A Modality Replacement Framework for the Communication between Blind and Hearing Impaired People

Konstantinos Moustakas[1], Dimitrios Tzovaras[1], Laila Dybkjær[2], and Niels Ole Bernsen[2]

[1] Informatics and Telematics Institute, Centre for Research and Technology Hellas, 6th km Charilaou-Thermi Road, 60361 Thermi-Thessaloniki, Greece
moustak@iti.gr, tzovaras@iti.gr
[2] NISLab, Copenhagen, Denmark
laila@hesselholm.dk, nob@nislab.dk

Abstract. This paper presents a multimodal framework for the communication between blind and hearing impaired people. The algorithms that are developed are based on the concept 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. Spatial information is conveyd in the blind user's terminal through the haptic modality utilizing an efficient haptic rendering scheme, while verbal information is presented in the hearing impaired user's terminal through sign language synthesis. All technologies are integrated in a virtual treasure hunting game, where a blind and a hearing impaired user have to collaborate so as to navigate in the virtual environment and solve the riddle of the game. Usability evaluation of this framework has shown significant impact of the proposed system for the disabled users.

Keywords: Modality replacement, multimodal interfaces, haptics, sign language, virtual reality.

1 Introduction

During the latest years there has been an increasing interest in the Human-Computer Interaction society for multimodal interfaces. Since Sutherland's SketchPad in 1961 or Xerox' Alto in 1973, computer users have long been acquainted with more than the traditional keyboard to interact with a system. More recently with the desire of increased productivity, of seamless interaction and immersion, of e-inclusion of people with disabilities, as well as with the progress in fields such as multimedia/multimodal signal analysis and human-computer interaction, multimodal interaction has emerged as a very active field of research.

Multimodal interfaces are those encompassing more than the traditional keyboard and mouse. Natural input modes are put to use (e.g. [3], [4]), such as voice, gestures and body movement, haptic interaction, facial expressions and more recently physiological signals.

C. Stephanidis (Ed.): Universal Access in HCI, Part III, HCII 2009, LNCS 5616, pp. 226–235, 2009.

As described in [8] multimodal interfaces should follow several guiding principles: multiple modalities that operate in different spaces need to share a common interaction space and to be synchronized; multimodal interaction should be predictable and not unnecessarily complex, and should degrade gracefully for instance by providing for modality switching; finally multimodal interfaces should adapt to user's needs, abilities, environment.

A key aspect in many multimodal interfaces is the integration of information from several different modalities in order to extract high-level information non-verbally 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 very important since the disabled user may now perceive information that was not previously available, using an alternate communication channel.

In the present work, a framework is developed that integrates efficient tools and interfaces for the generation of an integrated platform for the intercommunication of blind and hearing impaired persons. It is obvious that while multimodal signal processing is essential in such applications, specific issues like modality replacement and enhancement should be addressed in detail.

The rest of the paper is organized as follows. In Section 2 the concept of the modality replacement framework is described. Section 3 elaborates on the specific components of the framework focusing mainly on haptic interaction and sign language synthesis, which are the major modalities that are needed for modality replacement, since spatial information has to be transferred to the blind user using the haptics modality, while verbal information has to be converted into sign language so as to be intuitively perceived by the hearing impaired user. Section 4 describes the edutainment application that has been developed. Finally, Section 5 presents the results of the preformed usability evaluation and the conclusions are drawn in Section 6.

2 The Modality Replacement Framework

The basic development concept of the proposed multimodal framework for the disabled 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 illustrates the architecture of the proposed system for computer-aided interaction of blind and hearing-impaired users with the environment as well as with each other. The left part of the figure refers to the blind user's terminal, while the right refers to the hearing impaired user's terminal. All actions are controlled by the CSCW (Computer Supported Cooperative Work) game system. Visual information about the environment has to be conveyed to the blind user via the haptic and/or the auditory channel, while communication and the acquisition of semantic information can be performed using natural language. On the other side the hearing-impaired user acquires visual information using vision and communicates with the system using sign language.

