

Evaluation of a multimodal interaction concept in virtual worlds

Jože Guna, Andrej Kos and Matevž Pogačnik

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška 25, SI-1000 Ljubljana, Slovenia
E-mail: joze.guna@fe.uni-lj.si

Abstract. In the presented paper we address the issue of multimodal interaction concept in virtual worlds. The topics of multimodal interaction and virtual worlds are presented in general. We present and apply a multimodal approach consisting of voice commands, motion and gesture based modalities to a Second Life virtual world. Detailed description of key pilot system elements and architecture is given. The proposed multimodal approach was tested both objectively and subjectively by the evaluation study. The study results show that the addition of voice and alternative navigation modalities significantly improve overall user experience.

Keywords: multimodal interaction, virtual world, voice command, gesture, HCI

Evalvacija večmodalnega sistema interakcije v navideznih svetovih

Povzetek. V pričujočem prispevku predstavimo problematiko večmodalnih vmesnikov na primeru virtualnih svetov. Predlagamo rešitev, ki klasični interakciji z računalniško tipkovnico in miško dodaja tudi modalnosti prepoznave govornih ukazov in uporabe alternativnih načinov navigacije. Rešitev implementiramo na primeru okolja Second Life ter preizkusimo z uporabniško evalvacijo.

Ključne besede: večmodalna interakcija, navidezni svetovi, glasovni ukazi, gesta, interakcija človek-stroj

1 Introduction

In our daily lives, mutual human communication is perceived as easy and natural, even though multiple modalities such as speech, facial expressions, vision, gestures, etc. are inadvertently used and the majority of information is communicated nonverbally. Communication among humans and machines, however, has always been a difficult and complex matter. This can be attributed to a fact that due to technological limitations humans had to adapt to machines and not vice-versa. A lot of research effort [1, 2, 3, 4 and 5] has been given to alleviate this problem which is one of the prime focuses of the interdisciplinary field of human-computer interactions (HCI).

A traditional human-computer interaction system [1] uses a graphical user interface (GUI) approach which is based upon the WIMP concept (Windows, Icons, Menus, Pointing device). Input is typically obtained from a single mode or modality, such as a computer keyboard or mouse. A modality in terms of HCI designates a method of communication between the human and the computer and is bidirectional. Output

modalities, such as vision, hearing or haptic modalities coincide with human senses and provide a sense through which humans receive the output of the computer. Input modalities allow the computer to receive the input from the human using various sensors and other devices, such as keyboard, mouse, accelerometer, etc.

To provide a better user experience in human-computer interaction multiple modalities are combined and a new class of interfaces emerges called "multimodal interfaces". A definition of the multimodal interfaces is given in [2] by Sharon Oviatt who states: "Multimodal systems process two or more combined user input modes — such as speech, pen, touch, manual gestures, gaze, and head and body movements — in a coordinated manner with multimedia system output." Compared to traditional keyboard and mouse interface, multimodal interfaces provide flexible use of input modes and allows the users a choice of which modality or their combinations to use and when. Multimodal interfaces are thus perceived as easier to use and more accessible [3], provide better performance and robustness of interaction [2] and are suited for critical industrial and clinical use [6].

The rest of the paper is organized as follows: in Section 2 multimodal interaction applied to virtual worlds with an emphasis on Second Life is presented; experimental environment, method and evaluation procedure are described in Sections 3 and 4, respectively; results are shown in Section 5 with key conclusions drawn in Section 6 including suggestions and motivation for future work.

2 Multimodal interaction in virtual worlds

Virtual worlds are a genre of online social-based community, usually implemented in a form of computer-based simulated environment. The term today has largely become synonymous with interactive 3D virtual environments with immersive graphics and sound effects. Users interact with environment and one another in the form of avatars which represent their personalities (i.e. alter ego). Because physical constraints do not apply in virtual worlds, the user's avatar can take any form they like and perform actions that are otherwise impossible (e.g. flying, teleportation). Though many virtual worlds are intended only for gaming and enjoyment (Massively Multiplayer Online Role Playing Games, MMORPG), not all are limited to games and can have strong emphasis on educational or social components.

