Font* and *Group* (F (3.4, 62.97) = 5.02, $p < 0.05$) on actual task completion time.
- An interaction effect of *Font* and *Group* (F (3.44, 63.6) = 3.75, $p < 0.05$) on predicted task completion time.

The main effect of *Font* and interaction effects between *Font and Group and Spacing*, *Font and Spacing* does not have significant effects on both actual and predicted task completion times.

Fig. 9 Effect size comparison in ANOVA

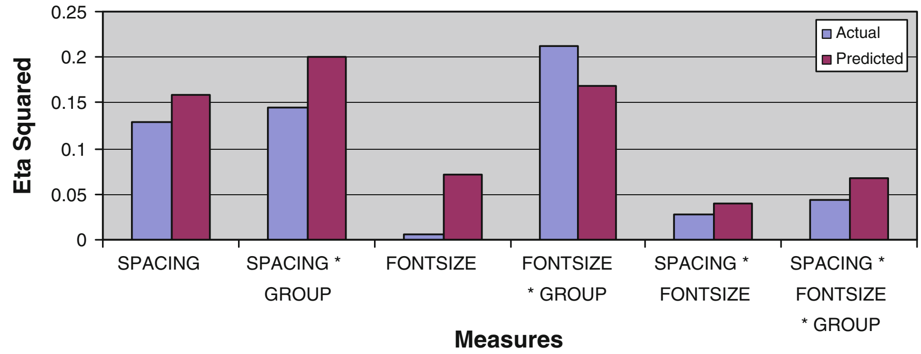
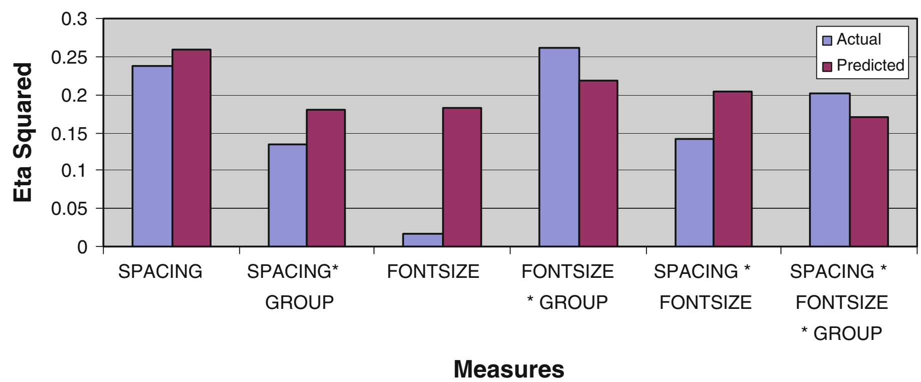


Fig. 10 Effect size comparison in MANOVA



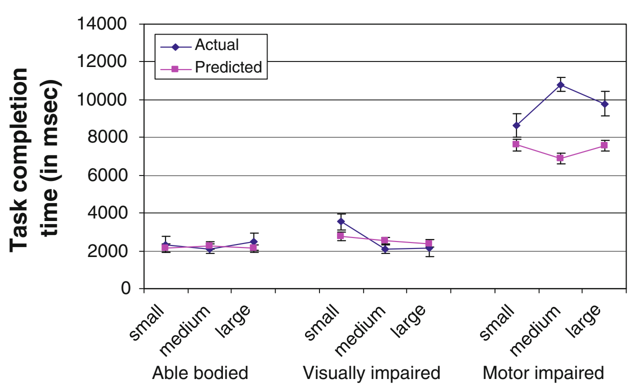


Fig. 11 Effect of *Font size* in different user groups

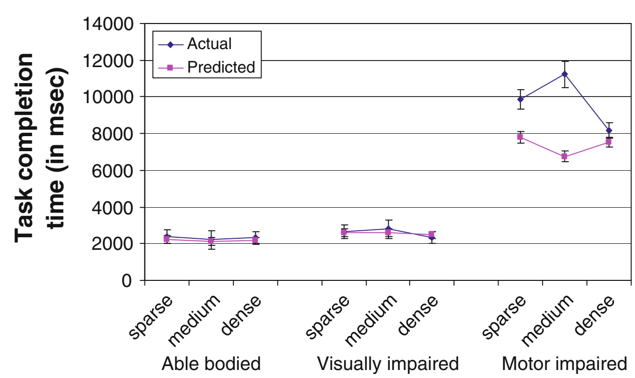


Fig. 12 Effect of *Spacing* in different user groups

We confirm these effects through a multivariate test, which is not affected by the sphericity assumption. The MANOVA shows the following significant effects (highlighted in bold in Table 3):

- A main effect of *Spacing* (*Wilks'* $\lambda = 0.762$, F (2, 36) $= 5.62$, $p < 0.05$) on actual task completion time.
- A main effect of *Spacing* (*Wilks'* $\lambda = 0.741$, F (2, 36) $= 6.28$, $p < 0.05$) on predicted task completion time.
- A main effect of *Font* (*Wilks'* $\lambda = 0.817$, F (2, 36) $=$ 4.05, $p < 0.05$) on predicted task completion time.
- An interaction effect of *Spacing* and *Group* (*Wilks'* $\lambda = 0.750$, F (4, 72) $= 2.78$, $p < 0.05$) on actual task completion time.
- An interaction effect of *Spacing* and *Group* (*Wilks'* $\lambda = 0.671$, F (4, 72) $= 3.97$, $p < 0.05$) on predicted task completion time.
- An interaction effect of *Font* and *Group* (*Wilks'* $\lambda = 0.545$, F (4, 72) $= 6.39$, $p < 0.05$) on actual task completion time.
- An interaction effect of *Font* and *Group* (*Wilks'* $\lambda = 0.610$, F (4, 72) $= 5.05$, $p < 0.05$) on predicted task completion time.

