Project ref. no.	FP6-507609
Project acronym	SIMILAR
Deliverable status	R
Contractual date of delivery	28 February 2007
Actual date of delivery	8 October 2007
Deliverable number	D98
Deliverable title	Report on Iterative Testing of Multimodal Usability and Evaluation Guide
Nature	Report
Status & version	Final
Number of pages	119
WP contributing to the deliverable	SIG7
WP / Task responsible	NISLab
Editor	N/A
Author(s)	Niels Ole Bernsen and Laila Dybkjær
EC Project Officer	Mats Ljungqvist
Keywords	Evaluation, test plan, test with users, test results
Abstract (for dissemination)	This deliverable documents two system usability tests with representative users that were carried out in Denmark in June 2007. The two systems – a Sudoku game with speech and 3D gesture input and a strongly multimodal treasure hunt game for collaboration between a blind and a deaf-mute partner - represent new multimodal technology. The purpose of the tests was partly to evaluate these new technologies and partly to test a draft of a book on Multimodal Usability written by the authors of this report Test planning, the tests themselves, and the subsequent analysis of results were based on recommendations from the book draft. The impact on the book of the usability tests reported in this report is described in Deliverable D100.

SIMILAR project (IST - FP6-507609)

SIMILAR

Deliverable D98

Report on Iterative Testing of Multimodal Usability and Evaluation Guide

6 October 2007

Authors

Niels Ole Bernsen[1] *and Laila Dybkjær*[1]

1: NISLab, Odense, Denmark


similar
network of excellence

Table of Contents

1 This Report

This report documents two system usability tests that were carried out in Denmark in June 2007.

The purpose of the tests were two-fold: (1) to test new multimodal technologies developed by partners in SIMILAR, and (2) to test the first version of the authors' book manuscript *Multimodal Usability* which has been written as part of the work on usability in SIMILAR. The present report addresses purpose (1) above. Purpose (2) is addressed in SIMILAR Deliverable D100 *Multimodal Usability Progress Report.*.

The systems tested were the following:

1. A speech and 3D gesture input system for playing Sudoku developed at ZGDV in Darmstadt, Germany.
2. A strongly multimodal computer game system which enables a blind and a deaf person to collaborate in a treasure quest. This system was developed at ITI-CERTH in Thessaloniki, Greece.

The Sudoku system was tested at NISLab. The blind-deaf system was tested at the Institute for The Blind in Copenhagen.

The present report is structured as follows:

The Sudoku test and the test of the treasure hunt system are documented in three chapters each, i.e., Chapters 2-4 and Chapters 5-7, respectively. Each set of three chapters has a similar structure: in the first chapter in a set, the system to be tested is presented and explained, followed by the test plan. The second chapter describes the actual test and the subjects, and the third chapter presents test results and conclusions. Both usability tests included structured post-trial interviews made immediately after the system trials. The interview script and interview results from the Sudoku and Treasure Hunt tests are presented in Appendix 1 and Appendix 2, respectively. Appendix 1 also includes the screening instructions for selecting subjects for the Sudoku test. The German versions of the Sudoku interview script and screening instructions were translated from the English versions by Cornelius Malerczyk, ZGDV.

2 Sudoku User Test Plan

2.1 This Section

This section describes the Sudoku system to be tested and presents the test plan for the Sudoku user test based on the test plan document by Bernsen and Dybkjær of 6 June 2007.

2.2 Description of the Sudoku System

The Soduko system is a prototype system developed as a student project at ZGDV in Darmstadt, Germany. The system requires at least a 2.5GHz Pentium 4, 512MB RAM, Firewire interface, two DCAM compatible firewire cameras, Windows XP, and a speech recogniser.

At NISLab we run the implemented prototype on an HP Deskpro workstation, 2.13GHz Intel Core2 DUO, 1GB ram, Firewire and Windows XP. The two cameras used are DMK 21F04 from ImagingSource (http://www.1394imaging.com/en/products/cameras/firewire_mono/dmk21f04/overview/) (640x480pixels, monochrome). Each camera has a Velleman CAML2IR 4mm objective with built-in IR-illumination. The speech recogniser we use is from Microsoft delivered for free as part of the SAPI 1.5 package ((http://www.microsoft.com/speech/download/old/sapi5.asp).

System output is provided on the 42" display shown in Figure 3 as a Sudoku game board consisting of a large 9 x 9 square containing a total of nine smaller 3 x 3 squares or a total of 81 input squares each of which may contain an integer between 1 and 9 (Figure 1). To successfully complete a game, each 3 x 3 square must include all integers between 1 and 9, and each row and each column in the large square must equally include all integers between 1 and 9.

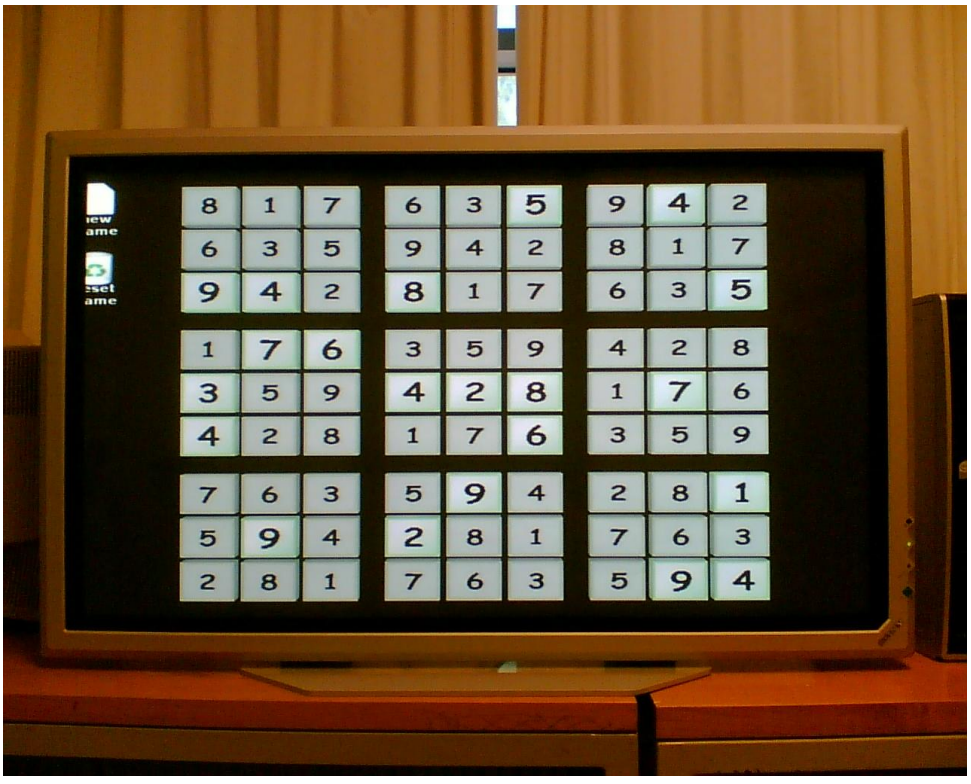

Figure 1. A correctly solved Sudoku game.

At the start of a new game, only a fraction of the 81 input squares are filled as illustrated in Figure 2. The difficulty of the game depends, among many other things, on how many squares are filled initially. The game is solved through logical reasoning rather than through applied mathematics. The evaluated version of the system contains three levels: easy, medium and hard, which have been set to correspond to an initial filling of 30, 45, and 60 squares, respectively.

When playing the game, the user stands in a designated (chalk-marked) spot on the floor in front of a large (42'') screen (Figure 3). User input is through camera-captured 3D pointing gesture and speech. Gaming can start when the cameras have located the user's index (or other) finger as evidenced by the appearance of a visible (highlighting) modification of the square the user happens to point at. The square gets highlighted no matter whether it already contains an integer or not. The user is expected to point at one of the input squares and say the word "number" followed by an integer between 1 and 9, e.g., "Number 2". If the square already contains an integer, that integer gets replaced by the new one uttered by the user. The attentive user is able to tell the difference between the integers that were present at the start of a game and those inserted during game play. The latter are slightly larger than the former and their background is lighter and greenish rather than grey (Figure 1). The user is only able to modify the integers which have been inserted during the game.

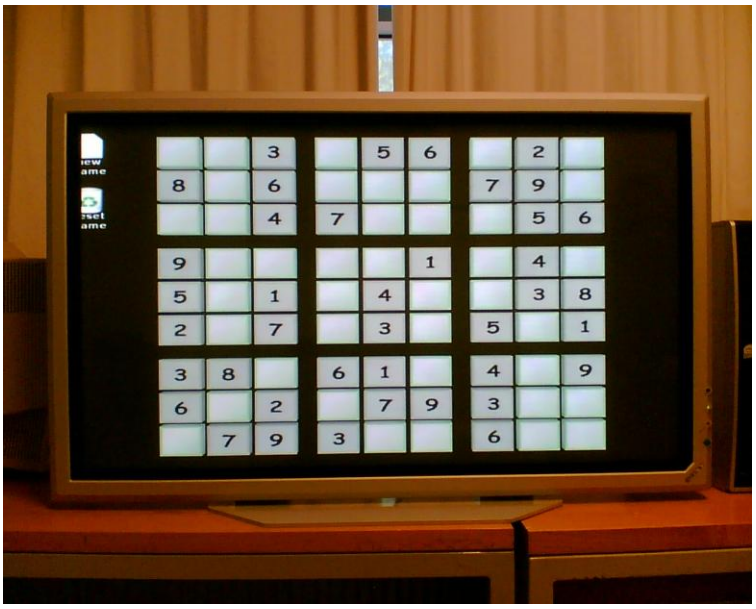

Figure 2. A new Sudoku game.

During the game, the system provides elementary on-line help, as follows. If the inserted integer is in conflict with the basic rules of the game (mentioned above), so that the same integer occurs twice in a row, column, or 3 x 3 square, that row, column, or 3 x 3 square turns red, cf. Figure 4. The red colouring remains until the error has been corrected. If the user proceeds to make another error before correcting the previous one, more of the game board may turn red. It should be noted that the help provided by the system is far from sufficient to guarantee a successful game, especially in medium and hard games. It is quite possible to insert a wrong number in a square without conflicting with the basic Sudoku rules, in which case the user gets no warning that something is wrong. If the user does not spot the error immediately, it may only be discovered later in the game when additional numbers have been inserted, which often makes it hard to locate the error.

Figure 3. The user stands inside the marked-up area.

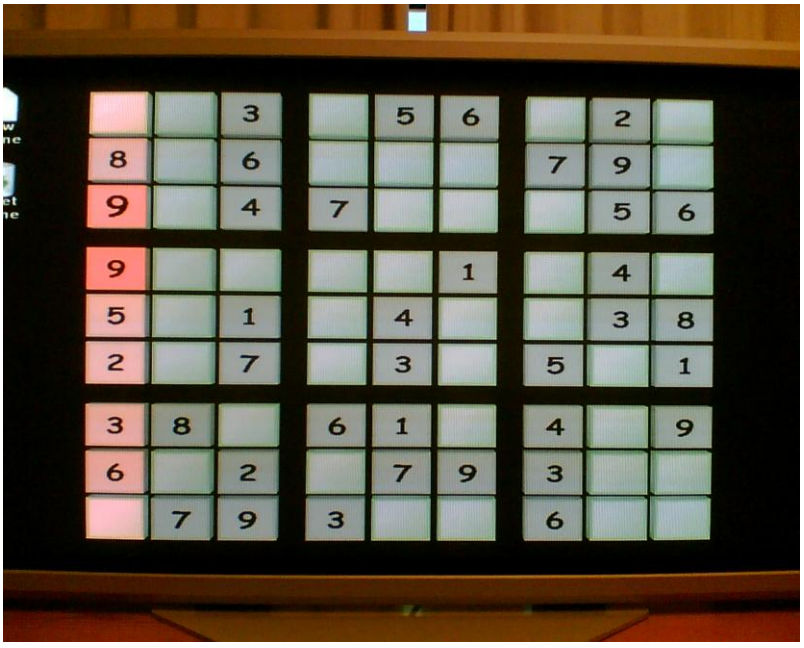

Figure 4. The red markup indicates a basic error in the column.

Speech recognition is grammar-based. In addition to spoken integer input, the following commands can be understood:

- delete that
- delete this
- remove that
- remove this

Any of these four commands may be given while the user points at one of the (user-) filled input squares on the game board. As a result the input square pointed at will become empty. Input speech and pointing gesture do not have to be synchronous or quasi-synchronous. It is possible to execute an input command by speaking both before and after pointing to a particular square. The system works by associating a pointing event with the temporally closest unused spoken command input event.

To start a new game one must say "Start new game" or point at the *new game* button in the top-left part of the screen (Figure 1). When starting a new game, the user must choose its level of difficulty by pointing to one of the three displayed options *easy*, *medium*, *hard* (Figure 5). It is also possible to restart an ongoing game by pointing at the *reset game* button below the *new game* button (Figure 1). When restarting a game, the user must confirm the choice made (Figure 6).

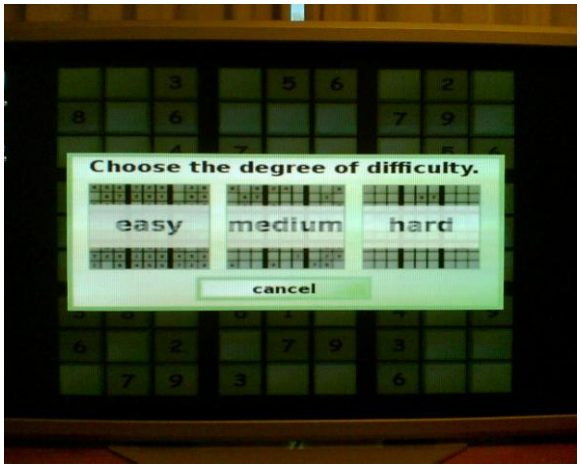

Figure 5. Levels of difficulty.

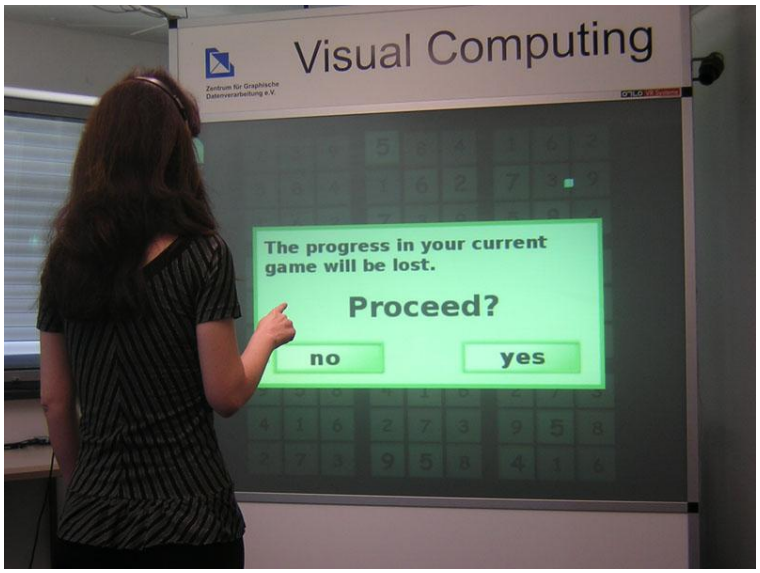

Figure 6. Restarting a game.

When a game has been completed correctly, a congratulation screen is shown (Figure 7).

2.3 Overall Evaluation Goals

The evaluation of the Sudoku system has two overall goals:

1. To provide input on usability aspects of the Sudoku game, in particular regarding appropriate use of modalities, offered functionality, ease of use, and user satisfaction.
2. To provide input on how well the relevant parts in [Bernsen and Dybkjær, in press] work with respect to evaluation with users as well as how to plan, carry out and analyse test results.

The second goal is addressed in SIMILAR Deliverable D100 *Multimodal Usability Progress Report*. This present document thus only concerns the usability evaluation of the Sudoku system with representative users invited to NISLab, following the present evaluation protocol

Figure 7. Congratulations when a game has been successfully completed.

2.4 Evaluation Criteria

Table 1 gives a brief overview of what to measure and how, based on data from the usability test. All 20 table items will be addressed in post-test interviews and 10 items (numbered 5 through 10, 12, and 15 through 17) will be topics for test data collection and analysis as well. Questions 1 through 4 will be closed Likert-scale questions whereas all other interview questions will be semi-open or open questions. The interview scripts in English, Danish and German is shown in the Interview Scripts in Appendix 1, Sections 10.13, 10.14 and 10.15, respectively.

What to measure	How to measure
Appropriateness of modalities used	
1. Appropriateness of pointing input	Closed interview question
2. Appropriateness of spoken input	Closed interview question
3. Appropriateness of screen output	Closed interview question
4. Appropriateness of modality combination for interaction	Closed interview question
Quality of interaction	
5. Quality of pointing input understanding	Interview question + data from the interaction
6. Quality of pointing input provision	Interview questions + data from the interaction
7. Quality of speech input understanding	Interview question + data from the interaction
8. Quality of speech input provision	Interview questions + data from the interaction
9. Quality of combined speech gesture input understanding	Interview question + data from the interaction
10. Quality of combined speech gesture input provision	Interview question + data from the interaction
11. Missing input modalities?	Interview question
12. Output interface intelligibility	Interview questions + data from the interaction
13. Missing output modalities?	Interview question
14. Naturalness of interaction	Interview question
15. Ease of interaction	Interview questions + data from the interaction

16. User in control	Interview question + data from the interaction
Functionality	
17. Sufficiency of functionality	Interview questions + data from the interaction
User experience	
18. User satisfaction	Interview question
19. Advantages and disadvantages	Interview question
20. Play again?	Interview questions

Table 1. What and how to measure in the Sudoku test.

2.5 Test Users and Their Profiles, Screening

Sudoku is a leisure game which enjoys world-wide popularity. A popular Sudoku website is http://www.websudoku.com/. Sudoku is played by all kinds of users, including (not too young) children, adults and elderly people, males and females, from novices through occasional players to very skilled gamers, and people with very different educational and professional backgrounds. It is notoriously hard to compose a representative user test group when the group cannot be a large one due to limited resources yet still has to be representative of virtually everyone. Although children also play Sudoku, we shall not include children among our test users. The primary reason is that the speech recogniser used (Section 2.2) is not special-tuned to children's voices, which means that it will most likely have some difficulties in recognising input from children and adolescents under some 16 years of age.

Having thus restricted our test user group to adults (or the over-16 years-old) only, we aim at a modest amount of representative spread among these, subject to the qualification that resources only allow for a relatively small user group of 12 users. We aim to have a reasonable balance regarding gender, i.e. at least 40% must be males and at least 40 % must be females, and age, i.e., approximately one third of the users must be under 30 years old, between 30 and 50 years, and above 50 years old, respectively.

Regarding Sudoku skills, we will divide test users into three groups each of which will be balanced with respect to gender, i.e.

1. subjects who have little or no experience in Sudoku but have an interest in trying it (again);
2. subjects with some experience in Sudoku who have managed easy-to-medium-difficulty games;
3. subjects who are used to manage difficult Sudoku games.

Each group will play games at a level corresponding to their skills and experience from among the three levels of difficulty offered by the system (Section 2.2). Each test group will include four users. We shall not consider their educational background and age in other respects than by ensuring the age spread described above and ensuring that there will be no more than two test persons with the same profession across all three groups.

Although it might be possible to recruit all subjects among employees (researchers, administrators and support staff) and students at the university and still be in line with the requirements above, we plan to recruit at least four subjects from outside the university.

In order to enable selection of test subjects according to the above criteria, potential test subjects must be contacted over the phone or face-to-face, and screened prior to recruitment. User screening instructions and questions in English, Danish and German are described in Appendix 1, Sections 10.16, 10.17 and 5.18, respectively.

2.6 Test Design

The usability test will take place in NISLab's usability lab at the University of Southern Denmark (SDU) on Thursday 7 June and Tuesday 12 June 2007. The system has been set up and is running in this room and the room is available on the mentioned dates. A camera for audio/video recordings is also available on those dates.

We plan to have six users per day and have set aside an hour per session. This should leave sufficient time for each test so that the next user does not have to wait.

A user session is expected to involve max. 5 minutes for introduction to system and tasks, followed by 25 minutes game playing with the system, followed by a 20 minutes interview in a separate room. The 25 minutes of game playing will normally be spent on 12 minutes for playing Game 1, 1 minute for changing game, and 12 minutes for playing Game 2, see also Section 2.9. The interview will be based on the script in Appendix 1, Section 10.14.

Each user will receive two cinema tickets as a reward for having participated.

Twelve test users will be recruited by Svend. Users will be recruited in accordance with the criteria described in Section 2.5. To minimize subjects' transportation costs we will try to find local subjects.

Before actually inviting subjects to participate, they will be informally screened regarding their Sudoku skills and interests by being asked which group they consider themselves to belong to (novice, intermediary, expert), how often they play Sudoku, for how long they have done it and informally at which level(s) of difficulty they prefer to play. They must also know the numbers from 1 to 9 in English. If a potential test user has a profile which is still missing according to Section 2.5 and this section (above), s/he will be invited to participate. If not, the person will be asked if we can keep him/her on standby in case there turns out to be a no-show during the test.

Each (potential) test user will be told that s/he is going to help us evaluate an electronic Sudoku game for about an hour, including about half an hour for introduction and game-playing and about 20 minutes for an oral interview immediately after the session with the system. The person will also be told that there will be a remuneration in the form of two open cinema tickets. No transportation or other costs can be covered. The exact location of the test will be described and date (cf. above) and time must be agreed upon if the screening test shows that the user has a profile that we still need at the time. All invited users will be given Svend's email address and telephone number in case they need more information or have to send their apologies for not coming. For ease of communication, Svend will ask the test users for their email address and phone number as well.

2.7 Roles

During the usability test of the Sudoku game, we need people for the following roles:

1. someone who receives and takes care of the user when the user is not in the test room;
2. an experimenter who has the contact with the user during the session;
3. a technician who ensures that the system is up and running;
4. a person who takes care of the video camera for recording test users during the sessions;
5. an observer;
6. an interviewer.

Role 1 will be taken care of by Svend, roles 3 and 4 by Torben, and the remaining roles (2, 5, 6) will be shared by Ole and Laila.

2.8 Test Environment and Equipment

The test sessions will take place in NISLab's usability lab where the Sudoku system is already up and running as described in Section 2.2, including calibration of the cameras needed for pointing input. It will be tested that the system is up and running before each session starts. The Sudoku game board is shown on a 42'' screen. The cameras are mounted at the ceiling. The user will use a Logitech headset microphone for spoken input. A chalk line on the floor will show the user where to position himself/herself to allow the cameras to capture his/her pointing gesture.

A webcam camera for recording the interaction (video and audio) is available. It will prior to each session be checked that it works appropriately.

As mentioned above, it is not expected that users will have to wait. However, they may arrive very early in which case they will be invited to sit in a room where there will be coffee and something to read. The interview will also take place in the relaxed environment of this room.

2.9 Tasks and Test Conditions

Each user will first be given a brief introduction to the system by the experimenter who will:

- tell that the expected use environment of the system is not at home but, rather, in places like a shop where there is a queue or while waiting for someone to decide on which clothes to buy, in airport terminals, train stations or any other place where people sometimes have to wait, at exhibitions, etc.
- provide a brief explanation of the rules if the user has never played Sudoku before;
- *briefly* explain and show what the buttons on the screen are meant for;
- demonstrate how to start a game, play it, and change to a new game;
- tell that the user is expected to try two games and start a third one;
- ask the user to select the easy, medium or hard game level (Section 2.2) depending on the group to which the user belongs (Section 2.5);
- explain and show how the user will be given signs by the experimenter during game play: the experimenter will slide a piece of paper in front of the user on the floor (in order not to interfere with camera and speech input), when they should change to a new game or game level;
- explain that the user will not necessarily have finished a game when signalled to start a new game;
- clearly emphasise that this is not at all a test of the user's Sudoku skills but a test of how good the system is for playing the game in special environments.

The purpose of asking the user to start a new game at the end of the session is to collect additional data on this particular action.

Exceptions: should a user go "cold" during game play so that little or no interaction happens for long periods of time, the experimenter has the possibility of signalling to the user to change to a new game right away or even to ask the user if the gaming level should be changed to an easier one. We don't get useful data if a user does not play. This will be done as follows: if the user seems to have run out of options and pauses for more than 3 minutes, or several times for longer than 2 minutes, the user will be asked to choose between selecting a new game or lowering the game difficulty level by one step.

NOTE on priming: In order to be able to collect spontaneous data on the use of combined speech and gesture, the experimenter must be *extremely careful* when demonstrating how to talk and point. In the demonstration before the user starts gaming, the experimenter must use

all three possible forms of temporal combination of speech and gesture, i.e., (i) speak and then point afterwards, (ii) speak and point at more or less the same time, and (iii) point, retract the hand, and then speak. The experimenter should train the demonstration in advance and make sure that it becomes natural to demonstrate the system in the way just described.

2.10 Data Collection

The data to be collected includes:

- video and audio recordings of user interaction with the system. The video will show the user's hand/arm and the screen contents, and thus will be taken from a position slightly to the left/ right of, and slightly behind, the user;
- observation notes produced by the observer during the sessions;
- interview notes written during the interviews with the users.

When user tests and data collection have been completed, the data will be validated to make sure that the data is, in fact, appropriate for the various kinds of data analysis planned (Section 2.4). A detailed plan for data markup and coding scheme creation will be specified at this stage.

2.11 Presentation of Results

An overview of results from the analysis of the collected data will be produced by augmenting the table from Section 2.4 with overall results per evaluation criterion. The results will then be explained in more detail per criterion with reference to the collected test data and, to the extent possible and relevant, accompanied by suggestions for system improvements. If relevant, results will also be discussed in relation to each of the three test user groups (Section 2.5).

Appendix 1, Sections 10.1 through 10.12 of the present report shows the results of the user interviews.

3 Sudoku User Test

The Sudoku system user test was conducted as planned over two days, 7 and 8 June 2007.

3.1 Contingencies

The following contingencies and aberrations from the plan in Section 2 should be noted:

1. One subject was planned for the first test day but came on the second day instead.
2. The average time for a full session, including user interview, turned out to be 55 minutes, which was longer than planned and too close to the 60 minutes scheduled for each user. This created some stress among the staff involved in the test.
3. In the before-gameplay instructions to subjects on how to speak and point, we opted for a solution different from the one described in Section 2.9. We simply did not show the subjects how to speak and point, as this was in the end felt to be the safer option for avoiding user priming.
4. For Subject 1, the interviewer forgot to ask questions 1-4 as closed questions.
5. For Subject 1, the interviewer forgot to ask the any-other-comments question number 27.
6. A bit more serious is the fact that the user population turned out to be less representative than planned, see Section 3.3.

