

Pekka Kallioniemi

Collaborative Wayfinding in Virtual Environments

Pekka Kallioniemi

**Collaborative Wayfinding in
Virtual Environments**

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Communication Sciences of
the University of Tampere, for public discussion in the Pinni Auditorium B1097
on June 15th, 2018, at noon.

Faculty of Communication Sciences
University of Tampere

Dissertations in Interactive Technology, Number 28
Tampere 2018

ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY

Supervisor: Professor Markku Turunen, Ph.D.
Faculty of Communication Sciences
University of Tampere
Finland

Opponent: Senior Lecturer Daphne Economou, Ph.D.
Faculty of Science and Technology
University of Westminster
United Kingdom

Reviewers: Professor Tassos A. Mikropoulos, Ph.D.
Department of Primary School Education
University of Ioannina
Greece

Assistant Professor Jayesh S. Pillai, Ph.D.
IDC School of Design
IIT Bombay
India

Acta Electronica Universitatis Tamperensis; 1889

ISBN 978-952-03-0753-0 (pdf)

ISSN 1456-954X

<http://tampub.uta.fi>

The originality of this thesis has been checked using the Turnitin originality check service in accordance with the quality management system of the University of Tampere.

Dissertations in Interactive Technology, Number 28

Faculty of Communication Sciences
FIN-33014 University of Tampere
FINLAND

ISBN 978-952-03-0752-3
ISSN 1795-9489

Juvenes Print—Suomen Yliopistopaino Oy
Tampere 2018

Abstract

Wayfinding is a complex process in which people orient themselves in the surrounding space and navigate from one place to another. The path selected may vary based on the purpose of the trip, but generally, people want to move from their origin to their destination as effortlessly as possible. Wayfinding often has collaborative aspects, for example, in situations where one person is guiding another. This dissertation evaluates aspects of collaborative wayfinding in virtual environments, answering the following research questions: *What strategies do people use to find their way in collaborative virtual environments?* and *What aspects affect collaborative wayfinding tasks?*

When sufficient realism is provided, human performance in virtual environments (VEs) is comparable to their real-world activities. For this reason, VEs have been suggested as a useful tool for measuring spatial ability. To find answers for the research questions, a collaborative virtual environment application called CityCompass with three evolutionary stages was designed, implemented, and evaluated. All these applications have the same approach for measuring spatial ability through collaborative wayfinding tasks, but they also have unique features, for example, regarding interaction. This work also introduces a model to highlight prominent landmarks that can provide further guidance in both virtual environments and real-world scenarios.

Besides spatial ability metrics, this work measured the effect of several factors, including immersion, video game experience, and gender on spatial ability and user experience in collaborative virtual environments. User experiments with the CityCompass application were conducted, and the findings suggest that people use strategies similar to real life when navigating in virtual environments. The collaborative aspects reduce effects like gender differences that are commonly detected with single-user experiments. In addition, immersion and user experience factors such as effortless use and clarity were found to be important aspects of collaborative VEs.

The results of this thesis suggest several factors that affect collaborative wayfinding in VEs. These should be considered when designing any applications with wayfinding aspects. Because of this work, I present guidelines for designing these applications to be clearer, more usable, and thus more enjoyable for the users. In addition, as a more constructive work, I present the three applications that are suitable for future experiments in various fields, for example, education.

Acknowledgements

First, I would like to thank my supervisor, Markku Turunen, for providing me with the possibility of working on my thesis even during my involvement in various projects. His guidance has immensely helped my work during these years. I would also like to recognize Daphne Economou for acting as an opponent during my public defense. In addition, I would like to say a big thank you to Tassos A. Mikropoulos and Jayesh S. Pillai for taking their time to review my work.

I would like to extend my gratitude to all my colleagues for giving me knowledge and support for all these years, especially Jaakko Hakulinen and Tomi Heimonen. Your motivation and expertise have helped me during those times of desperation that are familiar to all doctoral students. All this work could not have been accomplished without the support of TEKES and my collaborators in the projects with which I have been involved through the years. I also want to express my appreciation to the Faculty of Communication Sciences for funding the last years of this project.