Table 4 ANOVA for each user group

Source	Actual				Predicted			
	df	*F*	Sig.	Eta squared	*df*	*F*	Sig.	Eta squared
Able-bodied								
Spacing	2	0.21	0.815	0.014	2	0.38	0.688	0.025
Error (*Spacing*)	30				30			
Font	2	0.72	0.495	0.046	2	2.73	0.081	0.154
Error (*Font*)	30				30			
Spacing × *Font*	4	1.78	0.144	0.106	**4**	**2.69**	**0.039**	**0.152**
Error (*Spacing* × *Font*)	60				60			
Visually impaired								
Spacing	1.4	0.52	0.54	0.034	2	0.81	0.453	0.051
Error (*Spacing*)	21.3				30			
Font	1.4	8.39	0.004	0.359	2	5.72	0.008	0.276
Error (*Font*)	21.5				30			
Spacing × *Font*	1.5	2.90	0.089	0.162	4	0.21	0.933	0.014
Error (*Spacing* × *Font*)	22.3				60			
Motor-impaired								
Spacing	2	2.93	0.087	0.295	**2**	**3.78**	**0.049**	**0.350**
Error (*Spacing*)	14				14			
Font	2	1.53	0.251	0.179	2	1.56	0.245	0.182
Error (*Font*)	14				14			
Spacing × *Font*	4	0.26	0.904	0.035	4	0.67	0.62	0.087
Error (*Spacing* × *Font*)	28				28			

Figures 9 and 10 show that the effect sizes (η^2) are also fairly similar in the prediction as in the actual. The maximum difference is below 10% in within-subject test and below 20% in multivariate test. This suggests that the simulator successfully explained the variance in task completion time for different factors. As these factors include both interface parameters and physical characteristics of users, we can infer that the simulator has successfully explained the effects of different interface layouts on task completion time for people with visual and motor impairment. Figures 11 and 12 show the effects of font size and spacing for different user groups. In Figs. 11 and 12, the points depict the average task completion time, and the bars show the standard error at a 95% confidence level. It can be seen from Figs. 11 and 12 that the prediction is in line with the actual task completion times for different font sizes and icon spacing.

However, the prediction is less accurate in one of the nine conditions—the medium font size and medium spacing for the motor-impaired users. We found that, in these cases, the model underestimates the task completion times and also fails to capture the variability in it. We have further analysed the effects of *Spacing* and *Font* size for each user group separately (Table 4).

It can be seen from Table 4 that in terms of significance at $p < 0.05$, the prediction deviates from the actual in the following two cases (highlighted in bold):

- Interaction effect of *Spacing* and *Font* for able-bodied users (F (4, 60) = 1.78, $p > 0.05$ for actual, F (4, 60) = 2.69, $p < 0.05$ for prediction).
- Effect of *Spacing* for motor-impaired users (F (2, 14) = 2.93, $p > 0.05$ for actual, F (2, 14) = 3.78, $p < 0.05$ for prediction).

Finally, we compared the mean and standard deviation of the actual and predicted task completion times for each condition. Table 5 lists the relative difference $\frac{\text{Predicted}-\text{Actual}}{\text{Actual}}$ in mean and standard deviations between actual and predicted task completion time.

It can be seen from Table 5 that only in four conditions (highlighted in bold), the average predicted time is different from the actual predicted time by more than ±40%. However, the standard deviation is predicted quite less than in actual in many occasions. The difference is less severe for visually impaired users than for the other two groups. One possible reason for the difference may be the effects of learning and fatigue as able-bodied users might work quickly due to learning effect and motor-impaired users

Table 5 Relative differences in mean and standard deviation

	Spacing	*Font*	% Difference in	
			Mean	SD
Able-bodied	Sparse	Small	−7.58	−55.65
		Medium	21.25	−27.62
		Large	−21.29	−83.04
	Medium	Small	−11.16	−66.72
		Medium	12.80	−45.72
		Large	−9.61	−24.4
	Dense	Small	−7.93	−76.85
		Medium	−4.94	−44.12
		Large	−7.34	−1.27
Visually impaired	Sparse	Small	−26.97	−69.06
		Medium	12.70	−1.19
		Large	32.51	17.21
	Medium	Small	**−41.33**	−78.7
		Medium	**58.14**	21.87
		Large	14.93	−20.11
	Dense	Small	28.62	7.59
		Medium	0.01	−19.48
		Large	−5.84	−13.97
Motor-impaired	Sparse	Small	6.19	−48.68
		Medium	−35.31	−76.38
		Large	−24.97	−80.72
	Medium	Small	**−40.10**	−67.8
		Medium	**−43.22**	−71.32
		Large	−36.46	−83.7
	Dense	Small	9.52	−61.08
		Medium	−29.82	−74.52
		Large	4.23	7.57

Table 6 Effect of usage time

Model		*B*	Std. Error	Beta
1	(Constant)	502.56	212.43	
	Predicted time	0.67	0.04	0.62*
2	(Constant)	372.16	214.53	
	Predicted time	0.59	0.05	0.55*
	Usage time	2.77	0.91	0.14*

$\Delta R^2 = 0.62$* for Model 1, $\Delta R^2 = 0.64$* for Model 2 (* $p < 0.005$)

might feel fatigue. So, we have analysed the effects of usage time through a regression model.

3.6.1 *Analysing effect of usage time*

We have considered the predicted task completion time and the usage time as independent variables and the actual task completion time as the dependent variable. The usage time for each trial measures the total time spent (in seconds) from beginning of the session to the end of the trial. Table 6 shows the regression coefficients.

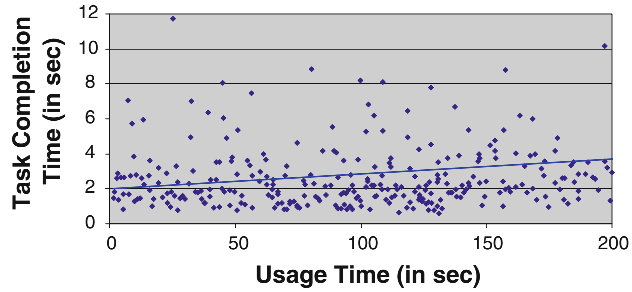


Fig. 13 Effect of usage time

It seems that usage time can significantly ($p < 0.005$) affect the actual time though the improvement in ΔR^2 is only 2%. The inclusion of usage time in the regression model also reduces the change in R^2 from 0.39 to 0.01, which means it increases the generalizability of the model [9]. The positive value of coefficient B indicates that the task completion time was directly proportional to the usage time. Figure 13 shows a weak positive correlation ($\rho = 0.42$) between usage time and task completion time. Perhaps it means that users felt fatigue or bored as the session went on and took more time to complete the task in later trials.

