Hindawi Publishing Corporation
Advances in Human-Computer Interaction
Volume 2011, Article ID 565689, 11 pages
doi:10.1155/2011/565689

Research Article

Expanding Interaction Potentials within Virtual Environments: Investigating the Usability of Speech and Manual Input Modes for Decoupled Interaction

Alex Stedmon, Victor Bayon, and Gareth Griffiths

Virtual Reality Applications Research Team, Faculty of Engineering, University of Nottingham, Nottingham NG7 2RD, UK

Correspondence should be addressed to Alex Stedmon, alex.stedmon@nottingham.ac.uk

Received 28 April 2011; Revised 11 August 2011; Accepted 7 September 2011

Academic Editor: Armando Bennet Barreto

Distributed technologies and ubiquitous computing now support users who may be detached or decoupled from traditional interactions. In order to investigate the potential usability of speech and manual input devices, an evaluation of speech input across different user groups and a usability assessment of independent-user and collaborative-user interactions was conducted. Whilst the primary focus was on a formative usability evaluation, the user group evaluation provided a formal basis to underpin the academic rigor of the exercise. The results illustrate that using a speech interface is important in understanding user acceptance of such technologies. From the usability assessment it was possible to translate interactions and make them compatible with innovative input devices. This approach to interaction is still at an early stage of development, and the potential or validity of this interfacing concept is still under evaluation; however, as a concept demonstrator, the results of these initial evaluations demonstrate the potential usability issues of both input devices as well as highlighting their suitability for advanced virtual applications.

1. Introduction

In the past, traditional virtual reality (VR) technology often sees single users interacting with their own dedicated applications; however, with developments in group-based technologies, collaborative virtual environments (CVEs) have emerged as a means to support cooperative work [1]. More recently with ubiquitous computing and mixed reality [2, 3], CVEs now support users distributed across physical boundaries and time zones who may be detached from their immediate interaction space and even the virtual environment (VE) they are interacting with [1]. This affords new potentials for collaboration as well as providing an enriched environment where users can exploit new interaction paradigms [4] and where user experiences can be "decoupled" from their original form [1].

As VR technologies advance, traditional desktop applications can be ported to run in new visualisation modes [5]. Speech-based and handheld technologies can be used to implement a subset of the graphical and interaction possibilities that can be incorporated within mainstream VR systems [1]. With this development of new interactions in VR, new challenges emerge in understanding which technologies might best support user requirements so that appropriate input devices are chosen which enhance the overall effectiveness of a virtual application as well as support the user experience [6].

As part of the Virtual and Interactive Environments for Workplaces: "VIEW of the Future" project, there was a clear emphasis on considering user requirements and application needs in developing novel interaction and visualization techniques [1], user-centred methods [7, 8], and modes of interaction [9].

With a focus on decoupled interaction and modes of interaction, this paper argues that there has been little recent development in understanding human factors issues of speech and manual input and even less in the specific area of their usability in virtual applications. Given the focus of this paper and the aim of presenting arguments that transcend specific technologies or trends in solutions, this paper does not set out to address issues associated with natural and spoken dialogue technologies [10, 11], dialogue and dialogue

management [12], graphical interaction devices for distributed VR systems [13], or recent work on embodied conversational agents in virtual applications [14]. Furthermore, the technologies and methods underpinning multimodal interaction [15, 16], tangible or mobile interfaces, [17, 18], or specific applications such as camera-equipped mobile phones [19] are not the primary focus of this paper. Rather than reviewing the current state of the art in interaction design, this paper addresses fundamental human factors issues by looking at the user first then seeking to develop usable virtual applications incorporating speech and manual input to support user interaction.

1.1. Representing 3D Concepts in 2D Interfaces. Multidisplay, multiscreen, VR systems are often used to visualise large 3D and computer-aided design (CAD) models [20]. These virtual applications have traditionally been used in conjunction with sophisticated 3D input tracking devices and stereo projection to provide users with an immersive experience within a VE [21]. Many 3D input devices are designed to perform specific tasks such as navigation, object selection, object manipulation, and system control, often only allowing one active user at a time to control the interaction space [21]. Information presented in 3D formats can enhance user experience in immersive situations; however, interaction can be hindered if data presented in a 3D manner would be better presented in a 2D format, such as text or graphical widgets [22]. In some cases the immersive experience is enhanced by constraining the interaction to a 2D representation [23]; however, most 3D interaction techniques and interfaces can be difficult to implement for specific input devices and uses [21].

Decoupled interactions develop some of the functionality in interactive VEs by exporting aspects of 3D manipulation tasks into the 2D interaction domain with three main objectives [5]:

(i) to provide an easier mechanism to trigger interaction and access functionality embedded within the VE,

(ii) to support multimodal and multidevice forms of interaction to perform the same actions,

(iii) to allow more than one user to participate in the interaction while using the VE as nonimmersive user.

1.2. Prototype VE for Decoupled Interaction. A prototype VE (Figure 1) was developed to evaluate the usability of different input devices [1]. The VE consisted of a vehicle model that provided a number of interaction opportunities. It was created using Newtek's Lightwave 7.5 and VR-Tools 2.1 software and was developed upon a user-centred design methodology with the input of VR experts focusing on the generic functionality of the VE [5].

The VE allowed users to open and close the doors, manipulate the vehicle bonnet/hood and boot/trunk, change specific properties of the vehicle (e.g., colour attributes, wireframe functions), navigate around the vehicle from an egocentric perspective, and activate a 2D menu interface using different input devices. The VE was designed so that there was a balance of navigation and object manipulation tasks.

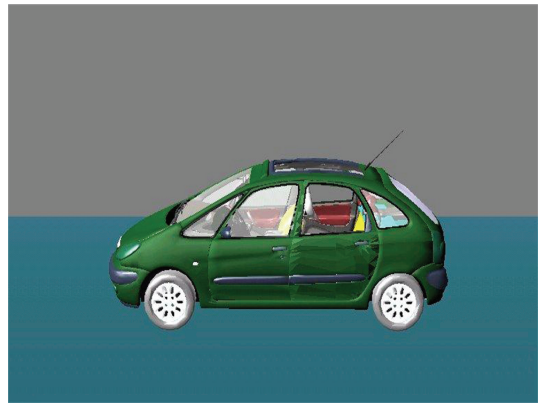

Figure 1: Prototype VE for decoupled interaction.

