Using Modality Replacement to Facilitate Communication between Visually and Hearing-Impaired People

Konstantinos Moustakas and Dimitrios Tzovaras
Informatics and Telematics Institute, Centre for Research and Technology Hellas

Laila Dybkjær and Niels Ole Bernsen
NISLab

Oya Aran
Idiap Research Institute

Using sign language, speech, and haptics as communication modalities, a virtual treasure-hunting game serves as an entertainment and educational tool for visually- and hearing-impaired users.

In recent years, there has been an increasing interest in human–computer interaction (HCI) for multimodal interfaces. Since Sutherland's SketchPad in 1961 and Xerox's Alto in 1973, computer users have long been acquainted with technologies other than the traditional keyboard for interacting with a system. Recently, with the desire for increased productivity, seamless interaction, immersion, and e-inclusion of people with disabilities, along with progress in fields such as multimedia, multimodal signal analysis, and HCI, multimodal interaction has emerged as an active field of research.

Multimodal interfaces are those encompassing more or other input and output devices than the traditional keyboard, mouse, and screen, and the modalities they enable. Various natural input modalities are now being applied.[1,2] These modalities include voice; gesture, face, and body movement; haptic interaction; and other physiological signals. A key aspect in many multimodal interfaces is the integration of information from several different modalities to extract high-level information nonverbally conveyed by users or to identify cross-correlation between the different modalities. This information can then be used to transform one modality to another. The potential benefits of such modality transformations to applications for disabled users are important because disabled users could access an alternate communication channel to perceive information that wasn't previously available.

Most existing aids for the visually impaired[3] detect visual cues captured from cameras and present them to the visually impaired using symbolic information in alternate communication channels, such as audio or vibrotactile feedback. Concerning aids for the hearing impaired,[4] most developments focus on the recognition and synthesis of sign language.[5] The aforementioned aids are designed to help disabled users perceive their environments and also potentially communicate with other people[2] by diminishing the effects of their disability. The importance of multimodal systems for the disabled has been outlined elsewhere.[6] Some approaches[7,8] tackle the problem of multimodal interfaces to provide accessible products and applications to the disabled. One method[7] uses haptics and aural feedback to provide navigation aids for the visually impaired, while another approach[8] uses sign language synthesis and enhanced imagery so partially sighted users can access multimedia content.

However, despite the proliferation of research in this area, the problem of the communication between visually- and hearing-impaired people is a special case of particular interest, where the aforementioned aids cannot be applied because these users don't share any common communication channel. This problem has been only theoretically dealt with in other work,[6] where a proposed scheme included vibrotactile sensors, cameras for 3D space perception, and

1070-986X/11/$26.00 © 2011 IEEE

Published by the IEEE Computer Society

algorithms for speech recognition and synthesis to aid communication between visually- and hearing-impaired people.

Our proposed framework consists of a situated communication environment designed to foster an immersive experience for the visually and hearing impaired. In situated communication, the partners interact in a real-time, shared environment.[9] Physical modalities, such as speech rather than haptic text, are used for the visually impaired, and sign language, rather than graphic text, is used for the hearing impaired because speech and sign language, unlike text, are situated communication modalities. Concerning the immersive experience of the environment and its objects, consistent object-to-object modality transformations are used instead of object-to-object-property transformations. For example, a visual path is transformed into a 3D grooved-line map instead of transforming the 2D path into its linguistic description. This framework performs situated and immersive modality replacement in the context of an interactive game to be played by visually- and hearing-impaired people with as much communicative bandwidth and immersiveness as possible.

To achieve the desired result, our situated-modality replacement framework combines a set of different modules, such as gesture recognition, sign language analysis and synthesis, speech analysis and synthesis, and haptic interaction, into an innovative multimodal interface for disabled users. Unlike this method, existing approaches focus mainly on the non-situated and symbolic presentation of information to the disabled.[6,7,10] In this article, the descriptions of the framework and platform are mainly used to contextualize our contribution.

The prospective value of such a system is huge due to the socioeconomic aspects of including people with disabilities in the information age, which modality-replacement technologies can make possible. It should be emphasized that the risk of exclusion is one of the largest threats for the disabled. The proposed system opens a new, previously unknown dimension for visually- and hearing-impaired people, and offers an interactive application where visually- and hearing-impaired users can entertain themselves while also getting familiar with novel interaction technologies. Potential applications include universally accessible phones, generic accessible computer interfaces, social networks, digital assistants, and so on.

System description

The basic concept of our system is the idea of modality replacement, which is the use of information originating from various modalities to compensate for the missing input modality of the system or the users. Figure 1 (next page) illustrates the envisioned system for computer-aided interaction of visually- and hearing-impaired users with the environment as well as with each other. Visual information about the environment has to be conveyed to the visually-impaired user via the haptic and/or the auditory channel, while communication and the acquisition of various semantic information can be performed using natural language. The hearing-impaired user acquires visual information using vision and communicates with other people using sign language.

A problem that stresses the high importance of modality replacement is that communication between a visually- and hearing-impaired user is not possible using physical means.[6] Ideally, as illustrated in Figure 1, a modality replacement system would be used to recognize all spoken language input of the visually-impaired user, convert it into sign language, and present it to the hearing-impaired user with an animated avatar. Similarly, sign language gestures would be recognized and converted into text and would then be synthesized into speech using text-to-speech synthesis techniques. The present work takes the first step toward the development of such interfaces. It's obvious that because multimodal signal processing is essential in such applications, specific issues such as modality replacement and enhancement should be addressed in detail.

Figure 2 presents the architecture of the proposed system, including the communication between the various modules used for integration of the system as well as intermediate stages used for replacement between the various modalities. The left part of the figure refers to the visually-impaired user's terminal, while the right refers to the hearing impaired user's terminal. All actions are controlled by the Computer Supported Cooperative Work game system.