3.7 Discussion

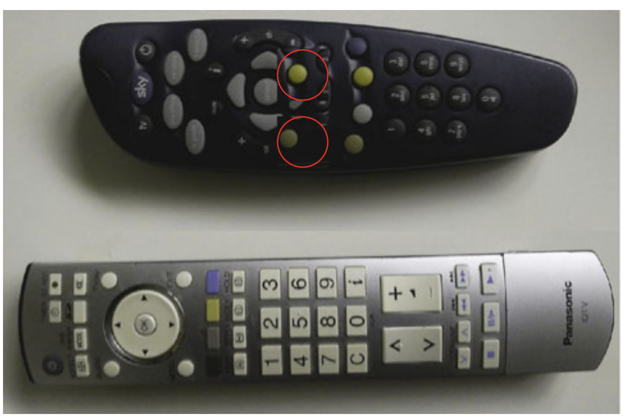
We have developed a simulator to help with the design and evaluation of assistive interfaces. Choosing a particular interface from a set of alternatives is a significant task for both design and evaluation. In this study, we considered a representative task, and the results showed that the effects of both factors (separation between icons and font size) were the same in the prediction as for actual trials with different user groups. The prediction from the simulator can be reliably used to capture the main effects of different design alternatives for people with a wide range of abilities.

However, the model did not work accurately for about 30% of the trials where the relative error is more than 50%. These trials also accounted for an increase in the average relative error from 0 to 16%. In particular, the predicted variance in task completion times for motor-impaired users was smaller than the actual variance. This can be attributed to many factors; the most important ones are as follows:

- *Effect of usage time—fatigue and learning effects:* The trial continued for about 15–20 min. A few participants (especially one user in the motor-impaired group) felt fatigue. On the other hand, some users worked more

(a)

(b)



(c)

Fig. 14 Comparing two TV remote controls with respect to visual impairments. **a** TV remote controls as perceived by people without visual impairment. **b** TV remote controls as perceived by people having less visual acuity. **c** TV remote controls as perceived by people having colour blindness

quickly as the trial proceeded. The model did not consider these effects of fatigue and learning. It seems from our analysis that the usage time can significantly affect the total task completion time. In future, we like to analyse the effect of usage time in more detail and plan to incorporate it into the input parameters of the model.

- *User characteristics*: The variance in the task completion time can be attributed to various factors such as expertise, usage time, type of motor impairment (hypokinetic vs. hyperkinetic), interest of the participant, etc. Currently, the model characterizes the extent of motor impairment of the user only by measuring the grip strength [5], in future more input parameters may be considered.

(a)

(b)

Fig. 15 Simulation of maccular degeneration for a TV remote control. **a** Perception of a person having early stage of dry maccular degeneration. **b** Perception of a person having early stage of wet maccular degeneration

4 Applications

We have investigated the principles of visual perception of visually impaired users and motor action of motor-impaired users and also compared those with their able-bodied counterparts. Our studies [4–7] provide the necessary knowledge about the relationship between physical ability and interaction, which will help designers of interactive systems to develop more inclusive systems. Our studies have already been used to design an inclusive accessible game [12] and a new assistive interaction technique [3].

The applications of our models can be extended to other digital devices (like digital television, ubiquitous devices and so on) beyond computers. For example, our models can be used to determine the optimum font size, contrast and colour of onscreen menu items used to select channels in a digital TV. Similarly, it can also be used to simulate the perception of visually impaired users [6] (like how a person having less visual acuity or colour blindness will view a remote controller) to make designers understand the

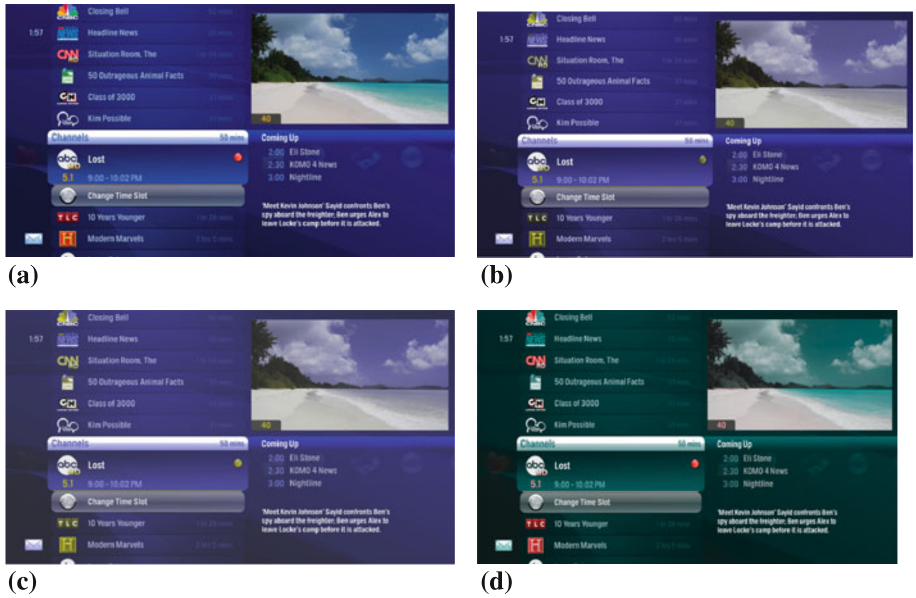


Fig. 16 Simulation of colour blindness for a digital TV interface. **a** No impairment. **b** Protanopia. **c** Deuteranopia. **d** Tritanopia

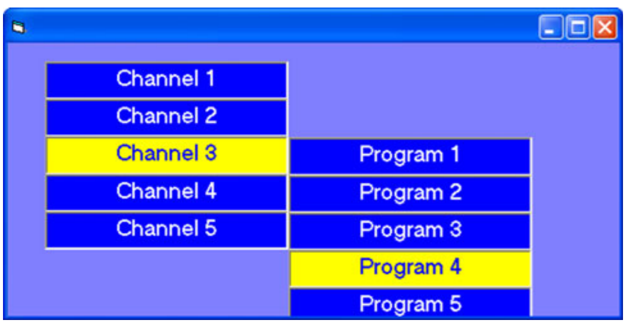


Fig. 17 Interface used in the study

Fig. 18 Representative of an actual interface

problems of visual impairment. Figure 14 presents a comparative analysis between two remote control interfaces for people having less visual acuity (resulting from age, myopia, retinopathy or other eye diseases) and red green colour blindness. It can be easily seen that the remote at the bottom is more legible than the top one for visual acuity loss. Similarly, for colour blindness, two buttons (marked with red circles) of the top remote look almost identical; clearly, it would be better to change the colour or shape of the buttons to cater people with colour blindness.