I also want to acknowledge my family for supporting my work. Finally, I would like to thank my dog, Freddie, for always being there for me unconditionally.

“If you accomplish something good with hard work, the labor passes quickly, but the good endures.” – Gaius Musonius Rufus, Fragment 51

Tampere, December 31, 2017

Pekka Kallioniemi

Contents

1	INTRODUCTION	1
1.1	Objective.....	1
1.2	Context of This Research	2
1.3	Methodology	3
1.4	Results	4
1.5	Structure.....	4
1.6	On terminology	4
2	APPLICATIONS	7
2.1	Berlin Kompass	7
2.2	CityCompass	11
2.3	CityCompass VR.....	13
2.4	Summary.....	16
3	WAYFINDING AND LANDMARKS	19
3.1	Human Wayfinding	19
3.2	Cognitive Aspects of Wayfinding	23
3.3	Route Directions.....	26
3.4	Collaborative Wayfinding	31
3.5	Landmarks in Wayfinding	34
3.6	Summary.....	37
4	VIRTUAL ENVIRONMENTS	39
4.1	Wayfinding in Virtual Environments	43
4.2	Landmarks in Virtual Environments	45
4.3	Immersion/Presence	50
4.4	Interactive Omnidirectional Videos (iODVS)	54
4.5	Summary.....	55
5	INTRODUCTION TO THE PUBLICATIONS	57
5.1	Model for Landmark Highlighting in Mobile Web Services	58
5.2	Evaluating Landmark Attraction Model in Collaborative Wayfinding in Virtual Learning Environments	60
5.3	Berlin Kompass: Multimodal Gameful Empowerment for Foreign Language Learning	62
5.4	Collaborative Navigation in Virtual Worlds: How Gender and Game Experience Influence User Behavior.....	64
5.5	CityCompass: A Collaborative Online Language Learning Application.....	66
5.6	User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays	67
5.7	Effect of Gender on Immersion in Collaborative iODV Applications	68
6	DISCUSSION	71
7	CONCLUSION.....	75
7.1	Future Work	76
8	REFERENCES	77

List of Publications

This dissertation is composed of a summary and the following original publications, reproduced here by permission.

- I. **Kallioniemi, P.**, and Markku Turunen. (2012). Model for landmark highlighting in mobile web services. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*. ACM, New York, NY, USA, Article 25, 10 pages. doi:10.1145/2406367.2406398
- II. **Kallioniemi, P.**, Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., Uusi-Mäkelä, M., Hietala, P., Okkonen, J, and Raisamo, R. (2013). Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 33, 10 pages. doi:10.1145/2541831.2541849
- III. **Kallioniemi, P.**, Pihkala-Posti, L., Hakulinen, J., Turunen, M., Keskinen, T., and Raisamo, R. (2015). Berlin Kompass: Multimodal gameful empowerment for foreign language learning. *Journal of Educational Technology Systems*. 43: pp. 429-450. doi:10.1177/0047239515588166
- IV. **Kallioniemi, P.**, Heimonen, T., Turunen, M., Hakulinen, J., Keskinen, T., Pihkala-Posti, L., Okkonen, J., and Raisamo, R. (2015). Collaborative navigation in virtual worlds: How gender and game experience influence user behavior. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST '15)*, Stephen N. Spencer (Ed.). ACM, New York, NY, USA, pp. 173-182. doi:10.1145/2821592.2821610
- V. **Kallioniemi, P.**, Sharma, S., and Turunen, M. (2016). CityCompass: A collaborative online language learning application. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, pp. 94-97. doi:10.1145/2818052.2874334

- VI. **Kallioniemi, P.**, Mäkelä, V., Saarinen, S., Turunen, M., Winter, Y., and Istudor, A. (2017). User experience and immersion of interactive omnidirectional videos in CAVE systems and head-mounted displays. In: Bernhaupt R., Dalvi G., Joshi A., K. Balkrishan D., O’Neill J., Winckler M. (Eds.) *Human-Computer Interaction – INTERACT 2017. INTERACT 2017. Lecture Notes in Computer Science*, Vol. 10516. Springer, Cham. doi:10.1007/978-3-319-68059-0_20
- VII. **Kallioniemi, P.**, Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., and Ronkainen, K. (2017). Effect of gender on immersion in collaborative iODV applications. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA, pp. 199–207. doi:10.1145/3152832.3152869