The visually-impaired user interacts with the computer using haptics and speech. Haptic

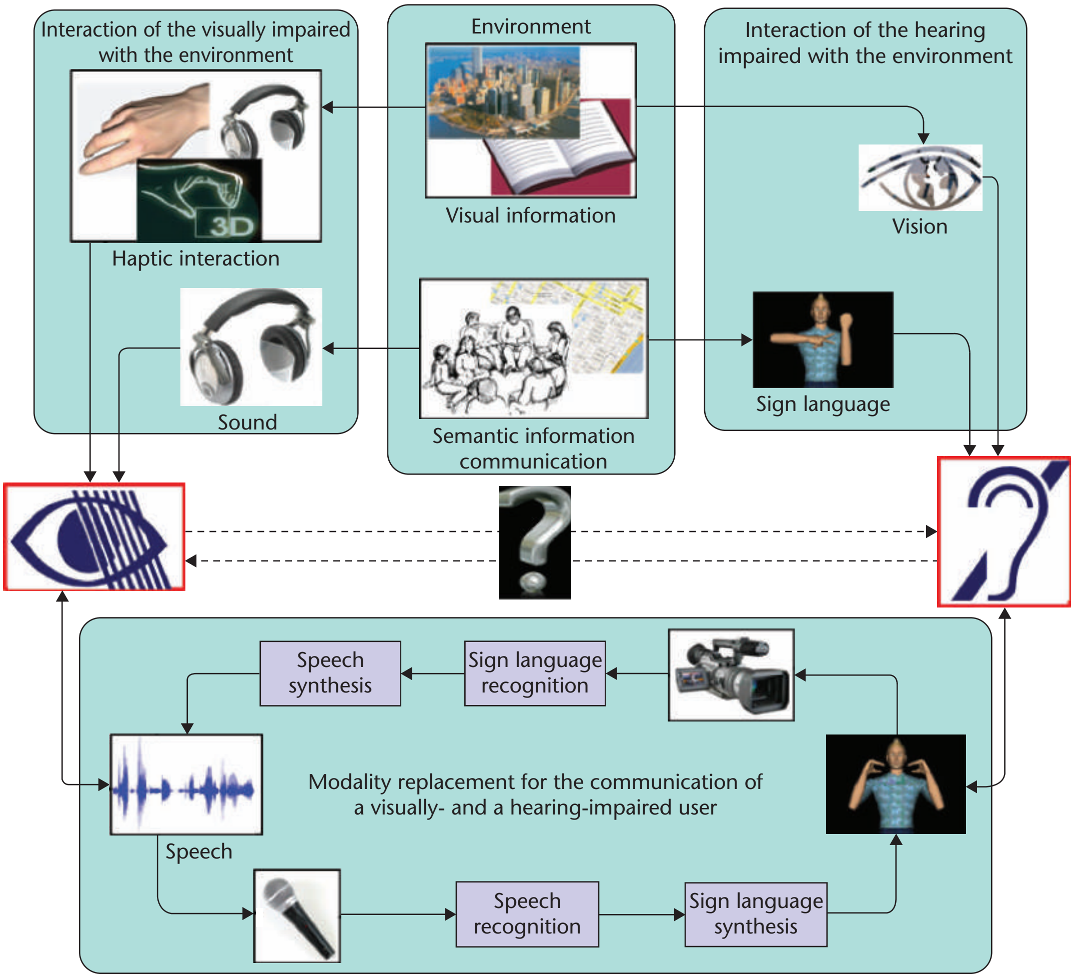


Figure 1. The proposed system for computer-aided interaction of visually- and hearing-impaired users with the environment as well as with each other, which until now has been physically impossible.

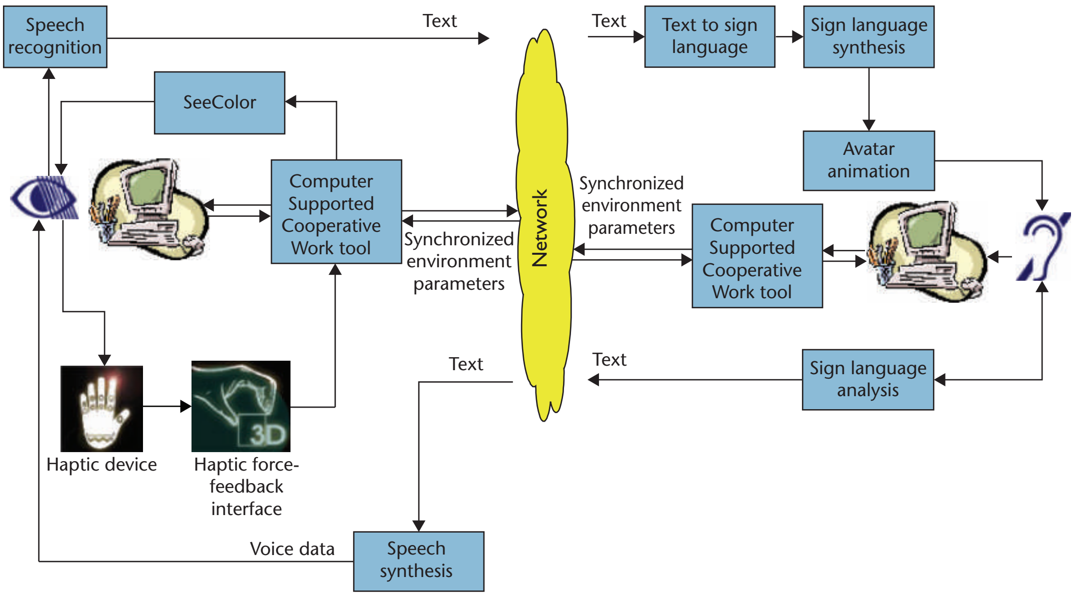


Figure 2. Architecture of the proposed system. The left part of the illustration refers to the visually-impaired user's terminal, the right to the hearing-impaired user's terminal.

interaction is used to navigate in the virtual environment and to process and recognize the objects of the virtual scene. System tasks and verbal information are presented to the visually-impaired user via speech synthesis, while verbal input from the visually-impaired user is perceived by the computer through speech recognition. The SeeColor utility enables the perception of colors through the sounds of musical instruments.[11]

At the hearing-impaired user's terminal, things are simpler because the user can use sight to navigate the virtual environment. To enable unobtrusive interaction, verbal information is presented to the user via sign language synthesis,[12] while the user can provide input to the system through the sign language recognizer.