This type of simulation can be extended to other diseases (Fig. 15) or interfaces (Fig. 16). More details about the simulation can be found in a separate paper [6]. In fact, the GUIDE project [16] will use the simulator to develop an adaptive toolbox to design and help in designing inclusive TV interfaces. In the following subsection, we have presented a study of applying the simulator for a representative TV application.

4.1 The study

In this study, we have investigated the accessibility issues of programme selection menus for a digital TV interface. We take help from our simulator [4–7] in identifying the accessibility problems of programme selection menu with respect to visually impaired and mobility-impaired users. Based on the results of the simulation, we have designed new interfaces. Our study consists of the following three stages:

- Problem identification through simulation.
- New interface evaluation through simulation.
- Validation of the simulation through a controlled experiment.

Initially, we have designed the following interface (Fig. 17), which looks similar to existing systems (Fig. 18).

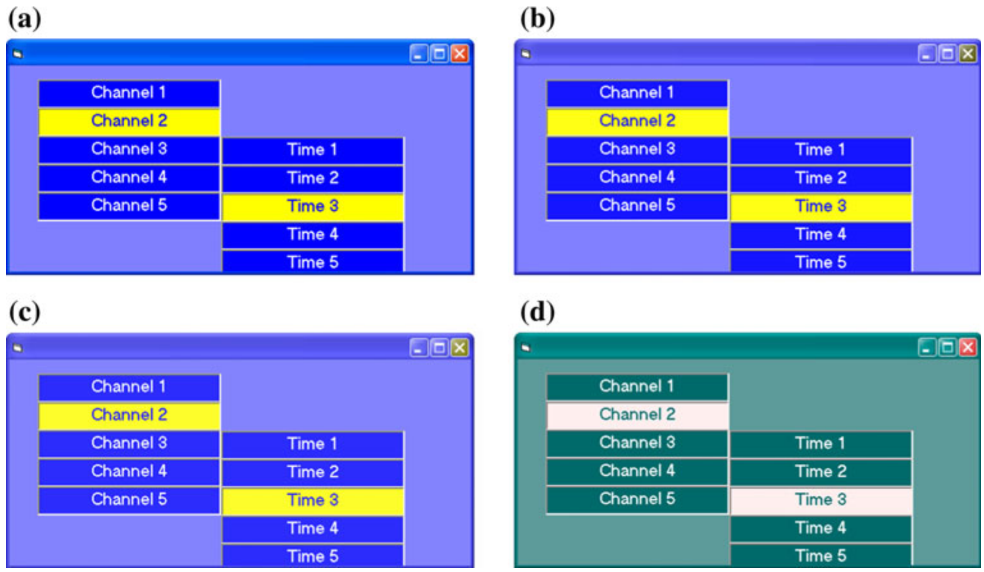


Fig. 19 Perception of people having colour blindness. **a** Original interface. **b** Interface perceived by protanopia. **c** Interface perceived by Deuteranopia. **d** Interface perceived by Tritanopia

The GUIDE project [16] explores accessibility issues of people with a wide rage of abilities (including visual, cognitive and motor impairment) using different modalities of interaction (like pointing, keypad, gesture, voice-based inputs and so on). In this particular work, we have investigated

- Sensory problems of
  - People with less visual acuity
  - People having colour blindness
- Interaction problems of
  - People with motor impairment using a pointing device

In this study, the simulator takes a sample task of selecting a menu item and the screenshot of the interface as input and shows the perception of visually impaired users and cursor trajectory of motor-impaired users as output. In the simulation study, we have not bothered about the particular words used as captions since the simulation results are not to be used by participants. We use captions like Channel 1, Program 1 or Time 1 as captions. However, in the validation study, we used different words as captions and discussed it in detail in a later section.

4.1.1 *Problem identification*

Initially, the output from the simulator is used to identify accessibility problems. Figure 19 shows the perception of the interface for three different types of colour blindness. It can be seen that though the colours look different, the particular colour combination of our interface does not reduce the legibility.

Figure 20 shows the perception of the interface of people with mild (less than approximately −3.5 Dioptre) and severe (more than approximately −4 Dioptre) acuity loss. It can be seen from the figure that the captions (which are of 14 pt) become illegible for severe acuity loss. Figure 21 shows a possible cursor trace of a pointing device (like mouse) operated by a person having motor impairment. The thin purple line shows a sample trajectory of mouse movement of a motor-impaired user. It can be seen that the trajectory contains random movements near the source and the target. The thick black lines encircle the contour of these random movements. The area under the contour has a high probability of missed clicks as the movement is random there and thus lacks control. It can be seen that as the buttons are closely spaced, there is a significant probability of missed click in a wrong button, which would surely frustrate any user.

Based on the simulation results, we identified the following two accessibility issues:

- Legibility of captions
- Spacing between menu items

4.1.2 *New interface*

Based on the previous discussion, we have redesigned the interfaces. We have increased the font size of captions for users with visual impairment. For people with motor impairment, we have changed the size of the buttons without changing the screen size such as no couple of buttons shares a common boundary. It should reduce chances of missed clicks. Figures 22 and 23 show the new interfaces. We have not designed anything new to cater colour-blind users as the present interface seems perfect for