A problem that stresses the high importance of modality replacement is that communication between a blind and a hearing-impaired user is not possible using physical

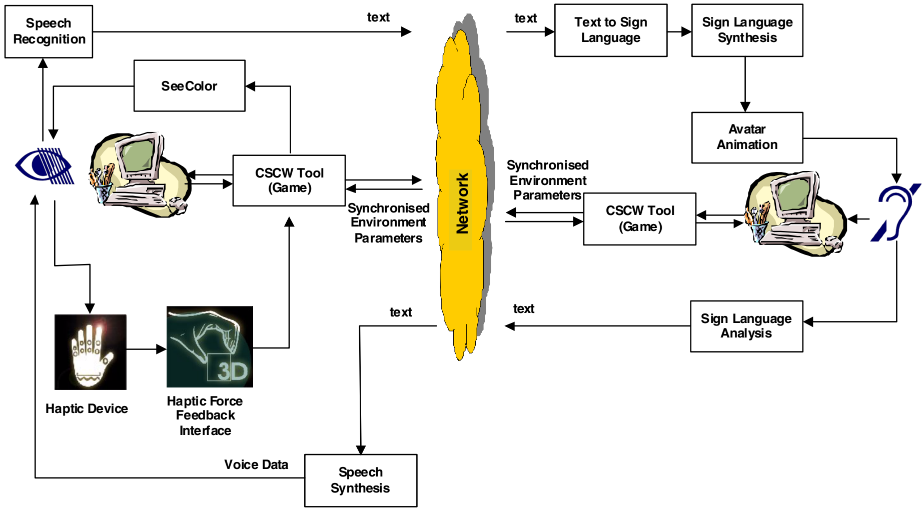

Fig. 1. Block diagram showing the intercommunication between the various modalities used to provide communication between visually impaired (left part of the diagram) and hearing impaired users (right part)

means. Ideally, as illustrated in Figure 1, a modality replacement system could be utilized to recognize all spoken language input of the blind user, convert it into sign language and present it to the hearing-impaired user using an animated avatar. Similarly, sign language gestures should be recognized and converted into text that should then be synthesized into speech using text-to-speech synthesis techniques.

3 Framework Modules

The main objective of the proposed system is the development of tools, algorithms and interfaces that will utilize modality replacement so as to enable the communication between blind or visually impaired and hearing-impaired users. To achieve the desired result the proposed system 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. Modality replacement was used in order to enable information transition between the various modalities and thus enable communication between the involved users.

Speech recognition and synthesis is performed using of-the-shelf algorithms, while sign language analysis is performed by a simplified version of the approach originally presented in [1]. Moreover, a system for converting scene colors into sound (SeeColor) originally developed in [2] has been integrated in the proposed framwework. In the following the major modules of the framework will be described in more detail since they constitute the backbone of the developed system.

3.1 Haptic Interaction

Haptic rendering is performed at every time step of the haptic loop using the extensively used spring-dumper model. The force feedback calculation is performed using directly the GHOST SDK library for PHANToMTM device. PHANToM desktop has 6 DOF for input (provides position and orientation) and 3 DOF for output (provides force feedback along the three axes). In particular, the force fed onto the haptic device is evaluated through the following formula:

$$F = k_s d - k_d v$$

where k_s, k_d are the spring and dumping coefficients and *d,v* the penetrating distance of the haptic probe into the grooved line map and its velocity.

In order to provide realistic force feedback it is important to ensure that force feedback loop runs at frequency equal or higher than 1KHz. Therefore, a sophisticated algorithm for collision detection and haptic rendering that utilizes implicit representations of the objects is applied [5].

In the context of the treasure hunting game, a tool for generating grooved line maps out of interactively sketched 2D drawings is developed. A grooved line map is a 3D terrain that is grooved in specific areas that represent streets or other meaningful areas that the blind user is able to perceive through a haptic device. Recently, a system for converting conventional 2D maps to haptic representations for the blind has been presented in [6].

The method presented in [6] has been extended so as to convert drawings that are interactively sketched by the user into haptic representations. Figure 2 (left) illustrates the sketched image, while Figure 2 (center) depicts the 3D grooved line map. Since haptic rendering is a very sensitive process and demands in every time step to perform the computationally intensive procedure of collision detection that performs slower for larger 3D meshes, the grooved line map is further processed so as to generate a multiresolutional grooved line map as illustrated in Figure 2 (right). It is obvious that this map is more detailed in the areas close to the path, thus reducing the redundant complexity of the initial 3D map.