Otherwise, the test proceeded as planned.

3.2 Interview Questions

The user interviews (Appendix 1, Section 10.13) included four Likert-scale questions (1 through 4). As an experiment, half of the subjects were asked these questions at the start of the interview whereas the other half were asked the questions after Question 16 in the interview script. We conclude that those four questions seemed hard to get across when asked at the start of the interview because they deal with "systems like the one you've just tried". At this point, just coming back from the system trial, people have a hard time abstracting from this particular system when trying to answer the closed questions. It would therefore seem preferable to ask those four questions after questions 5 through 16 have been asked. At this point, people have off-loaded their comments on the trial and are ready to think more abstractly about multimodal game-playing.

It also appears that several interview question pairs (5+6, 7+8, 9+10) may represent too fine-grained and research-oriented distinctions which are not shared by the subjects, so that several subjects tended to answer one when asked the other.

During the interviews and due to comments made by one subject in particular we realised that an additional question would have been productive. This was whether subjects found that they learnt anything during game-play that made them change their interaction style. It was too late to include the question at the time.

3.3 User Statistics

Potential test subjects were screened over the phone or face-to-face in order to compose a user group that met the requirements described in Section 2.5 and to some extent following the user screening instructions in Section 10.16.

Table 2 shows the actual composition of the subject population for the Sudoku system usability test. The table shows that regarding gender there is a fine balance with 50% of the subjects being males and 50% females. Regarding proficiency (at least according to the

subjects themselves) the balance is also as planned, i.e. one third beginners, one third medium experienced players, and one third experienced players.

Subject #	Age	Gender	Occupation/ education	Sudoku proficiency
1	24	male	medical student	beginner
2	23	male	medical student	beginner
3	60	male	lecturer in computer science	experienced
4	76	female	school teacher, retired	beginner
5	33	male	economist	medium
6	30	male	biomechanics/physical education student	experienced
7	23	female	mathematics/physical education student	experienced
8	23	female	mathematics/religion student	medium
9	31	female	PhD student in biology	medium
10	22	female	1 year science student	medium
11	+50	female	medicine	beginner
12	31	male	engineering student	experienced

Table 2. The test users.

The age distribution is less in accordance with our requirements. The average age is around 36 because there is an overrepresentation of young users in their early twenties and thirties. Only three users are over 50 and none are in their late thirties or forties. The unbalanced age distribution is probably related to another, more serious, inadequacy of representativeness in terms of occupation/education. No less than eight subjects are university students, including one PhD student, and most of them study natural science, medicine or engineering. And if we include the remaining four subjects, three of these are academics as well. In fact, the shortest education represented among the subjects is that of the schoolteacher. In summary, this is a very poor representation of a user population of Sudoku players that appears to comprise large numbers of people of a very wide range of educational backgrounds and professions.

The recruitment of subjects was handled by the person who normally does NISLab's subject recruitment following a screening script such as the one used in the present case, and does it well.

The required age division has to some extent been met. The problem is that the first age group (17 through 29 years) includes five users aged 22-24 and the second age group (30 through 49 years) includes four users aged 30-33 which is not a good spread within either age interval. The third age group (50+ years) includes three users with a much better age distribution (+50, 60 and 76, respectively).

The overrepresentation of young subjects may be ascribed to the fact that so many students were recruited. Actually, it was not the intention to recruit more than at most two students no matter what they are studying. However, our formulation "no more than two test persons with the same profession" was clearly not interpreted in the sense we intended. Furthermore, only three subjects came from outside the university. It turned out that the majority of subjects were found by announcing the Sudoku usability test in the daily university newsletter which is

distributed the university canteens and elsewhere and which is therefore being read almost uniquely by students and university staff. This was an easy way to recruit subjects but also a way which is likely to lead to the unrepresentative sample of subjects we had. The lessons learnt are that we should have supervised the recruitment process more closely than we actually did and that your subject requirements are not always as clear to others as they are to ourselves.

Game data / Subject No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Games started												
Total per player: * + ** below	**3**	**4**	**2**	**2**	**3**	**2**	**3**	**2**	**2**	**3**	**4**	**3**
Grand total: 33												
Games completed												
Easy	**xx**	**xx**		**x**						**x**	**x**	
Medium					**x**	**x**	**x**	**xx**		**xx**		**x**
Difficult						**x**						**x**
Total per player*	**2**	**2**	**0**	**1**	**1**	**2**	**1**	**2**	**0**	**3**	**1**	**2**
Grand total: 17 = 51,5%												
Uncompleted games: problems												
Game level too difficult: subject changes to easier level	**x**	**x**					**x**					
Game level too difficult: subject asked to select easier game			**x**									
Game problems: gives up and selects new game at same level					**x**				**x**		**xx**	
Game problems: resets game (same level)												**x**
Game reset by mistake					**x**							
Game stopped due to crash			**x**									
Total per player	**1**	**1**	**2**	**0**	**2**	**0**	**1**	**0**	**1**	**0**	**2**	**1**
Grand total: 11 = 33.3%												
Uncompleted games: time's-up												
Total per player	**0**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**0**
Grand total: 5 = 15.2%												
Uncompleted games: totals												
Total per player**	**1**	**2**	**2**	**1**	**2**	**0**	**2**	**0**	**2**	**0**	**3**	**1**
Grand total: 16 = 48.5%												

Table 3. Game statistics.

3.4 Game Statistics

Table 3 shows the user test game statistics based on test video analysis. The table shows in detail the causes of the 16 (48.5%) of the games that were not completed. Otherwise, these stats quite nicely match the user self-declared Sudoku game profiles in Table 2: the self-declared beginners stuck to the easiest game level, at least for a start. The self-declared medium-level players stuck to the medium game level. And the self-declared experienced gamers all tried to play at the hardest level but only two of them managed to complete a difficult game. The two others (Subjects 3 and 7) retreated to the medium level after some

struggle. In seven cases, subjects gave up because the system gave them problems. These problems are discussed in Section 4.

3.5 Data Validation

The test data collected was as planned, including observation notes from the subjects' gameplay with the system, 12 complete test videos recorded on webcam, and two sets of interview notes, made by the interviewer and the observer, respectively.

One test video has noisy background throughout but everything said by the subject can be heard.

In the test videos, it is sometimes hard to distinguish certain numbers on the recorded Sudoku game board. While this is not an obstacle to video analysis in other respects, it does make it difficult or impossible for the analyst to do one particular thing, i.e., to "play ahead" of the user and keep track of the non-elementary gameplay errors made by users. However, this is outside the scope of the present system usability evaluation and hence has no adverse effects on the evaluation presented in this report.

4 Sudoku Test Results

4.1 This Section

In this section we present and discuss results from the Sudoku user test. The discussion is structured as follows: identified technical issues are presented in Section 4.2. Section 4.3 discusses modality appropriateness. Actual test game play using the modalities available, i.e., speech, 3D pointing input and 2D static graphics output, is discussed in Section 4.4, information appropriateness in those modalities in Section 4.5, and the functional issues identified in Section 4.6. The Likert-scale questions from the user interviews are discussed in Section 4.7. Section 4.8 presents overall conclusions on the Sudoku usability test.

4.2 Technical Issues

Research prototypes are rarely technically perfect. Even in a user test report, it is important to list the main technical problems encountered during the test as these affect user performance and tend to "colour" their experience with the system as recorded in the post-trial interviews.

4.2.1 Robustness

The system behaved robustly in the test, with not a single crash during 5-6 hours of testing. Windows crashed once, thereby ending the test with Subject S3. The crash was probably due to overheating which was then countered by increasing ventilation of the machine.

The only other technical problem during the test was a microphone placement which hampered speech recognition of S8 until the microphone was adjusted after about 5 minutes of game-play.

4.2.2 Pointing precision

Pointing precision is the ease of placing the cursor on a screen object, such as a game board square, and keeping it there for as long as necessary for performing some action, such as inserting a number. For the Sudoku system, the requirements to pointing precision are determined by the facts that: (1) some users want to play fast and need a corresponding amount of cursor control; (2) most users get tired in their arm/hand if they play at length with their arm and index finger stretched out toward the screen, as a result of which they may lower the arm or the hand/finger may start shaking; (3) all users need an acceptable minimum of pointing precision and absence of cursor jitter in order to feel in control during game-play; and (4) the size of the screen objects pointed to.

Several subjects expressed difficulty with pointing precision in the post-test interviews, referring to the factors mentioned above (finger/hand shaking (S1, S11), arm lowered (S2), cursor jitter (S4, S5), or more generally (S6, S9), and a single user (S7) was observed to fail to insert a number in the intended square due to imprecise pointing. As for the size of the squares, S2 remarked that the squares should not be smaller than those used in the test.

In conclusion, given that the game is intended for public locations and users who are only expected to play for limited amounts of time, such as ½ hour, the pointing precision at present, given adequate camera calibration, would appear to be minimally acceptable for the general user. Both during our observation of the subjects playing and when playing ourselves, it seems that pointing can be done with reasonable precision and a negligible amount of jitter provided that the arm/hand/index finger is stretched out towards the screen. Also, there is always the possibility of increasing the size of the game board. This can be done quite easily while waiting for further improvement in pointing precision.

4.2.3 Number Ends Up in the Wrong Square

The number-in-wrong-square issue where a spoken number unintendedly ends up in a wrong square, probably is the most serious technical problem in the present version of the system.

A spoken number can end up in a wrong square in a variety of situations. The typical situation is one in which the user points to a square, makes sure that it highlights, and speaks a number which, however, fails to appear in the square. Most users then repeat the number one or more times – we observed up to 6-7 repetitions in several cases – while keeping the square highlighted. At some point the user gives up temporarily and moves the cursor out of the square in order to lower the arm, relax, and try again. In this phase, either before the arm/hand/finger is lowered sufficiently for the cursor to disappear from the screen or when the arm/hand/finger are raising to point to the square again, the cursor passes through some other square which highlights and shows the number spoken earlier. Now the user must remove this (in most cases) wrongly inserted number before getting back to the original square to retry to insert the number there. Even worse, the square into which the number was wrongly inserted was not empty but contained a previously inserted number which has now been overwritten. Sometimes the user could not remember the previously inserted number and had to try to work out what that number might have been in order to start getting back on track. Worse still, one user failed to notice the wrong insertion and suffered later on in the game. A variation of the scenario is when a pointed-to square fails to highlight while the user speaks the number that should be inserted in it. Since, we assume, the square is inactive but the speech recogniser remains active, the number spoken can end up in the wrong place as above. Another variation is when the cursor inadvertently errs into an adjacent square into which the spoken number gets inserted.

The causes of the problem remain somewhat obscure. One part clearly is the design decision to enable spoken number entry in any temporal relationship with pointing. If a number were enterable only as long as a square is highlighted and active, the problem could not arise in the first place. The number spoken would be erased from memory as soon as the square became inactive. However, the reason we classify this issue as a technical one rather than as a consequence of a design decision, is that there seems to be another problem in the implementation as well. If the failed insertion of a spoken number into a highlighted square is due simply to the fact that the number is not being recognised –it is not being misrecognised because no other number is gets inserted in the highlighted square – then this exact number would not end up in a different square later on. For that to happen, the number must have been correctly recognised and something else must be responsible for the failed insertion of the number in the highlighted square. One possibility is that, although the square pointed to actually does highlight, it fails to *activate* so that a number can be inserted in it. Another possibility is that there is a general problem in inter-module communication in the system.

The number-in-wrong-square problem happened to all subjects, often several times. None of the subjects seemed to find a way to avoid the problem, which is also hard to do: you might try to "sneak out" the cursor from the game board by moving it through a series of fixed numbers, avoiding all squares containing inserted numbers, but this is hard to control and isn't always possible; or you might fold the index finger and retract the hand/arm along an axis perpendicular to the screen so as to avoid any further cursor movement, but this is a complex, unnatural and not necessarily successful way to avoid the problem. The problem disrupts game-play and causes frustration, annoyance and incomprehension to all.

Since we don't know the exact cause of the problem it is difficult to propose a solution. One possible solution that would work has already been pointed out, i.e., to erase all received speech input from memory the moment the cursor leaves a square and the square de-highlights and de-activates. However, this would imply dropping the design decision to

enable any temporal combination of speech and pointing, and *this* would mean going against an observation often made in the literature on multimodal speech and pointing input, i.e., that users often speak before or after they point, thus curtailing users' natural speech-and-pointing behaviour. Moreover, as we shall see in the next section, the problem would still exist in cases in which the user speaks to a square which fails to highlight and activate.

A peculiar variation of the number-in-the-wrong-square problem was when the subject gave up inserting a spoken number into square S(n), lowered the arm, raised it again and pointed to S(n) anew, whereupon the number got inserted before the user managed to repeat it. Three subjects commented on this issue and we estimate that about 2/3s of the users experienced it at least once.

4.2.4 Issues involving the squares

Observation of the game board squares during game-play raises two issues, both of which seem to reflect simple bugs: (1) an empty square does not seem to activate (i.e., highlight) or it takes some struggle to activate it, for instance by moving the cursor back and forth over it. However, the same square may work fine later on in the game, so the problem always seemed to be a temporary one; and (2) a square seems to be highlighted without being pointed to.

Issue (1) occurred around a dozen times in the tests and always caused frustration because it was never possible to insert a number into a non-highlighted square. When you play Sudoku, it disrupts your game if you are unable to insert the number you are presently focused on inserting. In most cases, the subjects would continue to try to insert the number they had in mind for quite some time rather than moving on to other squares and numbers.

Issue (2) occurred quite frequently, one or several squares appearing to be permanently highlighted. This could happen both with squares which had not been pointed to during the game at all and with squares into which a number had been inserted by the subject. While this problem is only a minor one, it may cause uncertainty in the user with respect to which square the user is actually pointing to, given that two neighbouring squares are highlighted at the same time. No subject commented on the issue.

4.2.5 Speech Recognition

We mentioned that all subjects experienced to fail to get a spoken number inserted in the square pointed to even after repeating the number several times, and that indications are that this cannot always be an issue of speech recognition failure because the number spoken ended up in an unintended and normally wrong square, or sometimes even in the intended square, later on.

On the other hand, it did happen some times for most users, and many times for several users, that they were misrecognised by the system which would insert a number into a square which was different from the number actually spoken. For the users who had most difficulty being recognised, there was often a pattern to the misrecognitions, so that, for instance, S4, S6 and S7 had some difficulty getting "number three" understood whereas S9 had severe problems inserting a "4" which was nearly always recognised as a "5". In fact, the test notes show that all numbers except "9" posed problems for some subject. In addition, several subjects had difficulty removing numbers by uttering "delete this/that" and/or "remove this/that". In a couple of cases, the subject went so far as to reset the game when having failed to remove a misrecognised number. The worst case, though, was probably S9 who almost inevitably failed to insert "4" and ended up avoiding squares that needed a 4 in order not to loose yet another struggle with the system. This was the main reason why S9 failed to complete even a single game. On the other hand, several subjects who had few speech recognition problems concluded that the system's speech understanding was "perfect" (S3) or that "The system understood me fine" (S10).

As just illustrated, misrecognition, especially when repeated, is disruptive to game-playing. It must be kept in mind, though, that all subjects were native Danish speakers and none of them were native bilinguals. One cannot off-hand blame an English speech recogniser for performing less than perfectly when exposed to more or less strong accents. We did not do any detailed phonetic evaluation of the quality of the subjects' English but our impression is that the number of misrecognitions made per subject is largely in proportion to the strength of their Danish accent. Moreover, none of the subjects listed, neither directly nor indirectly, the system's speech recognition quality as a reason for not playing Sudoku with the system again. Even the unfortunate S9 who failed to complete any game mainly because of recognition problems, said afterwards that she just had to practice a more correct pronunciation (the experimenter demonstrated how the system would respond correctly to a properly pronounced "4" and she then repeated the pronunciation with success herself).

We conclude that the system recognises English reasonably well, even if the English is slightly accented, and that the presence of a more pronounced accent doesn't necessarily deter a user from continuing to play using the system.

4.2.6 Sudoku Generation Algorithm

The system uses a simple algorithm for generating a new game. The algorithm starts from a simple completed game and permutes the game board rows while preserving correctness. What the algorithm does not do, or at least does not do perfectly, is to check if the generated game actually does have a unique solution, which all Sudoku players are used to expecting. Thus, S5 ended up with multiple possible solutions and didn't seem to know what to do until the experimenter suggested that he simply select one of the possible solutions. Subject S6, the best Sudoku player among the subjects, ended up in a similar situation. He realised after a while what was the matter and then chose one of the possible solutions without experimenter interference. S6 confirmed this after the test and remarked that he had never met a Sudoku which did not have a unique solution.

The solution to this problem is to include a unique-solution test in the game generation algorithm.

4.3 Modality Appropriateness

In this section, we present an in-depth qualitative analysis of the test data in order to evaluate the appropriateness of the system's modalities and the information which can be exchanged by system and user in those modalities. So we ask, in effect: (i) are the modalities appropriate for this application? and (ii) is the information exchanged in those modalities appropriate?

4.3.1 Modality Appropriateness

This question concerns the appropriateness of using input speech, input 3D pointing gesture, and output graphics for the application. So the issues arising are whether:

1. existing modalities should be replaced by others; and/or
2. additional modalities are required.

Adapting an expression aimed at analysing modality appropriateness, an expanded version of these questions is whether:

> *Combined keyword-based speech input, 3D pointing gesture input, and various forms of static graphics output modalities is [useful or not useful] for:*

*[**generic task** GT and/or **speech act type** SA and/or **user group** UG and/or **use environment** WE and/or **generic system** GS and/or **performance parameter** PP and/or **learning parameter** LP and/or **cognitive property** CP]*

and/or are: [preferable or non-preferable] to: [alternative modalities AM1, AM2 and/or AM3 etc.]

and/or are: [useful on conditions] C1, C2 and/or C3 etc.

An earlier version of this expression was established in two studies of multimodal functionality claims made in the literature, i.e., claims about what particular modalities or modality combinations were good or less appropriate for [Bernsen 1997, Bernsen and Dybkjær 1999]. The present, revised version is in the process of being prepared for [Bernsen and Dybkjær, to appear]. The expression is an approximation to the complete set of parameters involved in evaluating modality appropriateness for a particular application. To exemplify the parameters, a *generic task* could be "inserting numbers"; a *speech act type* could be "commands"; a *user group* could be "all Sudoku players who can use 3D pointing and speech input, and graphics output"; a *use environment* could be "an airport"; a *generic* system could be "computer games using 3D gesture and speech input"; a performance parameter could be "fast"; a *learning parameter* could be "relaxed game-style"; and a *cognitive property* could be "difficult to remember". As we shall see, virtually all the parameters in the expression are actually being invoked in the evaluation of the Sudoku system.

4.3.2 Keyword-based Speech Input

The Sudoku system uses a particular sub-modality of input speech, i.e., keywords and key phrases which have been fixed by the developers. Any other spoken input will either not be recognised or will be misrecognised. This general approach suffers from various well-known problems and limitations [Bernsen et al. 1998]. Thus, in order to play the game at all, the users must first learn, or otherwise be informed about, which keywords are allowed. Secondly, the more keywords, the more difficulty users will have learning what to say to the system. It follows that spoken keyword-based technology is not suited for walk-up-and-use systems because such systems should be usable when a user walks up to them without receiving any introduction on what to do other than what the interface itself provides. But then again, the Sudoku system has not been claimed to be a "pure" walk-up-and-use system, it is primarily the intended use environment which seems to require something close to a walk-up-and-use system. In particular, it is not clear at this point how prospective users would be informed about the keywords that must be used during game-play. What is clear, however, is that the users must be told somehow because it would seem unlikely that they would be able to figure out on their own which keywords to use. The keywords might, for instance, simply be listed in small-font text on the main screen (Figure 1). In the user test, we simply told the users what the keywords were in order to let the test suggest how easy it was for the subjects to remember them.

Speech input, despite the recognition problems caused by Danish accents, was broadly regarded by the subjects as being useful for the Sudoku game and other, comparable, games, such as chess, in the intended use environment of public locations.

Regarding the use of keywords, the subjects managed quite well to stick to the vocabulary they had been told to use. The few exceptions were that: S5 forgot several times to say "number" before the integer; S10 tried once to say "erase this" to no effect; the disfluency in S11's "remove ... ahmm ... this" may have been due to difficulties in remembering what to say; and several subjects forgot at one time or other to add "this" or "that" to the "remove" or

"delete" commands. No user complained that the spoken keywords were difficult to remember despite the fact that each of the subjects had only been told them once before interacting with the system.

In comparative terms, one subject (S10) found that speaking was easier than using drag-and-drop-numbers selected from a palette on the Internet.

In the user test, input speech was captured by a headset microphone. Two subjects (S1, S10) pointed out that this was not an optimal solution because the headset is cumbersome to wear. In a public setting it would probably be advisable to use a lapel microphone instead as it is likely to be more acceptable to users.

It may be concluded that it is acceptable to use the small number of spoken keywords needed for the application.

4.3.3 Alternatives to Speech Input

For very different reasons, three subjects entertained ideas of removing speech input from the application or replacing speech with a different modality. Thus, S6 found it "funny", "daft", "strange" and "difficult to get used to" speak to a game machine. This seems to be a clear case of a "too exotic for me" user preference. As an alternative, and in order to make the game more physical and active than it already is, S6 suggested that users point to a square and then select the number to insert by jumping onto a numbered field in a palette on the floor. Similarly, S9 suggested to replace speech by pointing to a palette and then pointing to the game board. The suggestion was made in order to overcome the lack of control in making the spoken number end up in the intended square. S6 also pointed out that speaking whilst gaming in public locations might irritate other people. For these two latter problems, however, there are ways of avoiding to throw out the baby with the bathwater. Regarding speech in public places, the challenge is to let users play the game in an environment which people are not expecting to be quiet whilst avoiding crowd noise that may hamper speech recognition. The people interested in the game-play may perhaps be expected to give the players the reduced noise level needed.

4.3.4 3D Pointing Input

Most subjects found 3D pointing useful for playing Sudoku and comparable games in public locations. Seven subjects pointed out that 3D pointing with the arm/hand/finger fully stretched is only suitable for playing for limited periods of time. It is hard, and requires concentration, to keep the arm/hand/finger still when pointing to a particular square; it is hard to keep the arm up for a longish stretch of time; and it's a problem keeping the arm straight.

These points are obviously correct. However, several of the players displayed from the start a style of calm and controlled gestural play, with full control of the cursor. While the other subjects tended to keep the arm up and pointed in the direction of the screen for long periods of time, these players tended to put down the arm/hand/finger while scanning the game board for the next empty square to fill. The difference, we suggest, is that some subjects intuitively do what the others can *learn* to do.

In addition to adopting a more relaxed style of gaming, there is another thing one can do to make playing less strenuous, namely to change pointing arm from time to time. S6 found that it is good to use either arm for pointing. Several subjects began to do so at some point during game-play. However, one subject (S9) pointed out that changing arm for pointing is not necessarily a relaxing thing to do. People typically has a "leading eye" for aiming at something. If the arm is changed, it gets in the way for the "leading eye" (you don't change eye when you change arm), whereupon the body starts leaning to one side to avoid arm/hand

occlusion and actually see what is being pointed at. This explains why S9 tended to adopt a somewhat contorted position when pointing with her left arm.

4.3.5 Alternatives and Additions to 3D Pointing

3D pointing input and its combination with speech input are the most innovative aspects of the Sudoku game. This means that these aspects are liable to go against established user habits and preferences. Moreover, some subjects felt it tiring to stand still, arm outstretched for a long time, and there seems to be a technical problem in pointing and speaking. These factors led to several proposals for alternatives or additions to 3D pointing.

Referring to the physical exertion, S7 preferred a different sub-modality of pointing gesture, i.e., *2D* (touch screen) pointing, to the present game style which he found annoying rather than relaxing. Similarly, S3 said that gaming would be more stable with a touch screen and S6 mentioned the touch screen alternative as well. 2D (touch screen) pointing is a perfectly viable alternative for the Sudoku game, i.e., to have a large touch screen for playing Sudoku in public. Moreover, like 3D pointing, 2D touch screen pointing is done without the need for additional input devices. So this argument may, in fact, be the strongest argument against the idea of installing 3D pointing Sudoku games in public locations.

S1 suggested to augment the number of input modalities by adding a haptic modality to pointing by using "a pen instead of the finger so that one could click". It is not clear whether this suggestion is merely a reflection of the subject's habit of using a mouse for pointing-and-selecting or whether it also reflects that the user does not feel in full control of number insertion through 3D pointing and speech. Similarly, S3 said that the system might have featured double-clicking instead of just pointing. However, when asked why, he simply said that he misses the clicking, suggesting that he is merely unaccustomed to a graphics output domain into which one doesn't have to click to make things happen. However, echoing S1, S7 said that s/he prefers a touch screen – which may also use haptic code input - to standing with the arm in the air, saying "Did I click or not?". This at least suggests a feeling of not being in control. Also S9 would like to be able to click on something when there is a problem.

Yet another haptic pointing input modality was suggested by S4, i.e., to use a long pointing stick. It may be noted that S4 was particularly affected by the cursor jitter following the system crash. So, her suggestion clearly seems to reflect a genuine control problem which, however, is not characteristic of the system in its normal state.

Suggesting to replace 3D pointing by a stick may well seem exotic but it still does address problems with the system at hand. Some other modality alternatives proposed by subjects fall into the different category of being irrelevant for the application. Thus S5 pointed out that it is less physically demanding to use a pencil or mouse rather than 3D gesture, and S9 found that it is better to use the mouse for pointing because 3D pointing is annoying. While the criticisms of the system should be taken note of, and have been discussed above, the alternative suggestions are irrelevant because the present game is not being proposed as an alternative to, or even a replacement of, of traditional ways of playing Sudoku.