In order that users could investigate properties of the VE, it was represented with a hypertext markup language (HTML) tree-structure menu that visualised the properties and highlighted the potential interaction points as links. As the HTML page was dynamic, the menu could be expanded or collapsed by the user according to his/her preferences. When a menu link was activated on the web browser, a message was sent to the VE to initiate a task (e.g., producing a "screen dump" of the current viewpoint or selecting an object or specific function). In this way it was possible to decouple the user interaction from the original input device, such as a 3D mouse or a 3D wand, by using a remote 2D interface such as a handheld device or speech input. This approach was chosen so that future interactions might be conducted via the web from distributed locations.

1.3. Speech and Manual Input for Decoupled Interaction. With recent developments in reality-based interfaces (RBIs) and new interaction styles that draw on users' knowledge, experience, and expectations of the real world, there is a move to develop human-computer interaction (HCI) metaphors in a digital world that are more intuitive and less constrained by technology [24]. Whilst there is considerable progress in developing multimodal interaction, such as gesture, video tracking, and electromagnetic sensing [16], there is little research into the human factors of manual and speech input [9].

Within the prototype, a file created by the VE could be transformed to generate a speech grammar file that was then interpreted as a speech input command [5]. By using speech input in this way, users could activate the 2D interface (e.g., a visible hierarchical menu in the VE) that could then be used to execute certain actions within the VE (Figure 2). For example, should the user wish to open one of the car doors, they could verbally instruct the application to call up the "door menu" and then specify the door to be opened (e.g., "door open" > "front left" > "open").

A handheld input device, integrated with wifi connection, was used to interact with the VE based on web browsers that were incorporated as standard [1]. A particular feature of mobile handheld devices is that, within a single CVE, it is possible for a number of dedicated users to use independent

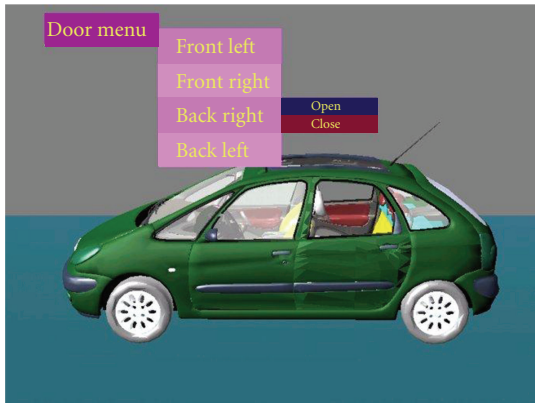

Figure 2: Decoupled menu structure.

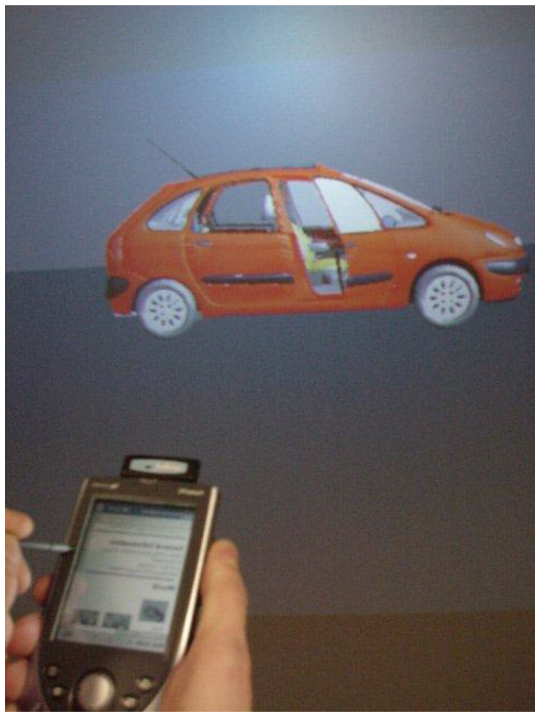

Figure 3: Decoupled interaction using a handheld device.

devices that allow them to share their interactions and experiences.

Figure 3 represents a user interacting with the prototype VE using a wifi-enabled handheld device. Selection and system control interactions, such as quick navigation to viewpoints, opening and closing doors, changing the model attributes, rendering objects visible and invisible, and activating the user interface, could be conducted through the handheld device in the same way that speech input was used.

In order to assess the potential of speech and manual input for decoupled interaction, a two-stage investigation was conducted. The initial activity examined the potential usability of speech within virtual applications that served as a basis for a more focused formative investigation of speech and manual input for independent and collaborative interactions.

2. Rationale

In order to assess the potential usability of speech input in VR, two research activities were conducted.

> Study 1: an evaluation of speech input across different user groups.
>
> Study 2: a usability assessment of independent-user and collaborative-user interaction modes using both speech and manual input configurations.

Study 1 was conducted with three different user groups. Two of the groups were taken from a previous RBI investigation into human-machine interaction (HMI) and human-machine interaction (HHI) principles of speech input for virtual applications [9]. The previous RBI study investigated differences in the perceptions of speech input based on users who believed they were talking to a machine (i.e., the HMI group) and users who were talking to another person (i.e., the HHI group). In Study 1 a third group of VR expert users assessed the potential of speech input in virtual applications independently of the RBI study population and from a more theoretical standpoint.

Data were collected using an Input Device Usability (IDU) Questionnaire, which contains fifteen questions designed to investigate user interaction, distraction, ease of use, user comfort, frustration, enjoyment, error correction, and overall usability. The questionnaire was developed from previous usability research at the University of Nottingham [25] and established sources [26] that were then formulated through expert review and developed specifically for input device usability issues within the VIEW project.

Following on from this, a series of expert evaluations were conducted to investigate the usability issues specifically associated with independent-user and collaborative-user interaction modes comparing both speech and manual input configurations. The prototype automotive CVE was used to conduct user trials, and, as with previous evaluations [1], this was a formative investigation.

Within formative evaluations and usability research, there is some discussion on approaches and effective sample sizes. Formative evaluation is often performed during the development or improvement of an application usually as part of an iterative cycle [27–29]. The aim of formative evaluations is to identify issues that may impact on future use of a product or application and highlight potential solutions as well as providing a design audit trail of planning implementation, monitoring, and progress of the evaluation. It is acknowledged that funding limitations often compromise the intensity of formative evaluations, and, whilst they do not necessarily meet the needs of most conflict resolution initiatives, they are an important first step in design improvement but not an end in itself [30].