3.2 Sign Language Synthesis

The sign language synthesis system [7] utilizes H-ANIM models to provide the animation and creates the animations using as input Sign Writing Markup Language

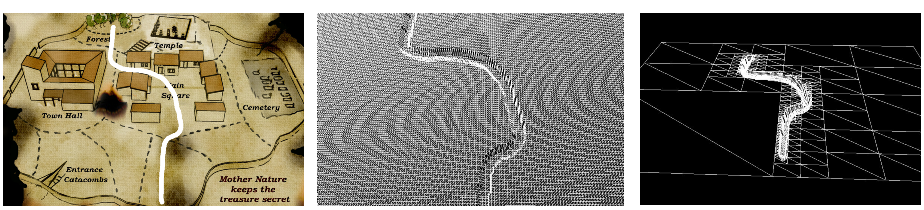

Fig. 2. Left: example of sketched image map, center: 3D grooved line map, right: 3D grooved map after polygon reduction

(SWML). Currently, symbols from the 1995 version of the Sign Symbol Sequence (SSS-1995) are supported. This sequence comprises an "alphabet" of the SignWriting notation system, while true images (in gif format) of each symbol contained in this sequence are available in [9]. The input for the sign synthesis system consists of the SWML entries of the sign boxes to be visualized. For each sign box, the associated information corresponding to its symbols is parsed.

The proposed technique first converts all individual symbols found in each sign box to sequences of MPEG-4 Face and Body Animation Parameters. The resulting sequences are used to animate a H-anim-compliant VRML avatar using MPEG-4 SNHC BAP and FAP players, provided by EPFL. The system is able to convert all hand symbols as well as the associated movement, contact and movement dynamics symbols contained in any ASL sign-box. Manual (hand) gestures and facial animations are currently supported. The proposed technique has significant advantages:

- Allows almost real-time visualization of sign language notation, thus enabling interactive applications,
- Avatars can easily be included in any virtual environment created using VRML, which is useful for a number of applications, such as TV newscasts, automatic translation systems for the hearing impaired, etc.

Figure 3 illustrates a snapshot of an avatar performing a sign.

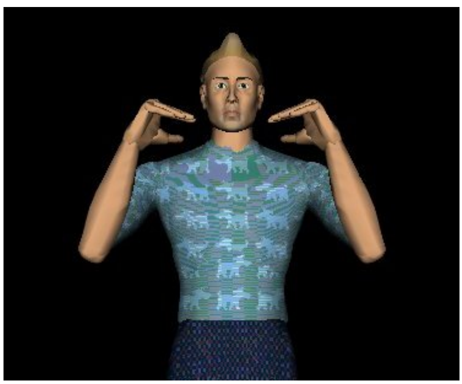

Fig. 3. After performing a synthesized sign

4 Application Scenario

The aforementioned technologies where integrated in the context of an entertainment scenario. The scenario consists of seven steps. In each step one of the users has to perform one or more actions in order to pass successfully to the next step. The storyboard is about an ancient city that is under attack and citizens of the city try finding the designs in order to create high technology war machines.

The first step involves the blind user. The user receives audio message informing him/her to find a red closed. The user starts from the initiating point at the entrance of the village and using the haptic device explores the village in order to find in one of the houses a red closet. In this step the blind user has to use the haptic device in order to explore the 3D Workspace. Furthermore audio modality replaces color modality, using SeeColOr module, and allows the blind user select the correct closet and thus receive an audio message. The audio message is then send to the second step of the application.

The second step involves the hearing impaired user, who receives an audio message. The message is converted from audio to text using the speech recognition tool and then to sign language using the sign synthesis tool. The user finally receives the message as a gesture through a 3D interactive avatar. The message guides the blind user to the town hall of the city where the mayor of the city assigns them a task.

The third step involves the blind user, who hears the message said by the Mayor and goes to the temple ruins. In the temple ruins the blind user has to search for an inscription. One of the columns in the destroyed temple has an inscription written on it that states, “The dead will save the city” (Fig. 4, left). The blind user is informed by an audio message whenever he finds this column and the message is sent to the hearing impaired user’s terminal.