4.3.6 Keyword-based Speech + 3D Pointing Input

Although there are many issues concerning the details of spoken and 3D pointing interaction, none of the subjects directly questioned the *combination* of speech and 3D pointing input. The overall acceptance of the speech and 3D pointing modality combination is hardly surprising because it is an extremely useful one which humans use naturally employ all the time. Speech is notoriously bad at disambiguating spatial reference [Bernsen 1997] and attempts to do spatial reference through speech tends to be replete with disfluencies as shown by Oviatt in numerous papers. It is possible, of course, to say, e.g., "Row number five column number three insert number seven" when playing Sudoku, thereby avoiding the use of pointing

altogether, but pointing is just so much more convenient for supplying the spatial reference part of this statement, which is why we use speech and pointing together as frequently as we do. Emphasising the naturalness of the modality combination, S3 said that the combination of speech and pointing input is a great help for a particular user group, i.e., those who are unaccustomed to the keyboard.

When discussing the number-in-the-wrong-square problem, we saw that the problem might be (almost) completely solved if the system would only enter a spoken number into a square that is (pointed to and) highlighted at the same time. The cost of this solution, however, is to disable the player's opportunity to speak and point in any temporal order. Since we know from experience and from the scientific literature that humans sometimes speak and point "out of sync" even when speaking and pointing are complementary parts of the same communicative intent, this cost might seem to constitute an serious sacrifice of interaction naturalness. On the other hand, human communication is generally of such an order of complexity that we should not take for granted that reported findings of different temporal orderings of speech and pointing during interaction with computers can be generalised to all interactive tasks. So how did the subjects actually speak and point?

In fact, the test videos show that all subjects consistently played the game by pointing and speaking *at the same time* despite the fact that they were clearly told in the introduction that they could speak and point in any temporal order. Even more specifically, they would *first* make sure that they received the highlighting feedback from the square they pointed, *then* make sure that the cursor remained stably pointed at the square in question, and *only then* speak the number to be inserted. And *then*, while still keeping the square highlighted, they would verify that the number got inserted and that it was the right one (no misrecognition) before they would finally move the cursor elsewhere of lower the arm in order to relax. In other words, speaking was invariably "temporally encapsulated" inside a pointing gesture. S12 gave words to the approach saying that he preferred to point and speak at the same time.

4.3.7 Static Graphics Output

In modality theory terms [Bernsen 2002], the system's static output graphics consists of four displays, a main screen showing a game board and two labelled abstract images or icons (Figure 1) and three separate text display for choosing the level of difficulty of the next game (Figure 5), resetting the present game (Figure 6), and congratulating the successful player (Figure 7), respectively. These output graphics were generally found to be simple, clear and easy to understand. The screen size was judged to be fine by one subject with no remarks to the contrary by other subjects. S2, S11 and S12 said that the red error messages are good and only a single subject (S7) mentioned the possibility of removing the error messages. One user (S2) remarked that the size of the squares should not be smaller than what was used in the test.

4.3.8 Adding Acoustic Output

For subjects familiar with computer games, it may seem rather obvious to consider adding acoustic output – speech as well as non-speech sound - to the system for various purposes. The subjects proposed a number of additions mostly in order to improve the system's entertainment value – or decreasing it, depending on the subject's preferences.

S1 mentioned the possibility of adding non-speech sound output but did not specify. S5 suggested that sound warnings might be used instead of, or together with, the red colouring for when a surface mistake has been made. S6 suggested an "AAUGH!" message when this happens. S6 and S12 found that it would be fun to have a fanfare or a "YES!" when a game has been successfully completed. Reflecting user uncertainty during interaction, S6 suggested to have "little sounds" for signalling that the system has recognised a spoken input number.

Some had other preferences, though. Thus, S7 said that non-speech sounds can be irritating and that it is good that no sounds are output, e.g., in case of errors. Similarly, referring to the system's output in general which, for the moment, is graphics-only, S8 remarked that it is "good that there is nothing else. Sound is irritating when you have to think, like in the "Who wants to become a millionaire" TV show."

These comments suggest the issue over using redundant acoustic output in computer games and entertainment systems is a controversial one.

4.3.9 Adding Dynamic Graphics Output

The Sudoku game's main screen requires static graphics for representing the game board in order to provide the players with the freedom of perceptual inspection required for planning their next move [Bernsen 2002]. This does not exclude using dynamic graphics representations for other purposes, and one subject (S5) suggested that, in particular, team gaming competition would benefit from adding a timer to the system for showing for how long the user has been playing so far. If the timer were to run on-screen, this would be a case of dynamic graphics.

4.4 Gameplay Using Speech, 3D Pointing Input and 2D Static Graphics Output

In this section we look at the system's modalities as a whole and how the subjects viewed their appropriateness for playing Sudoku and similar games. The subjects had many different points and views on these issues. To keep focused, we start by noting that, in our view, the key issues are: (i) how suited is the system's modality combination for the system's primary purpose, i.e., to enable Sudoku gaming in public locations, and (ii) how subjects view the general suitability of the modality combination for enabling not only Sudoku gaming in this kind of environment but also other, comparable, kinds of board gaming, such as chess.

It is less interesting, we submit, to compare the system with playing Sudoku on the Internet or on paper because the system is *not* meant to replace, or compete with, these ways of playing the game. Unsurprisingly, the subjects, and especially the avid Sudoku gamers among them, had lots to say about the more or less obvious differences between the system and traditional ways of playing the game as well as about which set of modalities they prefer for playing Sudoku, but this is essentially comparing apples and carrots. We summarise the subjects' remarks on these issues below but the remarks remain tangential to a functional evaluation of the system on its own premises. Still, some comparative claims actually do reveal a subject's views on the prospects of the system, for instance when the subject grants that the system is as good as using pencil and paper. This is revealing because we take for granted that playing pencil-and-paper Sudoku is a well-tested and successful way of playing the game, as testified by its popularity among the millions of players across the world who play every day.

Despite the technical and functional issues discussed elsewhere in this paper, several subjects were quite happy with the way the system worked during their test sessions. S1 (not a regular Sudoku player) and S4 found playing "relatively easy" and S1 said that all the functionality needed was there: for deleting numbers etc. S2 appreciated the error correction functionality as well. S3 found game-playing easy and said that it was his own mistake if there were problems. It is easy to get a new game and easy to delete in case of mistakes (S7).

Let us first look at some points made about the suitability of the system's modality combination for gaming in public.

4.4.1 Playing in a Public Location

The subjects provided substantial information on gaming in airports, train stations, shops and other public locations. Four subjects said that they might use the system if they came across it in a public location and had time to spare, and S3 said that, although he wouldn't play Sudoku in this way, others might well want to. In fact, he felt that the combination of speech and pointing input (and, we assume, static graphics output) is good for chess and many other games. Similarly, S9 found that the system's input/output modality combination can be used for many different purposes. S9 pointed out that you tend to sit a lot during air travel and that it would be fine to stand up and point in an airport where you otherwise sit a lot. S5 said that if there are spectators, they can better follow the game on the screen than if it's being played on a newspaper page. Given the fact that three of the subjects were not really interested in playing Sudoku in any form, this data provides evidence that there might be considerable interest in using the system in public locations.

Several also found that the system has potential for social entertainment. S5 said that the system might be used for team competition in game arcades, or people could play together in, e.g., a family competition. S12 compared Sudoku gaming in public with popular games, such as dart or billiard, and wished to use the game for competing in a similar way. S1 felt that the system might be used as a party game, S2 that it would be better if two people could play against each other, and S10 found that the system was better suited for entertainment with several people present. Adding the numbers, eight subjects adopted the idea of using the system in public locations.

The subjects however also had various concerns with respect to playing the game in public. S1, not a Sudoku player, would not like to play Sudoku in public but would be happy to play Trivial Pursuit instead. S4 would not play in public because spectators might interfere with her game. While S5 had nothing against onlookers, he did not want them to take too much interest in his game-play. S12 noted the risk that people will talk over when the game is played in public. This is true and raises the issue of appropriate microphone setup for public gaming. One might perhaps assume that, e.g., use of a lapel mike might work sufficiently well for the spectators to provide the other half of the bargain and not speak too loudly during someone else's game in order to avoid that the microphone captures the background speech.

4.4.2 Comparisons

Here follow those of the subjects' comments which, although they compare the system with other ways of playing Sudoku, include general points of interest for public game-play.

Optimal combination. S1 found the system's interactive modalities an optimal combination for playing Sudoku.

As good as on paper. To S4, using the system is more or less as good as solving Sudokus in the newspaper.

Children. S5 remarked that children might prefer the tested game over pencil and paper. This point might be related to the following:

Gets the body active. Like S9, S6 said that it's real fine to involve the arm and the body. "It's good to get the body active." S8 found the game more physical and more immersive or engaging than playing in a book. "Adrenalin increases a bit when your body is involved in this way. This game is more fun than the Internet. Great fun." However, before we get too exited about the game's physical training aspects, it's worth noting the comment from S9 who argued that the game setup requires a more action-oriented game – it's more relevant for more movement-oriented games. Similarly, S6 found the technology well suited for more physically active games than Sudoku.

Larger screen than on the Internet. S5 found that, compared to playing on the Internet, it's good to have a larger screen "so that you don't need glasses".

In the following comments, subjects compare gameplay with the system to more traditional ways of playing Sudoku.

S2 found the system better than paper in that errors are shown immediately. S4 pointed out that the system, unlike paper Sudoku, does not require pencil and eraser because you can automatically delete a wrongly inserted number. Moreover, you never end up with paper completely filled with numbers and notes (which damages the overview necessary to complete the game successfully). S2 found gaming a bit clumsy and slow compared to paper gaming and S7 concurred with respect to the slowness. However, she found system gaming fun as entertainment, confirming the system's potential for public gaming. S2 also pointed out that touch screen pointing + number selection from palette would be faster. S3 insisted that Sudoku requires paper and pencil and doesn't like playing on a screen. S4 preferred to do crossword puzzles.

4.5 Information Appropriateness

In this section, we assume the modalities actually used for interacting with the system and ask whether the information exchanged with the users in those modalities is necessary and sufficient for playing Sudoku. In other words:

1. should information in any of the modalities be supplemented by additional information; or
2. should information actually provided in any of the modalities be removed?

It should be noted that the information in question includes both input information from users and system output. In the following, we make no distinction as to how strongly a subject insisted on some information issue. There is a continuum in the way subjects phrase their suggestions, of course, and such nuances of conviction are sometimes important. However, given the small and not very representative test subject population, it matters more to list the ideas the subjects came up with in order to evaluate their merits than scrutinising the individual subject's conviction in proposing a particular idea. Still, when an idea is brought forward by several subjects independently, this fact might warrant serious consideration.

Most of the subjects' information remarks concerned gaming support functionality.

4.5.1 Help and support during game-play

Undo, backtracking as a spoken command with static graphics text feedback. Sudoku players are familiar with the problems arising when, later in the game, you discover that an earlier number insertion has been wrong. At this point, it is often impossible to remember the sequence of steps taken after the wrong number insertion, which means that one's game has effectively been ruined. For this reason, and given the electronic nature of the tested game, four subjects (S3, S6, S7, S12) suggested adding an *undo* function which could help the user backtrack to the point at which the wrong number had been inserted, erasing the numbers inserted since then. This function, three of them added, should be executed solely through speech, without any pointing.

Error messages as static graphics text elicited by speech and/or pointing. *Undo* is just an example among many help and support functions which could be included in electronic Sudoku games. The present game provides one of these. It signals - by colouring red the relevant row, column, or 3x3 field – that an inserted number is in conflict with an already existing identical number in that row, column, or 3x3 field. Subjects were divided as to whether the system should provide additional help to identify wrongly inserted numbers. S5 said that it would be useful to be able to switch on and off a function which signalled the non-obvious mistakes (the system would use its knowledge of the correct solution to evaluate each

inserted number), and S11 wanted "more information on errors". However, the other subjects who addressed this type of function warned against it. Thus, S9, while happy with the red error messages, said that the system should not identify "deep" errors because that would make the game too easy, and S12 felt that even if a "deep" error function was optional, its very existence might tempt him and others to use it, destroying the real challenge of game-playing. Instead, S9 suggested a *help* function which would offer to insert the next number when a player got stuck. S6, on the contrary, warned against a *help* function because it is easy to be tempted to use the function (too much or too early). S7 even argued that, if the game had worked better it shouldn't even highlight surface errors (in red colour). She is probably referring to the annoying bug that makes spoken numbers end up in the wrong square. Her remark is supported by the fact that very few genuine surface errors were observed during 5-6 hours of game-play in the test. Several subjects did make "deep" errors, though, and these caused quite some grief later on in their respective games.

Game notes as static graphics text (numbers) inserted through pointing and speech. Sudoku players do not all play the game in the same way but tend to develop personal strategies for success. A common problem for most players is working memory load, especially when the game difficulty level goes up. When this happens, you find yourself doing all manner of calculations over which numbers might fit into a certain row, column, 3x3 field or block of 3x3 fields. Having done those calculations and concluded that none of the numbers can be placed right away, it becomes difficult to remember the results for use later in the game when more numbers have been inserted. For this reason, many players use *external memory* support, writing the calculated numbers into the squares as possibles, writing them in the margins for later use, etc., and often suffer later on when all these writeups start obscuring perception of the already inserted numbers on the game board, as well as when they have to erase or overwrite the numbers which didn't work or which have been inserted. For this reason, top players tend to use external memory sparingly, if at all. If you play on paper, you are free to invent any which external memory system you want, but also some Internet Sudoku sites provide some kind of external memory functionality.

The present game version doesn't offer any note-taking functionality, and it is therefore to be expected that some of the players wish to have such functionality added to the system, partly because they are used to having the functionality and partly because it actually does off-load working memory. S7 remarked that "Paper makes it possible to insert possible numbers into the fields. Here I need to keep everything in my head." S7, S9 and S12 were missing the opportunity to insert possible numbers in the squares, while S10 wanted both this opportunity and the option to insert numbers in the margins for playing games more difficult than those she played during the test. S11 needed the margin support for difficult gaming as well. S3, who strongly prefers paper-and-pencil gaming, mentions that he writes possible numbers in the squares and erases them later on.

From a technical point of view, both in-square notes and margin notes are easy to add to the system. From a usability point of view, however, the former may seem preferable because it would take a considerable amount of screen real-estate to impose some clear structure onto the margin notes which users might want to make. This is probably why we haven't seen any margin note functionality on Internet Sudoku sites whereas several sites feature in-square note functionality.

4.5.2 Other acoustic input information

The subjects had various wishes concerning the addition of spoken commands to the system.

S1 pointed out that the system's input command language vocabulary is quite limited and said that it would be more fun if more words could be used. S4 mentioned the possible addition of a *move* command for moving an inserted number somewhere else. S6 mentioned the obvious

point that it would be nice to have a version of the system with Danish speech recognition. Reflecting the system's long delay when loading a new game, S6 suggested to use a spoken command for choosing a new game instead of having to wait so long. This is a mistake because the loading time will not be positively affected by a spoken command.

4.6 Functional issues

In this section we discuss the functional issues identified in the user test. A functional issue is not a technical problem, it's a design solution or a combined result of design decisions made which turns out to be questionable for one reason or another. This does not mean that every functional issue has to be resolved through re-design and further development or technical maintenance, but it does mean that the developers should take a hard look at their system and ask if could be made better.

4.6.1 Slow Response Time

Even though one subject felt that the system reacts sufficiently fast on speech and pointing, with no disturbing delays, several others had remarks on delays in how the system responded to input. In a system like this, "real-time behaviour" is a fuzzy notion and you have to build in delays, so that's why we classify slow response time as a functional issue.

According to S3, the slow speed requires some getting-used-to. "One has to get used to the fact that the system takes some time to discover that the finger has moved". Even S6 (cf. above) said that pointing is a bit slow due to the time it takes for the system to respond. The system cannot follow when you have spotted a pattern and have 4-5 numbers to put in. S6 clearly attributes the delayed response to delays in pointing processing since he also proposes to accelerate *new game* loading time by speaking instead of gesturing. He is right that the pointing gesture has a built-in delay but wrong in believing that speech would speed up the loading time. S7 said that the system is very slow in reacting to pointing input, S11 that the system was slow sometimes. S10 diplomatically said that it "sometimes took longer than others" for the system to understand speech + pointing.

A user's experience that the system's response time is too slow either generally or for certain input actions, can be due to many different observations made during interaction. To mention some: (i) several subjects tried repeatedly to insert a number in a particular square without success; (ii) some squares did not highlight (almost) immediately; (iii) when they selected a new game, it took an estimated 10-20 seconds for the system to load; (iv) the subjects often spent longer than strictly necessary to make sure that the intended square was highlighted and that the cursor was stable inside the square before speaking a number; and (v) a fast-playing user may feel hampered by the fact that there has to be a certain minimal activation delay for any pointable screen object lest it becomes to easy for a user to accidentally activate the object just by passing the cursor over it.

Of these factors, (i) is related to the problem of numbers ending up in the wrong square, so this is not necessarily a speech recognition or pronunciation problem, and even if it were a problem of one of these two kinds, it's not a response time problem. (ii) is not a response time problem either. (iii), however, *is* a response time problem although a very particular one since it only occurs when loading a new game. (iv) is not a response time problem. Moreover, (iv) can be alleviated through training in using the system efficiently. For instance, S5 said that, in retrospect, he might have spoken earlier once he had highlighted a square, rather than re-checking that the highlighting stayed in place before speaking. Finally, (v) is a response time problem if a user feels it is. However, it is a necessary one and the only real issue is to calibrate the response time so that it becomes as small as possible without enabling users to activate anything by accident just by passing the cursor over it.

In conclusion, it is not clear to us that the system does have a real response time problem except for the game loading time, and the uncertainty effects of this one can be mitigated, cf. next section.

4.6.2 Uncertainty and Lack of Control During Interaction

Removing user uncertainty is an important goal of interaction design. Uncertainty about what to do at a certain point during interaction, whether to do anything at all, whether the system actually did get the latest input, etc., militates against the user's feeling of being in control of interaction and negatively affects user experience. Several features of the current version of the Sudoku system tend to create user uncertainty.

User background. Sometimes user uncertainty may be a function of the user's background. Thus, it is only to be expected that technology which users have never been exposed to before creates uncertainty in and of itself. In many cases, we believe, nothing can nor should be done about uncertainty in these cases. The technology is new and people are not yet used to it, and that's that. You don't put a horse in front of a functioning car even if some drivers miss a horse up front.

Missing the click. In this particular case, however, S1, S3, and S7 all "miss the click" and at least S7 comes close to expressing uncertainty. as S7 also asks: "What makes it choose – the pointing, speech, both?", which may be interpreted in context as yet another expression of uncertainty (rather than intellectual curiosity). S7 also asked if it is necessary to keep the hand pointed at a square for some time. S9 expresses the lack of control and its effects directly, saying that she would like to be able to click on something when there is a problem, adding that "one feels helpless, and this is something one is not used to". The situation S9 is referring to is when the spoken number has failed to be inserted in the intended square and the subject is out of control with respect to (a) what is going on, (b) whether it will be possible to insert the number in the next attempt, and (c) whether the number will end up in some other square. The helplessness she expresses serves to emphasise how important it is to remove this problem.

Loading a new game. S9 also says that she doesn't know what to do when waiting for a new game to be loaded, S7 became visibly impatient, and S11 became uncertain when nothing happened in this case. Their uncertainty in this situation may be compounded by another factor, i.e., that the subjects don't know – despite being told in the introduction to the system - if they should speak in addition to pointing when choosing a new game level (Figure 5) as well as when choosing a new game or resetting the current game. After all, they have to speak *and* point to insert a number, and it may not be obvious to them that those other functions *do not* require both pointing and speech and that the developers *therefore* have chosen pointing-only for activation – *except* for choosing to play a new game where speech and pointing are equivalent alternatives. Especially when loading a new game but also when resetting a game or choosing a new game, we observed several subjects trying to speak the contents of the labels that were visible at the same time. So they would say "New game" when pointing at the *new game* icon or "Yes" when confirming that they wish to proceed to load a new game and acknowledge that the game just played will be lost (Figure 6). The levels-of-difficulty screen (Figure 5) is less amenable for the user to simply read aloud the chosen game difficulty level because it lacks verbs, so we observed several users (e.g., S6) start talking free-style to the system or even mumbling comments and questions in Danish.

Inconsistencies. Is the problem just noted, i.e., not knowing when to speak in addition to pointing, an *inconsistency* in the interaction design of the system? It is probably to early to tell for all the cases because we first need a paradigm, or a set of guidelines, for what is *consistent* design of pointing and speech input before it is possible to judge. Intuitively, however, this is not inconsistent in so far as users are only expected to speak in addition to pointing when

speech-only or pointing-only cannot do the job. Moreover, the test videos indicate that subjects mostly did understand that because many of them would first start to speak to the *Proceed?* screen when nothing happened in response to their pointing. However, it clearly is inconsistent design to selectively enable alternative speech and pointing for choosing to play a new game but not for other similar commands.

Icons. Another potential cause of user uncertainty concerns the labelled icons *new game* and *reset game* in the top-left corner of the screen (Figure 1). There are two problems with them: (1) they are partly hidden behind the screen frame, making the labels only partly visible; and (2) when you move the cursor up there, you might easily hit the lower one (*reset*) before getting to the *new game* one. Issue (1) is a relatively simple matter of ensuring that the game display fits standard screens. (2) is more interesting. Thus, S5 mistakenly pointed at *reset game* instead of *new game* and got the game he had just played rather than a new game. And when observing the players, we noted that some went overboard trying not to make the cursor pass through *reset* on its way to *new game*. Neither the user nor we know if there is an activation delay when pointing at these icons (ensuring free passage for a cursor which just passes by the icon), but a user who has seen his numbers end up in the wrong place is likely to be vary of passing the cursor through *reset* on its way to *new game*. The same applies to the selection of game level (Figure 5). As S7 asked, what happens if, in trying to select a particular game level, the pointing hand (i.e., the cursor) passes over some other active field in the menu?

Remedies. Some modifications that might help the user uncertainties noted above are, first, to solve the number-in-the-wrong-square problem which seems to be the most important cause of user uncertainty during gameplay by far. Secondly, the game loading time issue should be solved by providing process feedback during the loading of a new game, effectively informing the user that a new game is being loaded: through an on-screen message, output speech, or otherwise. Thirdly, it might be considered to provide low-volume non-speech sound feedback on (i) successful square activation in addition to the current highlighting whose reliability as indicator of square activation (i.e., the readiness of a square to have a spoken number inserted into it) we are not sure about; and/or, as, in fact, suggested by S6, (ii) on the fact that the system has received and is trying to recognise spoken input. Whether or not one or both of these latter feedback types should be added might be made dependent on how successfully the number-in-the-wrong-square problem gets solved. Even if the problem gets solved, uncertainty will remain in some cases, though, such as when a user fails to be recognised in many attempts in succession when trying to input a spoken number. Eventually, if the problem cannot be removed, an error message may have to be included, such as the system saying or displaying "I'm afraid that I have difficulty recognising you". Fourthly, it must be ensured that activation-through-pointing of squares, icons, and text fields can only be done when the cursor has pointed at any of these for a certain amount of time, so that subjects will not accidentally activate anything in the graphics output domain.

We don't know what S6 was referring to when remarking that the screen (graphics) quality "could be improved" but there are several candidates. One is the *reset game* and *new game* icons, cf. above. Another, the "falsely highlighted" squares. A third candidate is the levels of difficulty screen which is fuzzy and blurred in parts (Figure 5).

4.6.3 Display functionality

The fixed numbers on the game board are smaller than the ones inserted by the user, which, in addition, appear on a slightly lighter background than the fixed numbers (Figure 1). We assumed that all subjects would notice. S8 found the size difference good because it is useful when one has to remove a number. However, several subjects (S4, S5, S11) failed to notice

the differences and tried to remove fixed numbers when in trouble. Moreover, it seems clear that at least S11 did not discover that this was impossible to do.

The evidence above suggests that *either* not all subjects discovered the differences between fixed and inserted numbers *or* that they failed to get the meaning of the differences. We suggest making the difference between fixed and inserted numbers so conspicuous that they will be immediately perceived by virtually all first-time players. In this particular case, we believe that almost every user will manage to interpret, e.g., a conspicuous difference in size or colour between fixed and inserted numbers correctly because one would probably have to search extensively to find a Sudoku user who is not familiar with the distinction between fixed and inserted numbers: the former *define* the current game and you just don't try to remove them when in trouble! After all, nearly all game-players know that one is not allowed to change the rules of the game when in trouble.

It is probably for reasons such as those just mentioned - i.e., if you perceive a clear visual difference you will understand its pragmatic meaning – that we expected all users to understand the pragmatic meaning of the clearly visible red coloured rows, columns, and 3x3 fields that appeared when a surface error was made. To our surprise, one subject (S11) did not get the meaning of the red colour even though she clearly perceived its sudden appearance. "[I was uncertain] when part of the board turned red. I postponed the red problem for later. [I] didn't understand the red in the beginning", she said. This observation could be a reason for introducing spoken output so that the system would say, for instance, "Uh-oh, you made a simple mistake." Alternatively, we might, in this particular case and even if S11 were representative of a relatively large number of users out there, which we don't know at present, leave it to the users' natural intelligence to discover the pragmatic meaning of the red colouring.

4.6.4 The Language Issue

We have seen that S6 requested a Danish version of the system. The only subject who had a problem understanding the on-screen text was S4 who didn't understand the word "proceed" (Figure 6) and was helped by the experimenter. A Danish system version would also remove many of the speech misrecognition problems caused by subjects' Danish accents. Clearly, if the system is for public use it should include dedicated speech recognition for the country in which it is being installed. A system for an airport might use English or, even better, a choice of the planet's major languages, in which case a user would have to start by selecting the language of interaction.

4.6.5 Learning and Walk-Up-And-Use

How close is the system to being walk-up-and-use, so that it is possible to play Sudoku with it without any instruction on how to use it? The Sudoku game system is not a walk-up-and-use system because it does not teach, or even list, the simple rules of Sudoku but assumes that users know them already.

Is the system walk-up-and-use for people who are familiar with the Sudoku rules? Clearly not, but before we look into the reasons, it may be noted that there probably never was a walk-up-and-use system which used new and unfamiliar technologies, such as 3D pointing and, to a somewhat lesser extent by now, input speech. The present system uses both of these, to many users, new and unfamiliar technologies. In 10 years, the situation will probably be very different but, for now, the subjects clearly need instruction in how to play: how to stand at a pre-defined distance from the screen; to point and how to point, i.e., arm/hand/index finger stretched; the language to use (English); what to say, i.e., that specific keywords must be use and that no other words should be used; to use a microphone and how to mount it. Since the users were stationary during game-play, they didn't have to be told not to walk too much

around lest the cameras would follow them and try to match anything they came near to a stretched index finger, following which it might lock onto an object somewhere and – worst case - have to be re-started. This may seem like a lot of instruction but, arguably, most of it is likely to become commonplace when this kind of system begin to be used in public locations, so that eventually users only need to be informed about the keywords to use. And if these are displayed on the screen, the system will have become walk-up-and-use.