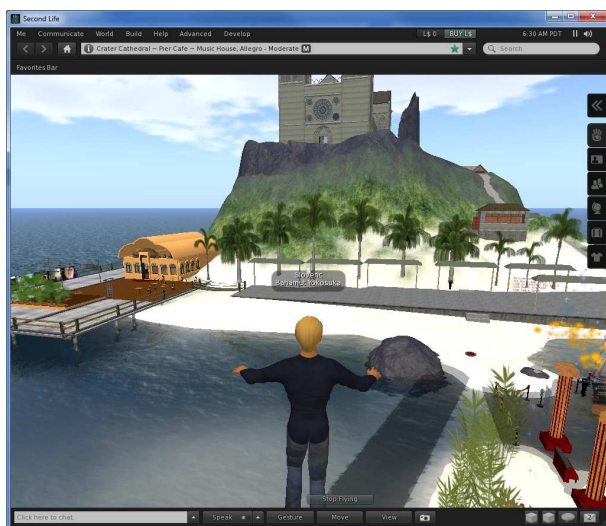


Figure 1. The Second Life environment. Evaluation location.

Slika 1: Okolje Second Life. Lokacija izvedbe poskusa.

Linden Labs Second Life [7] virtual world, as shown in Figure 1, is such an example. In contrast to MMORPG there is no strict storyline apparent in Second Life. Instead, a lot of emphasis is given to interaction and social activities among players. Second Life can be described as a form of highly advanced socio-chat and presentation platform and has the potential to become a 3D representation of the present World Wide Web. A peculiarity of the SL is also a virtual currency called Linden L\$, which represents the basis for the embedded monetary system. The monetary flow is bidirectional and thus virtual property actually has real value. The Second Life currently has approximately 1 million of active users and has great potential but has not yet fully it [8].

As physical limitations do not apply to Second Life, many users can coexist at the same location and share ideas and knowledge in a multimedia rich environment.

Educational component is therefore particularly strong in Second Life, with Open University [9] initiative as an example.

Some issues remain, however. The original Second Life interaction is based on usage of a computer keyboard and mouse. This, we believe, is unnecessarily complex due to complicated keyboard shortcuts, required for initiation of basic actions such as sitting, jumping, emotion gestures, etc. [10]. According to our own research [11, 12] and that of HCI experts [5] we believe that other modalities should be added. To achieve greater acceptance, we therefore propose to add voice commands, navigation using joystick and gesture based commands. A combination of Nintendo WiiMote advanced remote controller and Nunchuk was selected which has already successfully been proven in several applications [13], including virtual reality systems [14], rehabilitation [15] and accessibility [10] in Second Life.

We state and evaluate the following hypothesis:

“Multimodal user interface using voice commands and simple joystick and gesture based navigation modalities will be perceived as more intuitive and easier to use in comparison to standard navigation modalities using only mouse and keyboard.”

3 Experimental environment

The experimental environment consists of hardware control devices and communications, processing and rendering software components.

As control devices, besides classical computer keyboard and mouse, a combination of Nintendo WiiMote and Nunchuk controllers was used. The WiiMote [16], shown in Figure 2 (right), represents a cost effective solution of a multimodal control device which combines various input and output modalities. As output, visual (LED lights), auditory (built-in speaker) and haptic (force feedback) modalities are available. As input, classical button interface, visual (IR camera module) and force sensing modalities are used. For the latter a built-in three-axis accelerometer ADXL330 device from Analog Devices is used. The accelerometer is used to sample and digitize user motion gestures.



Figure 2. Nintendo WiiMote (right) and Nunchuk (left) control devices.

Slika 2: Kontrolna naprava Nintendo WiiMote (desno) in Nunchuk (levo).

The Nunchuk, as shown in Figure 2 (left), represents an additional controller to be used in combination with the WiiMote in user’s secondary hand. It consists of an analog joystick, two buttons and additional built-in three-axis accelerometer device. This allows for execution of more complex user gestures.