The fourth step involves again the hearing impaired user. The user receives the written text in sign language form. The text modality is translated to sign language symbols using the sign synthesis tool. Then the hearing impaired user has to understand the meaning of the inscription "The dead will save the city" and go to the cemetery using the mouse. There he/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 a sign in sign language in order to allow the blind user understand that the key opens a box in the catacombs. The hearing impaired user has to perform the sign "Cave". This sign is recognized be the sign language recognition tool and the text result is send to the next step of the scenario.

In the next step the blind 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 (Figure 4, center) enter in them and find the box that contains a map as illustrated in Figure 2 left. The map is then sent to the next level.

The hearing impaired user receives the map, and has to draw the route to the area where the treasure is hidden (Figure 2, left). The route is drawn on the map and the map is converted to a grooved line map (Figure 2, right), which is send to for the last level to the blind user.

The blind user receives the grooved line map and has to find and follow the way to the forest where the treasure is hidden. Although the map is presented again as a 2D

Fig. 4. Left: 3rd step, temple ruins, center: 5th step, catacombs, right: 7th step, forest map

image the blind user can feel the 3D grooved map and follow the route to the forest. The 2D image and the 3D map are registered and this allows us to visualize the route that the blind user actually follows on the 2D image. Finally, after finding the forest a new grooved line map is displayed (Figure 4, right) where the blind user has to search for the final location of the treasure.

5 Usability Evaluation

The overall goal of the usability evaluation of the blind user's terminal was to explore the use of modalities, offered functionality, ease of use, and user satisfaction. Table I shows what was measured and how. All table items were addressed in post-test interviews, while some information was also acquired through real-time observation of test subjects' interaction with the system and from analysis of the video-recorded interaction sessions. The video shows the user's hand/arm, the haptic interaction and the screen contents.

Table 1. What was measured during evaluation of the proposed framework, and how

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

The test took place at the Institute for the Blind in Copenhagen. Six blind users, three males and three females aged between 21 and 40 years participated. Five of them had solid computer experience, one being an IT consultant and four in IT education. None of them had used a 3D haptic device for exploring a virtual environment before. One was a consummate computer games player whereas the other five played computer games rarely (two) or never (three). This was evidently not a representative user group. What we wanted for this first, exploratory usability test was a group of target users who were computer literate, curious about new technology, and likely to rovide ample feedback on their interaction with the system.

The average test session duration was about 90 minutes. Before playing a single game, each subject was given an introduction to the system, the game scenario, and the modalities used, and received training in interacting by means of the haptic device. It was emphasized that this was not a test of the user's skills at all, but a test of how good and how interesting the proposed multimodal framework might be.

Based on analysis of the usability data collected in the test, key points are:

1. *Navigation in City and Landscape*: All except one subject, who found it easy, found it medium difficult to navigate in the city and landscape: difficult to begin with, easier towards the end. Several subjects required more support from the experimenter than anticipated in order to complete the game within the allocated time frame.
2. *Haptic Interaction*: Haptic interaction was positively received by all subjects. Three were thrilled by this technology and the 3D world experience it gave them, describing it as “wonderful and astonishing”, “good, positive, surprising, impressive”, a “new sensation to be in a room/space that felt real”.
3. *SeeColor*: Colour recognition by association of each colour with a particular musical sound was not considered very useful. Only one subject seemed to find it easy to apply the musical colour codes. The others found it difficult to remember which instruments were associated with which colour.
4. *Spoken Output*: All subjects found the system’s use of output keywords and phrases, rather than complete sentences, sufficient. Some missed having more information about what you bump into or getting a hint if you seem to be lost.
5. *Grooved Line Path*: Opinions diverged about how easy it was to locate and follow the grooved path as part of the final task. Most subjects found the grooved-line representation of paths, and eventually maps, very helpful.
6. *Ease of Interaction*: One subject found interaction difficult, three found it reasonably easy, and two found it easy. Answers mostly refer to haptic interaction and orientation in 3D space.
7. *Learning*: Asked if they learned anything during gameplay, all subjects confirmed and commented: more systematicity in uncovering areas; built a sort of map in the head during the game; became more aware of the surroundings; started to explore the area more in 3D; learned patience; learned not to move so fast. These comment all refer to 3D exploration by haptic interaction. It is considered acceptable that this is something that users have to learn by experience.
8. *Like the Game*: In general, the subjects liked the game and would play again (a good idea, fun, challenging, exciting). Only one subject, the passionate game player who learned faster than anyone else, found it boring and without enough action.
9. *Advantages and Disadvantages*: Asked about advantages of the game, four mentioned the 3D experience and the fantastic sense of space afforded by the haptic interaction. One saw potential in the techniques used for enlarging the pool of potential game partners for the blind. She also proposed using the techniques to explain objects to the blind and enable them to create objects as a kind of 3D images.