Walk-up-and-use, however, does not imply that users don't have to learn to become good at using a system. This need for training-during-use may be a nuisance in utility systems but may actually be an asset for an entertainment system. The test subjects, none of whom had used combined speech and 3D gesture before, began to learn how to use the system well when they began to play. This process is not just one of discovering exactly how to point and speak but may also involve abandoning, or modifying, preconceptions about what computers can and cannot do or understand, habits from using GUIs (standard Graphical User Interfaces) and from playing Sudoku on paper or on the Internet, etc.

Let us look at the evidence collected on the subjects' continued learning process during game-play. As S6 remarked, playing Sudoku in this way took some getting-used-to.

Speaking to the system. S4 found that it was fine to talk to the system once you got into it. Since subjects were not told *how* to speak to the system, they had to learn by themselves starting from whichever assumptions they might have had initially about how one speaks to a machine. Thus, S5 learned during the game that it wasn't necessary to speak so loudly. It would appear that this subject initially believed that machines need "special treatment" in terms of speaking differently compared to speaking to humans. S5 then discovered that this was not necessary and, presumably, relaxed more when speaking to the system from then on. S9 observed that "One has to learn to speak in the right way to the system". After having spoken to herself in Danish once and, as a result, turned a correct "7" into a "4" and getting red colour as well, S11 (mostly) stopped speaking to herself during game-play.

Pointing. S8 noted that it takes a bit of time to learn that there is some latency time when pointing. At some early point in the game, S5 starts using both arms interchangeably for pointing. S6 noted that it takes some getting-used-to to lower the arm when it's not needed. S8 would have liked to be able to both point and click on the numbers but found that the designed way to play the game is OK when you get used to it.

Speaking and pointing. S5 believed, in retrospect, that he might have spoken earlier once he had highlighted a square, rather than re-checking that the highlighting stayed in place before speaking.

Graphics output. By the end of the game, S11 may or may not have learned what the red error colouring means, we don't quite know.

Gaming as a whole. At the end of the session, S8 felt that she was in control "after having gotten used to it". S9 felt less of a routine user by then, saying that speech + pointing "would work better if I became a routine user". S11 "realised very soon how it worked". S12 said that it was simple enough when you first understood the system. He had to get used to the absence of mouse or pencil. After that it was OK.

4.7 User Interviews: Closed Questions Overview

The 4 Likert-scale questions in the user interviews concerned not only the specific Sudoku game which the subject had just tried, but were asked in a more general way to cover the appropriateness of speech and pointing gesture input and graphics output in similar games they might be able to think of.

Subjects were asked to answer the questions on a scale from 1 to 5 with 1 = unsuited, 2 = rather unsuited, 3 = neither/nor, 4 = rather suited, and 5 = well suited. We forgot to ask the first subject the questions in this way. The answers from this subject are therefore not included in Table 4.

In half of the cases the four questions were asked as the first four questions in the interview while in the other half of the cases (grey rows in Table 4), subjects were only asked these questions after interview question number 16. As remarked in Section 3.2, the Likert-scale questions seemed hard to get across when asked at the start of the interview because they deal with "systems *like* the one you have just tried". At this stage, just coming back from the system trial, people seem to have a hard time abstracting from this particular system when trying to answer the closed questions. It seems better to ask such questions after questions 5 through 16 have been asked, when subjects have off-loaded their comments on the trial and are ready to think more abstractly about multimodal game-play.

Subject #	**Pointing input**	**Spoken input**	**Screen output**	**Combination of the three**
1	-	-	-	-
2	4	3 (slow)	5	3
3	1 (4-5 for chess)	4	4	1 (4-5 for chess)
4	4 (stick instead)	4	5	4
5	4	3	5	3
6	4	4	4-5	5 (3 in concrete game)
7	4 (in public locations)	2 (funny)	5	3 (speech is funny)
8	3	4	4 (useful that inserted numbers look different)	4
9	2-3 (imprecise, annoying with outstretched arm, mouse better)	4 (when it works)	4	3 (4 if more action-oriented)
10	4 (missing note function)	3	5	4
11	4 (requires holding hand straight)	4	3 (sometimes slow)	4 (missing note function)
12	4	4-5 (for Sudoku, 3-4 for chess)	5	4 (2-3 in noisy environments)
Average (first)	3.1	3.4	4.2	2.8
Average (later)	3.83	3.75	4.75	4.0
Average (total)	3.5	3.6	4.5	3.45

Table 4. Appropriateness of the Sudoku game modalities. Grey indicates subjects who were asked these questions after some other questions relating to the concrete system they tried. All other subjects were asked the above questions at interview start (white background).

Table 4 shows that those subjects who were asked the modality appropriateness questions first on averaged scored all four questions lower than subjects who were asked the questions later on in the interview. It is possible that letting users talk about the experience with the concrete game first influenced the way they answered the more general evaluation questions.

4.8 Conclusions

Possibly the main conclusion to be drawn from the test is that computers that can see have become ready for board gaming. The system's main technical problem, i.e., the number-in-the-wrong-place issue, can be solved by abandoning the goal of enabling pointing in all temporal combinations with speech (before, simultaneously with, after). The test evidence strongly suggests that users are not likely to be missing the options of pointing before or after speaking because none of the test subjects used these options. Some other technical problems were identified in the test but these can all be fixed without having to consider system re-design. The basic technologies involved for ensuring pointing precision and correct speech recognition are at least minimally acceptable for game-play.

Functionally, the test suggests that the system uses an appropriate modality combination for its purpose, something which became even clearer during the discussion which showed that the nearest competitor the system might have is one which replaces 3D pointing by 2D touch screen pointing but otherwise preserves the advantages of the tested system. That is, the advantages of being based on the natural human communication ability of speaking and pointing, and of not requiring any haptic input devices. The suggested replacement makes sense because the Sudoku system does not really *require* capturing of 3D arm/hand action since all that the cameras have to process is a pointing gesture that produces 2D screen coordinates. It is only when pointing gestures become replaced by other 3D user gestures and, even more generally, arm/hand actions, such as grasping a chess piece and putting it down on the game board, and innumerable others – that the 3D gesture/action technology moves beyond what can be done with standard 2D touch screen gesture.

The tested system also sits on another fence, i.e., its physical all right but does not really qualify as a *physical game*. In this, however, stand-up Sudoku gaming resembles many popular games, such as dart or billiard, so there does not seem to be any obvious reason why gaming with the system should be made more physically demanding than it already is. Moreover, more than half of the subjects said that they might use the system if they came across it in a public location and had time to spare. These subjects were all Sudoku players as opposed to the three subjects who weren't really interested in playing this game, and there were only one or two of the Sudoku players among the subjects who did not envision using the system if they were to come across it. Arguably, there is a good chance that a more representative user test than the present one might produce a similar pattern, i.e., that most Sudoku players would like to use the system for playing this game. On the other hand, it should also be noted that the system did not manage to change the minds of those who were not already Sudoku enthusiasts.

As for the instruction and learning requirements of the system, the subjects were given minimal instruction in how to play and, despite its unfamiliar interaction technologies, playing and improving game-play turned out to be well within the abilities of users, such as those selected for the test. The only exception seems to be the English language which the native Danish speakers had to use in the test. A potentially important qualification should be noted, however, and that is the lack of representativeness of the test user population as discussed in Section 3.3.

5 Treasure Hunt User Test Plan

5.1 This Section

This section describes the system to be tested and presents the test plan for the Treasure Hunting Game usability test based on the test plan document by Bernsen and Dybkjær of 20 June 2007.

5.2 Description of the Treasure Hunt Game System

The Treasure Hunting Game system has been developed at ITI-CERTH in Greece. It is a game for two people with special needs for access to information and communication technologies, i.e., a blind person and a deaf and mute person.

The system requires two ordinary PCs, one for the blind and one for the deaf-mute. The blind interacts via haptics (input and output) and also receives audio output as (i) speech (in English) and (ii) musical sounds acting as codes for different colours. The haptic device used is a PHANToM^{TM} force feedback device. Audio messages are provided via loudspeakers. The deaf-mute has a GUI interface for graphics output and mouse-drawing input (the keyboard is not used) and provides sign language input via a webcam as well. The system recognises five different words in the Greek sign language and displays an animated agent that provides sign language output. Audio messages are converted to sign language for the deaf-mute.

The overall story-line in the Treasure Hunting Game is that the users are inhabitants of an ancient Greek city that is under attack and that the users must try to locate some hidden designs that will enable them to make high-technology war machines in order to defend the city. The city and its environment is shown on the map in Figure 8.

Figure 8. Map of the city and its environment with a path to the treasure area sketched by the deaf-mute.

Completion of the Treasure Hunting Game involves seven steps, some of which must be performed by the blind and some by the deaf-mute [Moustakas et al. 2006].

Figure 9. The red closet.

In summary, ping-pong fashion, the game steps are as follows:

1. The blind user is told to find a red closet (Figure 9). Using the haptic device, the user must search for the closet by first finding houses in the city which can be entered into, and then those houses one by one in search of a red closet. Houses and a small number of their properties are signalled by a voice whispering, e.g., "house" or "wall" when the user gets force-feedback from those properties. Colours are signalled by musical sounds, so a blue closet, for instance, emits a different sound on impact than does a red closet. The contents of the message in the red closet, when found and successfully clicked upon, is sent to the deaf-mute.
2. The deaf-mute user receives the message as converted to Greek sign language displayed by an animated on-screen agent. The message is sent to the blind user.
3. The blind user is told to go to the city's town hall where the mayor tells the user to go to the temple ruins (Figure 8). The user haptically searches for the ruins supported by whispers, goes there and must then find an inscription on a broken column. The message is sent to the deaf-mute who must decode the enigmatic inscription to find the instruction hidden in it.
4. The deaf-mute user goes to the cemetery (Figure 8) and must find a key on a grave. The user reads the location inscribed on the key and performs sign language to tell the blind to go to the catacombs (Figure 8).
5. The blind receives the sign language message converted to speech and must haptically – still also supported by whispers - search for the catacombs, enter them, find a box containing a map, and get the map from the box by clicking on the box while in haptic contact with the box. The map is sent to the deaf-mute user.
6. The deaf-mute user must resolve the riddle on the map into an instruction as to where the blind user should go, and then draw a route on the map towards the location, as illustrated by the white path in Figure 8. When sent to the blind user, the route line is converted into a 3D groove on a map, i.e., *not* a groove in the actual landscape depicted in Figure 8.
7. The blind user must haptically follow the groove, by-passing any obstacles that it might contain, such as large stones, and enter the forest area (Figure 8) where a new groove must be followed (Figure 10) until the treasure is found.

Whenever the blind user has successfully accomplished a step, the system emits a beep which is meant to inform the user that now it is the partner's turn to advance the game. Another beep tells the blind user that now it's this user's turn to take the next step in the game.

Figure 10. Grooved line map of the forest area visible in the top part of Figure 8.

5.3 Overall Evaluation Goals

The evaluation of the Treasure Game system has two overall goals:

4. To explore game usability and provide input on usability aspects of the game, in particular regarding appropriate use of modalities, offered functionality, ease of use, and user satisfaction for the blind.
5. To provide input on how well the relevant parts in [Bernsen and Dybkjær, in press] work with respect to practical evaluation with users as well as how to plan, carry out and analyse test results.

The second goal is addressed in SIMILAR Deliverable D100 *Multimodal Usability Progress Report*. The present document thus only concerns the usability evaluation of the Treasure Game system. With respect to the first goal, focus will be on testing the part of the game which is for blind users whereas the part for the deaf and mute users will not receive any particular attention in the present test since that part remains relatively limited and needs to be extended to enable a balanced amount of gameplay for both players. The evaluation will take place with blind users at the Institute for the Blind in Copenhagen, Denmark, following the present evaluation protocol. One of the testers will act the part of a deaf and mute person.

5.4 Evaluation Criteria

Table 5 gives a brief overview of what to measure and how, based on data from the usability test. All table items will be addressed in post-test interviews and 12 items (numbered 1 through 8, and 10 through 13) will be topics for test data collection and analysis as well. The phrasing of interview script and questions in English and Danish is described in the Interview Scripts in Appendix 2, Sections 11.8 and 11.9.

What to measure	**How to measure**
Quality of interaction	
1. Navigation in city and landscape	Interview question + data from the interaction
2. Use of haptic device	Interview questions + data from the interaction
3. Colour recognition via sound	Interview question + data from the interaction

4. System output understanding	Interview questions + data from the interaction
5. Ease of individual tasks	Interview question + data from the interaction
6. Ease of achieving game goal	Interview question + data from the interaction
7. Ease of following path	Interview question + data from the interaction
8. Quality of partner communication	Interview questions + data from the interaction
9. Missing modalities?	Interview question
10. Ease of interaction	Interview questions + data from the interaction
11. User in control	Interview question + data from the interaction
12. Learning	Interview question + data from the interaction
Functionality	
13. Sufficiency of functionality	Interview questions + data from the interaction
User experience	
14. Tried something similar before	Interview question
15. User satisfaction	Interview question
16. Advantages and disadvantages	Interview question
17. Play again?	Interview question
18. Other comments	Interview question

Table 5. What and how to measure in the test of the Treasure Hunting Game system.

5.5 Test Users and Their Profiles

The Treasure Hunting Game is an entertainment game which integrates a number of novel interactive technologies and modalities that are presumed to be particularly suited to blind and deaf-mute users. The goal is to enable the players to collaborate in finding the treasure. Since we are only focusing on the part of the game meant for the blind, our subjects must all be blind or strongly visually impaired. Furthermore, they must have some English skills since part of the output consists of audio messages presented in English, and they must be capable of using the haptic device that comes with the game. We have therefore decided not to include smaller children in the user population. Furthermore, to avoid issues related to obtaining permission from parents we have also decided not to include any user younger than 18 years old. Given the highly innovative nature of the technology and the many open issues it raises, we consider the present test a very first exploration to be carried out with a limited and not necessarily representative user population.

Test users will be recruited by our contact person at the Institute for the Blind who will find at least six persons who are willing to participate. The above requirements are the only ones we have given to the contact person. Thus, we have not imposed particular requirements to gender balance, age distribution, or educational background. Since the population from which we can recruit users is limited to the blind and visually impaired who regularly visit the Institute for the Blind, we have found it more important to focus on recruiting users who fulfil the above requirements and who have an interest in trying the game, rather than trying to achieve an imagined balance within a very small user population at a stage where it is impossible to predict who might be the users who would actually take an interest in playing the game in real life.

5.6 Test Design

The usability test is scheduled to take place at the Institute for the Blind on Wednesday 20 June and Thursday 21 June, 2007, between 9 am and 3 pm. A room with tables and chairs will be available at the Institute. We will bring all equipment needed.

Given the highly exploratory nature of this test, it is hard to plan with confidence in any detail. In fact, nearly everything in the test is of an exploratory nature: how the users will handle the technology, the questions to ask them in the post-trial interviews, the instructions to give each subject prior to the system trial, whether and how much to support them during the test, etc.

One hour has been set aside per session. This should leave sufficient time for each test so that the next user will not have to wait. How many users we will have on each of the two days depends on the users recruited and when they have time to participate.

A user session is expected to involve max. 20 minutes for introduction to the game and training in recognising colours by audio and using the haptic device, followed by 15-20 minutes gameplay with the system, followed by a 15 minutes post-test interview. The interview will be based on the script in Appendix 2, Section 11.8.

Each user will receive two cinema tickets as a reward for having participated.

At least six test users will be recruited by our contact person at the Institute for the Blind. Users will be recruited in accordance with the criteria described in Section 5.5.

Each test user will be told that s/he is going to help us evaluate a treasure hunting game which can be played by a blind and a deaf-mute. S/he will be told that the entire session will last about an hour, including some 20 minutes for introduction and training, 15-20 minutes for playing a single game, and about 15 minutes for an oral interview immediately after the session with the system. The person will also be told that there will be a remuneration in the form of two open cinema tickets. No transportation or other costs can be covered since the users are expected to participate as part of their regular visits to the Institute for the Blind. Our contact person will agree on the exact time for each session with the users he contacts. The sessions will be on the dates and during the time intervals mentioned above.

5.7 Roles

During the usability test of the Treasure Game, we need people for the following roles:

1. someone who receives and takes care of the users when the users are not in the test room;
2. an experimenter who has the contact with the user during the session;
3. a person who trains the user in recognising colours by sound and using the haptic device;
4. a technician who ensures that the system is up and running, including any logging software;
5. a person who plays the part of the game meant for a deaf-mute;
6. a person who takes care of the video camera for recording test users during the sessions;
7. an observer;
8. an interviewer.

Roles 3, 4, 5 will be taken care of by Kostas, and the remaining roles (1, 2, 6, 7, 8) will be shared by Ole and Laila.

5.8 Test Environment and Equipment

The test sessions will take place in a room made available at the Institute for the Blind in Copenhagen. All equipment will be brought there on the days for the test. It will be tested that

the system is up and running before each session starts. Two days before the test the equipment will be set up and run in the test room to ensure that nothing is missing and that everything works as intended.

The system runs on two laptops connected to each other via a local net and sharing a CSCW workspace displaying the treasure hunt map. The laptop for the blind user will have two loudspeakers and the haptic device connected, whereas the laptop for the deaf-mute will have a webcam camera attached.

A webcam camera will be used for recording the interaction (video and audio). Prior to each session it will be checked that it works properly.

The interview will also take place in the test room.

5.9 Tasks and Test Conditions

Each blind user will first be given an introduction to the system and receive training in interacting with it. This involves to (in the following order):

- tell that the subject is about to use a research prototype system which demonstrates a new way of playing computer games for the blind and deaf-mute;
- tell what the game is about, cf. the general storyline in Section 5.2; tell that there is a deaf-mute game partner at another computer, this part being acted by Kostas; the two users will take turns hunting for the treasure; the blind user starts the hunt; when the blind user has found an item crucial to the hunt, the user clicks on the haptic device and receives a beep in response; this means that (i) this step in the hunt has been completed, (ii) a message about the result has been sent to the partner, and (iii) the next task will be done by the deaf-mute partner; the blind user is not told about the problem to be solved by the partner, but will get the result from the partner in due course; while the partner works on the problem, the blind user must wait for the result to arrive; this procedure is repeated for each task or step in the hunt;
- explain and demonstrate how to use the haptic device;
- explain, demonstrate and train the use of audio for colour identification;
- explain how to start the game;
- clearly emphasize that this is not a test of the user's skills at all but a test of how good and how interesting the treasure hunting game might be.

Exceptions: should a user go "cold" during gameplay, the experimenter should provide help to continue. We are not likely to get sufficiently useful data unless there is a reasonable amount of progress with the game and our goal is to collect data on completed games for all users. It must of course be noticed when and which help was given since that forms part of the test data.

5.10 Data Collection

The data to be collected includes:

- video and audio recordings of user interaction with the system. The video will show the user's hand/arm, the haptic device, and the screen contents, and thus will be taken from a position slightly to the left/ right of, and slightly behind, the user;
- observation notes produced by the observer during the sessions;
- interview notes written during the interviews with the users.

When user tests and data collection have been completed, the data will be validated to make sure that the data is, in fact, appropriate for the various kinds of data analysis planned. A detailed plan for data markup and coding scheme creation will be specified at this stage.

5.11 Presentation of Results

An overview of results from the analysis of the collected data will be produced by augmenting the table from Section 5.4 with overall results per evaluation criterion. The results will then be explained in more detail per criterion with reference to the collected test data and, to the extent possible and relevant, accompanied by suggestions for system improvements.

6 Treasure Hunt User Test

The Treasure Hunt Game system user test was conducted as planned over two days, 20 and 21 June 2007.

Throughout the test, Ole acted as experimenter, Kostas demonstrated the devices and acted as deaf-mute partner, and Laila conducted the interviews whilst Ole took interview notes as well. During the test, observation notes were made by Ole and Laila.

Interview results and observation notes are shown in Appendix 2, Sections 11.1 through 11.6.

Subjects 1 and 2 interacted with the system and were interviewed on 20 June 2007 while subjects 3-6 interacted with the system and were interviewed on 21 June 2007. While waiting for a third subject on 20 June, , we gave a demonstration of the game to a seeing person who works with the blind on assistive technologies. The observation notes on this person's gameplay are shown in Appendix 2, Section 11.7.

Figure 11 shows a snapshot from the user test.

Figure 11. Snapshot from the user test showing a subject, the haptic device, the laptop screen in front of the user, the video camera, and the laptop used for audio/video recording.

6.1 Contingencies

As this was a first exploratory user test of a highly innovative system for a special user group, contingencies were to be expected. None of us had clear notions of how much instruction the subjects would need during gameplay nor about what the issues would be, nor did we have any clear idea about how the subjects would respond during the post-test interviews. Even bearing all that in mind, we were surprised by the tests.

The following contingencies and aberrations from the plan should be noted.

One subject never turned up but we managed to find another. One subject only had 45 minutes to spare for the test.

Two separate cameras were used to record the user trials. We had only planned to use a single camera but since we ended up having three available, we decided to use two of them. The camera we originally planned to use was not used because the other two cameras were easier to mount in the right positions - left and right of the user and slightly behind the user - given the environment.

During the test, Kostas handled one camera while Laila handled the other, and our contact person at the Institute for the Blind took care of new subjects as they arrived. Otherwise, Laila and Ole had plenty to do with the current user throughout.

However, the main contingency was a general and serious lack of time during the test sessions.

The planned 60 minutes per user session turned out to be insufficient for two reasons: (i) most users needed far more instruction during gameplay than anticipated, and (ii) the post-test interviews were generally richer than anticipated. 90 minutes per user session would have been preferable. Several users had to leave at the scheduled time, forcing us to either increase gaming instructions even more in order for the user to complete the game, hurry up in the post-test interview, or both.

The interaction with the system was audio/video recorded. However, we regret that we didn't also record the introduction given to subjects before their game-play. This would have allowed us a more detailed analysis of how much information was actually given to each subject and in which way the information differed across subjects and changed from one session to the next. As it was, we kept discovering new and unfamiliar interaction issues with each new user, in response to which we continuously revised the pre-test instructions as well as the game-play instructions.

The first subject probably received the least information, in particular during gameplay. However, it soon became clear that the information provided by the system to the blind user on what to do during gameplay is generally highly insufficient. *We* can see the city and its environment in 3D on the screen and can easily correlate what we see with the tactile operations done with the robot arm, and we are also familiar with the game's story-line and purpose. The blind subject knows little of these things and often didn't have a clue about what to do without getting supplementary information from the experimenter, for instance about where to search for the red closet in the first place, how to enter a house, or how and when to click on other objects to find out if they would respond in some way relevant to the subject's quest. Moreover, we gradually realised that the haptic orientation/action metaphor by which the subjects were introduced to the haptic device was very likely seriously flawed in a way which hampered game performance with the device. The result was that increasing amounts of information crept into the introduction to the system and into commentaries on the subjects' gameplay without waiting for the subjects to more or less give up and ask, as was rather the case with the first subject. Moreover, the contents of the information gradually changed as well.

Subject 5 probably received the most help. We only had 45 minutes with this subject and, to get through all tasks within the available time frame, substantial help was provided.

Subject 6 was extremely fast at getting the idea of the game and in learning how to use the haptic device. This subject's English skills were also very good.

All subjects except for the last (Subject 6), were introduced to the haptic device by first being given a pen and told that the force feedback they were about to experience is like touching objects lying on the table with the pen whilst writing. So the pen is touching the table as if it were a piece of paper, and the pen then accidentally encounters an object lying on the table, touching the object on the side as it were. The assumption was, of course, that blind users are thoroughly familiar with operating hand-held objects, and the underlying assumption was that

this is an apt analogy for operating the haptic device during gameplay. The grooved paths which the subject was going to meet in the game were explained by putting two objects on the table quite close together and let the user feel the "path" between them with the pen.

The reason for eventually abandoning this explanation-by-pen-analogy was that it appeared to make subjects think in 2D, as a result of which they tended not to be prepared to lift the robot arm and make it move in the third dimension. Rather, they tended to navigate the 3D gamescape as if it were a 2D map, which it is not, because it is quite possible in the game to stumble down a steep hillside or encounter obstacles in a groove, which must be climbed over. To be sure, the subjects were also introduced to the haptic device itself by having to move a ball around in a confined 3D space, but the 2D metaphor seemed to have stuck with some of them nevertheless. It might have helped if we had done more gameplay ourselves before the user tests than we actually did, but it is not clear that this would have guaranteed against the surprises just described.

Another problem in providing adequate gameplay instruction to subjects was the following. As long as the subjects move within the landscape displayed on the screen there is a white dot showing their current location. However, following the subjects' haptic search around the landscape, the white dot would frequently move outside the visible screen area because the virtual world is larger than the landscape as graphically displayed. When this happened, it became impossible to advise a user who had gotten lost in 3D virtual space on where s/he was and where to go.

6.2 Interview Questions

The questions worked well but, as already mentioned, it would have been useful to have more time for each interview. We often felt that we had to hurry on to the next question and that having waited a little longer for the user to think about a question might have elicited additional information of importance.

6.3 User Statistics

Tables 6 and 7 show some user stats information. The first three columns of Table 6 are repeated in Table 7 to support the reading. All subjects except Subject #1 are completely blind. Subject #1 kept his eyes closed throughout the gameplay session.

Subject #	**Age**	**Gender**	**Occupation/ education**	**Computer experience**	**Use**	**Input/output**
1	40	male	IT consultant	every day, every hour	all things possible	GUI, spoken output, no haptics
2	32	female	about to start call centre education	only at the institute for the blind	email, internet, a bit of everything	keyboard, spoken output, no haptics
3	21	male	about to start as a practician	every day at school	internet looking for film, music, …	keyboard, spoken output
4	23	male	attends IT-service education	every day also at home	internet, news, email, net banking, software installation, etc.	keyboard, key-board-controlled mouse, spoken output, Braille output, no haptics
5	21	female	attends IT-service education	every day also at home	email, internet, school things	keyboard, spoken output, no haptics

6	25	female	attends IT teacher education	every day also at home	internet, email, everything	keyboard, spoken output

Table 6. User stats, including computer experience and use.