The pilot setup architecture and corresponding layer model are shown in Figure 3 and Figure 4, respectively. All software components were executed on a PC computer with Windows 7 operating system. For communication between the WiiMote and Nunchuk control devices and processing terminal, Bluetooth wireless technology using BlueSolei protocol stack was used. Multimodal control was implemented in a specialized script which was executed in Glove Programmable Input Emulator [17] software. Microsoft speech api [18] was used for voice recognition and synthesis. Official Second Life 2 [7] and experimental Kirstens S20 viewers [19] were utilized for interaction and rendering of the Second Life virtual world.

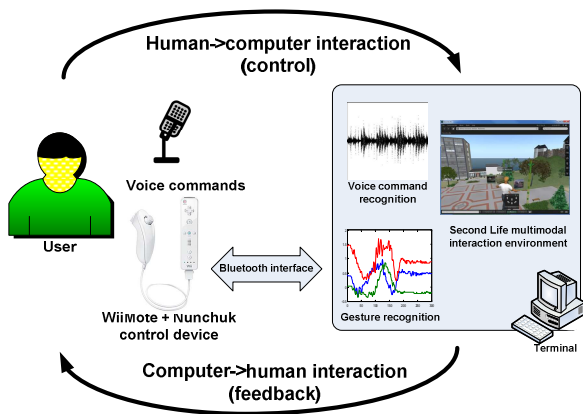


Figure 3. Pilot setup architecture.

Slika 3: Arhitektura pilotske postavitve.

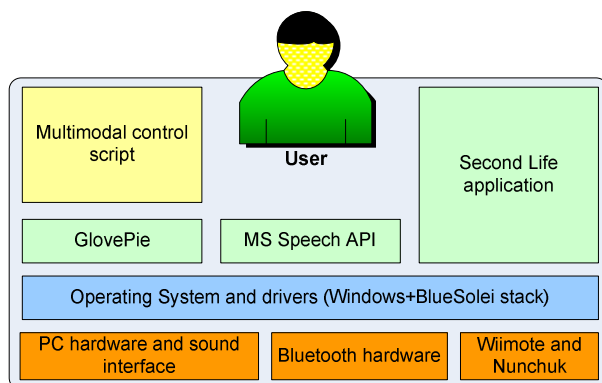


Figure 4. Pilot setup layer model.

Slika 4: Slojni model pilotske postavitve.

4 Usability evaluation

The evaluation study included five individuals of various ages ranging from 26 to 46 years, of both sexes and dexterity. Participants had different experience with computer interaction in 3D virtual worlds and input devices especially with joysticks, ranging from no experience to very experienced users. Previous experience of users was measured on a five-level Likert scale (1: no experience; 2: little experience 3: medium experience; 4: some experience; 5: very experienced) and was given as a subjective participant’s estimation. The study participants’ demographic data is summarized in Table 1.

Sex	Dexterity	Age	Experience	
male	3 left-handed	1 21 - 30	3 no experience	2
female	2 right-handed	4 31 - 40	1 medium experience	1
		41+	1 very experienced	2
SUM				5

Table 1. Study participant's data.

Tabela 1: Podatki o testnih uporabnikih.

Three different combinations of modalities were tested. The first (1) modality combination using computer keyboard and mouse relied on arrow keys for navigation and specific keyboard shortcuts for avatar actions. The second (2) modality combination added an option of triggering specific avatar actions through voice commands. System feedback was given by repeating the commands by speech synthesis. The voice command interface recognized English phrases and did not require special learning for each participant. The voice recognition was not optimal but was sufficient for the experiment. The third (3) modality combination added control using WiiMote controller for gesture input and Nunchuk for joystick navigation. Some actions were also triggered using motion gestures, such as “swing up”, “swing down” and “horizontal shake” motions, as shown in Figure 5. In this case user feedback was given by tactile vibration feedback and specific led lights pattern on the WiiMote controller.