The Author's Contribution to the Publications

This work was conducted as part of several research projects and was made possible by my project colleagues. The papers introduced in this dissertation were co-authored. However, I have been responsible for designing and developing the elements of the applications presented herein, and I was involved with designing and conducting the experiments. My publication-specific responsibilities for and contributions to each publication are as follows:

Publication I: Kallioniemi, P., and Turunen, M. "Model for landmark highlighting in mobile web services"

For this article, I implemented a model for the landmark highlighting of pedestrian route guidance. With the guidance of Markku Turunen, I designed a user experiment to evaluate the model. In addition, I recorded and reported the results from this experiment.

Publication II: Kallioniemi, P., Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., Uusi-Mäkelä, M., Hietala, P., Okkonen, J, and Raisamo, R. "Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments"

For this article, I implemented a Berlin Kompass application with Jaakko Hakulinen. The pedagogical and language learning concepts of the application were developed by Laura Pihkala-Posti from the research center Plural (the doctoral researcher of language pedagogy of the project). The experiment, including its methods, user questionnaires, and metrics, was designed by me, Laura Pihkala-Posti, Tuuli Keskinen, Mikael Uusi-Mäkelä, Pentti Hietala, and Jaakko Hakulinen. I organized and observed the actual evaluations with Laura Pihkala-Posti, Mikael Uusi-Mäkelä and Sanna Kangas. Sari Yrjänäinen also assisted in the evaluation sessions. Markku Turunen, Jussi Okkonen and Roope Raisamo were supervising the work and working on the administrative aspects of the project.

Publication III: Kallioniemi, P., Pihkala-Posti, L., Hakulinen, J., Turunen, M., Keskinen, T., and Raisamo, R. “Berlin Kompass: Multimodal gameful empowerment for foreign language learning”

The application implementation and the pedagogical aspects (also see, Pihkala-Posti et al., 2014) of the project were carried out by the same people as in Publication II. The experiment, including its methods, user questionnaires, and metrics, was designed by me, Laura Pihkala-Posti, Tuuli Keskinen, and Jaakko Hakulinen. I organized and observed the actual evaluations with Laura Pihkala-Posti and Mikael Uusi-Mäkelä. Pihkala-Posti was responsible for the pedagogy related results of the publication. Markku Turunen and Roope Raisamo were supervising the work and working on the administrative aspects of the project.

Publication IV: Kallioniemi, P., Heimonen, T., Turunen, M., Hakulinen, J., Keskinen, T., Pihkala-Posti, L., Okkonen, J., and Raisamo, R. “Collaborative navigation in virtual worlds: How gender and game experience influence user behavior”

The application implementation and the pedagogical aspects of the project were carried out by the same people as in Publication II. The experiment, including its methods, user questionnaires, and metrics, was designed by me, Laura Pihkala-Posti, Tuuli Keskinen, and Jaakko Hakulinen. I organized and observed the actual evaluations with Laura Pihkala-Posti. Markku Turunen and Roope Raisamo were supervising the work and working on the administrative aspects of the project.

Publication V: Kallioniemi, P., Sharma, S., and Turunen, M. “A collaborative online language learning application”

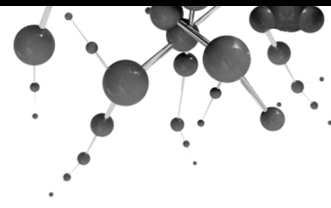
For this study, I implemented a CityCompass application (based on the pedagogical and language learning concepts of Berlin Kompass introduced in publications II, III and IV) together with Jaakko Hakulinen, and co-designed the Amaze 360 framework together with Santeri Saarinen and Ville Mäkelä for presenting omnidirectional video content with a head-mounted display (HMD). I collaborated with Sumita Sharma and Markku Turunen to report the results of this paper. Laura Pihkala-Posti worked as a language pedagogy expert in the project. Mark Barratt from Sanako acted in a consultative role in the planning of the application. Markku Turunen, Roope Raisamo and Olli Koskinen were supervising the work and working on the administrative aspects of the project.