In relation to sample sizes for usability testing, there is no single solution for all investigations, and so invariably there is a tradeoff between research objectives, available resources (e.g., time, money, and users) as well as the strategic importance of the research within a given project [31]. To some degree there is a law of diminishing returns with the numbers of users involved in usability testing and the issues

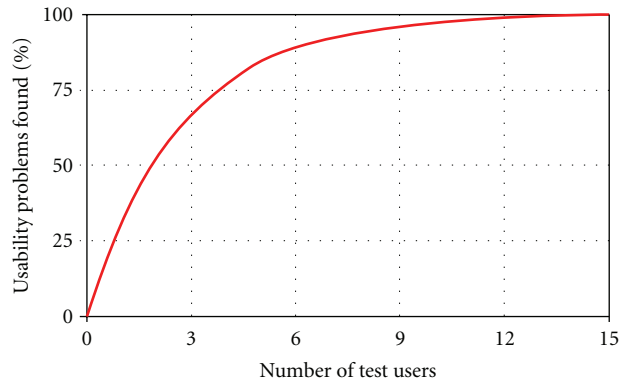

Figure 4: Rational for small samples in usability testing (adapted from [37]).

they might identify. In many cases a large proportion of issues (typically between 80% to 90%) can be identified with only five users and the most severe issues identified by three users [31, 32]. Furthermore, given the emphasis on iterative design cycles, it is often more prudent to employ 3×5 users at different stages of a design process than 15 users at a single point in the design cycle [31]. However, in trying to identify all the usability issues, there are those who argue for larger sample sizes of more than eight users; however, these should not be tested all at the same time [33]. There is considerable discussion over test validity and reliability, criticisms of the assumptions of small sample paradigms on methodological and empirical grounds, and important issues associated with user variability which can influence the decisions for different sample sizes (for an in-depth commentary, see [34]); however, where the sample is largely homogenous smaller samples of between three to five users can work well to identify key issues [35] although, with more variance in the user group or to ensure the highest capture rate of issues, larger samples are more appropriate [36].

Given that small samples with between three and six participants offer an effective and resource-efficient method of identifying a large proportion of initial and perhaps more obvious issues (Figure 4), this provided the approach for this early evaluation.

For this research, based on established usability testing protocols, the formative evaluation was conducted to provide an insight into early usability and interaction design issues associated with speech and manual input devices. By incorporating the views of a homogenous group of expert users, it was possible to gain an insight into the usability of different interaction modes and configurations.

3. Method

3.1. Participants. In the evaluation of speech input usability (Study 1), the same number of participants were used in each group as follows.

User group 1—speech recognition evaluation group: 12 participants (six men and six women) took part in the trial. All were staff or students from the University of Nottingham with English as their first language. Age ranged from 20 years to 52 years (mean age = 31.5 years).

User group 2—instructing another person evaluation group: 12 participants (six men and six women) took part in the trial. All were staff or students from the University of Nottingham with English as their first language. Age ranged from 21 years to 53 years (mean age = 32.5 years).

User group 3—expert user group: 12 participants (seven men and five women) took part in the evaluation. All were staff from the University of Nottingham. Age ranged from 26 years to 42 years (mean age = 33.4 years).

In the usability assessment of independent-user and collaborative-user interaction modes using both speech and manual input configurations (Study 2), four expert participants (two men and two women) took part. Age ranged from 24 to 31 years (mean age = 26.3 years). All participants had English as their first language, normal or corrected to normal vision, and were human factors experts from the University of Nottingham with VR, speech recognition, and handheld device experience.

3.2. Apparatus. In the evaluation of speech input usability (Study 1), the IDU questionnaire was administered. In the usability assessment of independent-user and collaborative-user interaction modes using speech and manual input configurations (Study 2), the VR system comprised a 800 MHz laptop PC, running VR-Tools 2.1 software, with a data projector and a 2.5 m $\times$ 3 m forward-projection screen to display the CVE in a dedicated usability laboratory at the University of Nottingham. Participants were free to move around the room and, therefore, had no fixed viewpoint of the VE. They typically stood approximately 2 m away from the screen for most of their time. User input was either via a Psion handheld device (for manual input) or a head-mounted microphone (for speech input). The software used for speech recognition was a standard version of "Microsoft Speech." The prototype automotive CVE was used for the evaluation trials. A selection of established VR questionnaires were administered to assess factors associated with user experience including a Simulator Sickness Questionnaire; Stress Arousal Checklist, Presence Questionnaire, Usability Questionnaire, Input Device Usability Questionnaire, Post Immersion Assessment of Experience Questionnaire, and an Enjoyment Questionnaire [38].

3.3. Design. The evaluation of speech input usability (Study 1) followed an intersubject design. The independent variable was the type of evaluation group as follows.

User group 1: this was the group where participants conducted a VR task using speech input to control interaction, and believed they were talking to a computer.

User group 2: this was the group who conducted a VR task using speech to instruct another person.

User group 3: this was the group where expert users conducted a stand-alone assessment for the potential of speech input for virtual applications.

The dependent variables were the responses to questions on the IDU questionnaire.

In the usability assessment of independent-user and collaborative-user interaction modes using both speech and manual input configurations (Study 2), the following comparisons were made:

(i) manual and speech input configurations (independent users and collaborative users),

(ii) independent-user and collaborative-user interaction modes (comparing single users and collaborative users and between the collaborative user groups).

The independent user trials were conducted first and were counterbalanced between the handheld or speech input devices so that any practice or learning effects did not bias the results. The independent user trials served to prepare participants for the collaborative-user trials when they worked with another user to complete tasks together. In the collaborative-user trials, pairs of participants used each of the input devices and were free to divide the tasks as they wished between them. Data from the VR questionnaires, based on measures of presence, usability and input device usability, experience, and enjoyment during immersion, allowed comparisons of each configuration to be made. Furthermore, objective performance data (task completion time), along with observational data and subjective remarks, helped identify the underlying issues of independent-user or collaborative-user VEs and manual or speech input devices.

3.4. Procedure. In the evaluation of speech input usability (Study 1), user groups 1 and 2 completed the IDU questionnaire at the end of a specific session of using a virtual application and speech input. The expert evaluation group completed the questionnaires offline, without using speech in a virtual application.

In the usability assessment of independent-user and collaborative-user interaction modes using both speech and manual input configurations (Study 2), participants trained the speech processor software for 30 mins prior to the trials. Instructions for completing the task were provided, along with familiarisation with the menu options that could be invoked by using the handheld device or speech input. Each participant conducted the trial three times according to the following configurations:

(i) single-user with handheld input device,

(ii) single-user with speech input device,

(iii) multiusers with manual and speech input devices.