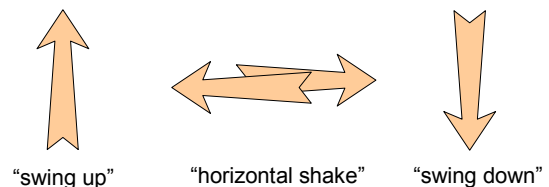


Figure 5. Gestures used.

Slika 5: Uporabljene geste.

All user avatar actions used in the evaluation scenarios are shown in Table 2.


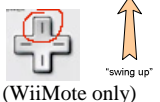




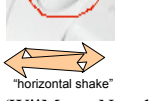
Action	Keyboard and Mouse	Voice	WiiMote and Nunchuk
basic navigation	arrow keys		
flying toggle	"f" key	"fly"	 (WiiMote only)
flying up	"PgUp" key		
flying down	"PgDown" key		
jumping	"e" key	"jump"	 (Nunchuk only)
sitting toggle	"ctrl+alt+shift+s" key	"sit"	 (WiiMote only)
run/walk toggle	"ctrl+r" key	"run" "walk"	 (WiiMote+Nunchuk)
clapping hands	"F2" key	"clap"	
muscle stretch	"F3" key	"showoff"	

Table 2. Avatar actions used in the evaluation scenarios.

Tabela 2: Uporabljene avatarske akcije v evalvacijskih scenarijih.

The three proposed combinations of modalities were tested in two scenarios, located on the "Allegro" island [20], shown in Figure 1 and were comprised of separate tasks. Tasks included basic movement (left, right, forward, backward), flying and avatar actions such as jumping, sitting, running/walking and two avatar emotion gestures (clapping hands, muscle stretch).

The first scenario, called "Beach walk", included all actions except flying and took place on a coastal area. The user's avatar started on a nearby sea island, explored the surroundings and then continued to the "Pier Caffe" structure. The aim was to test basic and more complex avatar movement including avatar actions jumping, sitting, running/walking, emoticons) inside and outside of structures. The second scenario, called "Sky dive" was flying oriented and took place in the sky. The user's avatar started on platform high in the sky and then flew to the vast cathedral like structure, explored the insides and then finished the flight at the bottom on the beach at the "Pier Caffe" structure. The main goal of the second scenario was to test the flying movement.

The main aim of the study was to determine the usability and suitability of all three proposed modality combination approaches for different tasks. Objective and subjective tests were performed. The objective test comprised of successful scenario completion times for

all three modality combinations while in the subjective test the study participants were asked to rate their experience in terms of simplicity, intuitiveness and general usability opinion. The term "simplicity" is related to the ease of use of the particular modality combination, the term "intuitiveness" to how hard it is to learn the proposed modality approach and the term "general usability opinion" to the level of user acceptance or usability. The subjective user opinion was rated on the five-level Likert scale (1:very poor; 2:poor; 3:neutral; 4:good; 5:very good), where rating one signified the most negative opinion, rating three neutral and rating five the most positive opinion.

All study participants completed both scenarios using all three combinations of proposed modalities. The scenario courses were predefined in order to assure comparable results. Firstly, the participants were only briefly introduced with the Second Life environment and control devices in order to gain better understating of differences between proposed modalities. Then they were shown video previews of both scenarios including all movements and avatar actions. Finally, both scenarios were performed and evaluated both objectively and subjectively. Study participants were asked to describe their experience in words and give opinion on what they particularly liked and disliked in each proposed modality combination.

5 Results

The objective test results show completion times (minutes:seconds) for all participants and scenarios. The subjective test results reflect user subjective opinion in terms of simplicity, intuitiveness and general usability measured using the five-level Likert scale. Average value (μ) and standard deviation (σ) are calculated for subjective score results.