Publication VI: Kallioniemi, P., Mäkelä, V., Saarinen, S., Turunen, M., Winter, Y., and Istudor, A. “User experience and immersion of interactive omnidirectional videos in CAVE systems and head-mounted displays”

In this study, I co-designed the Amaze 360 framework with Santeri Saarinen and Ville Mäkelä for presenting the omnidirectional video content with HMD. I also designed the cCAVE application with Andrei Istudor for providing omnidirectional video content in a CAVE-like system. I had the main responsibility for planning, conducting, and reporting the user evaluations presented in this publication. This project was administrated by York Winter and Markku Turunen.

Publication VII: Kallioniemi, P., Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., and Ronkainen, K. “Effect of gender on immersion in collaborative iODV applications”

For this work, I co-designed the CityCompass VR application (based on the pedagogical and language learning concepts of Berlin Kompass application presented in publications II, III and IV) that was then developed by Jussi Karhu and Kimmo Ronkainen. Together with Tuuli Keskinen and Jaakko Hakulinen, I was also responsible for planning, conducting, and reporting the user evaluations presented in this article. Markku Turunen was supervising the work and working on the administrative aspects of the project. This research was done as part of the PhD work funded by the University of Tampere.



1 Introduction

1.1 OBJECTIVE

The objective of this dissertation is to provide scientifically valid, novel research on *the process of human wayfinding in collaborative virtual environments*. It both offers new information on the topic and complements (and sometimes contradicts) the old research on the subject. The main research questions answered in this work are as follows:

RQ1: What strategies do people use to find their way in collaborative virtual environments?

RQ2: What aspects affect collaborative wayfinding tasks?

The answers to these questions are found in the publications presented in this thesis. For these experiments, many multimodal applications containing wayfinding tasks were implemented. Accordingly, the goal of this dissertation is to provide useful information for researchers and practitioners who work with virtual environments (VEs) or study human wayfinding. Each application presented contains different interaction methods and presentations of virtual environments. These applications also have similarities: each contains a collaborative element and utilizes panoramic images or videos. In addition, each of these applications and the tasks they provide rely heavily on landmark-based wayfinding. Using landmarks in wayfinding is a common strategy on which most people rely when navigating through space. Landmarks are often used as mental representations of space (Siegel and White, 1975; Hirtle and Heidorn, 1993), and people use them to communicate route directions (Denis et al., 1999; Lovelace, Hegarty and Montello, 1999).

1.2 CONTEXT OF THIS RESEARCH

The research reported in this dissertation is related to two main themes: collaborative human wayfinding and virtual environments. The focus of this work is the field of human-technology interaction (HTI), especially the design aspects of virtual environments that include collaborative wayfinding. To understand the methods of human wayfinding, we must rely on research regarding human cognition and spatial ability. HTI-related results provide guidelines for designing virtual environments that offer a better user experience and usability in general. This background provides the basis for the design of the applications used in the case studies of this dissertation and the context for the results presented in them. By conducting human-technology interaction analysis, we then attempt to understand further how humans find their way in virtual environments with multimodal applications. In addition, this analysis informs us of how parameters such as age and gender as well as landmarks affect our wayfinding tasks.

The three collaborative virtual environment applications (Figure 1) presented in this research were designed, implemented, and evaluated over the years 2013–2017. Subsequently, we created a theoretical framework for landmark-based wayfinding that was then used as the basis for the development of these applications. The applications were designed and developed in a range of interdisciplinary research projects. They were also used extensively for educational studies in the context of language learning. The results from these experiments are mostly outside the scope of this work, but findings related to this dissertation are reported in publication III. For further insights on this topic, see Pihkala-Posti and Uusi-Mäkelä (2013), Pihkala-Posti et al. (2014), and Pihkala-Posti (2014).



Figure 1. The three collaborative virtual environment applications presented in this dissertation, from left: Berlin Kompass, CityCompass, and CityCompass VR.