Subject #	Age	Gender	Computer game experience	Game input/output	Computer games mentioned
1	40	male	played a lot from time to time	mouse and keyboard (GUI)	-
2	32	female	never plays	N/A	-
3	21	male	plays but not very often	mouse, acoustic output; has tried a haptic mouse for three fingers	car race, dart, alien outbreak
4	23	male	never plays	N/A	old DOS games
5	21	female	never plays	N/A	
6	25	female	plays very often	keyboard, joystick, mouse, screen reader, acoustic output; has tried force feedback	topspeed2, counter strike

Table 7. User stats, including computer game experience.

6.3.1 Age and Gender Representativeness

We had not asked for any particular age or gender distribution, cf. Section 5.5. The important thing for us was to get six subjects from the target group, i.e. blind users. It was less important, given the development stage of the system, to obtain a representative spread in age and gender. The average age of our subjects is 27 years and the gender distribution is fifty-fifty. Actually, we don't know if this is a good representative spread or not since we don't know if equally many blind men and women are playing computer games or could be interested in the tested technology, and we don't know if young people, in particular, such as those under 20, or 25, would appreciate the kind of game tested, although we suspect so.

6.3.2 Computer Experience and Use

All subjects are used to using a computer although one or two of them does not have a computer at home. Their primary use of the computer is for information search on the internet and for email. Their primary means of communication with the computer is via keyboard and spoken output. Only 1 subject (#1) is able to use a screen with strongly enlarged text. Only two subjects (#3, #6) have tried force feedback before – both in connection with computer games. Overall, this user group may be described as strongly computer-oriented with no less than four in six subjects having IT as ongoing education or current profession. This fact strongly limits their representativeness of blind users in general.

6.3.3 Computer Game Experience

Only three (#1, #3, #6) in six subjects had played computer games before. One doesn't play very often (#3), one does it from time to time (#1), and one does it very often (#6). Two of them (#3, #6) have tried a haptic force feedback device with a game but never anything remotely like the haptic device tested. Acoustic output is important for those who cannot see anything. Actually, all subjects mention keyboard and spoken output as their (primary) way of interacting with the computer for non-game activities, apart from Subject 1 who can see a bit

and who uses a full GUI with spoken output. This is in contrast to interaction with computer games where all three subjects who have tried to play before, mention the mouse among the devices they use for interaction.

6.3.4 User Profile Deviations

The only minor deviation from the profile description in Section 5.5 was that a couple of the subjects had very limited English skills. However, this was not felt to be a major obstacle since there were only few words in the game to explain, and since the explanation of what to do was in all cases given in Danish.

6.4 Data Validation

The test data collected was as planned, including rather copious observation notes from the subjects' gameplay with the system, six complete test session videos recorded via a Logitech QuickCam camera and a video camera (brand not noted), and two sets of interview notes, made by the interviewer and the observer, respectively. However, as mentioned, we regret that we did not record the introduction given to subjects.

7 Treasure Hunt User Test Results

7.1 This Section

The results reported in this section are based partly on the input we received from users during the post-trial interviews and partly on our own observations during the sessions. Section 7.2 describes the few technical issues observed while Sections 7.3 through 7.5 go through subjects' answers to the questions asked in the interview as well as the related observations made during each trial. Section 7.3 addresses game quality aspects, Section 7.4 functionality, and Section 7.5 user experience. Section 7.6 concludes on the results.

7.2 Robustness and Other Technical Issues

The technical robustness of the blind user's part of the game was generally acceptable for the user test.

The main technical problem experienced was that the spoken output would loop from time to time, the system annoyingly repeating the same output keyword over and over. In several user sessions the system had to be restarted once or twice, and in one case the sound had to be switched off so that the subject had to be told the colour of the closet by the experimenter.

A comparatively much smaller technical problem was that the conversion of sign language to spoken output takes several minutes.

7.3 Game Quality Aspects

The game part for the deaf-mute partner was not finished and was partially simulated by one of the developers. This was not a problem in the test where only the part for the blind user was considered.

7.3.1 Navigation in City and Landscape

All but Subject 6 found it medium difficult or difficult to navigate in the city and landscape. Two of them (#2 and #3) mention explicitly that they had to be patient. No doubt the first task (finding the red closet) was difficult for them – probably the most difficult one, in fact. It took up to 20 minutes to get through this task and the only reason that it didn't take longer for some subjects was that considerable help was provided.

There are probably several reasons why the first task is the most difficult one:

First, at game start the user needs to become more familiar with the haptic device than the familiarity achieved during the introduction to the system. The device is new to all subjects and the 3D world experience it gives them is new to all as well.

Secondly, very little information is provided on game and game step purpose and structure. Subjects were told the story-line of inhabiting an ancient Greek city in war times and that they were to find a treasure consisting of documents showing how to build war machines. However, the first subjects were not given any description of the structure of the system up front in terms of seven tasks to be solved, four of these by the blind user. Eventually the later subjects received more and more of this information. When the first task starts, the user is told to find the red closet. A first problem here is that the user doesn't know where to look for the red closet. It was perhaps not evident that they had to look for it inside a house and it was not obvious that they could actually enter structures in the city and landscape related to the current task. They were not told this.

Thirdly, the subjects were missing spoken feedback information, not because this information is provided through keywords, which is fine, but because there isn't enough of it. For example, among the many houses that one cannot enter, there are three houses which one actually can enter, but the spoken feedback doesn't provide any means of distinguishing them. The system just says, or whispers, "house" whenever you touch any house, including those that cannot be entered. This would not be realistic for a seeing person and probably also isn't realistic for a blind person. A seeing person can both tell from just looking whether or not a house has a door in it and can also visually tell different houses apart in most cases. Arguably, the blind need feedback information with similar functionality.

In general, the user interviews show that the system provides too little information during the game for an average blind user to be able to complete his/her part of the game in the 20 minutes planned. If left with only the information provided by the system, two hours would probably be a more realistic time-frame for playing a single game in the case of most users.

The interesting question here is whether the lack-of-information problem is an artefact created by our own test design or whether this is a real problem for the game as it stands. It might be argued that it is a fundamental point of the game to have fun and learn while exploring the game landscape, painstakingly building a mental map of it, and gradually solving all the problems starting from sparse information. We cannot express any firm opinions on this issue since we did not test the system under those one- or two-hour conditions and hence have no user data to back up our views. The final user (#6) actually did complete the game in about 20 minutes, but she was unusually bright and highly used to computer gaming. Moreover, she benefited from all the learning about how to instruct users that we ourselves went through during the user tests. So it is impossible to guess how fast she could have completed the game if left only with the information the system currently provides, nor can we guess how she would have evaluated the experience afterwards. In addition, since this game is new in its kind, there does not appear to be any standards or practices as to how much information to provide to blind users in this kind of game.

For what it is worth, however, we may have some doubts as to whether the average blind user would have enough fun from struggling through the game completely on his/her own given the information and feedback it currently provides and given our observations of their difficulties and need for help during gameplay. It seems to us likely that future games of this kind for the blind will come to include various kinds of start-of-the-game support which the present game lacks, such as a very small tactile map which provides a condensed overview of the landscape and the relevant structures in it and some instructions about when and how to click on objects. Otherwise, the risk is that a large fraction of the intended users might find the frustration of being lost to outweigh the fun of discovery.

7.3.2 The Haptic Device

The haptic device was positively received by all subjects. None of the subjects had used such a device before. Three subjects (#1, #4, #6) were actually thrilled by this technology and the 3D world experience it gave them. They describe it as a "wonderful and astonishing experience", "good, positive, surprising, impressive", and a "new sensation to be in a room/space that felt real". The three other subjects were also positive but modified their statement a bit by saying it worked "reasonably well" once you got used to it. One (#3) had a problem with the houses which he believed were located at the boundary of the landscape, so that he had a problem finding the temple ruins that were located behind the houses. Another subject (#5) mentioned that the neck became a little stiff from using the device and that holding the device in the right way was a bit difficult.

The haptic (force) feedback was generally felt to provide the information required for gameplay. The problem is rather in the missing oral/acoustic information as mentioned by

Subject 1. Subject 5 mentions a problem with the grooved path which she had trouble feeling. It felt as if it were a bit up in the air. Subject 6 mentions that sometimes the force feedback gave way (or yielded) too much when operated by hand.

7.3.3 Colour Recognition by Sound

Colour recognition by association of colours with musical sounds was no success. None of the subjects had tried colour recognition by sound before. Closest came Subject 6 who once tried a program that painted a sound picture of the colours in a painting, which she found difficult to follow. Only five subjects tried the colour recognition because in one case (#2) the sound was switched off when the subject came to the closet due to repeated problems with a sound loop. It should be mentioned that none of the subjects tried colour recognition very much.

Only one of the five remaining subjects seemed to find it easy to apply the musical colour codes (#3). The system only uses four different musical sounds, produced by four different musical instruments, for coding the four colours relevant to the game. However, even this small number of sound-colour associations are likely to be hard to remember if you have only been introduced to them once before playing the game. Moreover, several subjects pointed out that two of the musical instruments made closely similar sounds, which made the sounds harder to distinguish from one another. It was also remarked that it might be possible to choose far more different sounds for the codes, sounds which, furthermore, if possible, might be chosen in such as way that the association between each of them and a particular colour is not entirely arbitrary as is the case at present.

7.3.4 Spoken Output

In general, the spoken output was found to work well. Some of the subjects noticed the quite strong Greek accent and mentioned that a more standard English pronunciation would be good to have. However, since the output was basically a single word at-a-time and since only a small number of different words were used, the difficulties were limited although a couple of the subjects with the least English skills needed a bit of help with some of the words.

Subjects' answers to the question regarding the sufficiency of what was being said by the system were rather surprising. All six subjects found it sufficient. We wonder if they understood the question in the way we intended. What we actually wanted to know was if they were missing information which could have been provided verbally during gameplay. However, we suspect that they rather interpreted the question in a narrow sense to mean if keywords were sufficient or if full sentences would have been preferable. This hypothesis is supported by the fact that some of the subjects mentioned that it would have been good to have more information about what you bump into or get a hint if you seem to be lost.

We believe that the considerable amount of help and instructions provided to the subjects during gameplay made them feel that they basically got the information they needed, and that this is the primary reason why they don't stress the scarcity of system-provided verbal information to a much higher degree.

7.3.5 Ease of Game Tasks and of Achieving the Goal

As already mentioned, the first task was for several reasons probably the most difficult one. It is characteristic that the two first subjects agree that the tasks were difficult while the following subjects found the level of difficulty okay. In particular, the first subject was left on his own for quite some time before receiving help while for the following subjects the time elapsing before the subject received help was gradually decreased. We notice that Subject #3 said that the task difficulty was okay "as long as you get instructions from the experimenter". Subject #5 had a similar comment. Without the considerable amount of help and instructions

provided throughout gameplay, we suspect that several of the subjects – if not all of them - would never have reached the goal.

The following list shows a number of problems mentioned as contributing to making a task difficult:

- There are too many details in the tactile landscape and the point you need to find and click on is too small.
- There is no indication of whether you are on your way in or out of an area or whether you have been there before (orientation is difficult).
- You don't know if a house has one or several rooms and there is no way to know if you have visited a house before (except by finding a closet and find out its colour).
- The information provided should correspond to what a seeing person would get.
- You are not told in any precise terms what to find.

On the question of how easy or difficult it was to achieve the goal of the game, some subjects understood the question as referring to the last task of the game, which was generally considered easy. However, this last task revealed a problem when a subject clicked at the wrong end of the path he had to follow. Apparently, you have only one click available in this task and the system does not advice the user of this limitation. If the click is spent before you have followed the path from beginning to end, you are not told when you reach the goal, even if you click in the right place. This turned out to be very frustrating for the subject who experienced it and for this reason did not get his congratulations for having completed the game. Thus, in the subsequent sessions the experimenter always warned the subjects against clicking in the wrong place, which again meant providing additional help to the subjects.

Actually, clicking does appear somewhat inconsistent and troublesome in the game. Sometimes you need to click once and sometimes twice without any apparent reason for the difference. Furthermore, coordinating force-feedback and haptic device clicks turned out to be difficult for several subjects. In several cases, subjects had actually found what they were looking for but because their touch/click coordination was not entirely right, the click(s) didn't have any effect and they went on to look in other places. Moreover, it seemed that in some cases the area in which clicking would have an effect, was quite small and difficult to hit, as already mentioned in the list above.

7.3.6 Following the Grooved Path

Although the final game task was in general considered fairly easy compared to, in particular, the first task, opinions were somewhat divided regarding how easy it was to follow the grooved path as part of the final task. The haptic force feedback was considered helpful by all subjects. However, the following problems were mentioned:

- You don't know what the path is supposed to look like – is it like a groove, a little ridge, or something else? Some verbal feedback when you have come across the path would be helpful.
- You don't know what to do when the path ends.
- You need to learn that you can jump over things cluttering the path.
- You need to learn that you actually move in 3D.

7.3.7 Partner Communication

Since the game is meant to be a highly social and collaborative one, it might be surprising that no subjects complained about the totally missing communication with the deaf-mute partner. However, this might be due to the fact that it was said in the introduction to the system that we would simply simulate the presence of a partner during gameplay.

In general subjects did not pay so much attention to the partner and felt that they got the information they needed. However, Subject #6 pinpoints why partner communication was felt to be okay when she says "You [the experimenter] explained what went on". This is exactly what happened and what was necessary because the system itself didn't give any other clue than a beep. When the blind had finished a task it was the deaf-mute's turn to carry out the next task. In the meantime the blind had to wait. Normally it didn't take long for Kostas to carry out the task of the deaf-mute. The only task that took a long time (a couple of minutes) was task 4 (cf. Section 5.2) in which sign language had to be translated into speech. This felt like a long time to wait and, as Subject #1 rightly noticed "The deaf person must become tired of waiting". This would definitely be a problem with the game in its present version. The partner is left in the dark as to what is going on when it is not his turn. Subjects #1 and #6 mentioned that it would be good to get an explanation of what the partner is doing and all except Subject #2 would like to have the possibility to communicate with the partner during gameplay via some kind of chat (e.g., via speech and sign language, speech-to-text and text-to-speech, or, simpler, via predefined messages). Communication needs mentioned include information about what the partner is currently doing, clarification of what was received from the partner, and requests for more information.

7.3.8 Ease of Use, Control during Game Play, and Learning

To the question of how easy or difficult the subjects found it to use the system for gameplay, only one subject (#2) said that it was difficult. Three subjects (#4, #5, #6) found it reasonably easy, and two subjects (#1, #3) found it easy. The answers mostly refer to the use of the haptic device and orientation in 3D space. Thus, the difficulties mentioned typically relate to moving around in a 3D space, finding out how the haptic device works, and remembering to click.

Despite the fact that the system was found easy to use, subjects did not feel much in control. The following list shows some of the problems that were mentioned as a reason for not feeling in control. However, sometimes the subject just expressed a general need for help during gameplay rather than being specific about what made him/her feel out of control.

- Being inside the ruins felt like floating in the air. A wall around the ruins would have been good.
- It was difficult to get past the houses and to get into them. You had to jump to go beyond them.
- Missing spoken feedback.

We believe that two major obstacles are (i) the missing, situated spoken feedback, such as "house with no door", "Corner House", "Corner House door", "inside Corner House", etc.; and (ii) the lack of other, more general information from the system of the kinds the experimenter had to provide. Another problem no doubt relates to the fact that it takes time to become really familiar with moving in 3D space if you are only used to 2D. This challenge may be fine as part of the game challenge. Subject #6 explicitly mentions that she eventually "began to explore the area more in 3D". The lack of information, on the other hand, is a more problematic aspect as also discussed in Section 7.3.1.

Subjects mostly became better at playing the game as they got used to the haptic device and found out what it felt like to be in a 3D space. When asked if they learned something during game play, all subjects had comments as listed below:

- More systematicity in uncovering areas (Subject #1).
- Built a sort of map in the head during the game (Subject #3).
- Became more aware of the surroundings (Subject #4).
- Started to explore the area more in 3D (Subject #6).

- Learned patience (Subject #2).
- Learned not to move so fast (Subject #5).

We notice that what is mentioned as being learned primarily refers to the 3D challenge and exploration by the haptic device. We find it acceptable that this is something that users have to learn by experience.

7.4 Functionality Aspects

Subjects were asked if the system offered them all the functionality they needed to play the game or whether they missed anything, e.g., regarding spoken output or what they do with the haptic device.

All subjects except #1 stated that they did not miss anything. Subject #6 added that she would like to get rid of the spoken output loop which she found annoying and which of course should be corrected. Subject #1 stated that he primarily missed information in terms of on-line help and feedback whereas the haptic device was fine.

It should be kept in mind that some of the questions asked prior to this one and in relation to game quality also addressed functionality aspects. It is often seen that subjects do not necessarily repeat themselves if they already believe that they have brought across what was on their mind. However, as explained earlier, it was exactly Subject #1 who received the least help and instructions, so we don't find it surprising that he is the one who misses more information.

7.5 User Experience Aspects

7.5.1 Previous Game Experience and Likeability of Game

As it appears from Table 7, three subjects (#2, #4, #5) are not used to playing computer games at all, one (#3) does it rarely, while two (#1, #6) play a lot. Only two of them (#3, #6) had tried force feedback before and never anything like the haptic device used in the treasure hunt game. On the question of whether they had ever tried something like the treasure hunting game before it became clear that this was not the case. In particular the haptic device was new and so was the real 3D experience.

In general the subjects liked the game (a good idea, fun, challenging, exciting). Only one (subject #6) found it boring and without enough action. She was probably the one with the most game experience among the subjects. She mastered the game faster than the other subjects and might have been the only subject who managed to take the step back it takes to plainly compare the contents of a game which offers a completely new gaming experience with other computer games of the same genre. However, the game techniques she liked very much and hoped to see in other games.

It is the authors' view that she is obviously right that the game contents are too thin. If the game were to be played by a seeing person, it wouldn't take long to complete. The only reason why it takes a lot of time for the blind is that precious little of the information a seeing person easily gathers via the graphical medium, is made available to the blind via information in other media that could support a similar fast perception of the game environment.

We need to add an obvious comment. It is that the current game is not supposed to be a complete game in any sense. It's a research prototype of how to combine a range of input/output modalities in new ways for game purposes for special user groups. As it stands, this game was never intended to be a competitor with commercial computer games.

7.5.2 Advantages, Disadvantages, and Trying the Game again

The fact that most subjects liked the game is, we believe, first of all due to their strong appreciation of the realistic 3D experience and the new way to interact with a computer, and has rather little to do with the game contents and their richness or lack of it. When asked about the advantages of the game, four of the subjects (#1, #4, #5, #6) mentioned the 3D experience and the fantastic sense of space afforded by the haptic device. Subject #6 saw further potential in the techniques used in the game. By enabling people with different disabilities to play together it enlarges the pool of potential game partners for the blind. Furthermore, she proposed that the game techniques might be used to present and explain new objects to the blind and perhaps also to create such objects as a kind of 3D images.

The mentioned disadvantages were few. Subject #1 mentioned the evident problem that not so many people has the kind of haptic device used with the tested system and that there are few games made for this kind of device. He also found the distance you move in the landscape when you move the arm of the haptic device, too small and proposed that one should be able to set a time factor for how far you travel per second. Subject #3 mentioned the annoying missing congratulations due to an already spent click (see Section 7.3.5). Subject #5 proposed that there might be more effects and sounds to, e.g., characterise the room you are in. Strikingly, the low level of system-provided information was not mentioned among the disadvantages. Again, we believe this to be due to the considerable amount of instructions and hints provided by the experimenter and to the fact that some of the subject had in fact already mentioned missing feedback or information as a problem in early replies.

All but Subject #6 would be happy to try the game again. Subject #6 found the game somewhat boring but was extremely interested in the techniques used and would be happy to try other games using the same techniques.

7.6 Conclusions

The main conclusions on the test of the treasure hunting game are that (i) the input/output modality combination, i.e., spoken keywords output, non-speech sound output, haptic 3D force-feedback output, haptic 3D navigation input, haptic click notation, is fine, except that spoken discourse is missing. Users were generally excited about the 3D experience they had from using the haptic device. (ii) The game itself is at a very early stage and still needs a lot of development to become a real usable and enjoyable game. We expand on the latter point in the following.

Despite a couple of technical problems, such as the annoying output speech loop, the robustness of the system is acceptable overall – not least for a system which in other respects is clearly at a very early stage.

Despite the appropriateness of the modality combination, there is plenty of room for improvement within individual modalities:

Haptic notation input, i.e., the ways in which the user needs to click (once or twice) to change the course of the game, should be made consistent across all tasks and coordination with the haptic device (haptic 3D force-feedback output, haptic 3D navigation input) should be reconsidered.

Haptic output, i.e., what the user feels when moving the arm of the haptic device, was fine. It gave a good 3D experience. However, it should be considered to delimit the haptically accessed 3D environment in some way in order that users do not get lost in virtual space.

Non-speech sound output, i.e., the sounds which represent colours, should be improved. Sounds which are difficult to distinguish should not be included. Since the sound-colour

matching is essentially arbitrary, the system should provide help regarding which sound denotes which colour.

Spoken keywords output was acceptable as a modality. The subjects felt no need for full sentences rather than keywords. However, there is definitely a need for much more spoken output information, and some of this output will have to be full sentences, i.e., spoken discourse for, e.g., explaining things about the game.

Keywords or key phrase feedback should be used to better inform the user during navigation regarding where the user is and what the user encounters. For example, houses could be numbered or labelled through keyword labels, so that the user gets far better opportunity to build a mental 3D map of the environment during exploration. Other examples of keywords and key phrases that might be useful include “door”, “ path”, “you have entered house N”, “you have left house N” where N is a number or label. Many more could be added easily but the issue needs careful consideration in order to be done to the right extent, so that there are still challenges left in the game.

There should be an (optional) spoken introduction to the system and how to use it. The user should be told, e.g., what the context and goal of the game is and that there are a number of tasks to be solved through structured collaboration with the partner. It should also be considered if a small tutorial on how to use the haptic device in 3D should be included.

Help should be available online at any time. It should include, e.g., explanation and illustration of the colour notation.

8 Acknowledgements

The work represented by this document was partly funded by EU Network of Excellence SIMILAR (www.similar.cc). We are grateful for the support.

As regards the Sudoku game user test, we would like to thank Svend Kiilerich, NISLab, for recruiting the test subjects, and Torben K. Madsen, NISLab, for technical support.

As regards the Treasure Hunt game user test, we would like to thank Kostas Moustakas, the prime developer of the system, for setting up the system and participating in the test. We would also like to thank the Institute for the Blind for hosting the user test and Mr. Flemming Berthelsen for recruiting the test subjects and making the test days a very pleasant experience.

9 References

Bernsen, N. O.: Towards a tool for predicting speech functionality. *Speech Communication* 23, 1997, 181-210.

Bernsen, N. O.: Multimodality in language and speech systems - from theory to design support tool. In Granström, B., House, D., and Karlsson, I. (Eds.): *Multimodality in Language and Speech Systems.* Dordrecht: Kluwer Academic Publishers 2002, 93-148.

Bernsen, N. O., Dybkjær, H. and Dybkjær, L.: *Designing Interactive Speech Systems. From First Ideas to User Testing.* Springer Verlag 1998.

Bernsen, N. O., and Dybkjær, L. (1999). A theory of speech in multimodal systems. In Dalsgaard, P., Lee, C.-H., Heisterkamp, P., and Cole, R. (Eds.). *Proceedings of the ESCA Workshop on Interactive Dialogue in Multi-Modal Systems,* Irsee, Germany. Bonn: European Speech Communication Association, 1999, 105-108.

Bernsen, N. O. and Dybkjær, L.: Multimodal Usability Progress Report, SIMILAR Deliverable D100, September 2007.

Bernsen, N. O. and Dybkjær, L.: Multimodal Usability. To appear.

Malerczyk, C., Dähne, P. and Schnaider, M.: Exploring Digitized Artworks by Pointing Posture Recognition. Proceedings of 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, Pisa, Italy, 8 – 11 November 2005.

Moustakas, K., Nikolakis, G., Tzovaras, D., Deville, B., Marras, I. and Pavlek, J.: Multimodal Tools and Interfaces for the Intercommunication between Visually Impaired and "Deaf and Mute" People. eNTERFACE'06, July 17th – August 11th, Dubrovnik, Croatia, Final Project Report, 2006.

10 Appendix 1. Sudoku User Interviews

10.1 Subject 1

USER id: Subject 1	**Sudoku proficiency: beginner**
Name: Frederik	**Start level: easy**
Age: 24	**Interview date: 7.6.2007**
Gender: male	**Interview time: 10.30**
Profession/education: medical student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	Fun alternative. Difficult to control at times.
2. How suited do you think spoken input is for games like the one you have just tried?	Clearly more fun to interact in this way. Fun to speak to the computer. Never spoke to a computer before.
3. How suited do you think screen output is for games like the one you have just tried?	OK, good idea.
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	Optimal combination. Would not be so fun if one only has to point, like with the mouse. Plays on multiple senses. More exciting with speech, more factors involved.
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Room for improvement. Impressive that it is possible. But as a game this should be improved.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	Hard to keep up one's arm for a long time.
7. To which extent did the system understand what you said?	Its understanding of what I said was fine but it didn't always hear what I said. It didn't always react.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Worked easy. But limited command language.
9. To which extent did the system understand combinations of speech and pointing input?	Problems if one pointed in a different square when the system hadn't understood what one said. Problems with focus if one's finger is shaking a bit.
10. How well did it otherwise work to use combined speech and pointing input?	It's negative that one has to stand still and is not allowed to move.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Cannot imagine other ways. Maybe a pen instead of a finger, so that one could click.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Easy. No problems.
13. To which extent did you miss other forms of output (than via the screen)?	None. Maybe sound. [Probably said after being mentioned by the interviewer.]

14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Fun to try.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Relatively easy. The only problem was to do perfect pointing.
16. To which extent did you feel in control when playing the game?	Frustrating to hit the wrong square. Not 100% control due to (i) pointing problems, (ii) that the system doesn't always hear what is said.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	All possibilities: deleting numbers etc. Enough command words but more fun if there were more.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Real fun to try.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	An experience, like being in a future movie. It requires quite some equipment.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	No. Sudoku games are not so exciting. Normally not playing Sudoku. But with a different game, like Trivial Pursuit, OK! Maybe as a party game. The headset is cumbersome, though.
21. Any other comments?	Not asked.
Observations on this user	
Points for 25 minutes without taking his hand down and while playing more than two games. Starts at level 1 and plays 2 games successfully. Then selects 3rd game at level 2 and plays for a while.	