In each trial, within the VE, participants were required to navigate to specific, numbered, waypoints and perform short tasks that became more complex as they progressed, such as changing the colour of a vehicle, opening a door, changing a representation to wire-frame, and opening the doors and changing the vehicle colour. At the end of each trial, questionnaires were administered, and participants were paid for their time.

4. Results

For Study 1, data were tested for normality and equality of variance and met the assumptions for parametric analysis. In both studies the IDU questionnaire was rated across a 5-point Likert's scale (1 = strongly agree through to 5 = strongly disagree). Data were collected and compared between the three groups using statistical package for the social sciences (SPSS) statistical software (version 16). Post hoc analyses, where applicable, were conducted using Tukey's Tests. With the small sample for Study 2, only summary observations are reported for manual and speech input configurations and independent-user and collaborative-user interaction modes.

Study 1: Overall Comparisons. Mean scores for participants in each evaluation group were obtained and analysed using a one-way ANOVA. No significant differences were observed ($P > 0.05$) illustrating that, even though the usability scores were higher for the group who believed they were talking to a machine (mean = 3.35; SD = 0.32), they were not significantly different to the group who believe they were talking to another person (mean = 3.05; SD = 0.50) or the expert user group (mean = 3.06; SD = 0.31).

Study 1: Individual Comparisons. When data for individual questions were obtained and analysed using one-way ANOVAs, significant effects were observed.

"*I found it easy to understand how to use the input device to interact with the virtual environment*". a significant main effect was observed for user group ($F(2,33) = 3.49$, $P < 0.05$ (2-tailed)). Post hoc analyses illustrated that the evaluation group who believed they were using a speech interface (mean = 4.00; SD = 0.74) rated the ease of use of speech input higher than the evaluation group who were talking to another person (mean = 3.08; SD = 1.08; $P < 0.05$).

"*The input device was complicated to use*". a significant main effect was observed for user group ($F(2,33) = 3.45$, $P < 0.05$ (2-tailed)). Post hoc analyses illustrated that the evaluation group who believed they were using a speech interface (mean = 1.88; SD = 0.43) did not think speech was as complicated to use as the expert evaluation group (mean = 2.67; SD = 0.98; $P < 0.05$).

"*I found it easy to correct any mistakes that I made when using the input device*". a significant main effect was observed for user group ($F(2,33) = 4.35$, $P < 0.05$ (2-tailed)). Post hoc analyses illustrated that the evaluation group who believed they were using a speech interface (mean = 3.75; SD = 0.45) found it easier to correct mistakes than the expert evaluation group expected to resolve mistakes (mean = 2.83; SD = 0.72; $P < 0.05$). No other significant effects were observed for any of the remaining questions ($P > 0.05$).

Study 1: Qualitative Statements. In addition to the quantitative analysis of the questionnaires, qualitative statements

were also collected. Participants were invited to contribute comments to illustrate some of the findings in more detail (Table 1).

Study 1: Summary. From the analyses, participants who believed they were using a speech recognition system, to control their interaction in the VE, rated the usability of speech input higher than the participants instructing another person or the expert user group rating the potential of speech as an input device. Whilst the results of overall usability were not significant, the data support an element of user perception and experience upon the significant effects that were observed for the specific usability issues of speech as an input device in VR applications. Participants who believed they were using a speech recognition system, therefore, felt usability was higher than the participants who instructed another person to perform the task on their behalf. Given the emphasis on actual use of a system, the follow-on study results are presented by comparing speech input with handheld interaction.

Study 2: Initial Impressions. In the usability assessment of independent-user and collaborative-user interaction modes using speech and manual input configurations, participants were asked about their initial impressions of the different interaction modes (Table 2).

Study 2: Task Completion Time. Time was recorded from the start of the evaluations to the end of the final task. In comparing independent-user data for manual and speech input, participants took longer to complete the task using speech input (mean = 3 mins 31 secs; SD = 47 secs) more than using the handheld device (mean = 1 min 58 secs; SD = 35 secs). When comparing the independent and collaborative user groups, users performed the tasks more quickly when collaborating (mean = 1 mins 30 secs; SD = 28 secs) than when they conducted the task independently (mean = 2 mins 41 secs; SD = 32 secs). It was not possible to compare collaborative user data for manual and speech input, due to the timings made for each collaborative group as a whole (rather than each member separately). However, comparisons between the two collaborative evaluation groups illustrated similar completion times (Group 1 mean = 1 min 23 secs, SD = 18 secs; Group 2 mean = 1 mins 37 secs, SD = 39 secs).

Study 2: Assessment of Experience. This questionnaire was designed to measure user experience of the virtual application along a 7-point scale (e.g., 1 = "much worse than I expected;" 7 = "much better than I expected"). Across all the comparison pairings, user experience was not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input mean = 5.00, SD = 0.82; manual input mean = 4.88, SD = 1.55; individual mean = 4.94, SD = 1.15; collaborative mean = 4.50, SD = 0.58; collaborative speech input mean = 4.50, SD = 0.71; collaborative manual input mean = 4.50, SD = 0.71; collaboration Group 1 mean = 4.00, SD = 0; collaboration Group 2 mean = 5.00, SD = 0).

Study 2: Assessment of Enjoyment. This questionnaire assessed user enjoyment of the virtual experience over 12 questions with six positive and six negative statements, all rated along a 5-point scale (e.g., 1 = "low;" 5 = "high"). Scores for both the positive and negative statements ranged from 6 to 30 and for the total score from −24 to +24. Across all the comparison pairings, user enjoyment was not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input mean = 7.25, SD = 4.27; manual input mean = 10.50, SD = 6.14; individual mean = 8.88, SD = 5.19; collaborative mean = 12.00, SD = 3.56; collaborative speech input mean = 11.00, SD = 5.66; collaborative manual input mean = 13.00, SD = 1.41; collaboration Group 1 mean = 10.50, SD = 4.95; collaboration Group 2 mean = 13.00, SD = 2.12).

Study 2: Presence Questionnaire. This questionnaire was designed to evaluate levels of perceived presence across two subscales: involvement and presence. The ranges of possible scores on the questionnaire were 4 to 20 for involvement (mid-point = 12) and 14 to 70 for presence (mid-point = 42).

For the involvement measure, the comparison pairings were not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input mean = 13.00, SD = 2.58; manual input mean = 13.50, SD = 2.52; independent mean = 13.25, SD = 2.38; collaborative mean = 14.13, SD = 1.93; collaboration speech input mean = 14.50, SD = 0.71; collaboration manual input mean = 13.75, SD = 3.18; collaboration Group 1 mean = 15.50, SD = 0.71; collaboration Group 2 mean = 12.75, SD = 1.77).