1.3 METHODOLOGY

Most of the studies presented in this dissertation follow the same pattern. First, an interactive application containing a wayfinding task was designed and implemented. These applications were often designed for several purposes, for example, as a language learning application. Second, the application was evaluated by a varying number of participants in a laboratory setting. Preceding these evaluations, the data collection and analysis process were carefully planned so the results obtained from the study were scientifically applicable. Existing evaluation methods were used where applicable. In some cases, these existing methods have been altered so they provide better results for the context of the study. Using these common methodologies in the studies also offers the future possibility of comparisons between the applications. Finally, the applications themselves result from constructive research and can be utilized in many fields, including education and language learning.

The laboratory-based user studies presented in this dissertation were traditional controlled experiments that examined the participant's wayfinding abilities or perception skills in virtual environments. The analysis mostly concentrated on quantitative metrics, such as total time spent on a task or the number of mistakes the user made during the task's completion. Since many of the applications contained collaborative aspects in which the users communicated via audio, some studies also include analysis in this context. Besides these metrics, the user experience questionnaires provided us with two types of information: what the user's initial expectations were for these applications and how the users experienced them. Individually, these results offered interesting observations on the user experience in multimodal applications in general and validated them for the wayfinding studies.

Some studies also contain qualitative results in the form of questionnaires or interviews. They provide supporting data for the quantitative results and in some cases explain the phenomena behind the quantitative numbers. The timeline of the publications and this introductory part can be seen in Figure 2.

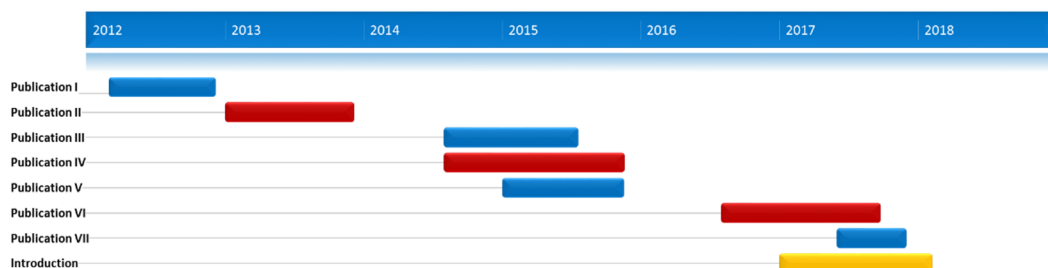


Figure 2. Timeline of the publications presented in this dissertation.

1.4 RESULTS

In this dissertation, I report on the strategies people use while performing collaborative wayfinding tasks in VEs. As collaborative wayfinding tasks conducted in VEs are novel research approaches, these results offer valuable insights into this subject. As previous research has shown, virtual wayfinding tasks are relatable to real-world situations (Witmer et al., 1996). Therefore, these results can also be applicable to these scenarios. For example, they can be used to improve existing wayfinding applications such as Google Maps or Apple Maps.

Publications I and II also introduce a model for landmark highlighting for wayfinding applications. The model is first introduced in Publication I, where it is evaluated in a laboratory setting with panoramic images, and it is later evaluated in a collaborative VE in Publication II. It is also evident in the following publications and applications where the use of landmarks is emphasized as a wayfinding strategy.

In addition, the results from user experience questionnaires offer more valuable data for researchers and developers who are interested in collaborative VEs. They also provide validation of the applications used in the experiments – positive overall results suggest that the users found the applications both useful and efficient.

1.5 STRUCTURE

This dissertation is a collection of scientific publications and a summary of the related work and backgrounds of these publications. The summary part is structured as follows: First, I introduce the three applications used in the experiments (BerlinKompass, CityCompass, and CityCompass VR). Second, I discuss the background of and previous research on VEs. Third, I will go through the basic concepts of human wayfinding supported by the previous research on the topic. Finally, I summarize my work and the results presented in my publications. The dissertation concludes with a discussion of the results and their relevance to the current state of the research in the field. In the conclusion, I also outline future work on the topic.