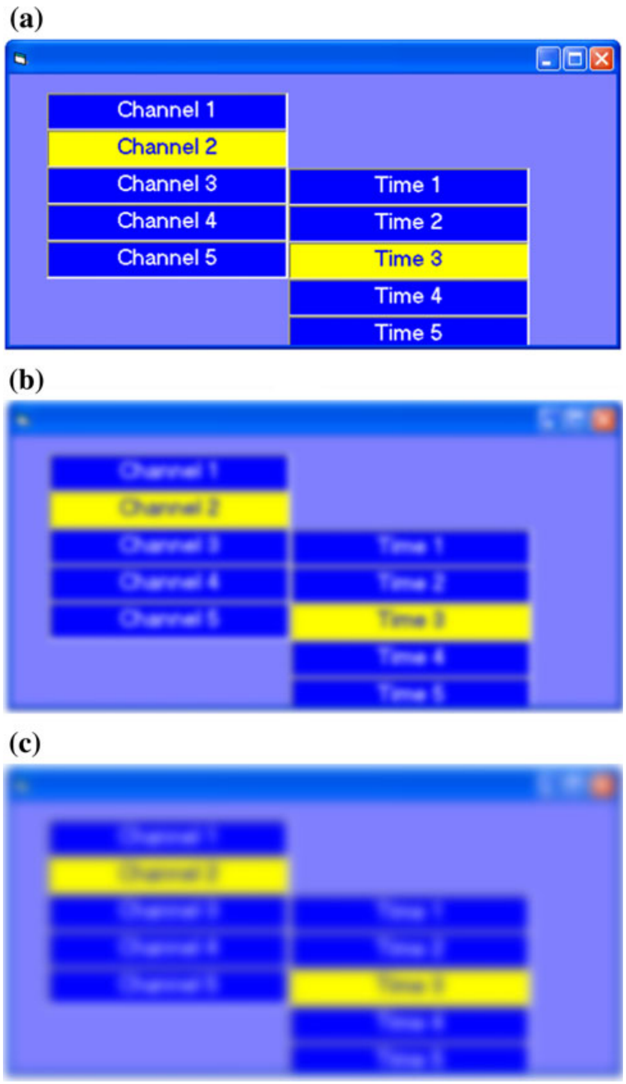


Fig. 20 Perception of people having less visual acuity. **a** Original interface. **b** Interface perceived by mild visual acuity loss. **c** Interface perceived by severe visual acuity loss

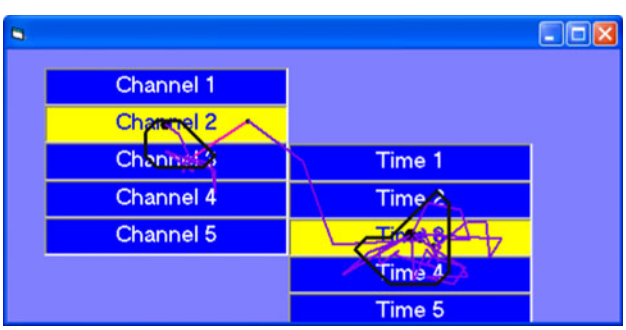


Fig. 21 Possible cursor trace of a mouse operated by motor-impaired person

them. Figure 24 shows the perception of the new interface for people with mild and severe acuity loss. It can be seen that the modified caption (now at 18 pt) has better legibility

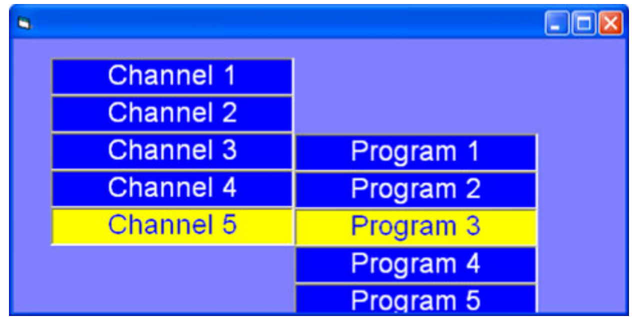


Fig. 22 Interface for people having less visual acuity

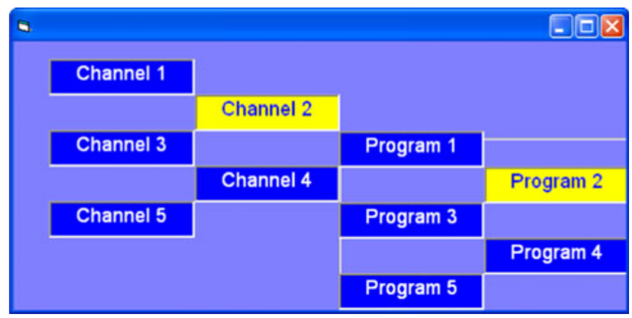


Fig. 23 Interface for people with motor impairment

than the previous case even for severe acuity loss. We have also investigated the effect of severe visual acuity loss for the following six font types (Fig. 25):

- Microsoft Sans Serif
- Veradana
- Arial
- Sabon
- Times New Roman
- Georgia

It can be seen in Fig. 25 that the legibility is not much different for different font types and nearly same for all, which also supports previous research [2].

Figure 26 shows the possible cursor trace of a person with motor impairment for the new interface. It can be seen that the contour covering the area of missed click does not contain more than one menu item now.

4.1.3 Validation

We have validated the new interface through a user study. In this study, we hypothesize the following:

- People with visual acuity loss and motor impairment will perform a task faster and with less number of errors in the new interface (Figs. 22, 23) than the unchanged version (Fig. 17).
- People with colour blindness will perform a task equally well with respect to people with no impairment