For the presence measure, the comparison pairings were not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input mean = 45.31, SD = 0.47; manual input mean = 47.31, SD = 2.39; independent mean = 46.31, SD = 1.92; collaborative mean = 44.13, SD = 2.59; collaboration speech input mean = 45.50, SD = 3.54; collaboration manual input mean = 42.75, SD = 0.35; collaboration Group 1 mean = 45.50, SD = 3.54; collaboration Group 2 mean = 42.75, SD = 0.35).

Study 2: VR Usability and Input Device Usability. Two questionnaires, designed to evaluate levels of perceived usability for the virtual application in general and also the specific input device used, were rated along a 5-point scale (e.g., 1 = "low;" 5 = "high").

For the general VR usability measure, the comparison pairings were not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input mean = 3.22, SD = 0.36; manual input mean = 3.74, SD = 0.61; independent mean = 3.47, SD = 0.52; collaborative mean = 3.66, SD = 0.64; collaboration speech input mean = 3.23, SD = 0.18; collaboration manual input mean = 4.00, SD = 0.82; collaboration Group 1 mean = 3.97, SD = 0.87; collaboration Group 2 mean = 3.26, SD = 0.23).

For the IDU measure, the comparison pairings were not affected by using the particular input devices or by conducting the tasks independently or collaboratively (speech input

Table 1: Qualitative comments from IDU questionnaire.

	Advantages of speech	Disadvantages of speech
User trial 1	(i) Good for interacting with objects within the environment (ii) More natural form of command than typing (or moving a mouse) (iii) Did not have to think as much when deciding how to interact with the environment (iv) Made control of VE more relaxing and less intense (v) It was quick once you knew the command (vi) It was simple to use	(i) Not the most natural method for moving in a VE (ii) Microphones can be too large, bulky and intrusive (iii) Commands can take time to perfect (iv) Joystick/keyboard easier for controlling movement (v) Felt self-conscious (vi) Not sure what commands to use
User trial 2	(i) Could handle several instructions in one command (ii) Less strenuous than using a mouse/joystick (iii) No real learning process required (iv) Natural language—the ultimate user interface	(i) Too easy to say one thing when you mean another (ii) Not good for trivial repetitive tasks (iii) Have to add instructions when initial instructions are not carried out (iv) Not sure what are acceptable commands (v) Microphones can be too large, bulky and intrusive
Expert evaluation group	(i) Good for selecting menus (ii) Good for simple instructions (iii) Good for menus and settings (iv) Natural provided it works! (v) Good for multitasking (vi) Adds to already available interactions when with existing input devices, especially in a "busy" VE (vii) Single word can initiate a complex automated procedure (viii) Hands free, allow other tasks to be performed	(i) Could be disturbed by other people (ii) Might feel self-conscious (iii) Fine adjustment manipulation may prove difficult (iv) Might be difficult for navigation (v) Interaction metaphors not as precise as using joystick or mouse (vi) Inaccurate if user loses concentration (vii) Dislike using for locomotion (viii) Could lead to side effects

Table 2: Initial impressions of manual and speech input.

Question	Interaction mode	
	Manual	Speech
What do you think of the general idea?	(i) Good, as long as it does not distract from the tasks or make it too complicated (ii) Fairly good, they are a standard easily available platform which is mobile (iii) Could be useful (iv) Think it is a good idea	(i) Good, has to be effective and for suitable tasks (if it does not work well, could do more harm than good) (ii) Allows the user to be hands free (iii) Could be useful in certain situations, in principle (iv) I like the idea
What do you think are the general advantages of this mode of input?	(i) More precise movement/control (ii) It is handheld, and it gives a physical interface to manipulate (iii) Wireless, allows for complex full colour interfaces (iv) Quick easy interaction	(i) When input devices not possible (hands using something else) or as additional input device (ii) It frees your hands to do other tasks; people with motor control problems could use the system (iii) Do not need to use complicated handheld devices which may be uncomfortable to use
What do you think are the general disadvantages of this mode of input?	(i) Could make interaction more complicated and navigation difficult (ii) It is a uniform platform, so it has not been designed specifically with this in mind. A tool developed purely for this may be better (iii) Could get overcomplicated (iv) Having to look away from the main display when using them/if presence is important, then this may be distracting	(i) Frustrating if does not work well, when not speaking to, it how does it know? (ii) The user has nothing physical to manipulate so accuracy may be less, and some people may prefer a physical interface (iii) Problems of lag, delay in recognising commands, if at all (iv) Possible inaccuracies might need a lot of training for good recognition
Which functions in virtual applications do you think would be suitable for a handheld device and why?	(i) Hard to say, perhaps group interactions or where rapid response is not necessary (ii) Navigation, interaction of menu items, these items probably easier to display on the handheld device (iii) Discrete tasks, making specific changes to objects or selecting them because the menus used on handheld devices similar to desktop and hence desktop type interaction	(i) Change view point, pull up menus, selecting menu options (ii) Where hand-free is a bonus, for example, surgical uses for surgeons (iii) Anything that requires the users to be using their hands for something else (iv) Discrete tasks, not general navigation

mean = 2.98, SD = 0.64; manual input mean = 3.93, SD = 0.83; independent mean = 3.45, SD = 0.85; collaborative mean = 3.58, SD = 0.80; collaboration speech input mean = 3.00, SD = 0.19; collaboration manual input mean = 4.17, SD = 0.71; collaboration Group 1 mean = 3.90, SD = 1.08; collaboration Group 2 mean = 3.27, SD = 0.57).

Study 2: IDU Statements. In addition to the questionnaire responses, subjective statements from the IDU questionnaire were obtained (Table 3).