Few disadvantages were mentioned. One mentioned the problems that few people have the kind of haptic device required and that there are few applications made for this kind of device. One proposed more sound effects for characterising the scenery.

The fact that most subjects liked the game is possibly 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. Subjects rated interaction difficulty as being moderate, while mentioning that they improved their skills over time and then felt more in control of the game. More research is needed of the factors that contribute to interaction control and creation of a reliable 3D mental model of the game space, including: how to "tune" subjects' mental models when introducing the game; the roles and amount of spoken feedback from virtual objects; how to do different actions on 3D objects by clicking the haptic device; haptically recognisable 3D modelling; and the importance of tightly bounding the 3D interaction space. The input/output modality combination used – output as spoken keywords, non-speech sound, sign language synthesis, and haptic 3D force-feedback, input as haptic 3D navigation, sign language, and device clicks – was positively commented.

6 Conclusions

This developments presented in the paper provide the means for unobtrusive human-computer interaction involving blind and hearing-impaired users, and also enables their intercommunication by proposing an efficient modality replacement scenario. The modalities that are integrated in the framework include haptics, sign language analysis and synthesis, speech recognition and synthesis. All these modalities and algorithms are integrated in a 3D virtual reality game where the blind and the visually impaired user have to cooperate to solve the riddle of the game. The game serves both as an entertainment tool and as an educational system that familiarizes the users with these new upcoming technologies for the disabled. The usability evaluation of the system has provided very positive comments from end-users and has triggered further research in this direction.

Acknowledgments. This work was supported by the EU funded project ACCESSIBLE, FP7-224145.

References

1. Aran, O., Akarun, L.: Recognizing two handed gestures with generative, discriminative and ensemble methods via Fisher kernels. In: Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B. (eds.) MRCS 2006. LNCS, vol. 4105, pp. 159–166. Springer, Heidelberg (2006)
2. Bologna, G., Deville, B., Pun, T., Vinckenbosch, M.: Transforming 3d coloured pixels into musical instrument notes for vision substitution applications. Eurasip Journal of Image and Video Processing, Special Issue: Image and Video Processing for Disability 24, 14–27 (2007)
3. Lumsden, J., Brewster, S.A.: A paradigm shift: Alternative interaction techniques for use with mobile & wearable devices. In: Proc. 13th Annual IBM Centers for Advanced Studies Conference CASCON 2003, Toronto, Canada, pp. 97–100 (2003)
4. Marsic, I., Medl, A., Flanagan, J.: Natural communication with information systems. Proc. of the IEEE 88(8), 1354–1366 (2000)
5. Moustakas, K., Tzovaras, D., Strintzis, M.G.: SQ-Map: Efficient Layered Collision Detection and Haptic Rendering. IEEE Transactions on Visualization and Computer Graphics 13(1), 80–93 (2007)

6. Moustakas, K., Nikolakis, G., Kostopoulos, K., Tzovaras, D., Strintzis, M.G.: Haptic rendering of visual data for the visually impaired. IEEE Multimedia Magazine 14(1), 62–72 (2007)
7. Papadogiorgaki, M.G., Grammalidis, N., Makris, L., Strintzis, M.G.: Gesture synthesis from sign language notation using MPEG-4 humanoid animation parameters and inverse kinematics. In: 2nd International Conference on Intelligent Environments (IE 2006), Athens, Greece, July 5-6 (2006)
8. Raman, T.V.: Multimodal Interaction Design Principles For Multimodal Interaction. In: CHI 2003, pp. 5–10. Fort Lauderdale, USA (2003)
9. SWML, `http://swml.ucpel.tche.br/`