Modalities and modules

Although the entire automated functionality of the system presented in Figure 1 would enable full, unobstructed communication between visually and hearing-impaired people, full automation is an ambitious goal. To our knowledge, this article presents the first attempt to develop a system where visually and hearing-impaired users cooperate in an interactive game, using advanced human–computer interfaces. Because the direct communication between visually and hearing-impaired persons is not possible, computer-mediated communications technologies attempt to translate modalities using modality-replacement technologies.

Haptic interaction

Haptic rendering is performed at every time-step of the haptic loop using the spring-dumper model.[13] Force feedback is provided using the Sensable Technologies Phantom Desktop haptic device; however, interaction using Immersion's CyberGrasp is also supported. The Phantom Desktop has 6 degrees of freedom for input (providing position and orientation) and 3 degrees of freedom for output (providing force feedback along the three axes of the Cartesian coordinate system). In particular, the force fed onto the haptic device is evaluated through the following formula:

$$F = k_s\mathbf{d} - k_d\mathbf{v}$$

where k_s and k_d are the spring and dumping coefficients, and $\mathbf{d}$ and $\mathbf{v}$ are the penetrating distance of the haptic probe into the 3D model and its velocity, respectively.

To provide realistic force feedback it's important to ensure that the force feedback loop runs at a frequency equal to or higher than 1 KHz. As a result, calculations for collision detection and force feedback are performed using the approach in Moustakas, Tzovaras, and Strintzis[14] that models the scene objects via analytic implicit surfaces and uses explicit formulae to compute the force feedback instead of calculating the force produced by triangle-to-triangle intersections in the virtual environment.

Navigation using grooved-line maps

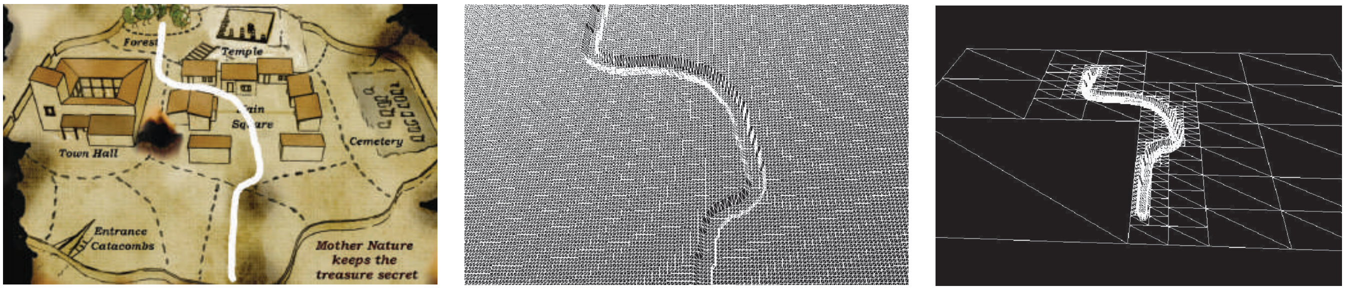
Apart from using haptics to recognize objects and navigate in specific scenes, the visually-impaired user can, in the proposed framework, navigate in road structures that are either originally included in the game or dynamically generated by processing[15] the hearing-impaired user's sketches. The road structures are implemented as grooved-line maps. A grooved-line map is a 3D terrain that is grooved in specific areas that represent streets or other meaningful areas that the visually-impaired user is able to perceive through a haptic device. Grooved-line maps are more efficient for haptic interaction than the traditionally used raised-line maps.[16]

In the context of the developed game application, the hearing-impaired user has to draw a path, during a particular phase of the game, to lead the visually-impaired user to the area where the treasure is located. The system uses as input the sketch made by the hearing-impaired user and converts it into a grooved-line map representation.[15]

Figure 3a (next page) illustrates the sketched image, while Figure 3b depicts the 3D grooved-line map. Because haptic rendering is a sensitive process and requires, for every time-step, computationally intensive collision detection that performs more slowly for larger 3D meshes, the grooved-line map is further processed to generate a multiresolutional grooved-line map as illustrated in Figure 3c. It is obvious that this map is more detailed in the areas close to the path and thus reducing the redundant complexity of the initial 3D map.

Speech recognition and synthesis

Speech recognition[17] and synthesis[18] is performed using off-the-shelf software. In the



(a) (b) (c)

Figure 3. (a) Example of a path sketched by the hearing-impaired user. (b) The 3D grooved-line map. (c) The multiresolutional 3D grooved-line map.

context of the proposed framework, the speech synthesizer can synthesize every word of the English dictionary, while the speech recognizer is trained to recognize a set of words and commands that are needed for the developed application scenario, that is, the interactive game.

SeeColor

The SeeColor utility is a limited version of the interesting system that is presented in Bologna et al.[11] It consists a stereo camera mounted on the user to capture the scene and recognize the color of the scene objects and their distance from the user. The color is presented to the user via sounds of different musical instruments, while the distance is rendered by altering the pitch of the synthesized sound.