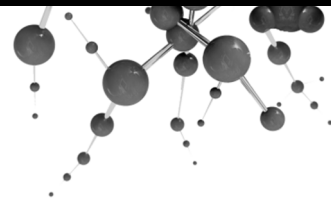
1.6 ON TERMINOLOGY

Presence/Immersion

Unlike in publications III, IV, VI, and VII, I suggest that for future studies the term *presence* should be used to refer to the phenomenon commonly called “the feeling of being there” and *immersion* as “an objective characteristic of a VE application,” as they are more commonly adopted. For a more extensive analysis of this terminology, see Skarbez, Brooks, and Whitton (2017).

Collaboration

There is no consensus in the scientific community on what constitutes *collaboration* and *cooperation* and the differences between these two concepts. Both terms are defined in many ways, often depending on the context of the research. Panitz (1996) defined *collaboration* as a “structure of interaction designed to facilitate accomplishment of a product or goal through people working together in groups.” In my research, there is no clear definition of either term, but in several experiments, the participants are working and interacting together to reach a common goal. From here on, the word *collaboration* is used, and it refers to this type of activity.



2 Applications

In this chapter, I introduce the applications that were used for the evaluations and studies presented in this dissertation. Each of the applications have similar characteristics and were developed back to back as an iterative process. One common denominator for these applications is that their tasks involve wayfinding in VEs. Two users also employed them collaboratively: one user acted as a *tourist*, trying to locate a local landmark (usually a tourist attraction), and the second one as a *guide*, helping the other user to find the goal. This concept was first introduced in Publication II.

The first application, Berlin Kompass, was evaluated in publications II, III, and IV. The second one, CityCompass, was rated in Publication V, and the third one, CityCompass VR (or Amaze360, as it was called in Publication VI) was assessed in publications VI and VII.

2.1 BERLIN KOMPASS

Berlin Kompass is a collaborative application that supports two simultaneous users. The first user takes the role of a tourist who has just arrived at a new city and needs to locate a local tourist attraction. Another user, acting as a guide, helps the tourist with the task. The application's content consists of 360-degree panorama photographs from various cities. This content is then ordered into sequential routes that the users can follow. Both users are located in different spaces and have their individual view of the application, which they can pan around freely. To go from one panorama (i.e., location) to another, the tourist must move along the route as per the guide's instructions. The tourist's view of the application can be seen in Figure 3.



Figure 3. The tourist's view of the Berlin Kompass, as used by the researcher. The panorama image is projected with three projectors (Publication IV).

Interaction Design

The Berlin Kompass supports embodied interaction—the application view can be panned by turning one's shoulders to either the left or right, which refreshes the panorama accordingly. This gesture was planned to emulate the natural way in which a human looks around his or her surroundings. This application has two kinds of user interface elements. The first is called an *exit*, which moves both users along a route. Exits are only visible in the tourist view, as only they can move along a route. When a tourist has a visible exit on the screen, he or she can activate this exit by walking towards the center of the projection (actually, the Kinect device located below the screen). This action moves both the tourist and the guide to the next panorama. The Kinect device is used to track both users' movements.

Users can also use pointing gestures at the *hotspot* objects found in the panoramas. These objects offer vocabulary and contextual information about the surroundings, and they are always overlaid over landmarks. Once one of these hotspots have been activated, a textual information box (e.g., "a modern office building") describing the object becomes visible. This content is also played audibly to support the language learning aspects of the application. These utterances are output via speech synthesis. The hotspot information varies from single nouns to longer descriptions of the target object (e.g., "a building" versus "a gray office building with a sign on

top”). This information can then be used to describe the surroundings to the other user.

Both users can see the panoramic view as a projection in front of them. The field of view (FOV) used by the application can vary based on the projection type and can be extended with multiple projectors and displays. For example, in Figure 3 three projectors were used to achieve a 160-degree view. Polys et al. (2005) stated that a larger FOV is more efficient with search-based tasks. Based on our results with Berlin Kompass (Publication II), it was also perceived as more satisfactory than interfaces with a lower FOV.

Berlin Kompass is a realistic collaborative VE that uses 360-degree panoramic photographs with embodied interactions. Sequential routes provide a good basis for both wayfinding and language learning experiments, and the collaborative aspects of the application encourage users to communicate and work together.