Study 2: Summary. Before participants began Study 2, they generally felt that manual input could provide a useful basis for decoupled interaction if it was not too complicated to use and that, as a standard platform, many people would have a wider experience of using such devices for other tasks (e.g., smartphones and tablet PCs). Speech input was also considered to be a useful interaction device although more caution was expressed if the system did not work effectively. Participants regarded the strengths of handheld devices being a precise control format that could be quick and easy to use. It also allowed for a wireless interaction process but still had the benefits of a colour visual display. For speech input, the benefits were considered in relation to simple, handsfree interaction, allowing the user's hands to do other tasks or as an additional input device to complement other, more conventional, interaction devices. It was also considered that speech input might assist users with motor control problems, allowing them to interact with VEs when traditional input devices might be too difficult to use or undermine their experience of the virtual application. Potential problems associated with handheld devices were that they could add to the complexity of interaction with the VE and would mean that the user would have to look away from the VE to view the visual display that could prove distracting. In addition, although handheld devices are ubiquitous, they are not designed specifically for VR use, and so there could well be hidden usability or technical interfacing issues. Speech input could also have problems associated with user frustration if the recognition rate was poor and that without a physical input device task accuracy could be undermined. Another problem that users were cautious of was the amount of time required to train a speech recognition system prior to use.

5. Disscussion

In Study 1 users who believed they were using a speech recognition system generally rated the usability more favourably than the other evaluation groups. Their comments related more to actual system use than participants in the other groups who instructed another person or provided their assessment independently. Users who believed they were using a speech recognition system felt it was easier to understand as an input device than the users who instructed another person. In addition, the expert evaluation group felt that speech would be more complicated to use and more difficult to correct any mistakes than the users who actually used speech input. This would indicate that the users with direct experience of using speech overcame some of the issues that the experts thought might impact on the usability of speech as an input device.

Participants were invited to contribute their own comments, which illustrated the following:

(i) users would enjoy using speech;

(ii) it would be comfortable to use speech;

(iii) speech would make it easy to interact with the VE;

(iv) using a different input device would not make it easier to move around the VE;

(v) it would be not be easy to move and position themselves in a VE using speech;

(vi) it would not feel natural to use speech to control movement in a VE.

The comments from user Group 1, who believed they were using a speech recognition system, illustrated that the speech interface made interaction easier and quicker than instructing another person or potentially using another interaction device. From the previous study [9], speech was an easy and enjoyable input device if it is used for appropriate interactions. Anecdotal evidences and suggests that speech may not be suited for specific actions such as navigation, and so the best use of speech interfaces might be in combination with other input devices for a more integrated approach [6]. The finding that using a different input device would not make it easier to move around the VE might indicate that it was a difficult environment to navigate around and thus highlights the need for careful integration of input devices into the VE design process [39].

From Study 2 it is apparent that each input device had its relative merits and that some of the initial perceptions were borne out or altered after using the devices. Time lags between the hyperlinks and subsequent changes in the VE caused frustration and confusion. Users also stated that it was easy to become disorientated with the handheld-to-VE interpretation, whereas they had initially thought it would offer a precise control device. That said, the handheld device was considered to be intuitive, easy to learn, and consistent with natural heuristics for navigation that initially were not thought to be the case. Speech input was considered easier to remain orientated in the VE rather than using the handheld device. Users liked the novelty factor of speech input but were frustrated at times by poor recognition rates and difficulties experienced in navigating around the VE. This may have been because navigation was a continuous process, and other research supports the notion that speech input is not well suited for this type of task [9]. From the collaborative evaluations, it is interesting that speech input was considered useful when combined with another input device where it was intuitively used for object manipulation whilst navigation was controlled by the handheld device.

From the single-user evaluation, questionnaire responses for involvement, presence, VR usability, experience, and enjoyment ratings did not illustrate any major differences between the handheld device or speech input. This was supported by similar observations for that IDU questionnaire that would have been more sensitive to input device

Table 3: Input device usability statements.

Interaction mode	
Speech	Handheld device
(i) I liked the novelty factor of using speech input, but, as it failed to recognise my voice, it became frustrating. Easier with someone else navigating so I did not have to use speech so much	
(ii) Unlike the handheld device, one was able to control looking up or down. Although it was more limited, it made it a lot harder to get disorientated	(i) I like everything about the device; it was comfortable, reasonably intuitive, but has the potential to be more so
(iii) Bad for navigation, good for discrete tasks	(ii) The only thing I found myself wanting to do is use the up and down "keys" on the control to move forward and backward
(iv) Not good for navigation, felt natural for tasks other than movement. Speech was really good for interacting with the VE, thought it carried out the wrong command at times which confused me	(iii) The navigation system was poor as it was based on rotating in two axes; once I had left the floor, it was hard to regain the orientation
(v) Did not respond quickly enough, had to repeat some commands several times, and sometimes wrong action was carried out	(iv) I liked the display; the only problem was a slight delay in the acceptance of links on the handheld device which lead to incorrect selection from the menu a couple of times
(vi) It was hands free; all the required interaction with the environment was possible	(v) Use seemed very natural and very easy to learn
(vii) It did not recognise my voice; small precise adjustments were not possible	(vi) It was consistent with natural heuristics to move. Performing the task was more frustrating as one had to navigate a menu hierarchy with no short cuts
(viii) I liked novelty, disliked errors, found it difficult to recall available functions (need a constant menu?)	

differences. The only difference was observed for the task completion time with participants taking nearly twice as long to complete the task using speech than using the handheld device. This was probably due to the poor recognition accuracy of the software (even though a speaker-dependent system was used), where participants often had to repeat commands a number of times. Even so, this did not affect data for the IDU questionnaire, user experience, or enjoyment during the trials, perhaps indicating that participants were not unduly affected by the longer completion times or repetitive interaction processes.

In considering the comparison of single users and collaborative users, the findings illustrated similarities across the general questionnaires data as well as for the IDU questionnaire. As the single user evaluation, the only difference observed was for task completion time where collaborative participants were quicker than those who completed the task alone. This was probably due to the division of tasks between participants and how the input devices were used for the tasks. In both collaboration trials, participants were given the choice of which input devices they used for which tasks (e.g., object manipulation and navigation). In both cases participants naturally used the handheld device for navigation and speech input for object manipulation. This may have seemed the most intuitive way of combining the input devices although it was possible to complete that tasks using either or both the devices for all or part of the trial.

As with the single-user evaluation, any problems associated with using speech input did not influence responses to the IDU questionnaire, user experience, or enjoyment during the trials, indicating that participants were not unduly affected by the longer completion times. However, trends in the data illustrate that there was a slight increase in the ratings of usability and enjoyment when participants collaborated than when they completed the task alone. This may have been due to the activity of collaborating masking any negative effects through mediation of tasks and the use of input devices. Furthermore, presence and the VR experience were higher when participants conducted the task alone, which may be due to participants not having to think about another user in the same task application.