In the context of the proposed framework, the camera is not needed because all the action takes place in a virtual environment. When the user touches a virtual object, a synthesized sound refers to the specific color of the object. This utility requires training so users can link the colors with the associated instrument sounds.

Sign language recognition

At the hearing-impaired user's terminal, the main unobtrusive modality for communication between user and system is sign language. The algorithm implemented in the present framework is based on that found in Aran et al.[10]

The first step in hand gesture recognition is to detect, track, and segment both hands. This is a complex task due to the frequent occurrence of interactions and occlusions between the two hands and the face. Moreover, other skin-colored regions, such as arms and neck, make the task even more complex. To make the detection and segmentation problem easier, users wear differently colored gloves on each hand.

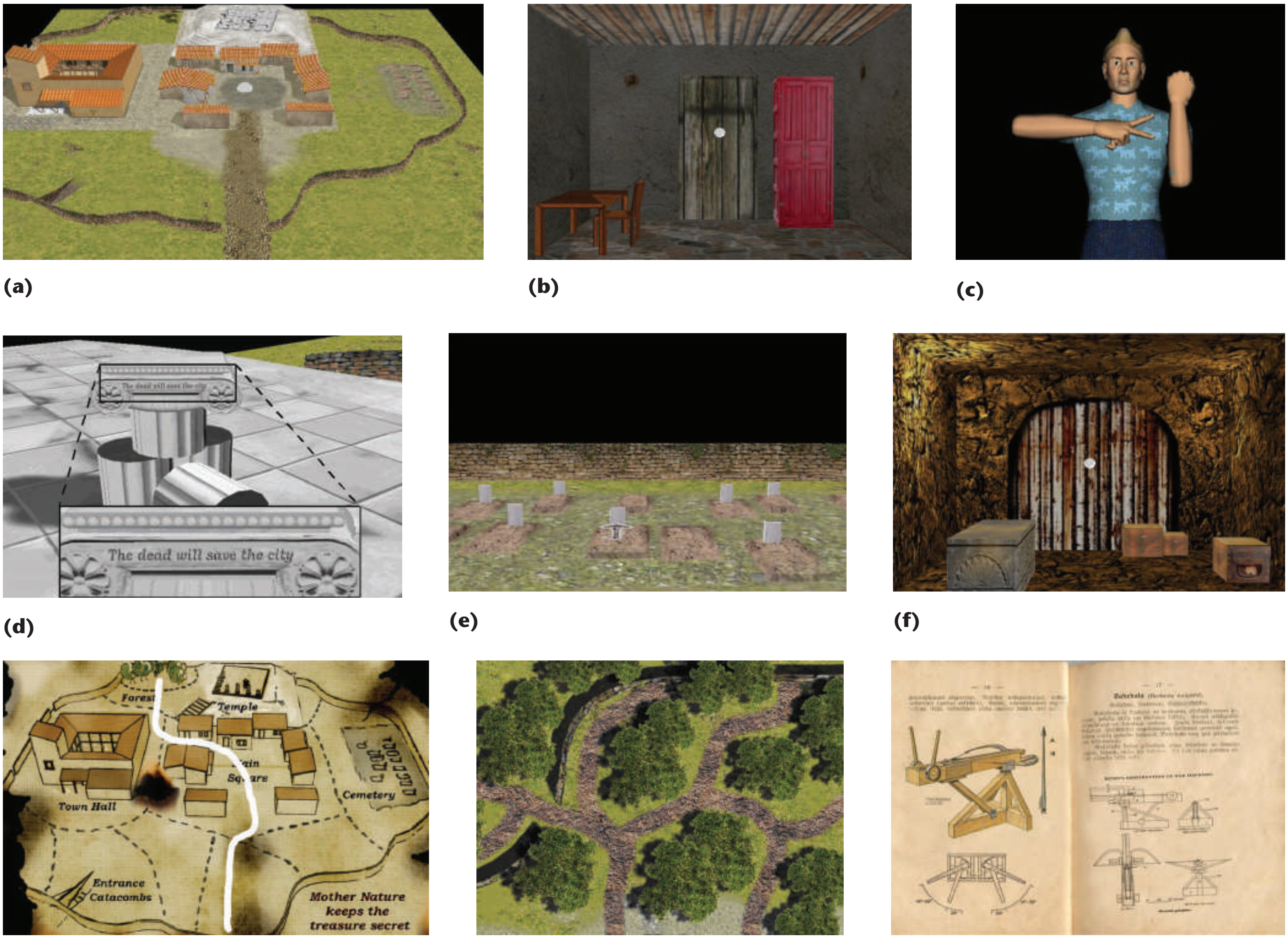
Once the hands are detected, simple hand-shape features are extracted and combined with hand motion and position information to obtain a combined feature vector. The classification of the signs is made via hidden Markov models (HMMs). A left-to-right continuous HMM model with no state skips is trained for each sign and the decision on the performed sign is made using the maximum likelihood criterion. For a test sign, the likelihood of each HMM is calculated and the sign class of the HMM that gives the maximum likelihood is selected as the class of the test example.

Five signs from American Sign Language are selected for use in the game application: map, exit, start, town, and cave. For each sign, 15 repetitions from two subjects are recorded. To measure the accuracy of the system, 10-fold cross-validation is performed. We achieved an average accuracy of 100 percent on the test set.

Sign language synthesis

The sign language synthesizer[12] uses humanoid animation models to provide the animation and creates the animations using as input Sign Writing Markup Language. The technique first converts all individual symbols found in each sign box to sequences of MPEG-4 Face and Body Animation Parameters (FAP and BAP). The resulting sequences are used to animate an humanoid-animation -compliant Virtual Reality Modeling Language (VRML) avatar using MPEG-4 FAP and BAP players. The system is able to convert all hand symbols as well as the associated contact and movement dynamics symbols contained in any Automatic Sign Language sign-box. Manual hand gestures and facial animations are currently supported.