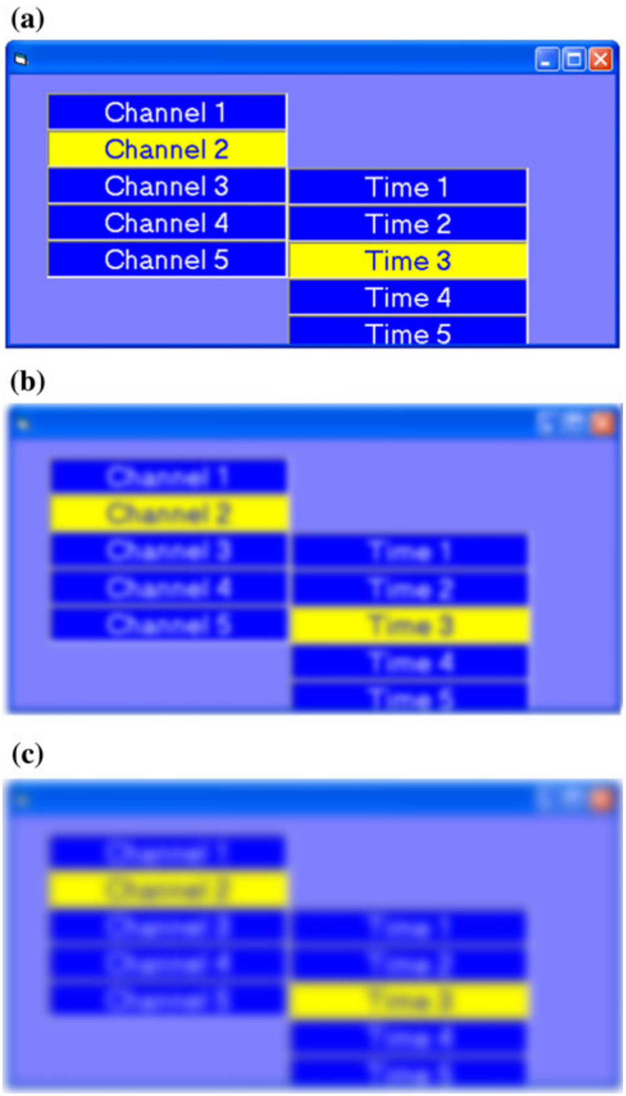


Fig. 24 Perception of the new interface of people having less visual acuity. **a** Original interface. **b** Interface perceived by mild visual acuity loss. **c** Interface perceived by severe visual acuity loss

(control group) in the unchanged version of the interface (Fig. 17).

We measured the task completion time as a measure of performance and the number of missed clicks as a measure of errors.

4.1.3.1 Procedure The procedure mimics the process of selecting a channel from a list followed by selecting a programme from a drop-down menu. Initially, the participants were shown a channel name and a programme name. Then, they made two selections matching the previously shown channel and programme names. We did not use real channel and programme names to avoid any biasness of users. The first two letters and length of all the captions were kept nearly same to avoid any pop-out effect of the captions during visual search. We used the Veradana font type due to its bigger *x*-height and character spacing than other conventional fonts. Each participant repeated the task ten times. All participants were trained before undertaking the study.

4.1.3.2 Material We used a standard optical mouse and an Acer Aspire 1640 Laptop with a 15.5″ monitor having 1,280 × 800 pixel resolution. We also used the same seating arrangement (same table height and distance from table) for all participants.

4.1.3.3 Participants We collected data from two institutes, National institute of Orthopedically Handicapped at Kolkata, India, and Papworth Trust at Cambridge, UK. All participants (Table 7) have some experience of using computers—they were either learning or using computers regularly. All of them volunteered for the study.

4.1.3.4 Results The average reaction time (total time needed to select the channel and programme) was less in the new design than in the control design (Fig. 27) though the difference was not statistically significant in an independent-sample two-tailed t test (t (120, 1) = 0.64, $p > 0.05$). The average number of missed clicks was also less (Fig. 28) in the new design than in the control design though the difference tends to statistical significance in a Wilcoxon ranked test (W (120, 1) = 163, $p = 0.1$). In the experimental condition (new deign), missed clicks occurred in 21 trials, while it occurred 31 times in control condition.

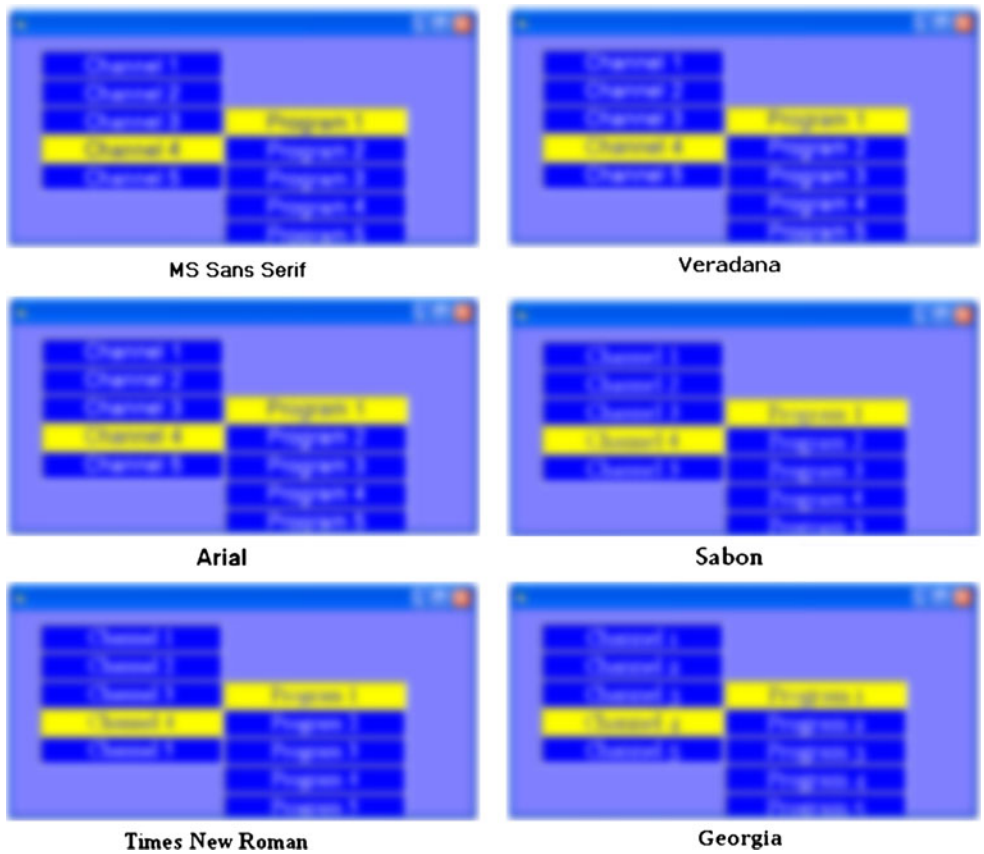
We have also analysed the reaction times and missed clicks for each individual participant. Table 8 and Figs. 29 and 30 show the average reaction time and total number of missed clicks for each participant. It can be seen that only 4 out of 12 participants (P4, P5, P8 and P9) have an average reaction time greater for the experimental condition, and only 2 out of 12 participants (P8 and P12) missed clicked more in the experimental condition than in the control condition.

Unfortunately, we did not get any participant with colour blindness. So, we have used a colour blindness filter (from Cambridge Research Systems, http://www.crsltd.com) to simulate the effect of dichromatic colour blindness. In this case as well, we do not find any significant difference in reaction times (Fig. 31) in an independent-sample two-tailed t test (t (20, 1) = 0.81, $p > 0.05$) and did not record any missed clicks as well.