System Architecture

Berlin Kompass has four distinct components: 1) central logic, 2) graphics and voice service, 3) Kinect service, and 4) audio transmission service (Publication II). The overall program logic is handled by the central logic component. This component is responsible for receiving and sending messages between other system services. In addition, the communication between the clients is handled by this component. It also controls the activation of exits and dead ends.

The graphics and voice service handles the visual and auditory aspects of Berlin Kompass, and it is implemented on top of a graphic engine called Panda3D.¹ This component handles the display of cylindrical panoramas, their FOV, and speech synthesis content. The Kinect service tracks the user’s physical location and skeletal joints. The data from these is then transformed into gestures, which are used to control the GUI of the system. This includes the panning of the screen and movements from one panorama to another. This service also handles the pointing gesture while utilizing the Microsoft Kinect SDK. The audio transmission service handles the communication between the two installations. The service sends audio between the clients in User Datagram Protocol (UDP) packages. The Berlin Kompass’s application architecture can be seen in Figure 4.

¹ <https://www.panda3d.org>

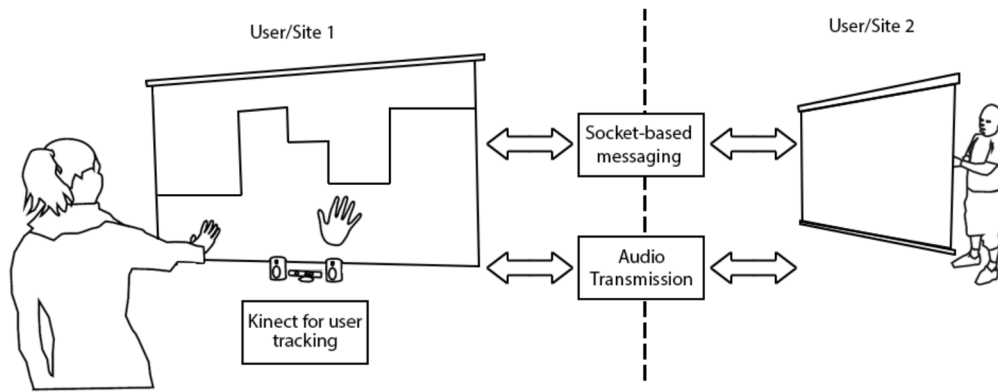


Figure 4. The Berlin Kompass’s system architecture. Remotely located users can interact using the built-in audio connection. The application’s statuses (tourist is moving, users have reached their goal, etc.) are transferred via socket-based messaging. (Publication II)

Wayfinding Scenario

Berlin Kompass supports two simultaneous users who must communicate and collaborate to complete a wayfinding task. Before starting this task, the user who acts as a tourist selects one of the three routes. All these routes start from the same location, but each has its own goal. The routes are based on real geographic locations around downtown Berlin. Once the route has been selected, both users are taken into the first panorama, and they can start communicating with each other. This communication is mediated via a headset. Because the application is designed for language learning, the users communicate in a predetermined language. In addition, contextual information is presented in this language. Currently, the application supports German, English, French, and Swedish.

As mentioned before, only the tourist can activate the exits and move along the route. Once an exit is activated, both users are transitioned to the next panorama. In each panorama, the tourist has three to four exits to choose from, and only one of these exits takes users closer to the goal. Therefore, it is crucial that users communicate with each other clearly.

Once an incorrect exit is chosen, both the tourist and the guide are transitioned to a dead end. After this, the tourist must describe the contents of an image to the guide, who then needs to pick the correct image from four options. This scenario is presented in Figure 5. If an incorrect image is selected, both users stay in the dead-end panorama. When the guide chooses the correct image, both users are transitioned back to the panorama where they got lost (i.e., where they activated the dead-end exit).



Figure 5. In dead-end panoramas, the tourist describes his or her location (above) so the guide can select the correct image from four options presented below (Publication IV).

2.2 CITYCOMPASS

CityCompass is the web-based successor to the Berlin Kompass application. It was developed for cross-cultural collaboration and language learning. Modern web technologies allowed the application to be used within the browser, something that could not be done with the previous version. CityCompass uses 360-degree panoramic cityscapes, or panoramas, just like its previous version.