This method has two significant advantages. It allows almost real-time visualization of sign language notation, thus enabling interactive



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 4. Screenshots of the treasure hunting game: (a) the virtual environment; (b) step 1, the red closet; (c) step 2, avatar performing sign; (d) step 3, in the temple ruins; (e) step 4, the cemetery; (f) step 5, the catacombs; (g) step 6, sketching the path to the treasure; (h) step 7, searching for the treasure in the forest; and (i) the treasure, design patterns for the war machines.

applications. In addition, avatars can easily be included in virtual environments using VRML, which is useful for a number of envisaged applications, such as TV newscasts, automatic translation systems for the hearing impaired, and so on.

Application scenario: the game

The aforementioned technologies are integrated in a multimodal edutainment game. We designed the game to be attractive but not take a long time to complete so that the players don't get tired. Moreover, to optimally use the communication between the players, the visually- and hearing-impaired users play in alternate steps. After each step, one user's terminal transmits a message to the other user's terminal and, whenever necessary, converts the message to another perceivable modality (for example, from sign language to speech). Moreover, the game is designed so that its tasks can be fulfilled only in the case of a proper communication between the users (map sketch, sign language recognition, and speech recognition) and a proper modality transformation of the conveyed messages into a perceivable form.

The game scenario consists of seven steps. For each step, one of the users has to perform one or more actions to pass successfully to the next step. The storyboard is about an ancient city (Figure 4a) that is under attack. Citizens of the city try to find the lost design patterns of high technology war machines to protect their city.

Table 1 (next page) summarizes the input and output information for each step of the game along with general actions that the users

Table 1. Summary of the gameplay.

Information type	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
Input	Audio message "Find red closet"	Animated sign "Go to the town hall"	Audio message "Go to the temple ruins"	Animated sign "The dead will save the city"	Audio message "Cave"	Map with riddle	Two grooved-line maps
Action	Search for the red closet in one of the houses	Go to the town hall and talk to the king	Find oracle inscription in temple ruins	Find the key in the cemetery. Sign its label	Find map in catacombs	Solve riddle and sketch path to treasure	Follow maps and find forest and treasure
Output	Audio message "Town hall"	Audio message "Go to the temple ruins"	Inscription (text) "The dead will save the city"	Key labeled "Catacombs"	Map	2D sketch on the map	Treasure
Modality replacement	Color into sound	Input audio into animated sign	3D spatial information into haptic representation	Input audio into animated sign. Recognition of user's sign	Input text to synthesized speech	Sketch on the map into grooved-line map	—

have to perform and the cross-modal transformations that have to take place. Note that haptic interaction with 3D space takes place in every step involving the visually-impaired user. The odd steps are played by the visually-impaired user and the even by the hearing-impaired user.

The first step involves the visually-impaired user, who receives an audio message informing him or her to find a red closet. The user starts from the initial point, which is the entrance of the village, and, using the haptic device, explores the village to find a red closet in one of the houses (see Figure 4b). In this step, the visually-impaired user has to use the haptic device to explore the 3D workspace. Furthermore, using the SeeColor module, the audio modality replaces the color modality and enables the visually-impaired user to select the correct closet and thus receive an audio message. The audio message is then sent to the hearing-impaired user and the second step of the game starts.

The second step involves the hearing-impaired user. The message comes in the audio modality and is converted from audio to text using the speech recognition tool and then to sign language using the sign synthesis tool. Finally, the user receives the message as a sign language sentence through a 3D interactive avatar as illustrated in Figure 4c. The message "Go to the town hall" sends the hearing-impaired user to the town hall of the city where the king provides additional information.

The third step involves the visually-impaired user who hears the message of the king "Temple ruins" and goes to the temple ruins, where the visually-impaired user has to search for an object that has an inscription written on it. One of the columns in the destroyed temple has an inscription written on it that states "The dead will save the city" as illustrated in Figure 4d. The visually-impaired user is informed by an audio message whenever he or she finds this column and the message is sent to the hearing-impaired user's terminal.

The fourth step again involves the hearing-impaired user, who receives the written text in sign language form. The text modality is translated into sign language symbols using the sign synthesis tool. Then the hearing-impaired user has to figure out the meaning of the inscription "The dead will save the city" and navigate to the cemetery (see Figure 4e). There he or she should search for a key that lies in one of the graves. The word "Catacombs" is written on the key. The hearing-impaired user has to perform the sign "Catacombs" to enable the visually-impaired user to understand that the key opens a box in the catacombs. This sign is recognized by the sign language recognition tool and the text result is sent to the system to trigger the next step of the scenario.

In the fifth step, the visually-impaired user receives the text, which is converted to audio using the text-to-speech tool. This step involves haptic and audio information. The user has to search for the catacombs, enter them and find the box that contains a map (see Figure 4f). The map is then sent to the next level.

In the sixth step, the hearing-impaired user receives the map and, after solving the riddle "Mother nature keeps the treasure secret," has to draw a route to the area where the treasure is hidden (see Figure 4g). The route is drawn on the map and it's converted to a grooved-line map representation, which is sent to the visually-impaired user.

In the final step, the visually-impaired user receives the grooved-line map and has to find and follow the path to the forest where the treasure is hidden. The visually-impaired user can feel the 3D grooved map and follow the route to the forest. The visually-impaired user is asked to press the key of the Phantom device after reaching the target. Finally, after finding the forest, the user gets a new grooved-line map (see Figure 4h) where he or she has to search for the final location of the treasure. After searching in the forest by following the paths, the visually-impaired user should find the treasure that is illustrated in Figure 4i.