Scenario 1: "Beach walk"				
Modality combination - completion time (min:sec)				
User	Experience	Keyboard and Mouse	Voice Commands	WiiMote and Nunchuk
1	5	2:57	1:50	1:50
2	1	4:40	2:40	4:20
3	5	2:45	2:15	2:05
4	1	3:00	2:00	3:50
5	3	3:24	2:20	2:40
Scenario 2: "Sky dive"				
Modality combination - completion time (min:sec)				
User	Experience	Keyboard and Mouse	Voice Commands	WiiMote and Nunchuk
1	5	1:12	1:02	1:00
2	1	2:30	1:40	1:40
3	5	1:25	1:15	1:25
4	1	1:55	1:30	2:40
5	3	1:59	2:00	1:50

Table 3. Modality combinations scenario completion times.

Tabela 3: Čas za izvedbo testnih scenarijev glede na uporabljene kombinacije modalnosti.

Scenario 1: "Beach walk"

User	Keyboard and Mouse			Voice Commands			WiiMote and Nunchuk			
	Experience	Simplicity	Intuitiveness	General opinion	Simplicity	Intuitiveness	General opinion	Simplicity	Intuitiveness	General opinion
1	5	4	3	4	4	5	4	5	3	
2	1	2	2	2	3	3	3	4	4	4
3	5	2	2	3	5	5	5	4	3	4
4	1	3	3	3	5	5	5	2	3	3
5	3	2	2	2	4	5	4	5	4	4
μ		2.6	2.4	2.8	4.2	4.4	4.4	3.8	3.8	3.6
σ		0.89	0.55	0.84	0.84	0.89	0.89	1.1	0.84	0.55
Average satisfaction:		2.6			4.3			3.7		

Scenario 2: "Sky dive"

User	Keyboard and Mouse			Voice Commands			WiiMote and Nunchuk			
	Experience	Simplicity	Intuitiveness	General opinion	Simplicity	Intuitiveness	General opinion	Simplicity	Intuitiveness	General opinion
1	5	4	3	3	4	3	3	4	5	4
2	1	3	3	3	3	4	3	4	3	4
3	5	2	2	3	4	5	4	5	5	5
4	1	4	3	3	4	4	4	3	3	3
5	3	3	3	3	5	5	5	5	5	5
μ		3.2	2.8	3	4	4.2	3.8	4.2	4.2	4.2
σ		0.84	0.45	0	0.71	0.84	0.84	0.84	1.1	0.84
Average satisfaction:		3			4			4.2		

Table 4. Subjective user opinion scores.

Tabela 4: Subjektivne uporabniške ocene.

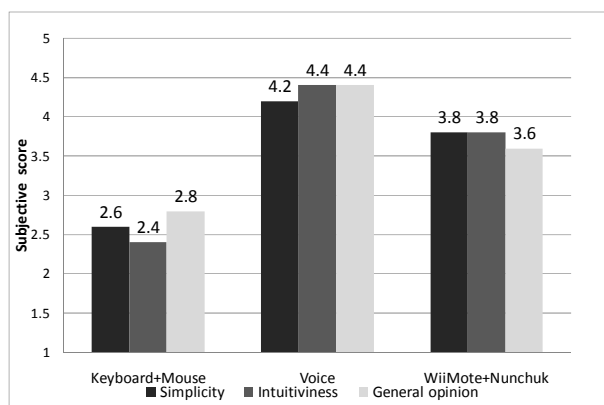


Figure 6. Subjective average user opinion scores for »Beach walk« scenario.

Slika 6: Subjektivne povprečne uporabniške ocene za scenarij »Beach walk«.

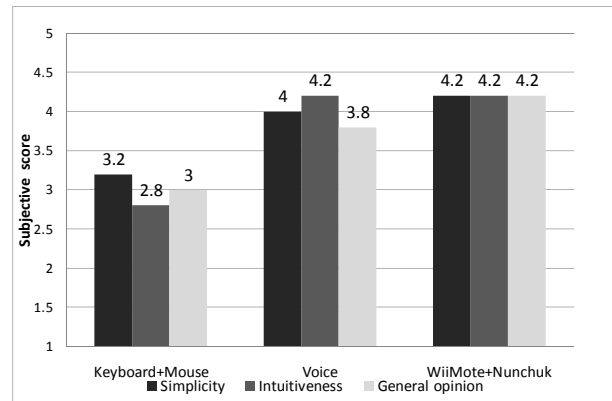


Figure 7. Subjective average user opinion scores for »Sky dive« scenario.