When the collaborative evaluation was assessed for mode of interaction, there was no difference between the use of speech or manual input. Trends in the data illustrate that involvement and presence were rated higher when using speech input. This was probably due to speech being less intrusive in the virtual application, allowing participants to become more involved in the task. General usability, IDU, and enjoyment were rated higher when using the handheld device perhaps because it mapped onto the navigation task more readily than speech input mapped onto object manipulation.

As the two collaboration groups should have been homogenous, they were compared to investigate any potential differences between them. Data for presence, VR usability, IDU, experience, and enjoyment were similar across the groups. Based on these findings the participants' perceptions of the VE and use of the input devices did not appear to have any effect on their collaborative behaviour. Furthermore, task completion times between the two collaboration groups were similar, indicating that both groups performed the tasks within similar time frames.

From the overall findings it would appear that each input device had its relative merits, supported by the subjective feedback from the trials. Given the small sample size for Study 2, it could be argued that the more obvious issues have been identified and that, with a larger sample or further iterations, more subtle usability issues might be highlighted. However, in terms of the questionnaire data,these merits did

not produce any clear differences and so speech input would generally appear to offer a viable mode of interaction within virtual applications, and both speech and manual input offer potentials for decoupled interaction.

5.1. The Potential of Speech Input in Decoupled Interaction. Whilst visualising menus in the VE as and when they are required might be an advantage, using speech input could eliminate the need for complex menu structures. This could reduce the overall interaction time within a VE and could cut down the amount of programming time required to build menus into the VE. Another possibility is that menus could be implemented at relative points in the VE or interaction process (rather than having them visible throughout using a VE), and they could be gradually faded as users become more proficient at using speech input [6]. With speech input, it would be possible to issue specific commands thereby reducing the time required in locating menu items and manually "clicking" on them. This could be beneficial for users completely immersed in a VE as the interaction could remain within the VE and continue uninterrupted. Using speech input, users are removed from cumbersome devices such as keyboards and joysticks [40] creating a more natural method of interaction as contact with the VE would be of a more intuitive nature [9]. Speech input also removes the need for any input calibration although it is arguable how much time might be required to train a speech interface before interacting with a VE [6].

Building on recent progress in understanding collaborative interactions, CVEs have often focused on enhancing the sense of presence within the VE in order to support collaborative activities [4]. However, a key purpose of decoupled interaction is supporting single users who are colocated rather than group-based distributed interactions [1]. Compared with the traditional approach of one active user in a particular application, where other users are often passive observers, this approach could generate new group dynamics and interaction potentials within CVEs. Several users could control the VE or query some of its properties using independent interaction devices at the same time, enabling colocated access to the CVE [5]. This has led to the development of interaction through multiple decoupled interaction (MDI) as it will be more common for an increasing number of users to carry small devices with advanced interactivity, connectivity capabilities, and functionality, opening up new possibilities for interaction design [1].

6. Conclusion

From Study 1, user perception would appear to be influenced by direct experience of using speech input. With respect to this, the findings highlight how some tasks (e.g., menu selection, object manipulation) might be suitable for speech input, whereas other tasks (e.g., navigation) might be better suited to other input devices. However, it is only when the underlying human factors issues are addressed that the usability of speech input can be enhanced. In order to develop these ideas further, it was important to investigate combining interaction devices whilst also considering advanced virtual applications such as the notion of "decoupled interaction" and single or multiple users. This led to the evaluation in Study 2 which presented ideas for decoupling interaction in VEs, where it is possible to translate interactions and make them compatible with other types of input devices such as handheld technologies or speech recognition processors. This approach to CVE interaction is still at an early stage of development and the potential or validity of this interfacing concept is still under evaluation; however, as a concept demonstrator the results of these initial evaluations demonstrate the potential of both input devices, highlighting their suitability for advanced virtual applications.

Acknowledgments

The work presented in this paper is supported by the IST Grant 2000-26089: "VIEW of the Future." The authors are indebted to the anonymous reviewers of this paper who offered valuable and constructive feedback.

References

[1] V. Bayon, G. Griffiths, and J. R. Wilson, "Multiple decoupled interaction: an interaction design approach for groupware interaction in co-located virtual environments," *International Journal of Human Computer Studies*, vol. 64, no. 3, pp. 192–206, 2006.

[2] S. Benford, C. Brown, G. Reynard, and C. Greenhalgh, "Shared spaces: transportation, artificiality, and spatiality," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, (CSCW '96)*, pp. 77–86, November 1996.

[3] H. Schnädelbach, B. Koleva, M. Flintham et al., "The augurscope: a mixed reality interface for outdoors," in *Proceedings of the Human Factors in Computing Systems, (CHI '02)*, pp. 9–16, ACM Press, April 2002.

[4] E. F. Churchill, D. N. Snowdon, and A. J. Munro, *Collaborative Virtual Environments. Digital Places and Spaces for Interaction*, Springer, London, UK, 2001.

[5] V. Bayon and G. Griffiths, "Co-located interaction in virtual environments via de-coupled interfaces," in *Proceedings of the 10th International Conference on Human-Computer Interaction, (HCI '03)*, C. Stephanidis, Ed., Lawrence Erlbaum Associates, 2003.

[6] A. W. Stedmon, "Developing virtual environments using speech as an input device," in *Proceedings of the 10th International Conference on Human-Computer Interaction, (HCI '03)*, C. Stephanidis, Ed., Lawrence Erlbaum Associates, 2003.

[7] H. Hoffman, O. Stefani, J. Deisinger et al., "Users' needs in terms of applications, required applications and recognition of gaps," Tech. Rep. IST-2000-26089, View of the Future Project Deliverable 2.2, 2002.

[8] H. Patel, S. Sharples, S. Letourneur et al., "Practical evaluations of real user company needs for visualization technologies," *International Journal of Human Computer Studies*, vol. 64, no. 3, pp. 267–279, 2006.

[9] A. W. Stedmon, H. Patel, S. C. Sharples, and J. R. Wilson, "Developing speech input for virtual reality applications: a reality based interaction approach," *International Journal of Human Computer Studies*, vol. 69, no. 1-2, pp. 3–8, 2011.

[10] M. H. Cohen, J. P. Glangola, and J. Bagola, *Voice User Interface Design*, Addison-Wesley, Boston, Mass, USA, 2004.

[11] A. Leuski and D. Traum, "Practical language processing for virtual humans," in *Proceedings of the 22nd Innovative Applications of Artificial Intelligence Conference, (IAAI '10)*, Georgia, Ga, USA, July 2010.

[12] O. Lemon, "Learning what to say and how to say it: joint optimisation of spoken dialogue management and natural language generation," *Computer Speech and Language*, vol. 25, no. 2, pp. 210–221, 2011.