Table 2. What was measured during evaluation of the proposed framework, and how.

	How was it measured	
What was measured	**Interview**	**Interaction**
Quality of interaction		
1. Navigation in city and landscape	Yes	Yes
2. Haptic interaction	Yes	Yes
3. Color recognition via sound	Yes	Yes
4. Use of sign language recognizer	Yes	Yes
5. Use of sign language synthesizer	Yes	Yes
6. System output understanding	Yes	Yes
7. Ease of individual tasks	Yes	Yes
8. Ease of achieving game goal	Yes	Yes
9. Ease of following path	Yes	Yes
10. Ease of interaction	Yes	Yes
11. Feeling in control	Yes	Yes
12. Learning	Yes	Yes
User experience		
13. Tried something similar before	Yes	No
14. Like the game?	Yes	No
15. Advantages and disadvantages	Yes	No
16. Play again?	Yes	No
17. Other comments	Yes	No

Usability evaluation

The terminal of the visually-impaired user is by definition more complex in terms of HCI utilities than the terminal of the hearing-impaired user, because interaction with the computer must be augmented with navigational tools that enable the visually-impaired user to search the 3D environment.

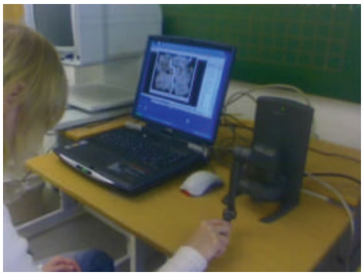
Therefore, the evaluation of the proposed framework, through the evaluation of the treasure hunting game, focused mainly on the visually-impaired user's terminal by analyzing and evaluating several aspects of the provided HCI tools. The usability of the hearing-impaired user's terminal has been evaluated independently in terms of the efficiency of the sign language recognition and synthesis methods. It should be emphasized that the functionality of the two terminals can be independently evaluated because the developed system is a network game, where the visually- and hearing-impaired users could be separated, playing from any location. There is no need for copresence.

Goals and setup

The overall goal of the evaluation of the proposed framework is to explore its usability and provide input on usability aspects, in particular regarding appropriate use of modalities, offered functionality, ease of use, and user satisfaction. Table 2, which gives a brief overview of what was measured and how, includes data from the usability test. All table items were addressed in post-test interviews. For some items, information was acquired from the collected interaction data and subsequent analysis.

The test regarding the visually-impaired user's terminal took place at the Institute for the Blind in Copenhagen. Six visually-impaired users participated in the evaluation that comprised introduction to and interaction with the treasure hunting game (see Figure 5a) followed by an interview. Details on the users and their familiarity with computers are presented in Table 3.

The average duration of the sessions lasted approximately 90 minutes. Each visually-impaired user was first given an introduction to the system and received training in interacting with it, in particular training in using the haptic device and information on the game scenario and the other modalities that are used. It was clearly emphasized that this was not a test of the user's skills at all, but a test of the



(a)

(b)

Figure 5. Using the system during usability evaluation. (a) Visually-impaired user's terminal: navigating the virtual environment with the Phantom desktop. (b) Hearing-impaired user's terminal: performing a sign.

goodness and interestingness of the proposed multimodal framework.

The collected data included the following:

- Video and audio recordings of user interaction with the system. The video shows the user's hand and arm, the haptic device, and the screen content.
- Observation notes produced by two observers during the sessions.
- Two sets of interview notes written during the interviews with the users.

The usability of the sign language recognition system was evaluated independently. The users perform the sign in front of a camera, wearing differently colored gloves on each hand, and the recognition result is given as feedback to the user. Moreover, the users see the synthesized versions of the signs that they perform.

The subjects testing the system part for the hearing impaired were six volunteer students taking an introductory Turkish Sign Language course given in Bogazici University (see Figure 5b). Two of the students (one male and one female) were from the computer engineering department and the rest were from the foreign language education department. Additionally, the lecturer of the course, who is a native signer, was also a subject in the user study.

Evaluation of the visually-impaired user's terminal

On the basis of observation of the visually-impaired users' interaction with the system, analysis of the video recordings made during interaction, and analysis of the subsequent interviews, the main conclusions are as follows. In general, the subjects liked the game and would play again if they had the chance. Only one found that gameplay did not offer a sufficient amount of action. She was probably the one with the most game experience among the subjects. However, all subjects were excited about the 3D haptic interaction, which none of them had tried before. On the basis of these interview comments, we believe that subjects liked the game due mostly to their strong appreciation of the realistic 3D experience and the new way to interact with a computer.

Basically, the input and output modality combination for the visually-impaired user—spoken keywords output, nonspeech sound

Table 3. Subject profiles.

Subject	Age	Gender	Occupation	Computer experience	Plays games
Visually-impaired subjects					
1	40	Male	IT consultant	High	Rarely
2	32	Female	Call center	Moderate	Never
3	21	Male	Practician	High	Rarely
4	23	Male	IT education	High	Never
5	21	Female	IT education	High	Never
6	25	Female	IT education	High	Often
Hearing impaired subjects					
1	27	Female	Lecturer	High	Rarely
2	25	Female	Student	High	Never
3	25	Male	Student	High	Rarely
4	21	Female	Student	Moderate	Never
5	21	Female	Student	Moderate	Never
6	25	Male	Student	Moderate	Rarely
7	23	Female	Student	Moderate	Never

Table 4. Performance times (in minutes) both cumulative (left) and on each game step (right) for the six subjects.