4.1.4 Discussion

The reaction time and number of missed clicks were both less in the new design though we failed to find any

Fig. 25 Comparing different font types



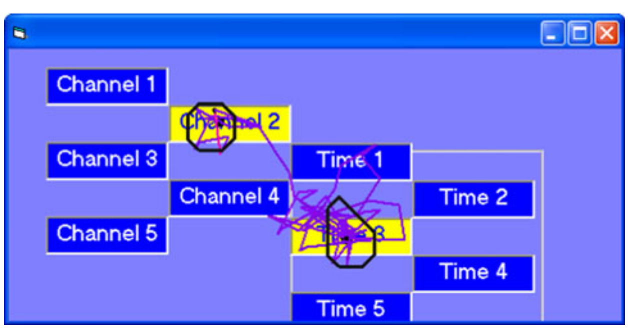


Fig. 26 Possible cursor trace of a mouse operated by a motor-impaired person for the new interface

statistical significance of the difference. Most of our participants did not have any problem in moving hands, and thus they could control the mouse movement pretty well. Except participant P1, the visual acuity loss was also not severe. Additionally, in the present experimental set-up, a missed click did not waste time, while in a real interface, a missed click will take the user to an undesired channel and getting back to the previous screen will incur additional time. So, the higher number of missed clicks in the control condition will also increase the channel selection time further in an actual scenario. However, in future, we plan to run the study with more cautious selection of participants. All of the visually impaired participants preferred the bigger font size. However, a few participants reported difficulty in reading the zigzag presentations of captions of the new interface. In future, we also plan to use an eye tracker to compare the visual search time for both types (linear and zigzag) of organizations of menu captions.

This study addresses a small segment of accessibility issues related to digital TV interfaces. Future studies will include more interaction modalities (like keypad or gesture-based interaction), devices (like remote control, set-top box and so on) and impairments (like cognitive impairments). However, the results of this study can be extended beyond programme menu interfaces of digital televisions. For example, the font size of captions in absolute terms (x-height $\approx$ 0.5 cm) indicates the minimum font size required for any text in an interface for serving people with severe visual acuity loss. Similarly, the particular colour combination of the screen (white text in blue background) can be used in any other interface as well to cater people with colour blindness. Finally, the modified menu structure can be used in computers or other digital devices to make the menus accessible to people with mobility impairment.

Table 7 Participants

Participants	Age	Sex	Impairment
P1	>45	M	No mobility impairment. Age-related hypermetropia (+3.75/+3.25 Dioptre)
P2	25–45	M	Difficulty in walking, right leg is shorter than left leg. Mild myopia (−2.75/−2 Dioptre)
P3	25–45	M	Right hand was cut in accident, no impairment in left hand. No visual impairment
P4	25–45	M	No mobility impairment. Lost vision in right eye, left eye is perfect
P5	25–45	M	Left arm is affected by polio, no impairment in right hand. No visual impairment
P6	<25	F	Lower body is affected by polio from birth, no impairment in hands, wheelchair user. No visual impairment
P7	<25	M	Difficulty in walking from birth. Slight myopia (−0.7/−0.7 Dioptre)
P8	44	M	Cerebral palsy reduced manual dexterity also some tremor in hand wheel chair user. Slight loss of visual acuity
P9	63	M	Left side (non dominant) paralysed after a stroke in 1973 also has tremor
P10	31	M	Cerebral palsy reduced manual dexterity wheel chair user
P11	>45	M	Reduced manual dexterity in limbs due to neurological problem, wheel chair user
P12	44	F	Did not mention disease restricted hand movement no tremor. Slight loss of visual acuity

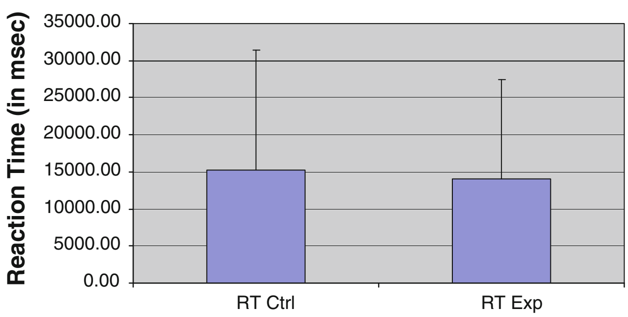


Fig. 27 Comparing reaction times

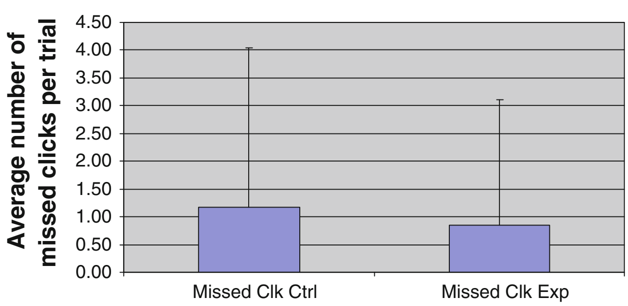


Fig. 28 Comparing number of missed clicks

Table 8 Result per participant

	Avg RT ctrl (in ms)	Avg RT exp (in ms)	Total MC ctrl	Total MC exp
P1	3,886	3,259	0	0
P2	5,755	5,033	0	0
P3	7,230	6,149	0	0
P4	21,777	26,838	72	56
P5	4,481	4,611	0	0
P6	12,195	11,739	11	4
P7	15,628	6,747	13	0
P8	15,394	18,628	20	28
P9	7,213	9,184	0	0
P10	36,160	25,084	11	0
P11	20,752	20,550	14	8
P12	32,228	30,223	0	6
Avg	15,225	14,004	11.8	8.5

5 Implications and Limitations

User trials are always expensive in terms of both time and cost. A design evolves through an iteration of prototypes, and if each prototype is to be evaluated by a user trial, the whole design process will be slowed down. Additionally, user trials are not representative in certain cases, especially for designing inclusive interfaces for people with special needs. A good simulation with a principled theoretical foundation can be more useful than a user trial in such cases. Exploratory use of modelling can also help designers to understand the problems and requirements of users, which may not always easily be found through user trials or controlled experiments.