Slika 7. Subjektivne povprečne uporabniške ocene za scenarij »Sky dive«.

6 Discussion and conclusions

The results obtained through the evaluation study show significant accordance with the initially stated research hypothesis. Although the pilot study included only five participants, the study group was diverse in terms of participant's sex, dexterity, age distribution and especially previous computer experience. The latter, particularly considering more advanced input devices such as joysticks, also had great impact on general user's opinion of proposed input modalities.

Considering the objective test results (see Table 3) measuring successful completion times, the most noticeable improvement compared to "Keyboard and Mouse" modality was observed when "Voice Commands" modality was added in both scenarios. Somewhat lesser but still apparent improvement was observed when "WiiMote and Nunchuk" modality was used.

The observed results are consistent with user experience criteria, where better results were achieved with users who were more experienced with advanced input devices, similar to those used in our study (i.e. joysticks, gamepad remote controllers). Differences in completion times regarding both scenarios can also be attributed to the nature and complexity of scenarios. Scenario "Beach walk" was more complex in terms of tasks with emphasis on walking type of avatar navigation, while the "Sky dive" scenario was flying oriented. Flying was also perceived as an easier task than walking, since there was less complex navigation required between objects outside and inside of structures.

Considering the subjective test results (see Table 4), the increase of "average user satisfaction" by 1 point in average is apparent in both scenarios when modalities "Voice Commands" and "WiiMote and Nunchuk" were added to the basic "Keyboard and Mouse" modality. When performing walking type of avatar navigation,

users subjectively liked most the addition of “Voice Commands” modality. The usage of joystick (Nunchuk), however, was better accepted when performing flying avatar navigation tasks. Users with more experience with advanced input devices also had fewer difficulties with joystick usage (Nunchuk) and gestures (WiiMote) than users with less experience.

In general, users liked the possibility of performing various tasks using different modalities with strong emphasis on simplicity. Avatar actions that required complex keyboard shortcuts (e.g. avatar sit command required CTRL+ALT+SHIFT+s) were easily performed using simple voice commands. Likewise, gestures such as controller swinging motion upwards for flying action complemented other modalities.

We experimentally introduced the usage of gestures on both Wii controllers. The “horizontal shake” gesture was mapped to the walking toggle action on both controllers, while the “swing up” gesture was mapped to different actions, which was slightly ambiguous (jumping – Nunchuk, flying – WiiMote). Although only three different gestures were used, the users preferred that one gesture was mapped to only one action (1:1 mapping). Consequently, only unambiguous gesture combinations were used - gestures for flying, sitting and running/walking on the primary WiiMote controller.

In general, the users liked most the usage of fewer but well defined and simple to use features. The predictability and stability of execution (e.g. gesture recognition, voice command recognition) was important as well.

The results obtained in our usability study show that the usage of multimodal interfaces and advanced input devices in 3D virtual worlds is reasonable, which is also compliant with findings in [5, 10]. The proposed multimodal approach will be upgraded in terms of additional functionalities and will be evaluated by a greater number of users.

7 Acknowledgements

The research and development work was supported by the Slovenian Research Agency.