Finds	User 1		User 2		User 3		User 4		User 5		User 6	
Closet	16	16	18	18	12	12	4	4	11	11	6	6
Inscription	22	6	23	5	18	6	10	6	12	1	9	3
Box	28	6	29	6	26	8	19	9	18	6	14	5
Forest	33	5	33	4	29	3	—	—	22	4	17	3
Treasure	38	5	35	2	34	5	24	2	23	1	18	1

output, haptic 3D force-feedback output, and haptic 3D navigation input—appears to be fine. This is positive because this modality combination, in particular the 3D haptics input and output, is central to the underlying idea that is being tested through the game.

However, there is still plenty of room for improvement regarding game content and its presentation to the user. The test subjects rated game difficulty (navigation and ease of achieving tasks and goal) as being moderate, and in particular the first two subjects didn't feel in control while later subjects felt reasonably in control. This difference probably reflects the increasing amount of guidance and help given to subjects during gameplay.

Table 4 shows how subject performance speeds up as a function of the amount of real-time help provided by the experimenter and other test staff. The table also shows how each subject tends to complete each subsequent step more quickly than the previous one, which probably reflects progress in learning haptic navigation and in building a mental map of the townscape. Notice that the left column for each subject and each action refers to cumulative time, while the right refers to absolute time. The missing number for subject 4's reaching the forest in which the treasure was hidden was due to the fact that subject 4 received an incorrectly drawn grooved map and had to be brought into the forest by short-cutting the game.

In general, the spoken output was found to work well and all subjects found it sufficient. However, some subjects nevertheless mentioned that it would have been good to have more information about what it is that you bump into or get a hint if you seem to be lost, clearly suggesting that the spoken output was not entirely sufficient after all.

Subjects needed help to recognize colors via sound. Only one subject seemed to find it easy to apply the musical color codes. The other subjects found it difficult to remember which instrument was associated with which color. In particular, two of the colors were represented by instruments that sounded very similar.

Haptic interaction was received positively by all subjects. However, because none of the subjects had used a similar force-feedback haptic device before, and despite trying it briefly before they started the game, it clearly took them some time to get reasonably used to it. Navigation is expected to become easy eventually, but only after several hours of practicing with force-feedback devices. It's symptomatic that when subjects were asked if they learned something during gameplay, they primarily mentioned issues referring to 3D exploration by haptic interaction. Similarly, when asked about ease of interaction, nearly all comments made by the subjects concerned the 3D haptics and orientation in 3D space, and the fact that it takes time to become familiar with how to move around.

Thus, some of the users' difficulties during gameplay might be ascribed to lack of practice. For example, opinions were somewhat divergent regarding how easy it was to follow the grooved path as part of the final task. Most subjects mentioned that the grooved-line representation of paths, and eventually of maps, is helpful and interesting. Others said that even if following the grooved-line path is easy, it is sometimes difficult to locate it in the map in the first place.

Overall, the conclusion is that the underlying idea, the technology, and the basic modality combination are well-suited for the purpose of the game, even if game contents and their presentation could be improved.

Evaluation of the hearing-impaired user's terminal

Concerning the hearing-impaired user's terminal, the subjects found the interaction through the camera fairly easy in general. They especially liked the idea of touchless interaction. An important requirement of the interaction is that the camera view should contain the upper body of the user. Some of the subjects, in their first interaction, had difficulties perceiving the bounds of the camera view but got used to it quickly in their second trial.

Concerning the synthesized sign animation, most of the subjects considered the possibility of receiving messages from computers via sign language promising, but they also commented that the speed of the animation could be increased.

To measure the learning effect, the experiments were performed in two sessions. In each session, subjects were asked to practice three different signs. The second session was conducted with a time lapse of at least one week after the first. Before the first session, the interface was presented to the subjects in terms of a small demonstration. In the second session, the subjects were expected to start using the system without prior presentation. The user study shows that the subjects are slower in their first session because they spend some time learning to use the system. In their second session, where the average completion time decreases by half, subjects recall the usage and perform tasks more quickly.

None of the subjects tried a similar system before, where the user's motions are recorded through a camera and automatically recognized and the computer responds using sign language synthesis. Although three of the subjects were expert computer users and the remaining subjects were intermediate level users, it became clear during the user study that the computer usage level has little effect on performance, as interacting through a camera is extremely different from interacting through a typical mouse and keyboard interface.

The subjects made positive comments about the system in general. They found the interface user friendly and the system easy to use. One of the subjects found the decisions of the system very sensitive to small variations in signing and she therefore noted that the decision-making system could be more tolerant to variations of a single sign.

Discussion

The main conclusions on the test of the system are that the input and output modality combination, that is, spoken keywords output, nonspeech sound, sign language synthesis output, haptic 3D force-feedback output, haptic 3D navigation, and sign language analysis input, is considered useful. Visually-impaired users were generally excited about the 3D experience they had from the haptic interaction, while the hearing-impaired considered the sign language analysis and synthesis a helpful tool for interacting with the computer. The system is acceptable overall. Haptic output, that is, what the user feels when moving the arm of the haptic device, was fine. It gave a good 3D experience.

More specific user comments include the suggestion to include a small tutorial on how to use the haptic device in the 3D environment and to make the system more self-explanatory via oral guidance for visually-impaired users. More oral feedback concerning the current position and what is being touched would also be helpful. Hearing-impaired users suggested that automatic feedback could be given by the system if the user is out of the camera view.