10.2 Subject 2

USER id: Subject 2	**Sudoku proficiency: beginner**
Name: Christoffer	**Start level: easy**
Age: 23	**Interview date: 7.6.2007**
Gender: male	**Interview time: 11.30**
Profession/education: medical student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	4
2. How suited do you think spoken input is for games like the one you have just tried?	3 Touch pointing + number selection from palette would be faster.
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	3
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	In the large majority of cases.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	It's fine for limited periods of time. It's reasonably precise. But the arm can get tired.
7. To which extent did the system understand what you said?	To a great extent.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Good. No wrong numbers were entered.
9. To which extent did the system understand combinations of speech and pointing input?	In the large majority of cases.
10. How well did it otherwise work to use combined speech and pointing input?	When it went wrong it was because the arm was lowered. But one has to be very stationary when playing.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Not really. But mouse or touch screen could be an alternative, otherwise it becomes tiring to play.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Was fine. No problems. Clear instructions.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Pointing in this way is clumsy if precise actions are required. Works good if the squares are large enough, like in this case, but they should not be smaller than they are.

15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Level 1 was easy enough. But then it became difficult. It becomes monotonous in the long run. But it's pleasant. But it's not for playing at home.
16. To which extent did you feel in control when playing the game?	To a reasonably large extent. When I made errors I could correct them.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Nothing missed. Everything you need was there. Good to be told about one's errors immediately.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Quite fun for a limited period of time. A bit cumbersome.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Has played on paper a single time. Has not played on the Internet. Better than paper: errors are shown immediately. A bit clumsy and slow compared to paper play. Fun as entertainment. Otherwise too clumsy.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Probably not. Maybe if I had 20 minutes to kill. Wouldn't stop to do it. Sudoku is better played for pleasure at home rather than in public. Better if two could play against one another.
21. Any other comments?	Fun to try. EyeToy is a fun feature but one got bored from playing in the end.
Observations on this user	
Starts at level 1 and plays 2 games successfully. Selects level 2, plays for a while, gives up, resets to level 1.	

10.3 Subject 3

USER id: Subject 3	**Sudoku proficiency: experienced**
Name: Edmund	**Start level: difficult**
Age: 60	**Interview date: 7.6.2007**
Gender: male	**Interview time: 12.30**
Profession/education: lecturer, computer science	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	1 Sudoku requires pencil and erasor! If the game had been chess: 4/5
2. How suited do you think spoken input is for games like the one you have just tried?	4 In general, not for Sudoku, maybe in the airport, e.g.
3. How suited do you think screen output is for games like the one you have just tried?	4
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	1 (for Sudoku) If the game had been chess: 4/5. The system didn't allow backtracking.
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Understood fine. But the slow speed requires some getting-used-to.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	Worked good. No other problems.
7. To which extent did the system understand what you said?	Perfect.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Fine.
9. To which extent did the system understand combinations of speech and pointing input?	Fine. But one has to get used to the fact that the system takes some time to discover that the finger has moved.
10. How well did it otherwise work to use combined speech and pointing input?	No other problems.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Maybe an 'undo' function during pointing. An undo function without having to point at the latest number
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	No problem. I know Sudoku well.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Fine. Great help for people who are unaccustomed to the keyboard. Good for many other games.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	It was easy.

16. To which extent did you feel in control when playing the game?	It was my mistake if there were problems. Excellent control. Errors could be corrected.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Undo would be useful. Backtrack would be useful. Otherwise OK.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Prefers paper and pencil. Does not like to do it on a screen. Used to think pencil-in-hand.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Used to play Sudoku on paper. This is pleasant. Did not try Sudoku on the Internet. Does not want to. Thinks in terms of recursion and wants the system to have the function "try, regret, try again". When he plays on paper, he writes something in with the pencil in order to be able to erase it later on. The game would be more stable with a touch screen.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Not Sudoku. But maybe chess. Easier to carry a piece of paper with a Sudoku game on it than to carry a chess game player. Hesitant as regards playing chess in public.
21. Any other comments?	Sudoku is maybe too recreative, it's a game. Would rather use this kind of system for goal-oriented activities.
Observations on this user	
Starts at level 3. Puts in some numbers but with increasingly long intervals in-between. Changes to level 2 at the suggestion of the experimenter. Then plays for a while until the system crashes. It was probably Windows that crashed, probably because of overheating. The disc ventilator was then turned up.	

10.4 Subject 4

USER id: Subject 4	**Sudoku proficiency: beginner**
Name: Rosa	**Start level: easy**
Age: 76	**Interview date: 7.6.2007**
Gender: female	**Interview time: 15.30**
Profession/education: school teacher, retired	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	4 But maybe a pointing stick instead?
2. How suited do you think spoken input is for games like the one you have just tried?	4
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	4
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Partly. Had to repeat a number of times.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	The cursor was shaky.
7. To which extent did the system understand what you said?	It didn't like my way of saying "three".
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Fine when you got into it. Never spoke to a system before.
9. To which extent did the system understand combinations of speech and pointing input?	Went fine, except for the jitter. Sometimes had to try several times to enter a number – then it suddenly appeared even if one didn't talk. Wouldn't it be better to point using a long pointing stick?
10. How well did it otherwise work to use combined speech and pointing input?	See above.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	I missed a pointing stick. Nothing else.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Did not understand the word "proceed". Otherwise no doubts on how to proceed.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Very much a new thing. Exciting.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Relatively easy. But prefers to play crossword puzzles.

	The first game was easy.
16. To which extent did you feel in control when playing the game?	OK
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Didn't miss anything. Maybe a "move" command to move a number somewhere else?
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	More or less the same as on paper.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	In total it's more or less as good as solving Sudokus in the newspaper. But you are not dependent on pen and eraser. There, the paper sometimes gets completely filled with numbers and notes, this doesn't happen with this game.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	No. Not in public – people might interfere. Rather at home. Prefers crossword puzzles.
21. Any other comments?	Fun invention.
Observations on this user	
76 year old woman in admirable shape both physically and mentally. Starts at level 1 and plays one successful game. Then selects level 2 where she starts making mistakes which are not being signalled by the system (which only signals an obvious kind of mistakes). She starts realising that she is in trouble and makes many unsuccessful attempts to erase one of the fixed initial numbers on the game board. The system's catching of the user's pointing was very jittery in the beginning and the system was re-started. This could be effects of the disc over-heating presumed for Subject 4. As the stability is still problematic, the system is re-calibrated twice for the lighting conditions. But it remains more unstable than for the 3 previous subjects. Speech recognition is worse than for the 3 previous subjects, probably due to this subject's pronunciation of, in particular, the number "number three". Looks at length at the screen "Proceed, yes/no" and then looks to the experimenter for advice on its meaning.	

10.5 Subject 5

USER id: Subject 5	**Sudoku proficiency: medium**
Name: Jørgen	**Start level: medium**
Age: 33	**Interview date: 7.6.2007**
Gender: male	**Interview time: 16.30**
Profession/education: economist	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	4
2. How suited do you think spoken input is for games like the one you have just tried?	3
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	3
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Most of the time. Learned by playing that a previously pronounced number might appear in a square when it was pointed to later on, without the need to speak the number again.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	A bit strenuous to try to keep the hand still all the time. Had to concentrate a lot. Subject thinks that he might have spoken earlier once he had highlighted a square, rather than re-checking that the highlighting stayed in place before speaking.
7. To which extent did the system understand what you said?	It generally understood the numbers but had problems with “delete”/”remove”.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Fine. Learned during the game that it wasn’t necessary to speak so loudly.
9. To which extent did the system understand combinations of speech and pointing input?	There was a 9 which wouldn’t go away. But it understood most number-pointing pairs. Sometimes the number appeared in the square at the second pointing attempt.
10. How well did it otherwise work to use combined speech and pointing input?	A bit of fun, and different.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	No.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	There wasn’t a lot to understand, you just had to speak and point. Mistakenly pointed at *reset* instead of *new game*. Some of the letters in these two options are invisible on the screen [COMMENT: the labelled buttons/icons sit in the top left corner and are partly

	hidden behind the screen frame].
13. To which extent did you miss other forms of output (than via the screen)?	There might be sounds, e.g., sound warnings instead of, or together with, the red colouring when a mistake has been made.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	A bit strenuous. Otherwise it worked in general. But it is probably possible to get used to. Wouldn't play for as long as if the mouse or a pencil had been used.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	A bit difficult. A bit irritating that the cursor jittered. And a bit irritating not always to succeed in deleting a number.
16. To which extent did you feel in control when playing the game?	In control most of the time, largely. Couldn't always enter the number on the first try. In one case couldn't delete a number.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Has played Sudoku on the Internet (Jyllandsposten's website). Would be useful to be able to switch on and off a function which signalled the non-obvious mistakes, like on Jyllandsposten's website.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	A bit o' fun, a bit strenuous. Different.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Less physically demanding to play with a pencil or a mouse. And in that case one is in full control. Fun to speak to the system. If there are onlookers, they can better follow the game on the screen than if it's being played on a newspaper page. Children might prefer the tested game. Compared to playing on the Internet it's good to have a larger screen. The game might be used for team competition in game arcades. For that, the game should have a timer for timing each game played.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. Although not if the spectators take too much interest. 1-3 people could play together, e.g. as a family competition.
21. Any other comments?	The game might have a timer added to it.
Observations on this user	
At this point, the system is back to normal and doesn't	

jitter more than usual. Starts by playing a level 1 game well which, unfortunately, ends up having multiple solutions (at least it seemed so the experimenter at the time). Then successfully plays a level 2 game. Subject quickly gets into a style of calm and controlled gestural playing, with full control of the cursor. At some point, the subject starts to use both arms interchangeably for pointing. When he gets a red line with two times 9 in it, he tries to delete one of the fixed numbers. Has difficulty being understood when saying “remove this” and “delete this” concerning a (removable) 9. Then chooses to reset the game. Forgets a couple of times to say “number” before the integer.	

10.6 Subject 6

USER id: Subject 6	**Sudoku proficiency: experienced**
Name: Jens	**Start level: medium**
Age: 30	**Interview date: 12.6.2007**
Gender: male	**Interview time: 10.30**
Profession/education: biomechanics/phys.edu. student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used Closed questions – asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	4
2. How suited do you think spoken input is for games like the one you have just tried?	4
3. How suited do you think screen output is for games like the one you have just tried?	4-5 - in the abstract. In this game, the screen quality could be improved, 3.
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	5
Quality of interaction (Open questions)	**Open questions**
5. To which extent did the system understand you when you pointed at something?	80-90% The pointed-to square failed to light up several times – annoying. Had to move the finger to back and forth to make it light up.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	No, except that the arm gets tired. It's good to be able to use either arm. Never did 3D pointing before. Used the stretched arm to also survey the numbers in the rows. It takes some getting-used-to to lower the arm when it's not needed.
7. To which extent did the system understand what you said?	Always. I had to repeat in 2-3 cases (to delete a number). I try to pronounce clearly when I know that I am talking to a machine. It's a bit artificial to have to say "number" first.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	See 7 above.
9. To which extent did the system understand combinations of speech and pointing input?	Fine. A spoken command for choosing a new game would be good – instead of having to wait [NOTE: subject didn't understand the reason for having to wait no matter which input is given.]
10. How well did it otherwise work to use combined speech and pointing input?	See 9 above.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Maybe "undo" – without having to point.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any	Easy enough.

problems with some of what was shown on the screen? Which?	
13. To which extent did you miss other forms of output (than via the screen)?	Fanfare when you win (said jokingly). Maybe add little sounds to signal that an input number has been received. Maybe an "Aaugh" sound when the red colour appears.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Fine.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Suitable. Had to realise that the second game played had multiple solutions – never tried that before.
16. To which extent did you feel in control when playing the game?	Most of the time. But not when encountering the multiple solution phenomenon.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	A 'help' function like on the Internet? He doesn't believe that this is an advantage: it may make playing too easy since it is easy to be tempted to use the function.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Different. Requires getting-used-to. It's fine with paper and pencil. But it is also real fine to involve the arm and the body.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Pros: good to get the body active, good to include the use of speech. Cons: pointing is a bit slow due to the time it takes for the system to respond. The system cannot follow when you have spotted a pattern and have 4-5 numbers to put in. Speed and precision of pointing is a cause for irritation.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. I cannot help it.
21. Any other comments?	The technology could be used for more physically active games than Sudoku. But one might choose a touch screen instead. The speech might annoy other people. In a more active version of the game, you point at a square and select the number to insert by jumping onto a numbered field in a palette on the floor. The palette might also include 'undo' and 'delete'.

	Would be nice to have a version with Danish speech recognition. The game idea is fine and could be taken further.
Observations on this user	
It takes rather long time to load a new game. This user, and several other subjects, don't seem to realise that and seem to start wondering if something is wrong. Due to the lack of process feedback at this stage, they tend to start talking to the game and/or mumble (Danish) comments and questions. Chooses level 2 the first time. Has a slight lisp in pronouncing "3" which seems to cause some difficulties for the recogniser. Manages the game fast and smooth. Then chooses level 3. It happens several times that an empty square doesn't light up – and probably doesn't activate – so a neighbouring square grabs the spoken number, resulting in red. The non-working square works fine later on. He plays well and turns out to be the best player of all 12 subjects. The game has multiple solutions, which he realises after a while. He then simply chooses one of the variants and completes the game. Starts new game at level 3.	

10.7 Subject 7

USER id: Subject 7	**Sudoku proficiency: experienced**
Name: Inger	**Start level: difficult**
Age: 23	**Interview date: 12.6.2007**
Gender: female	**Interview time: 11.30**
Profession/education: math./phys. edu. student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	4 for a system in public locations prefers pen and pencil
2. How suited do you think spoken input is for games like the one you have just tried?	2 speech is a bit funny ("åndet" meaning: not very smart, daft)
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	3 funny/daft ("åndet") to talk to a game machine
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	OK But the system is very slow in reacting to pointing input.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	Choosing a new game: what makes it choose – the pointing, speech, both? Is it necessary to keep the hand pointed at the field for some time? What happens if, in trying to select a game level, the pointing hand passes over some other active field in the menu?
7. To which extent did the system understand what you said?	Some times better than others. It varied a lot. Did not try to speak particularly clearly, just spoke normally. It seemed to confuse 5 and 9 once. 3 was also a problem.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Nothing else.
9. To which extent did the system understand combinations of speech and pointing input?	Sometimes. Problem: speak number - move finger – the number pops up in a different field. For instance when I did not coordinate speech and pointing. Speech and pointing were not always coordinated.
10. How well did it otherwise work to use combined speech and pointing input?	Nothing else.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Might have used a double click instead of just pointing. Misses the clicking. Prefers a touch screen to standing with the arm in the air. Did I click or not?

12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Fine.
13. To which extent did you miss other forms of output (than via the screen)?	No, sounds can be irritating.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Is somewhat similar to playing on the Internet – except for the speech. Difficult to get used to speaking to a screen.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Annoying to have to stick one's arm forward. But I took the arm down when I had to stand and think. It's not relaxing to use the arm. Playing is not so difficult, it's OK.
16. To which extent did you feel in control when playing the game?	Not 100%. For instance, when it does not insert the number I said. It didn't always obey.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Undo is clearly missing – without pointing. A function for inserting notes (= possible numbers) in the squares would be a good thing.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Prefers paper. Strange to speak to a machine. Felt a bit under pressure in the level 3 game – it's usually faster for me to play one of those. Might also play Sudoku on a computer using a mouse.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Paper makes it possible to insert possible numbers into the fields. Here I need to keep everything in my head. I felt it was hard to establish and keep an overview of the game, I cannot explain why. Maybe it was because I felt under pressure. Slow. Easy to get a new game. Easy to delete in case of mistakes. Doesn't waste paper.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes, if I had time to kill, like when waiting someplace, and if the game were improved. Like playing a game of table soccer.
21. Any other comments?	Good that no sounds are output, e.g. in case of errors. If the game worked better it shouldn't even

	highlight errors in red.
Observations on this user	
2 or more empty squares are highlighted simultaneously, one by user pointing, the other(s) just like that. Starts at level 3. Gets stuck. The experimenter proposes medium. She selects a medium game. Gets impatient due to the new game loading time. Has difficulties with no's 3, 6 and 1. Her accent is rather strongly Danish. Relaxes her arms in-between the pointing actions. Opens with a coloured field lit up [?? CHECK video] Finishes the game at level 2 [?? CHECK video] Starts third game as medium.	

10.8 Subject 8

USER id: Subject 8	**Sudoku proficiency: some experience**
Name: Signe	**Start level: medium**
Age: 23	**Interview date: 12.6.2007**
Gender: female	**Interview time: 12.30**
Profession/education: math./religion student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	3
2. How suited do you think spoken input is for games like the one you have just tried?	4
3. How suited do you think screen output is for games like the one you have just tried?	4 It's good that the numbers one inserts look different. This is useful when one has to remove a number.
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	4
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	An outstretched arm is necessary. Takes a bit of time to learn that there is some latency time when pointing. It sometimes inserted the spoken number in an adjacent field which was annoying.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	See above.
7. To which extent did the system understand what you said?	Difficulty with no. 9. When the initial mike problem was solved it became fun to play.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	OK to talk to the system in English.
9. To which extent did the system understand combinations of speech and pointing input?	Mainly well. But there were lapses leading to red and the question: what was the number which just became overwritten?
10. How well did it otherwise work to use combined speech and pointing input?	'Remove' worked better than 'delete'.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Would like to be able to point and, e.g., click on a number. But OK when you get used to it.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	OK. User-friendly.
13. To which extent did you miss other forms of output (than via the screen)?	Good that there is nothing else. Sound is irritating when you have to think (refers to a TV show where sounds are being played

	while people have to think in order to answer questions and become millionaires).
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	This is how it is in many games. Speech is just a variation. It's fine.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	Difficult at the start because of the mike problem. Tricky that the number might appear in a square later on without one's having to repeat it in speech, i.e. that the order of speech and pointing does not matter.
16. To which extent did you feel in control when playing the game?	At the end I was in control, well and good. After having gotten used to it and after the mike problem was solved.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Fine.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Fun. Different from playing in a book. More immersive/engaging. Mentions Nintendo: one holds a thing in one's hand and plays tennis. More physical.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Has tried paper (a Sudoku book) and the Internet. This game allows more physical activity. Adrenalin increases a bit when your body is involved in this way. Sudoku is good for relaxing by using the brain in a different way from, e.g., doing one's homework. This game allows more immersion and engagement. The book is better than the Internet (TV2) where the Sudoku board is small and requires glasses. This game is more fun than the Internet. Great fun. But it's more relaxing playing in a book.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. It's fun. I'm a bit addicted to Sudoku.
21. Any other comments?	No.
Observations on this user	
2 empty squares are permanently highlighted. Difficulty with no. 9. Difficulties getting the speech understood, leading to red	

rows, columns and 3x3 fields – sighs and groans. Many of the problems are in the middle square in the left-most column. Difficulty with “remove this”. The experimenter re-adjusts the mike – this improves things. Plays 2 level 2 games both of which have multiple solutions. Succeeds with both of them. Starts third game at level 3.	

10.9 Subject 9

USER id: Subject 9	**Sudoku proficiency: some experience**
Name: Pia	**Start level: medium**
Age: 31	**Interview date: 12.6.2007**
Gender: female	**Interview time: 14.00**
Profession/education: PhD student biology	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	2-3 Mouse is better. Doesn't feel that it is possible to point precisely enough. Annoying with out-stretched arm.
2. How suited do you think spoken input is for games like the one you have just tried?	4 When it works.
3. How suited do you think screen output is for games like the one you have just tried?	4
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	3 Fun to try. But prefers to take Sudoku with her in the sofa. The setup requires a more action-oriented game. She would score such games higher than 3 (suggests 4).
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Understood all of it.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	Problems keeping the arm straight. A problem changing arm: people typically has a "leading eye" for aiming at something. If the arm is changed, it gets in the way for the "leading eye" (you don't change eye when you change arm) – whereupon the body starts leaning to one side to be able to see what is being pointed at.
7. To which extent did the system understand what you said?	To a large extent except for problems with no. 4.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Problems with delete. Worked otherwise surprisingly well [!!].
9. To which extent did the system understand combinations of speech and pointing input?	She tended to forget that the number which failed to get inserted can pop up in an empty field somewhere else next time one points. System reacts sufficiently fast on speech and pointing. No disturbing delays.
10. How well did it otherwise work to use combined speech and pointing input?	Worked fine. Would work better if I became a routine user. One has to learn to speak in the right way to the system.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Would like to be able to click on something when there is a problem – one feels helpless, and this

	one is not used to. It would help to replace speech by pointing to a palette and then pointing to the board.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Easy.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Fine. Can be used for many purposes.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	80% nemt. 20% = 4 and delete. Otherwise very easy.
16. To which extent did you feel in control when playing the game?	80%.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Help if stuck: get the next number. The system should not find "deep" errors. Then it would be too easy. It's fine that it catches the evident errors. Misses the opportunity to insert possible numbers in the squares. Is used to doing this.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Rather fun. But mostly because it's new. Would hardly become addicted to it. The setup is more relevant for more movement-oriented games.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Plays only on paper. Pro: finds the evident errors. Con: paper Sudokus are easily portable and one can insert notes.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. Was quite fun. It's fun to play Sudoku. Would be fine to stand up and point in an airport where you otherwise sit a lot. More suited to fill out waiting time than for pure active entertainment.
21. Any other comments?	None. Would like a more physical action-oriented game.
Observations on this user	
Strong Danish accent. First game at level 2. 4 misrecognised as 5.	

Uses both arms – several subjects do that. Inserts a 1 by mistake and without noticing (the usual problem of numbers getting inserted after the subject has given up inserting them in the highlighted square) and suffers later on. Gets experimenter help to remove the problem. 4 misrecognised as 5 – cannot get the 5 removed. Problems with no. 1. Again problem with 4 – system inserts 5 (4 was wrong, actually). Works hard. The arm is up for long stretches of time. Asked to start new game. Aborts game and chooses new one. Doesn't know what to do when waiting for a new game to be loaded. New game: medium. People often points to a square constantly while repeating a number up to 6 times. Mostly they don't succeed in getting the number inserted. But it often pops into a different square when they have given up on it. Problems with delete this/that. Problem with no. 8. 6 and 2 end up in the wrong squares by mistake – the usual problem. This subject never completes a game.	

10.10 Subject 10

USER id: Subject 10	**Sudoku proficiency: some experience**
Name: Isa	**Start level: easy**
Age: 22	**Interview date: 12.6.2007**
Gender: female	**Interview time: 15.00**
Profession/education: 1st year science student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	4 Misses the opportunity to insert numbers in the margins.
2. How suited do you think spoken input is for games like the one you have just tried?	3
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	4
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Almost too good: it went too fast choosing to end the game.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	The arm tires when pointing at a square without succeeding to insert a number. When the arm is moved, the number ends up in the wrong place.
7. To which extent did the system understand what you said?	Fine. It was difficult two times.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	No.
9. To which extent did the system understand combinations of speech and pointing input?	OK. Sometimes it took longer than others.
10. How well did it otherwise work to use combined speech and pointing input?	See above.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Is used to writing in the margins when the Sudoku is difficult. Needs that for difficult games but not for the games played today.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Reasonably easy.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Quite fun. Reminds of EyeToy, which misses the speech, though. Fun to talk to the system.
15. How easy or difficult was it to play the game? Did	Rather easy.

you have any problems playing? Which?	
16. To which extent did you feel in control when playing the game?	Annoying that there was a square into which one couldn't insert a number. It resulted in wrong numbers in other squares. Otherwise I was in good control.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Margin number writing opportunity.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Fun to try. Not easy to bring with you in the pocket. More for entertainment when several are present.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Only tried the Internet a few times. You cannot take it with you. Cumbersome to drag and drop numbers on the Internet. It's easier to talk.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes, from curiosity. To kill time in a shop or at a train station.
21. Any other comments?	Would be good to avoid having to use the headset.
Observations on this user	
Chooses level 1. Plays slick and nice. Fast. Wins the game. Then chooses level 2. Tries "erase this". A field refuses to highlight. Red a couple of times – always a side-effect of moving the pointing hand. Problems with no. 1 which becomes no. 5. Wins the game. Then chooses level 2. Several numbers light up permanently. Problems with no. 7. Wins the game. Only user to complete 3 games.	

10.11 Subject 11

USER id: Subject 11	**Sudoku proficiency: beginner CHECK**
Name: Hanna	**Start level: beginner**
Age: +50	**Interview date: 12.6.2007**
Gender: female	**Interview time: 16.00**
Profession/education: medicine	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked first)	
1. How suited do you think pointing input is for games like the one you have just tried?	4 Requires that one holds the hand straight.
2. How suited do you think spoken input is for games like the one you have just tried?	4
3. How suited do you think screen output is for games like the one you have just tried?	3 Was slow sometimes.
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	4 Misses to be able to insert possible numbers in the squares.
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	Not every time. A bit annoying.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	My hand shook so the cursor flew into the wrong square. Tiring to stand with a straight arm for a long time.
7. To which extent did the system understand what you said?	Sometimes it didn't.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	See above. It didn't always react.
9. To which extent did the system understand combinations of speech and pointing input?	Reasonably. I succeeded to complete a game. I postponed the red problem for later.
10. How well did it otherwise work to use combined speech and pointing input?	See above.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	If game is difficult I need to be able to insert possible numbers in the margins.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Didn't understand the red in the beginning. Otherwise OK.
13. To which extent did you miss other forms of output (than via the screen)?	None.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	Pretty much OK. Realised very soon how it worked.
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?	The hardest part was to get the cursor into the intended square. A problem was to wait until the spoken number appeared in the square.

	The size of the screen is fine.
16. To which extent did you feel in control when playing the game?	Uncertain when nothing happened. Or when part of the board turned red.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	More information about errors. The possibility to add numbers in the margins.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Rather fun to try. But prefers paper.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Only tried Sudoku on the Internet a couple of times. Pro: red error message. Pro: easy to fully remove a wrong number. Con: no possibility to use the margin for writing numbers.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. I might. Fun to try.
21. Any other comments?	None.
Observations on this user	
Chooses level 1. Talks to herself in Danish -> turns a 7 into 4 -> red. Doesn't try to remove the error but plays on. Discovers the error rather late. Then chooses new game level 1. Problem with no. 2. Inserts 7 with imprecise pointing -> red. Does not fix the error immediately. Tries "remove ahmm this" – doesn't work. Tries to remove a fixed number and doesn't discover the impossibility. Gives up again and chooses new game level 1. Cannot insert no. 5. Points, like several other subjects, to a square for a long time trying repeatedly to state the number to be inserted. They mostly don't succeed. This time she completes the game. Then chooses level 2. Makes one of the few genuine surface errors observed in the trials: a no. 4 in a 3x3 which includes a 4. Tries to remove a fixed number and then manages to get rid of the wrong 4.	