We have shown that it is possible to develop engineering models to simulate HCI of people with a wide range of abilities and that the prediction is useful in designing and evaluating interfaces. According to Allen Newell's time scale of human action [11], our model works in the cognitive band and predicts activity in millisecond to second range. It cannot model activities outside the cognitive band like micro-saccadic eye gaze movements, response characteristics of different brain regions (in biological band [11]), affective state, social interaction, consciousness (in rational and social band [11]) and so on. Simulations of

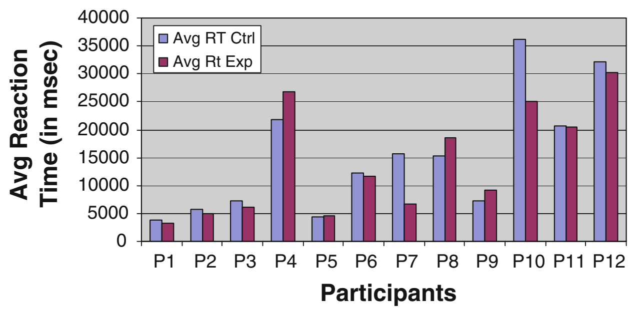


Fig. 29 Comparing Avg RT per participant

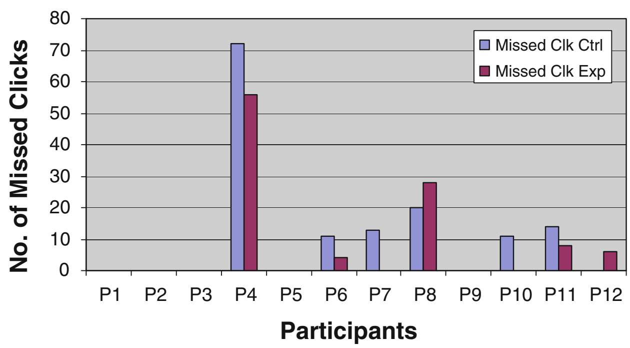


Fig. 30 Comparing total number of missed clicks per participant

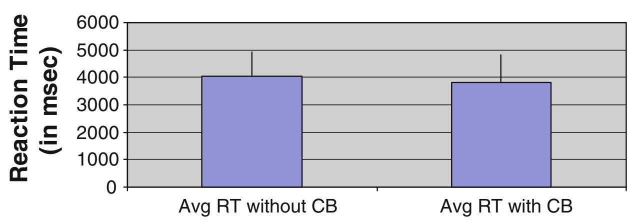


Fig. 31 Comparing RT for colour blindness

each individual band have their own implications and limitations. However, the cognitive band is particularly important since models working in this band are technically feasible, experimentally verifiable and practically usable. Research in computational psychology and more recently in cognitive architectures supports this claim. A new dimension is added in cognitive modelling by including users with disabilities. This work will also be useful for able-bodied people to address situational impairment. For example, an interface suitable for people having less visual acuity will be useful for small-screen devices that have high pixel density. Similarly, a good interface for a hyperkinetic motor-impaired user will also be suitable for a handheld device during its use in a moving vehicle. However, our simulator should not be used in isolation, rather it should complement existing qualitative techniques [14] of assessing users' experience.

6 Conclusions

Following the development of a simulator to help in designing and evaluating inclusive interfaces, this paper has demonstrated the use of the simulator for a representative task of evaluating an interface layout. The simulator predicted the task completion time with statistically significant accuracy for people with a wide range of abilities. It is also found to correctly predict the main effects of different layout options for people with a wide range of abilities. The simulator has been used in investigating accessibility requirement of a representative digital TV interface. It is hoped that our work will help to understand the effect of task and devices on human cognition in more detail, which will also be of interest to researchers in other disciplines besides computer science.

Acknowledgments We would like to thank the Gates Cambridge Trust and EU GUIDE project for funding this work. We would also like to thank Dr. Pat Langdon of Engineering Design Centre, University of Cambridge, for his help in designing the studies. We are also grateful to our participants for their time and effort in taking part in our studies.

References

1. Benyon, D., Murray, D.: Applying user modeling to human computer interaction design. Artif. Intell. Rev. **7**(3), 199–225 (1993)
2. Beymer, D., Russell, D.M., Orton P.Z.: An eye tracking study of how font size, font type, and pictures influence online reading. 22nd British Computer Society Conference on Human-Computer Interaction, pp. 15–18 (2007)
3. Biswas, P., Robinson, P.: A new screen scanning system based on clustering screen objects. J. Assist. Technol. **2**(3), 24–31 (2008). ISSN: 1754-9450
4. Biswas, P., Robinson, P.: Automatic evaluation of assistive interfaces. In: Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI), pp. 247–256 (2008)
5. Biswas, P., Robinson, P.: Modelling perception using image processing algorithms. 23rd British Computer Society Conference on Human-Computer Interaction (HCI 09) (2009)
6. Biswas, P., Robinson, P.: Modelling user interfaces for special needs. Accessible Design in the Digital World (ADDW) (2008)
7. Biswas, P., Langdon, P.: Developing multimodal adaptation algorithm for mobility impaired users by evaluating their hand strength. Int. J. Hum. Comput. Interact. Taylor & Francis, Print ISSN: 1044-7318
8. Card, S., Moran, T., Newell, A.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale (1983)
9. Field, A.: Discovering Statistics Using SPSS. SAGE Publications Ltd, Beverly Hills (2009)
10. John, B.E., Kieras, D.: The GOMS family of user interface analysis techniques: comparison and contrast. ACM Trans. Comput. Hum. Interact. **3**, 320–351 (1996)
11. Newell, A.: Unified Theories of Cognition. Harvard University Press, Cambridge (1990)
12. Phillips, N.: Graphical Modification for Partially Sighted Gamer Accessibility. Computer Laboratory, Tripos Part II project University of Cambridge (2009)

13. See it Right, RNIB (2006)
14. Springett, M.: Assessing user experiences within interaction: experience as a qualitative state and experience as a causal event. In: Proceedings of VUUM (2008)
15. Stephanidis, C., Paramythis, A., Karagiannidis, C., Savidis, A.: Supporting interface adaptation: the AVANTI web-browser. 3rd ERCIM Workshop on User Interfaces for All (1997)
16. The GUIDE Project. http://hcim.di.fc.ul.pt/wiki/GUIDE (2009). Accessed 2 Oct 2009
17. WHO Website. http://www.who.int/ageing/en (2009). Accessed 18 Sept 2009
18. World Bank Website. http://web.worldbank.org (2009). Accessed 18 Sept 2009