Conclusions

We consider it important that even if the application scenario (game) can be further enriched, the interaction framework and all its novel technologies was well received by all users and could be used as the basis for further development of such games or interaction systems. Concerning the applicability and future potential of the proposed system, the users were asked about their impression of communicating with other users to elicit information. Users were asked how it was to communicate with their partner and how they imagined that communication could be extended. All users said they found the advancements fostered by this proposed framework exciting. In particular, a visually-impaired user mentioned that the blind and the deaf tend to socially interact with similarly impaired people, while further development of the proposed system could open entire new possibilities for the visually impaired to have social interaction with the hearing impaired. **MM**

References

1. J. Lumsden and S.A. Brewster, ''A Paradigm Shift: Alternative Interaction Techniques for Use with Mobile & Wearable Devices,'' *Proc. 13th Ann. IBM Centers for Advanced Studies Conf.,* IBM Press, 2003, pp. 97-100.
2. O. Lahav and D. Mioduser, ''Exploration of Unknown Spaces by People Who Are Blind, Using a Multisensory Virtual Environment (MVE),'' *J. Special Education Technology,* vol. 19, no. 3, 2004, pp. 15-24.

3. T. Pun et al., ''Image and Video Processing for Visually Handicapped People,'' *Eurasip J. Image and Video Processing,* vol. 2007, article ID 25214, 2007.
4. A. Caplier et al., ''Image and Video for Hearing Impaired People,'' *Eurasip J. Image and Video Processing,* vol. 2007, article ID 45641, 2007.
5. S.C.W. Ong and S. Ranganath, ''Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning,'' *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 6, 2005, pp. 873-891.
6. N. Bourbakis, A. Exposito, and D. Kobraki, ''Multi-Modal Interfaces for Interaction-Communication Between Hearing and Visually Impaired Individuals: Problems and Issues,'' *Proc. 19th IEEE Int'l Conf. Tools with Artificial Intelligence,* IEEE Press, 2007, pp. 522-530.
7. D. Tzovaras et al., ''Design and Implementation of Haptic Virtual Environments for the Training of Visually Impaired,'' *IEEE Trans. Neural Systems and Rehabilitation Eng.,* vol. 12, no. 2, 2004, pp. 266-278.
8. H. Petrie et al., ''Universal Interfaces to Multimedia Documents,'' *Proc. 4th IEEE Int'l Conf. Multimodal Interfaces,* IEEE CS Press, 2002, pp. 319-324.
9. N.O. Bernsen and L. Dybkjær, *Multimodal Usability,* Human–Computer Interaction Series, Springer, 2010.
10. O. Aran et al., ''Signtutor: An Interactive System for Sign Language Tutoring,'' *IEEE MultiMedia,* vol. 16, no. 1, 2009, pp. 81-93.
11. G. Bologna et al., ''Transforming 3D Coloured Pixels into Musical Instrument Notes for Vision Substitution Applications,'' *Eurasip J. Image and Video Processing,* special issue on image and video processing for disability, vol. 2007, article ID 76204, 2007.
12. M. Papadogiorgaki et al., ''Gesture Synthesis from Sign Language Notation Using MPEG-4 Humanoid Animation Parameters and Inverse Kinematics,'' *Proc. 2nd Int'l Conf. Intelligent Environments,* IET, 2006.
13. G.C. Burdea, *Force and Touch Feedback for Virtual Reality,* Wiley-Interscience, 1996.
14. K. Moustakas, D. Tzovaras, and M.G. Strintzis ''Sq-map: Efficient Layered Collision Detection and Haptic Rendering,'' *IEEE Trans. Visualization and Computer Graphics,* vol. 13, no. 1, 2007, pp. 80-93.
15. K. Moustakas et al., ''Haptic Rendering of Visual Data for the Visually Impaired,'' *IEEE MultiMedia,* vol. 14, no. 1, 2007, pp. 62-72.
16. R. Ramloll et al., ''Constructing Sonified Haptic Line Graphs for the Blind Student: First Steps,'' *Proc. ACM Conf. Assistive Technologies,* ACM Press, 2000.
17. S. Young, *The HTK Hidden Markov Model Toolkit: Design and Philosophy,* tech. report CUED/F-INFENG/TR152, Cambridge Univ. Engineering Dept., Sept. 1994.
18. B. Bozkurt et al. ''Improving Quality of Mbrola Synthesis for Non-Uniform Units Synthesis,'' *Proc. IEEE Workshop Speech Synthesis,* IEEE Press, 2002, pp. 7-10.

Konstantinos Moustakas is a postdoctoral research fellow in the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece. His research interests include virtual reality, haptics, computer vision, and stereoscopic image processing. Moustakas has a PhD in electrical and computer engineering from the Aristotle University of Thessaloniki, Greece. He is a member of the IEEE and the IEEE Computer Society. Contact him at moustak@iti.gr.

Dimitrios Tzovaras is a senior researcher in the Informatics and Telematics Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece. His research interests include multimodal interfaces, virtual reality, and 3D content-based search. Tzovaras has a PhD in electrical and computer engineering from Aristotle University of Thessaloniki, Greece. He is an associate editor of *IEEE Signal Processing Letters, Eurasip Journal of Advances in Signal Processing,* and *Eurasip Journal on Advances in Multimedia*. Contact him at: tzovaras@iti.gr.

Laila Dybkjær is a test manager at the Danish Ministry of Taxation. Her research interests include human–computer interaction, usability, and evaluation. Dybkjær has a PhD in computer science from DIKU, KØbenhavns Universitet. Contact her at ld@skat.dk.

Niels Ole Bernsen is former professor and director of the Centre for Cognitive Science at Roskilde University, Denmark, and of the Natural Interactive Systems Laboratory (NISLab) at the University of Southern Denmark. His research interests include natural interactive systems for future WWW generations, modality theory, and the usability of multimodal systems. Contact him at nob@nislab.dk.

Oya Aran is a postdoctoral researcher and a Marie Curie fellow at Idiap Research Institute, Switzerland. Her research interests include pattern recognition, computer vision, and social computing. Aran has a PhD in computer engineering from Bogazici University, Istanbul. She is a member of the IEEE. Contact her at oya.aran@idiap.ch.