10.12 Subject 12

USER id: Subject 12	**Sudoku proficiency: experienced**
Name: Henning	**Start level: medium**
Age: 31	**Interview date: 12.6.2007**
Gender: male	**Interview time: 17.00**
Profession/education: engineering student	**Interviewers: LD, NOB**
Interview question	**User answer**
Appropriateness of modalities used (Closed questions – asked later)	
1. How suited do you think pointing input is for games like the one you have just tried?	4
2. How suited do you think spoken input is for games like the one you have just tried?	4-5 – for Sudoku 3-4 - for chess
3. How suited do you think screen output is for games like the one you have just tried?	5
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?	4 Apart from the risk that people will talk over one another when the game is played in a public location, and interfere with the gaming. In such environments maybe 2-3.
Quality of interaction (Open questions)	
5. To which extent did the system understand you when you pointed at something?	80%. For the rest, it executed my input into the wrong square.
6. How well did it otherwise work to use pointing input? Were there any problems? Which?	The arm gets tired.
7. To which extent did the system understand what you said?	90% perhaps closer to 100%. Sometimes needed to repeat delete/remove this.
8. How well did it otherwise work to speak to the system? Were there any problems? Which?	Simple enough when you first understood the system. Most problems were with pointing.
9. To which extent did the system understand combinations of speech and pointing input?	Preferred to point and speak at the same time. This worked well enough.
10. How well did it otherwise work to use combined speech and pointing input?	No other problems.
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?	Misses to be able to insert possible numbers into the squares.
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?	Quite easy after the introduction. No problems.
13. To which extent did you miss other forms of output (than via the screen)?	No. Would be fun with a fanfare or YES when a game was successfully completed.
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?	OK. You have to get used to the absence of mouse or pencil. After that it's OK.
15. How easy or difficult was it to play the game? Did	The first game was quite easy.

you have any problems playing? Which?	Second game somewhat more difficult but went well when I first got going.
16. To which extent did you feel in control when playing the game?	Full control – apart from the numbers which ended up in the wrong place due to problems with pointing.
Functionality	
17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance: • anything missing from what you could do by pointing? • anything missing from what you could do by speaking to the system; • missing information on the screen; • other missing information. This is important for us to know in order to be able to improve the game.	Possible numbers in squares. Possibility to backtrack to a certain point, erasing the numbers inserted since then. Red help is OK. More help is likely to destroy game-playing.
User experience	
18. What do you think of solving Sudoku games in the way you just tried?	Quite fun. But I miss my newspaper with cartoons. I like to draw by hand. Fun to compete with others using this game, like an alternative to Dart or billiards.
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?	Played 5-6 times on the Internet. Pro: errors are marked immediately. Con: No possibility to insert numbers in squares/margins.
20. If you were to come across the system somewhere in a public space, would you play again? If yes: why? If no: why not?	Yes. Fun to try. Likes a quick competition. Loves chess. Is a happy gamer.
21. Any other comments?	When is the next test round?
Observations on this user	
Chooses level 2. Problem: delete this. Immediately chooses a new game. Thinks with his arm stretched. Changes arm. Proceeds rather well. Problems with 6 (lisps). Square doesn't light up, so 4 overwrites a correct number which was just put there. Happens several times. 5 ends up in a wrong square. Again problem with 6. Wins game. Chooses level 3. Disturbing with several highlighted squares next to one another: the one pointed at and a permanently highlighted one: where will the number end up?	

The game seems to get screwed up and he starts to change several numbers. He guesses (confirmed afterwards). And wins the game. During the games the majority of problems seem to be with the middle square in the left-most column.	

10.13 Interview Script, English Version

10.13.1 Input and Output

You should answer the following questions by simply picking a word or number from the following list: 1 = unsuited, 2 = rather unsuited, 3 = neither/nor, 4 = rather suited, 5= well suited. 1 (unsuited) is worst, 5 (well suited) is best, and 3 (neither/nor) is right in between. [**NOTE**: type this metrics on paper and hand it to the subject, or write the metrics on a blackboard which is clearly visible to the subject].

The questions are all about your input to the system and the system's output to you.

1. How suited do you think pointing input is for games like the one you have just tried?
2. How suited do you think spoken input is for games like the one you have just tried?
3. How suited do you think screen output is for games like the one you have just tried?
4. How suited do you think the combination of pointing and spoken input, and screen output, is for games like the one you have just tried?

10.13.2 Quality

The following questions are all about how you think it was to play the game and communicate with it.

5. To which extent did the system understand you when you pointed at something?
6. How well did it otherwise work to use pointing input? Were there any problems? Which?
7. To which extent did the system understand what you said?
8. How well did it otherwise work to speak to the system? Were there any problems? Which?
9. To which extent did the system understand combinations of speech and pointing input?
10. How well did it otherwise work to use combined speech and pointing input?
11. To which extent did you miss other ways of inputting information (other than pointing and speech)?
12. How easy or difficult was it to understand what to do from looking at the screen? Did you have any problems with some of what was shown on the screen? Which?
13. To which extent did you miss other forms of output (than via the screen)?
14. In general, how was it to use combined pointing and spoken input, and screen output, for interacting with the system?
15. How easy or difficult was it to play the game? Did you have any problems playing? Which?
16. To which extent did you feel in control when playing the game?

10.13.3 Functionality

17. Do you think the system offered you all the functions you need for playing Sudoku, or did you miss anything? If yes, what? For instance:
    - anything missing from what you could do by pointing?
    - anything missing from what you could do by speaking to the system;
    - missing information on the screen;
    - other missing information.

This is important for us to know in order to be able to improve the game.

10.13.4 User Experience

18. What do you think of solving Sudoku games in the way you just tried?
19. Comparing with traditional Sudoku games on paper, or possibly with games on the Internet, what do you think are the advantages and disadvantages of the game you just tried?
20. If you were to come across the system somewhere in a public space, would you play again? *If yes:* why? *If no:* why not?

10.14 Interview Script, Danish Version

10.14.1 Input og Output

De følgende spørgsmål skal du simpelthen besvare ved at sige et ord eller tal fra følgende liste: 1 = uegnet, 2 = temmelig uegnet, 3 = hverken/eller, 4 = nogenlunde velegnet, 5= velegnet. 1 (uegnet) er altså dårligst, 5 (velegnet) er bedst, og 3 (hverken/eller) er lige midt imellem. [**NOTE**: skriv metrikken ud på et stykke papir og give det til personen, eller skriv metrikken på en tavle som er klart synlig for personen].

Spørgsmålene drejer sig alle om dit input til systemet eller systemets output til dig.

1. Hvor velegnet finder du pegeinput til spil som det du lige har prøvet?
2. Hvor velegnet finder du talt input til spil som det du lige har prøvet?
3. Hvor velegnet finder du output på skærm i spil som det du lige har prøvet?
4. Hvor velegnet finder du kombinationen af pegning og talt input og skærm-output i spil som det du lige har prøvet?

10.14.2 Kvalitet

De følgende spørgsmål har alle noget at gøre med hvordan du synes det var at spille spillet og kommunikere med det.

5. I hvilket omfang forstod systemet det når du pegede på noget?
6. Hvor godt fungerede det derudover at bruge pegeinput? Var der problemer? Hvilke?
7. I hvilket omfang forstod systemet hvad du sagde?
8. Hvor godt fungerede det derudover at tale til systemet? Var der problemer? Hvilke?
9. I hvilket omfang forstod systemet kombinationen af tale og pegeinput?
10. Hvor godt fungerede det derudover at bruge en kombination af tale og pegeinput?
11. I hvilket omfang savnede du andre måder at indgive information på (dvs. at kunne bruge andet end pegning og tale)?
12. Hvor nemt eller svært var det at forstå hvad man kunne gøre når man så på skærmen? Havde du problemer med noget af det der blev vist på skærmen? Hvilke?
13. I hvilket omfang savnede du andre former for output (end via skærm)?
14. Hvordan var det generelt at bruge kombinationen af pegning og talt input og skærm-output i interaktion med systemet?
15. Hvor nemt eller svært var det at spille? Havde du nogen form for problemer undervejs? Hvilke?
16. I hvilket omfang synes du at du havde kontrol over spillet?

10.14.3Funktionalitet

17. Synes du at systemet gav dig alle de muligheder, du havde brug for for at kunne spille Sudoku, eller var der noget du savnede? I så fald hvad? For eksempel:
    - mangler mht. hvad du kunne gøre ved at pege;
    - mangler mht. hvad du kunne sige til systemet;
    - manglende information på skærmen;
    - anden information du savnede.

Dette er vigtigt for os at vide for at kunne forbedre spillet.

10.14.4 Brugeroplevelse

18. Hvad synes du om at løse Sudokuer på den måde du lige har prøvet?
19. Sammenlignet med traditionelle Sudokuer på papir eller evt. på Internettet, hvad synes du så er fordele og ulemper ved det spil du lige har prøvet?
20. Hvis du stødte på systemet et eller andet sted i det offentlige rum, ville du så spille igen? *Hvis ja:* Hvorfor? *Hvis nej:* Hvorfor ikke?

10.15 Interview Script, German Version

10.15.1 Eingabe und Ausgabe

Zum Beantworten der folgenden Fragen sollten Sie einfach die Nummer oder das Wort aus der folgenden Antwortliste verwenden: 1 = ungeeignet, 2 = eher ungeeignet, 3 = weder noch, 4 = eher geeignet, 5= sehr geeignet. 1 (ungeeignet) gilt als schlechteste Wertung, 5 (sehr geeignet) als beste, und 3 (weder noch) liegt genau dazwischen. [**HINWEIS**: Diese Liste sollte auf ein Blatt Papier geschrieben werden, das der Testperson geben wird, oder auf eine Tafel geschrieben werden, die für den Anwender gut sichtbar ist].

Die folgenden Fragen beziehen sich alle auf Ihre Eingaben in das System und die Ausgaben des System an Sie.

1. Wie geeignet finden Sie die Verwendung der Zeigegeste als Eingabe für Spiele wie gerade ausprobiert?
2. Wie geeignet finden Sie die Verwendung von Spracherkennung als Eingabe für Spiele wie gerade ausprobiert?
3. Wie geeignet finden Sie die Bildschirmausgabe für Spiele wie gerade ausprobiert?
4. Wie geeignet finden Sie die Kombination aus Zeigegeste, Sprache und Bildschirmausgabe für Spiele wie gerade ausprobiert?

10.15.2 Qualität

Die folgenden Fragen beziehen sich auf die Bedienung und die Kommunikation mit dem Spiel.

5. Inwieweit hat das System Sie verstanden, wenn Sie auf etwas gezeigt haben?
6. Wie empfanden Sie ansonsten die Verwendung der Zeigegeste? Gab es Probleme? Welche?
7. Inwieweit hat das System Sie verstanden, was Sie etwas gesagt haben?
8. Wie empfanden Sie ansonsten die Verwendung der Spracherkennung? Gab es Probleme? Welche?
9. Inwieweit hat das System die Kombination aus Zeigen und Sprechen verstanden?
10. Wie empfanden Sie ansonsten die Kombination von Zeigen und Sprechen?
11. Inwieweit haben Sie andere Eingabemöglichkeiten vermisst (außer Zeigen und Sprechen)?
12. Wie einfach oder schwierig war es, den Bildschirm und was zu tun ist zu verstehen? Hatten Sie irgendwelche Probleme mit dem, was auf dem Bildschirm gezeigt wurde? Wenn ja, was?
13. Inwieweit haben Sie andere Ausgabeformen vermisst (außer über den Bildschirm)?
14. Wie haben Sie im Allgemeinen die Kombination aus Eingabe durch Zeigen und Sprechen und der Ausgabe auf dem Bildschirm empfunden?
15. Wie einfach oder schwierig war es, das Spiel zu bedienen? Hatten Sie Probleme beim Spielen? Welche?
16. Inwieweit hatten Sie das Gefühl, das Spiel zu kontrollieren?

10.15.3 Funktionalität

17. Glauben Sie, dass das System Ihnen alle Funktionen bereit gestellt hat, um Sudoku zu spielen oder haben Sie etwas vermisst? Wenn ja, was haben Sie vermisst? Beispielsweise:
    - Funktionen, die man durch Zeigen auslösen kann;

- Funktionen, die man durch Sprechen auslösen kann;
- Fehlende Informationen auf dem Bildschirm;
- Andere fehlende Informationen

  Die Antworten hier sind für uns wichtig, um zu erfahren, wie man das Spiel verbessern könnte.

10.15.4 Anwender-Erlebnis

18. Was halten Sie davon, Sudoku so wie gerade ausporbiert zu spielen?
19. Was glauben Sie sind die Vor- und Nachteile dieses Sudoku-Spiels gegenüber dem klassischen Sudoku auf Papier oder der Möglichkeit, Sudoku im Internet zu spielen?
20. Würden Sie wieder spielen, wenn Sie das System an einem öffentlichen Ort vorfinden würden? Wenn ja, warum? Wenn nein: warum nicht?

10.16 User Screening, English Version

The questions below will be asked to potential test subjects in order to determine if we need their participation and in which test user category.

Name What is your name?

Age How old are you?

Gender Are you male or female?

Education What is your education??

Occupation Which kind of work do you do?

Knowledge of English Is your knowledge of English = none, modest, average, good, very good? **Game experience** Have you played Sudoku before?

If yes:

For how long have you played?

How much have you played?

Do you play regularly? *If yes:*

How often do you play?

Have you tried to play Sudoku over the Internet?

Game strength Describe the degree of difficulty of the games that you play and normally solve.

If no:

What makes you interested in participating in testing our Sudoku game?

Experience with similar systems

Have you tried systems which understand spoken input before?

If yes: Have you tried systems which understand spoken and pointing input before?

NOTE: Users with little or no Sudoku experience are characterised by only having played a little, if at all, not playing regularly, and not being good at solving Sudoku games other than quite easy ones.

10.17 User Screening, Danish Version

Spørgsmålene nedenfor stilles til potentielle testdeltagere for at kunne afgøre om der er brug for deres deltagelse og i hvilken kategori.

Navn Hvad hedder du?
Alder Hvor gammel er du?
Køn Hvad er dit køn?
Uddannelse Hvilken uddannelse har du?
Beskæftigelse Hvilket arbejde har du?
Engelskkundskaber Er dit kendskab til engelsk = intet, lidt, middel, godt, særdeles godt?
Spilerfaring Har du spillet Sudoku før?
Hvis ja:

I hvor lang tid?
Hvor meget har du spillet?
Spiller du regelmæssigt? *Hvis ja:*

Hvor ofte?

Har du prøvet at spille på Internettet?
Spilstyrke Beskriv sværhedsgraden af de spil du spiller og, som regel, løser.

Hvis nej:

Hvad er det der gør dig interesseret i at deltage i denne test af et Sudoku spil?

Erfaring med lignende systemer
Har du tidligere prøvet systemer som man taler til?
Hvis ja: Har du tidligere prøvet systemer som man taler og peger til?
NOTE: Brugere med lidt eller ingen Sudoku erfaring kendes på kun at have spillet lidt, om overhovedet, ikke at spille regelmæssigt, eller ikke at være gode til at løse Sudoku opgaver ud over de letteste.

10.18 User Screening, German Version

Die nachfolgenden Fragen werden potentiellen Testpersonen gestellt um zu ermitteln, ob ein Kandidat für den Test benötigt wird und in welcher Benutzer-Kategorie.

Name Wie ist Ihr Name?

Alter Wie alt sind Sie?

Geschlecht Sind Sie männlich oder weiblich?

Ausbildung Was ist Ihr Ausbildungsstand?

Beruf Was ist Ihr Beruf?

Englisch-Kenntnisse Beschreiben Sie Ihre Englisch-Kenntnisse: keine, wenig, durchschnittlich, gut, sehr gut?

Spielerfahrung Haben Sie schon einmal Sudoku gespielt?

Wenn ja:

Für wie lange haben Sie gespielt?

Wie häufig haben Sie gespielt?

Spielen Sie regelmäßig? W*enn ja:*

Wie oft spielen Sie?

Haben Sie schon einmal Sudoku im Internet gespielt?

Spielstärke Beschreiben Sie den Schwierigkeitsgrad der Spiele, die Sie normalerweise lösen.

Wenn nein:

Was interessiert Sie daran, unser Sudoku-Spiel testen?

Erfahrungen mit ähnlichen Systemen

Haben schon Systeme mit Spracheingabe ausprobiert?

Wenn ja: Haben Sie schon einmal ein System mit Spracheingabe und Gestenerkennung verwendet?

HINWEIS: Anwender mit wenig ohne ohne Sudoku-Erfahrung werden als Anwender eingestuft, die nur wenig gespielt haben, nicht regelmäßg spielen und höchstens sehr einfache Sudoku-Spiele lösen können.

11 Appendix 2. Treasure Hunt Game User Interviews

11.1 Subject 1

User details and background information	
1. What is your name?	Flemming.
2. How old are you?	40.
3. Gender?	Male.
4. How much can you see?	Weak-sighted. He closed his eyes throughout the interaction.
5. What is your education/occupation?	Datamatiker (a 2-years informatics education), IT consultant.
6. How often do you use a computer?	Whenever awake, all day
7. How often do you play computer games? How do you, e.g., do that?	Has played a lot periodically, using mouse and keyboard.
8. What (else) do you use a computer for?	All kinds of things.
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Interacts via GUI and speech synthesis. Has not tried haptic feedback.
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	Medium difficult. Bumped into things that he didn't know what was and wasn't told by the system: is this house 1, 2, 3, 4? "red house", "blue house", etc. would be useful. Would be good with zones like in "tampen brænder". Left the room because he (mistakenly) believed that the sound was right, so he clicked. Missed the white cane and more detailed information.
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) Surprisingly good 3D world. Wonderful and astonishing experience. Has not used such a device before. b) The haptic device gave the feedback it should but information was seriously missing on what it was that was touched (what the feedback meant) and what to do.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	Worked badly. The sounds for blue and red are very close. The sound differences should be more marked, like a bell, a wailing sound, etc. [NOTE: something one can conceptualise] Has not tried colour recognition by sound before.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which? b) How sufficient was it what was said?	a) Fairly Greek accent but there were only single words so it worked well enough. However, the spoken feedback should come faster, like immediately when you touch something. b) Sufficient and comprehensible.

	But there should be more information on what it is that one bumps into.
14. How difficult or easy was each of the tasks in the game? Please describe.	Difficult. There are too many details in the landscape. The point you need to hit in order to click is too small.
15. How easy or difficult was it in total to achieve the goal of the game?	Medium difficult. For example, you cannot tell whether you are on your way in or out of an area or whether you have been there before. There is a lot of surfing about for no particular reason. You don't know if a house has more than one room, nor whether you first visited a "red" house, then a "blue" one, then a "red" one. You need to get the information you would get if you were able to see. It may be an idea to introduce a kind of sonar that would work some way ahead.
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	Not well/badly. Didn't know what the path looked like: a dam, a ditch, or what. Feedback like "You are on the path" would be helpful when you plunge into the path. The haptic force feedback was helpful.
17. How was the communication with your partner? How could you imagine that the communication might be extended?	It's well indicated who has the initiative. However, the deaf person must become tired of waiting. Brief descriptions of what is going on at the other end would be good.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	Didn't miss other ways of communicating. Unless one knows what the partner is doing. Then one might, e.g., chat with the partner.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	It was easy.
20. To which extent did you feel in control when playing the game?	Not very much. For example, being inside the ruins felt like floating in the air. A wall around the ruins would be good.
21. Did you learn anything during the game which made you change the way in which you use the system?	Yes, after navigating randomly I became more systematic in uncovering areas.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Primarily information was missing. What is most needed is on-line help and feedback. The haptic device provides what it should.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	Has tried First Shooter and Half Life that border on this game quite well. Difficult to play these games due to the requirements to your eyesight.

24. How did you like the treasure hunting game?	The idea is good. The game play is a bit difficult due to the lack of information. Maybe it could be graduated into "baner", the first one including only boxy things, the second including other shapes. Then you could have levels.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	Advantages: the 3D world, the limited area, and the fantastic sense of space. Disadvantages: Not so many has a such a haptic device. The radius is limited. This could be fixed by adding a time factor on the movement so that you can decide how far and fast one can travel in a second. There are not many games for the robot right now.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Yes. It is a fun world. It is a different world with totally different demands on the sense of space compared to the traditional games.
27. Do you have any other comments?	None.

Observations on this user

Task 1 – Find red closet

Has problems understanding temple ruins.

Is told that he must go into a house to find the red closet.

Searches for a long time. Finds houses but no doors. Doesn't stick to the houses.

It turns out that he thinks that he must hear a beep before he clicks on a house. He is told-to just click to enter and that there are 4 houses one can enter and 4 one cannot enter. [In fact there are only 3 houses one can enter.]

Enters the right house but gets out through the door after clicking on it and after having clicked around and not found the closet. He only searched in one side of the room (where there is no red closet).

[Strangely, the system sometimes says "house" when the wall is touched from the inside, and sometimes it emits the tone for grey colour instead.]

Enters a new house – this time the one with the blue closet.

Now he sticks to the houses.

Exits and finally enters the house with the red closet again. Finds the closet and is told he needs to click.

This task took approx. 10 minutes.

Task 3 – Go to temple ruins

Goes to temple ruins and clicks (1 minute).

Gets a new scene. Asks after a while what he is supposed to find. Is told that he is looking for an obstacle.

Waits for something to talk to him. Searches for 5 minutes. Seems often to be searching beyond the visible part of the screen, which makes it difficult to help him navigate because one cannot tell where he is in virtual space.

Gets navigation advice until he finds the "inscription".

Task 5 – Go to catacombs

Explores the landscape while he waits for the message from the partner. He has to wait again after the beep signalling a partner message because the system takes 1-2 minutes to convert the partner's sign language sign to speech output (using Embrola synthesis).

At first there is no message on what he is supposed to do because the initial beep is not accompanied by any message.

After a long while (about a minute) the system says "catacombs" which he is told is a message.

Finds the catacombs after some time and clicks. He clicks a number of times before he manages to enter.

Again he gets some help to find a box. His clicking on, or near, it does not work to begin with.

Task 7 – Follow grooved paths

Moves the cursor around and doesn't at first discover the groove when passing over it. Is helped: try to find something you can follow and which goes in a certain direction. Seems to follow the groove but looses it again. Is helped, finds the groove and its end.

Gets the forest map and is told that he must imagine that the map is hanging on a wall (i.e., this is a – mostly – 2D object). He is told not to click this time but to find the path, go to its end, maybe then go to the other end (its start).

Finds the starting point but has difficulties following the path. Is told that the path has obstacles on it. Subject: this is when one thinks that the path ends.

Finds it after some time and with some help and then finds the treasure.

[The test with this first subject made us realise that we probably had to do a lot of coaching and explanation before and, mainly during, the game-play.]

11.2 Subject 2

User details and background information	
1. What is your name?	Betina.
2. How old are you?	32.
3. Gender?	Female.
4. How much can you see?	Nothing.
5. What is your education/ occupation?	Attending the Youth Group at the Institute for the Blind. Will probably start on a call centre education after the summer holidays.
6. How often do you use a computer?	Only while at the institute. Does not have a computer at home.
7. How often do you play computer games? How do you, e.g., do that?	Never before. This is the first time.
8. What (else) do you use a computer for?	Email, Internet, documents, etc.
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Uses keyboard and speech synthesis. Has never tried haptic feedback.
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	Medium difficult. Easy towards the end. Difficult among the houses. Tended to make too big movements with the haptic device. Learned that one has to be patient.
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) It worked reasonably WELL once you got used to it. Happy to try. Got gradually easier. b) The sufficiency of the haptic feedback was okay.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	[Audio output was switched off at the time she got into the houses due to a problem with repeated spoken output: the system starts repeating its latest spoken output indefinitely. This happened several times later. Lets call it the SPOKEN OUTPUT LOOP] Has not tried colour recognition by sound before.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which? b) How sufficient was it what was said?	a) She could understand it although she says that she is not good at English. b) The sufficiency of the spoken output was okay.
14. How difficult or easy was each of the tasks in the game? Please describe.	The first task was the most difficult one: to find the red cupboard. The last one was the easiest one: to find the end of the path.
15. How easy or difficult was it in total to achieve the goal	It was difficult. I made it difficult by not

of the game?	clicking fast. Fun to try.
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	That went well, didn't it? The haptic force feedback was helpful.
17. How was the communication with your partner? How could you imagine that the communication might be extended?	Fine. As long as you get to know what is going on at your own end, it is fine. Doesn't need to know what the partner is doing. Received enough information about her side of the game.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	Didn't miss anything.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	Everything is difficult the first time. The most difficult thing was to learn how to use the haptic device. It was not difficult to remember to click.
20. To which extent did you feel in control when playing the game?	Not so much. Had to have a little help.
21. Did you learn anything during the game which made you change the way in which you use the system?	Easier towards the end. Learned patience and to be more calm. Believed at the start that it was possible to be done better/faster.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Didn't miss anything.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	Hadn't tried any games before.
24. How did you like the treasure hunting game?	Fun to try.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	Doesn't know what to say.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Yes, I could become better.
27. Do you have any other comments?	None.
Observations on this user	
Task 1 – Find red closet Doesn't understand much English. Is told that she is in a village with houses in the centre and other things surrounding the village. Searches around for a long time without bumping into anything. Gets help: stick to the houses and try to get into one by clicking. The sound starts looping, repeating "cemetery" over and over. The system is restarted. The same happens again, now with "temple ruins" over and over, and the system is restarted once again. Hasn't understood that she needs to click to enter a house and probably also doesn't realise that a red closet must be found inside a house, so she is told. Gets some navigation advice to find the house. Enters the house with a red closet but clicks on the door and exits. Enters the house with the blue closet. The sound has been switched off since it continued to start looping so she is told about the colour.	

Gets help to exit the house.

Re-enters the house with the blue closet. Exits again.