8 References

- [1] M. Turk: Multimodal Human-Computer Interaction, Real-Time Vision for Human-Computer Interaction, Springer US, 2005, Pages: 269-283
- [2] S. Oviatt: Breaking the Robustness Barrier: Recent Progress on the Design of Robust Multimodal Systems, *Advances in computers*, vol. 56, 2002, Pages: 305-341
- [3] A. Jaimes, N. Sebe: Multimodal Human Computer Interaction: A Survey, IEEE International Workshop on Human Computer Interaction, Beijing 2005
- [4] J. Sodnik, C. Dicke, S. Tomazic M. and Billinghurst: A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, vol. 66(5), 2008, Pages: 318-332.
- [5] P. Barthelmess, S. Oviatt: Multimodal Interfaces: Combining Interfaces to Accomplish a Single Task, *HCI Beyond the GUI*, 2008, Pages: 391-444
- [6] C. Krapichler, M. Haubner, A. Löscher, D. Schuhmann, M. Seemann, K. Englmeier: Physicians in virtual environments — multimodal human-computer interaction, *Interacting with Computers* 11, 1999, Pages: 427-452
- [7] Linden Labs Second Life, accessed at <http://secondlife.com/> (last seen 24.8.2010)
- [8] A. M. Kaplan, M. Haenlein: The fairyland of Second Life: Virtual social worlds and how to use them, *Business Horizons* 52, 2009, Pages: 563—572
- [9] LindenLabs Second Life E-learning, accessed at <http://education.secondlife.com/?lang=en-US>, http://secondlifegrid.net.s3.amazonaws.com/docs/Second_Life_Case_OpenU_EN.pdf (last seen 24.8.2010)
- [10] S. Hansen, P. Davies, C. Hansen: Addressing accessibility: emerging user interfaces for Second Life communities, IADIS International Conference on Web Based Communities 2008
- [11] J. Guna, M. Pogačnik, A. Štern, I. Humar, J. Bešter: Inovativni vmesniki in načini interakcij, Zbornik osemnajste mednarodne Elektrotehniške in računalniške konference ERK 2009, 21. - 23. September 2009, Portorož, Slovenija
- [12] J. Guna, J. Žilavec, M. Pogačnik, S. Tomažič, A. Kos, J. Bešter: Večmodalna interakcija v navideznih okoljih, Zbornik devetnajste mednarodne Elektrotehniške in računalniške konference ERK 2010, 20. - 22. September 2010, Portorož, Slovenija (in press)
- [13] J.A. Vicaria, J.M. Maestre, E.F. Camacho: Academical and research wiimote applications, IADIS International Conference Interfaces and Human Computer Interaction 2008
- [14] A. B. Craig et al.: *Developing Virtual Reality Applications*, Elsevier, 2009, Pages: 33-57
- [15] P.J. Standen, C. Camm, S. Battersby, D.J. Brown, M. Harrison: An evaluation of the Wii Nunchuk as an alternative assistive device for people with intellectual and physical disabilities using switch controlled software, *Computers & Education*, 2010
- [16] Lee Johnny C.: Hacking the Nintendo Wii Remote, *IEEE Pervasive Computing*, Volume 7, Issue 3, 2008
- [17] GlovePie Software, accessed at <http://glovepie.org/glovepie.php> (last seen 24.8.2010)
- [18] Microsoft Speech API, accessed at <http://www.microsoft.com/speech/default.aspx> (last seen 24.8.2010)
- [19] Kirsten S20 unofficial viewer, accessed at <http://www.kirstensviewer.com/> (last seen 24.8.2010)
- [20] Allegro Second Life island, accessed at <http://maps.secondlife.com/secondlife/Allegro/121/242/23> (last seen 24.8.2010)

Jože Guna (joze.guna@fe.uni-lj.si) received his B.Sc. and M.Sc. degrees from the Faculty of Electrical Engineering, University of Ljubljana, in 2002 and 2005 respectively. His research focuses on IP multimedia services, human-computer interaction, networking and Internet technologies, digital rights management, and smart home technologies.

Andrej Kos (andrej.kos@fe.uni-lj.si) graduated from the University of Ljubljana, Slovenia in 1996 and was awarded his Ph. D. degree in telecommunications in 2003. His current work and research focus on managed broadband packet switching and next generation intelligent converged services.

Matevž Pogačnik (matevz.pogacnik@fe.uni-lj.si) graduated from the University of Ljubljana, Slovenia in 1997 and obtained his Ph.D. in 2004 at the University of Ljubljana. His main R&D fields are interactive multimedia services and applications on heterogeneous devices. Currently, he is focused on development of interactive services for remote education and IPTV systems. He is a member of IEEE organization.