[13] M. De Paiva Guimarães, B. B. Gnecco, and M. K. Zuffo, "Graphical interaction devices for distributed virtual reality systems," in *Proceedings of the ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry, (VRCAI '04)*, pp. 363–367, ACM Press, June 2004.

[14] D. Traum, "Talking to virtual humans: dialogue models and methodologies for embodied conversational agents," in *Modelling Communication*, I. Wachsmuth and G. Knoblich, Eds., Springer, Heidelberg, Germany, 2008.

[15] P. R. Cohen, D. McGee, S. L. Oviatt et al., "Multimodal interaction for 2D and 3D environments," in *Proceedings of the IEEE Computer Graphics and Applications*, pp. 10–13, 1999.

[16] M. Lee and M. Billinghurst, "A wizard of Oz study for an AR multimodal interface," in *Proceedings of the 10th International Conference on Multimodal Interfaces, (ICMI '08)*, pp. 249–256, October 2008.

[17] E. Farella, D. Brunelli, M. E. Bonfigli, L. Benini, and B. Riccò, "Multi-client cooperation and wireless PDA interaction in immersive virtual environment," in *Proceedings of the 8th Annual Scientific Conference on Web Technology, New Media Communications and Telematics Theory Methods, Tools and Applications, (EUROMEDIA '03)*, pp. 177–184, University of Plymouth, England, UK, April 2003.

[18] M. Billinghurst, H. Kato, and S. Myojin, "Advanced interaction techniques for augmented reality applications," in *Virtual and Mixed Reality*, R. Shumaker, Ed., Springer, Heidelberg, Germany, 2009.

[19] S. Jeon, J. Hwang, G. J. Kim, and M. Billinghurst, "Interaction with large ubiquitous displays using camera-equipped mobile phones," *Personal and Ubiquitous Computing*, vol. 14, no. 2, pp. 83–94, 2010.

[20] H. Bullinger, R. Blach, and R. Breining, "Projection technology applications in industry. Theses for design and use of current tools," in *Proceedings of the 3rd International Immersive Projection Technology Workshop*, Springer, Stuttgart, Germany, 1999.

[21] D. Bowman and L. Hodges, "Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments," *The Journal of Visual Languages and Computing*, vol. 10, no. 1, pp. 37–53, 1999.

[22] R. W. Lindeman, J. L. Sibert, and J. Hahn, "Hand-held windows: towards effective 2D interaction in immersive virtual environments," in *Proceedings of the IEEE Virtual Reality, (IEEE VR '99)*, pp. 205–212, March 1999.

[23] G. Smith, T. Salzman, and W. Stuerzlinger, *3D Scene Manipulation with 2D Devices and Constraints*, Morgan Kaufman, 2001.

[24] R. J. K. Jacob, A. Girouard, L. M. Hirshfield et al., "Reality-based interaction: a framework for post-WIMP interfaces," in *Proceedings of the 26th Annual CHI Conference on Human Factors in Computing Systems, (CHI '08)*, pp. 201–210, Florence, Italy, April 2008.

[25] A. Barone, *A usability comparison of two virtual environments and three modes of input*, M.S. Dissertation, University of Nottingham, 2001.

[26] R. Kalawsky, *Exploiting Virtual Reality Techniques in Education and Training: Technological Issues*, Advisory Group on Computer Graphics, Advanced VR Research Centre: Loughborough University, 1996.

[27] M. Scriven, "The methodology of evaluation," in *Perspectives of Curriculum Evaluation*, R. W. Tyler, R. M. Gagne, and M. Scriven, Eds., pp. 39–83, Rand McNally, Chicago, Ill, USA, 1967.

[28] C. Weston, L. McAlpine, and T. Bordonaro, "A model for understanding formative evaluation in instructional design," *Educational Technology Research and Development*, vol. 43, no. 3, pp. 29–48, 1995.

[29] J. Earthy, B. Sherwood Jones, and N. Bevan, "The improvement of human-centred processes—facing the challenge and reaping the benefit of ISO 13407," *International Journal of Human Computer Studies*, vol. 55, no. 4, pp. 553–585, 2001.

[30] S. A. Nan, Formative evaluation, 2003, http://www.beyondintractability.org/essay/formative_evaluation/.

[31] J. Nielsen, "Why you only need to test with 5 users," *Alertbox*, 2000, http://www.useit.com/alertbox/20000319.html.

[32] R. A. Virzi, "Refining the test phase of usability evaluation: how many subjects is enough?" *Human Factors*, vol. 34, no. 4, pp. 457–468, 1992.

[33] C. Perfetti and L. Landesman, "Eight is not enough," 2001, http://www.uie.com/articles/eight_is_not_enough/.

[34] C. W. Turner, J. Nielsen, and J. R. Lewis, "Current issues in the determination of usability test sample size: how many users is enough?" in *Proceedings of the Usability Professionals' Association, (UPA '02)*, Chicago, Ill, USA, 2002.

[35] J. R. Lewis, "Testing small system customer set-up," in *Proceedings of the 26th Annual Meeting of the Human Factors Society*, pp. 718–720, Human Factors Society, 1982.

[36] A. Woolrych and G. Cockton, "Why and when five test users aren't enough," in *Proceedings of IHM-HCI (IHM-HCI '01)*, J. Vanderdonckt, A. Blandford, and A. Derycke, Eds., vol. 2, pp. 105–108, 2001.

[37] J. Nielsen and T. K. Landauer, "A mathematical model of the finding of usability problems," in *Proceedings of the ACM Conference on Human Aspects in Computing Systems, (CHI '93)*, pp. 206–213, Amsterdam, The Netherlands, April 1993.

[38] S. V. G. Cobb, S. C. Nichols, A. R. Ramsey, and J. R. Wilson, "Virtual reality-induced symptoms and effects (VRISE)," *Presence*, vol. 8, no. 2, pp. 169–186, 1999.

[39] M. D'Cruz, A. W. Stedmon, J. R. Wilson, P. J. Modern, and G. J. Sharples, "Building virtual environments using the virtual environment development structure: a case study," in *Proceedings of the 10th International Conference on Human-Computer Interaction, (HCI '03)*, Lawrence Erlbaum Associates, Crete, Greece, June 2003.

[40] W. A. Lea, "Speech recognition: past, present, future," in *Trends in Speech Recognition*, W. A. Lea, Ed., Prentice Hall, New York, NY, USA, 1980.