Gets help to find the house with the red closet. Finds it with some help.

This task took about 20 minutes. The subject appears a bit tired from the effort.

Task 3 – Go to temple ruins

Finds the temple ruins pretty fast and goes there. Touches and clicks but does not respond by clicking fast enough. This must be done simultaneously.

Enters the ruins and gets some "tampen brænder" navigation help.

Finds the inscription but doesn't click.

After some attempts manages to find the inscription again and click.

Task 5 – Go to catacombs

Waits for quite some time while the input sign language from the deaf-mute is being processed.

Gets the message "catacombs".

Searches for a while and finds them with some help. She seems to remember from previous exploration that the catacombs are in – what a seeing person would call - the bottom-right corner of the screen.

Again it's difficult to help when you cannot see the cursor.

Enters after three clicks. The first two times she clicks too late.

Finds the box fast with some "tampen brænder" help and clicks.

Task 7 – Follow grooved paths

Gets a map.

Gets an explanation to think of the map as if it hangs on a wall. Is told that there is a path to follow.

Finds the path immediately but floats up in the air. Gets advice. Gets the forest map.

Finds the path rather easily with a little help. Is told the end-beginning story.

Goes to the end first, then to the beginning, and then to the end. Jumps obstacles easily.

11.3 Subject 3

User details and background information	
1. What is your name?	Nikolaj.
2. How old are you?	21.
3. Gender?	Male.
4. How much can you see?	Only light and dark.
5. What is your education/occupation?	May be going into a traineeship.
6. How often do you use a computer?	Every day at school (at the Institute for the Blind)
7. How often do you play computer games? How do you, e.g., do that?	Mostly when visiting somebody else but not that often. Likes to play. Uses keyboard and sound output.
8. What (else) do you use a computer for?	Internet looking for film, music, etc.
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Uses keyboard and sound. Has tried a mouse for three fingers and with a button. The mouse gave haptic feedback. Tried it with a car racing game where one could feel the surface differences, and a dart game.
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	Difficult in the beginning but one got used to it quite soon. You try patiently and find a way. Had to get used to clicking fast on the button.
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) Fine, except for the houses. Believed that these were at the top of the village. Had difficulty finding the temple ruins. Hasn't tried this kind of haptic device before. b) It was sufficient.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	Worked OK. But misses an introduction in English about what the game is about, so that you don't need to be helped during gameplay. Has not tried colour recognition by sound before but found it easy.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which? b) How sufficient was it what was said?	a) Depends on your English skills. Did not understand all the words in Greek English. But it was okay easy. Noticed background noise when the system said "tomb". b) It was okay. [NOTE: it's an error to only have a single click available in the final game step.]]
14. How difficult or easy was each of the tasks in the game? Please describe.	Okay, as long as you get instructions from the experimenter. But you don't know what to find. It should be said that you have to find "the box", so that you don't click on, e.g., the tomb.
15. How easy or difficult was it in total to achieve the goal	Not so difficult as one might have thought..

of the game?	
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	Good, except from where it stopped. You go from the end to the beginning and back again. Not clear what to do when the path ends. The force feedback was good.
17. How was the communication with your partner? How could you imagine that the communication might be extended?	The information from the partner was sufficient. It might be good if the partner could expand on the message sent: what was sent, etc., and if one could send messages to the partner asking for more information.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	Didn't miss anything.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	It was easy. Strange to get outside the plane by the temple ruins.
20. To which extent did you feel in control when playing the game?	The most difficult was to get past the houses and into them, and that you had to jump them to go beyond them. Had expected that the system would say "door".
21. Did you learn anything during the game which made you change the way in which you use the system?	Not really. Just walked around. Built a sort of map in the head during the game.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Didn't miss anything. The output is okay brief. It's good it says "house" and not "the right house" because then it would be too easy.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	The closest would be the car racing game and the dart game.
24. How did you like the treasure hunting game?	Okay, challenging.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	Very annoying that clicking at the starting point of the path means that you won't hear any congratulations when you find the treasure because you have already used your one click.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Yes, to see if you find the treasure again and get congratulations
27. Do you have any other comments?	Correct the errors [in particular the problem of not hearing any congratulation] and get the system out on the market.
Observations on this user	
Gets forceful instruction in clicking on everything following spoken system output. **Task 1 – Find red closet** Searches around. The output "temple ruins" start to loop and the system is restarted. Searches again. Works a lot outside the visible area on the screen. This should be delimited more strongly.	

Gets a lot of help to navigate to the area with houses.

Enters the right house.

Finds quite soon the red closet and clicks.

Does not seem to find it easy to follow instructions in going “up”, “left”, “right”, or “down” on the map. [Experimenter realised why later in the test.]

Identifies the instrument names himself.

Task 3 – Go to temple ruins

Looks for the temple ruins. Has probably forgotten where they were.

Gets navigation help. Must click many times before it works. Is told to hold the button down when clicking.

Enters the temple ruins.

Gets an explanation of what he is supposed to do.

Finds relatively soon the inscription.

Task 5 – Go to catacombs

After a while the message catacombs is returned.

Finds them quickly and enters.

Searches around.

Must click very many times on or around [it’s impossible to tell from looking at the screen] the right box before it opens. Is told to push towards the object while clicking. [Clicking gradually turns out to be quite complicated!]

Task 7 – Follow grooved paths

Gets a map.

Finds the path immediately and follows it with strong help.

Is told that he needs to click at the end of the path even though there is no sound output.

Gets the forest map.

Finds the path after searching a little.

Clicks at the beginning which results in silence when he reaches the treasure at the end, i.e. it doesn’t say “congratulations”. He was not warned that, with the forest map, you only have a single click available. But he did get the beginning-end story.

The game took 45 minutes.

11.4 Subject 4

User details and background information	
1. What is your name?	Anders.
2. How old are you?	23.
3. Gender?	Male
4. How much can you see?	Nothing. Was born blind.
5. What is your education/occupation?	Attends IT service education. His first time with a 3D game. Incredibly exiting and instructive.
6. How often do you use a computer?	Several times every day, also at home.
7. How often do you play computer games? How do you, e.g., do that?	Never – only back in the DOS days, with keyboard commands.
8. What (else) do you use a computer for?	Internet, news, email, net banking, installing software, installing speech programs for people, etc. Homepage maintenance.
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Keyboard + software enables mouse control from the keyboard. Sometimes needs the mouse button to enter. Output is speech synthesis. Uses Braille but is not happy to because his Braille input machine is old and doesn't work well with Windows. Has never tried haptic feedback. Incredibly exiting.
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	A little difficult to begin with. Has never tried to see using 3D. Had to get used to it but just needed a little time to explore the arm.
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) Good, positive, surprising, impressive. Would like to use the robot arm more. Has not tried to use such a haptic device before. Fascinated by its use for surgery training. b) It was okay.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	It was quite fun but a repeat function would have been good for reminders about the colour codes, if you forget them, e.g. via an additional button. Has not tried colour recognition by sound before.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which? b) How sufficient was it what was said?	a) A somewhat Greek accent. Better English pronunciation would be good. The most difficult to understand was "catacombs". b) Fine.
14. How difficult or easy was each of the tasks in the game? Please describe.	It was difficult to follow the walls to the catacombs. The others were fairly easy.
15. How easy or difficult was it in total to achieve the goal of the game?	Very easy. There is only one way to go and you cannot die. Imagines a long game version: consider to make it possible to save the game so that you can

	return later to where you were.
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	Okay once you found out. The 3D was totally new. The force feedback was very helpful. It should not shake at times.
17. How was the communication with your partner? How could you imagine that the communication might be extended?	Okay. Didn't pay so much attention to him. Communication is important, would be good to be able to communicate with the partner. It might be good to introduce speech recognition which could then be converted to sign language.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	Doesn't miss anything.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	Easy once you found out about the 3D. Difficult when you haven't tried 3D before.
20. To which extent did you feel in control when playing the game?	Reasonably.
21. Did you learn anything during the game which made you change the way in which you use the system?	Yes, once he found out there were walls, boxes, etc., he became more aware of what was in the surroundings.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Didn't immediately miss anything. It had the needed possibilities.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	Never.
24. How did you like the treasure hunting game?	Incredibly exciting and instructive with the 3D). Would like to act as subject again.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	No immediate comments.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Yes, it is just so exciting to experience that you could walk around in 3D and see things rather than just reading about them. 3D would make one much more eager to play. It's much better than to navigate with the arrow keys.
27. Do you have any other comments?	The SPOKEN OUTPUT LOOP must be corrected.
Observations on this user	
Seems to have a strong technical interest. "Cool VR for the sense of touch". "When will it be on the market?" Asks to have repeated the colour/sound instructions. Is told that there are houses in the centre of the village, that he must push and click, and that not all houses can open. **Task 1 – Find red closet** Searches around.	

Doesn't seem to understand that you can jump an obstacle by lifting the robot arm and put it down again.
Seems to get a bit impatient.
"Temple ruins" loop.
Starts at the houses but doesn't manage to enter.
Searches some more and then enters the right house.
Finds very soon the red closet and clicks.

Task 3 – Go to temple ruins

Goes quickly to the temple ruins.
Is told about how to use the robot arm.
Searches around and soon find the inscription but does not click.
Searches again for the inscription.
Falls off the plane.
Gets navigation help and finds it again.
Says that he has never used 3D haptics.

Task 5 – Go to catacombs

Gets the catacomb message after a while.
Searches for the catacombs.
Irritating that we cannot see the cursor for lengths of time.
Finds them after a while and enters.
Is told what he is supposed to look for: box not tomb.
Searches around and finds the wrong box.
Gets substantial navigation help to find the right box.

Task 7 – Follow grooved paths

Receives a map. However, the path is convoluted, does not lead to the goal and cannot be used. [Kostas says that it was a mouse problem when drawing the path.] This is explained to him and the experimenter clicks to continue to the forest map.
Gets help to pass obstacles.

11.5 Subject 5

User details and background information	
1. What is your name?	Nadia.
2. How old are you?	21.
3. Gender?	Female.
4. How much can you see?	Nothing.
5. What is your education/occupation?	Attends IT service education.
6. How often do you use a computer?	Every day, also at home.
7. How often do you play computer games? How do you, e.g., do that?	Does not play games.
8. What (else) do you use a computer for?	Email, Internet, school things, …
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Keyboard and spoken output. Has not tried haptic feedback before.
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	Somewhat difficult to begin with. Otherwise reasonably. It was weird that one could walk on the walls and the ceiling. It felt strange as if the temple ruins were above the houses. [It is!]
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) Worked reasonably well. The neck got a little stiff from holding the robot arm in mid-air. There was a little problem holding the arm so that the button was where you needed it. Has never used such a device before. b) Fairly good. Difficult to feel that the path was there. It felt as being a bit up in the air.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	This is just a question of learning and remembering. Couldn't remember which was which. Has not tried colour recognition by sound before.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which? b) How sufficient was it what was said?	a) Is not so good at English. Felt a little like a lego city. Misses people and some life in the city. b) Okay.
14. How difficult or easy was each of the tasks in the game? Please describe.	Not that difficult. Was introduced to how to use the game.
15. How easy or difficult was it in total to achieve the goal of the game?	Medium difficult. Was surprised when the game ended – already!?
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	It was difficult. The force feedback was helpful. Also this path was a bit up in the air.
17. How was the communication with your partner? How could you imagine that the communication might be	Got the information needed.

extended?	The communication could be extended by enabling the partners to write together if there is something they don't understand.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	Didn't miss anything, except chat functionality.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	Fairly easy when first you find out how the haptic device works.
20. To which extent did you feel in control when playing the game?	Could have been better.
21. Did you learn anything during the game which made you change the way in which you use the system?	Found out that she should not move so fast. Could use some more information during about what one has to do. [Response to a leading question: you want to know what the partner is doing?] yes.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Missed nothing.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	Never.
24. How did you like the treasure hunting game?	Quite fun and exiting. Totally new way in which to play.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	You look at the haptic device because you think the space is there. It is a funny experience. There might be more effects and sounds that, e.g., indicate the room you are in.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Yes, it is very exciting, a totally new way in which to play. Often the second time you try something it becomes easier and more fun.
27. Do you have any other comments?	None.
Observations on this user	
Due to circumstances beyond our control, we only had a total of 45 minutes with Subject 5. Due to observations made on the previous subjects, we not tell the subject that the robot arm is not a pen [i.e., something associated with a plane surface] and strongly stress spatiality and 3D in the use of the arm. **Task 1 – Find red closet** Searches around in the landscape. Gets navigation help. "Temple ruins" starts to loop and the system is restarted. Has great difficulty doing systematic search of the space, especially wrt. moving foreground-background. Gets quite a lot of navigation help and manages to enter the house with the red cupboard. Gets search help but clicks on the door and exits the house. [This happens to several subjects despite the fact that the door says "door". But if you search fast within the room there is a lot of spoken and sound output – grey sound + "house" from walls, ceiling and floor – so it is easy to miss it when the door says "door".]	

Gets help to enter the house again.

Eventually finds the red closet and gets help to click on it.

Task 3 – Go to temple ruins

At the start of this task, the subject expects to still be in the room with the red closet but no, miraculously, she is out in the open without having moved there herself.

Quickly finds the temple ruins.

Also soon finds the inscription after having received an explanation of what she is supposed to find.

Task 5 – Go to catacombs

Receives the catacomb message after a while.

Gets help to find the catacombs and enters.

Finds fairly soon the right box.

Task 7 – Follow grooved paths

At the start of this task, it is emphasised that the subject should not click until at the very end.

Gets a map.

Finds the groove fairly soon.

Exits the grooved path a number of times.

Eventually finds the path and follows it to the end.

Gets the forest map.

Finds the path quite soon with a little help.

[There is a deep mismatch between the landscape shown on the screen and what the subject experiences through haptics. The graphics and haptics output domains don't correspond at all. Look at this as an issue of perspective. In haptics, you are on the ground all the time unless you move upwards. One couldn't usefully show the landscape in this way in an image because all you would be able to see is the immediate foreground. So the image perspective is looking down onto the landscape at an occlusion-avoiding angle.]

11.6 Subject 6

User details and background information	
1. What is your name?	Susanne.
2. How old are you?	25.
3. Gender?	Female.
4. How much can you see?	Nothing.
5. What is your education/occupation?	Under education as IT teacher.
6. How often do you use a computer?	Every day, also at home.
7. How often do you play computer games? How do you, e.g., do that?	Very often. Plays many games. Uses keyboard, sound, many have spoken output from screen reader, joystick, mouse (rarely, only for experimental games). Has not tried spoken input for gaming.
8. What (else) do you use a computer for?	Internet, email, everything.
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?	Uses keyboard and spoken output. Has tried a car game (Topspeed2) with force feedback (gamepad).
What to measure	**How to measure**
Quality of interaction	
10. How was it to navigate in the city and landscape? Were there any problems? Which?	Actually rather easy. Should have used more structured search, like they tell you to. Needs to develop more structure [inside head]. It was clear when there was something you bumped into, including the boundaries of the space. Didn't understand what went on when it fell into nothing, e.g., in the temple ruins. This also happened with the map. Had to ask: Where am I – outside or inside the environment?.
11. Two questions concerning the haptic device: a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before? b) How sufficient was the haptic feedback?	a) Fine. New sensation to be in a room/space that felt real. Loves to move around and find things, learn new things. Has never tried such a haptic device before. b) Good, but sometimes it gave way too much.
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?	Tried colour recognition only a little because she found the cupboard quickly. Too little difference between red (oboe) and blue (flute). Use, e.g., drum and guitar instead. The closest to trying this before was a program that painted a sound picture of the colours in a painting so that you could "hear" the painting. Difficult, isn't good at this.
13. Two questions concerning spoken output: a) How was it to understand what the system said? Were there any problems? Which?	a)Could understand it (has an American boyfriend). b) Okay, no use for more.

b) How sufficient was it what was said?	If you are looking for something for a long time it would be good to get a hint, like "try that".
14. How difficult or easy was each of the tasks in the game? Please describe.	Not so difficult. Relatively simple and clear and in a not-so-large environment.
15. How easy or difficult was it in total to achieve the goal of the game?	Not so difficult. [When there were obstacles on the path] one just had to jump over [them] to see if there was a path. Loves this.
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?	Fairly well. Had to get used to being able to jump over things. It was clear where the boundary was. The force feedback was helpful.
17. How was the communication with your partner? How could you imagine that the communication might be extended?	OK. You [the experimenter] explained what went on. An explanation of what the partner is doing would be good. Would like to have a chat window to the partner.
18. To which extent did you miss other ways in which to communicate with the system or with your partner?	There might have been a chat window to the partner. There could be predefined messages about what the partner is doing and vice versa. This would not work in Counter Strike. There it would be irritating for the partner to both communicate what one is doing *and* play the game.
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?	Reasonably easy after the introduction with the cube. Gives a clear sense of 3D space. Did not try to turn the arm (anti) clockwise – what will happen?
20. To which extent did you feel in control when playing the game?	Reasonably good control. But then there was only one button. It would have been more difficult if there had been more buttons.
21. Did you learn anything during the game which made you change the way in which you use the system?	Is used to 2D navigation. Eventually got used to being able to move up and down and started to explore the area a bit more in this dimension.
Functionality	
22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance: • - anything missing from what you could do with the haptic device? • - anything missing regarding spoken output; • - other missing information. This is important for us to know in order to be able to improve the game.	Didn't miss anything apart from getting rid of the SPOKEN OUTPUT LOOP.
User experience	
23. To which extent did you try any similar games before or just something in the same direction?	Topspeed2 has force feedback. The genre reminds of Last Crusade with the computer as partner. Quake for the blind has up – down direction via sound but no joystick - is still mostly a 2D experience. This system is extremely good. Develop it further!

24. How did you like the treasure hunting game?	A bit boring. Not enough action. It depends on what you are used to playing.
25. What do you consider advantages and disadvantages regarding the way in which the game is played?	No disadvantages. It is good that two persons with different disabilities can play together. This makes the potential pool of game partners larger. It's something the blind have been missing because they tend to only meet other blind people and not to meet people with other disabilities.
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?	Maybe not exactly this game, but others, happily. This technique is really cool. You get a totally different experience and sense of space. Really exciting to try.
27. Do you have any other comments?	Is this technology being used for other purposes? [Yes, for, e.g., dental surgery.] Really cool. Hopes the technology development will continue so that this will not only be for games. Could perhaps be used to sense a graphical object and perhaps to draw with. Might be used to explain to a blind what something looks like. Or for reading pictures. There might be a menu on double clicking, e.g., with a choice to skip. You might use the cube [demonstrates before the game began] as an environment for creating things inside it. I can see images in my mind. I want one of these. The system could be really good for presenting visual information for the blind: to explain things by showing them to people.

Observations on this user

With this final subject we had a total of 75 minutes.

No pen introduction was given. Only the haptic device was shown. [All other subjects had started the practical part of the introduction by trying to move an ordinary pen among obstacles on a table.]

The subjects happily speaks English. She seems quite bright.

Gets a careful introduction to 3D. Gets the scene/landscape described.

"Cool".

[The sounds of the flute and the oboe are too similar.]

Task 1 – Find red closet

"Town hall" starts to loop and the system is restarted.

"Temple ruins" starts to loop and the system is restarted, and again.

Searches a little around.

Is helped to find a house. Enters the right house.

Finds very soon the red closet.

Task 3 – Go to temple ruins

Immediately finds the temple ruins, moves to cemetery, immediately goes back, and enters.

Quickly finds the inscription.

Task 5 – Go to catacombs

Receives the catacomb message after a while.

Quickly finds the catacombs and enters.

Is told that she is supposed to find a box.

Looks around and finds the right one, fast.

Has problems in making the clicking work. [like several other subjects]

Task 7 – Follow grooved paths

Gets a map.

Quickly finds the path and follows it. Discovers that one can jump obstacles and does it. Find the turn in the path immediately [several other subjects ran off the path there.]

Is told that she must click at the end of the path.

Gets the forest map.

Finds the path easily and follows it to the end, then follows it to the start and back to the end. She does this fast. "I love games".

11.7 Demo person

After session 2, we were waiting to see if we would get a third subject on this first day (we didn't). While waiting, we gave a demonstration of the game to a seeing person who works with the blind on assistive technologies.

The following records some observations made on this person's game-play.

It seems difficult for subjects to click in the right way so that something happens. Are they clicking too late relative to the sound/speech output that makes them click, or what?

Loops: "temple ruins", "house".

Finds red closet.

System is re-started.

Forgets to click upon "inscriptions" output in temple ruing.

"But I don't know what I am supposed to find".

Subject is told to just click on everything he meets.

Again "house" loop.

We realise that we need to tell the subjects that the map with the drawn path is "only slightly enhanced 2D) and must be searched in a very different way from the 3D landscape. The former is like a 2D map on the wall.

Subject likes the idea of collaborative gaming.

Has never tried haptic robot arm feedback.

Remarks that the cursor has delays in responding [it also jumps about erratically]. The blind cannot see this but the delay is due to the fact that the cursor receives minimal computer power in order to spend the power on the rest of the game. The 2.4 Giga computer is at its limits running the game.

Subject finds the idea of ubiquitous communication across disabilities "incredibly exiting".

11.8 Interview Script, English Version

Note that the following questions are meant for the blind user only. Some of the questions would have to be different when being designed for a deaf and mute user.

11.8.1 Background

1. What is your name?
2. How old are you?
3. Gender?
4. How much can you see?
5. What is your education/occupancy?
6. How often do you use a computer?
7. How often do you play computer games? How do you, e.g., do that?
8. What (else) do you use a computer for?
9. How are you used to interact with a computer? How are you used to input information and how are you used to receive information from the computer? Did you ever try haptic feedback?

11.8.2 Quality

The following questions are all about how you think it was to play the game and communicate with it.

10. How was it to navigate in the city and landscape? Were there any problems? Which?
11. Two questions concerning the haptic device:
    a) How did it work to use the haptic device? Were there any problems? Which? Did you use such a device before?
    b) How sufficient was the haptic feedback?
12. To which extent did it work to recognise colours via sound? Were there any problems? Which? Did you try such a thing before?
13. Two questions concerning spoken output:
    a) How was it to understand what the system said? Were there any problems? Which?
    b) How sufficient was it what was said?
14. How difficult or easy was each of the tasks in the game? Please describe.
15. How easy or difficult was it in total to achieve the goal of the game?
16. How did it work to follow the path to the treasure? To which extent was the physical resistance in the haptic device of any help?
17. How was the communication with your partner? How could you imagine that the communication might be extended?
18. To which extent did you miss other ways in which to communicate with the system or with your partner?
19. How easy or difficult was it to use the system to play the game? Did you have any problems when playing? Which?
20. To which extent did you feel in control when playing the game?
21. Did you learn anything during the game which made you change the way in which you use the system?

11.8.3 Functionality

22. Do you think the system offered you all the functions you need for playing the game, or did you miss anything? If yes, what? For instance:
    - anything missing from what you could do with the haptic device?

- anything missing regarding spoken output;
- other missing information.

This is important for us to know in order to be able to improve the game.

11.8.4 User Experience

23. To which extent did you try any similar games before or just something in the same direction?
24. How did you like the treasure hunting game?
25. What do you consider advantages and disadvantages regarding the way in which the game is played?
26. If you got the opportunity, would you then play this or a similar game again? *If yes:* why? *If no:* why not?
27. Do you have any other comments?

11.9 Interview Script, Danish Version

Note that the following questions are meant for the blind user only. Some of the questions would have to be different when being designed for a deaf and mute user.

11.9.1 Baggrund

1. Hvad hedder du?
2. Hvor gammel er du?
3. Køn?
4. Hvor meget kan du se?
5. Hvad er din uddannelse/beskæftigelse?
6. Hvor ofte bruger du en computer?
7. Hvor ofte spiller du spil på computeren? Hvordan foregår det fx?
8. Hvad bruger du computeren til (i øvrigt)?
9. Hvordan er du vant til at interagere med computeren? Hvordan er du vant til at indgive information og hvordan er du vant til at modtage information? Har du prøvet haptisk feedback?

11.9.2 Kvalitet

De følgende spørgsmål har alle noget at gøre med hvordan du synes det var at spille spillet og kommunikere med det.

10. Hvordan gik det med at orientere sig i by- og landskabet? Var der problemer? Hvilke?
11. To spørgsmål vedrørende robotarmen:
    a) Hvordan fungerede det at bruge robotarmen? Var der problemer? Hvilke? Har du brugt sådan en før?
    b) Hvordan var tilstrækkeligheden af den haptiske feedback?
12. Hvordan fungerede det at genkende farver ved hjælp af lyd? Var der problemer? Hvilke? Har du prøvet den slags før?
13. To spørgsmål vedrørende taleoutput:
    a) Hvordan var det at forstå hvad systemet sagde? Var der problemer? Hvilke?
    b) Hvordan var tilstrækkeligheden af det der blev sagt?
14. Hvor lette eller svære var de enkelte opgaver? Beskriv.
15. Hvor svær eller let var det som helhed at nå målet i spillet?
16. Hvordan fungerede det at følge stien til skatten? I hvilket omfang var den fysiske modstand i robotarmen til nogen hjælp?
17. Hvordan var kommunikationen med din partner? Hvordan kunne du evt. forestille dig kommunikationen udbygget?
18. I hvilket omfang savnede du andre måder at kommunikere med systemet på, eller med din partner?
19. Hvor nemt eller svært var det at bruge systemet til at spille? Havde du nogen form for problemer undervejs? Hvilke?
20. I hvilket omfang synes du at du havde kontrol over spillet?
21. Lærte du noget undervejs som fik dig til at ændre måden at bruge systemet på?

11.9.3 Funktionalitet

22. Synes du at systemet gav dig alle de muligheder, du havde brug for for at kunne spille spillet, eller var der noget du savnede? I så fald hvad? For eksempel:
    - mangler mht. hvad du kunne gøre med robotarmen;
    - mangler mht. talt output;

- anden information du savnede?

Dette er vigtigt for os at vide for at kunne forbedre spillet.

11.9.4 Brugeroplevelse

23. I hvilket omfang har du prøvet lignende spil tidligere eller bare noget i retning af dette?
24. Hvad synes du om skattejagtsspillet?
25. Hvad synes du er fordele og ulemper ved måden man spiller på?
26. Hvis du fik muligheden, ville du så spille dette eller lignende spil igen? *Hvis ja:* Hvorfor? *Hvis nej:* Hvorfor ikke?
27. Har du andre kommentarer?