Interaction Design

CityCompass has two interaction methods: 1) a traditional mouse and keyboard, and 2) a touchscreen for monitors with touchscreen support, smartphones, and tablets. Like the previous version, the routes in CityCompass consist of 360-degree panoramic images of real-world cityscapes that can be panned freely by the user. Like Berlin Kompass, CityCompass also has two types of user interface objects: *hotspots* and *exits*. The former offers contextual information to the user about locations in the panorama. This information is presented as both text and audio. For the audio, a speech synthesis service was used. The latter is used to activate transitions from one panorama to another.

Like Berlin Kompass, CityCompass has separate views for each user. The basic interaction with the application is similar to the previous version, and the user acting as a tourist can move along per the guide's instructions. Both users can activate hotspots for contextual information and guidance for their collaboration and communication. The tourist's view of the application can be seen in Figure 5, and the guide's view, in Figure 6. The application also has a small dictionary for its navigational vocabulary.



Figure 6. A tourist's view in CityCompass with a hotspot activated and one green arrow (exit).



Figure 7. A guide's view of CityCompass with a hotspot activated and a blue line that indicates the direction where the tourist should be led.

System Architecture

CityCompass was implemented with web technologies. The application is based on a client-server architecture. The client-side panorama view was created with JavaScript and three.js², a JavaScript library that enables the creation and display of 3D graphics in a web browser. In addition, WebRTC³ was used for audio and video transmissions between users.

For the server side, a Node.js JavaScript component was used alongside Express and MongoDB. These components transmitted the necessary data between the clients. There is also a CityCompass implementation that provides the same embodied interaction as Berlin Kompass. This version uses Microsoft Kinect SDK and a custom module called Skeleton Server for tracking the user's location and skeletal joint data.

² <https://www.threejs.org>

³ <https://webrtc.org>

Wayfinding Scenario

CityCompass has the same premise as Berlin Kompass. One user acts as a tourist, and the other takes the role of guide. The users then collaborate to find a local tourist attraction. The application has three different cities to choose from: Tampere, Delhi, and Berlin. Hotspots contain contextual information about the environment (e.g., an old white office building) that can be used as assistance for the wayfinding task. The tourist can move along the route by activating exits. The dead ends in CityCompass work the same way as in Berlin Kompass. The biggest difference between Berlin Kompass and CityCompass are the contents of the routes – Berlin Kompass had routes from only one city, whereas CityCompass has routes from other countries and cultures. The landmarks and contextual information between these routes differs vastly, allowing users to take virtual tours to different cultural environments.

2.3 CITYCOMPASS VR

The latest version of this application stack is CityCompass VR. This collaborative iODV application offers the same premise in a more immersive environment. Instead of photographs, CityCompass VR uses omnidirectional videos as content.

Interaction Design

To use the application, both users wear the HMD device seen in Figure 8. This application uses a head-position-based interaction technique presented in Publication VI, meaning that the user's viewport is refreshed based on the position of the HMD. For creating a stereoscopic effect, the application view is divided into two views, one for each eye. This viewport division can be seen in Figure 8. This presentation is accomplished by overlaying the video on a virtual sphere. CityCompass VR has the same user interface (UI) elements as the two previous versions, but their activation differs a little bit – this was done by using a dwell timer, meaning that these elements are activated after the user has focused the desired UI object on the center of the screen for a pre-defined duration of time. This activation method has been utilized with interfaces that use gaze or mid-air gestures, for example, in Mäkelä et al. (2013).



Figure 8. Top: A researcher wearing the CityCompass VR headset (Publication VI).
Bottom: Two viewports for the CityCompass VR application.

System Architecture

CityCompass VR was implemented on Unity, and it uses Unity's native video for video playback. It can be used with Samsung Galaxy S7 and S8 smartphones together with the Samsung GEAR. Like its previous version, CityCompass VR deploys a client-server architecture. The application also has a separate observer view that shows video content and clients' UI elements and logs all the users' necessary actions. This logger also supports the recording of gaze data, making it feasible for gaze tracking-related experiments. The observer view can also play back the audio from both clients, thus allowing the recording of communication between the two users. All messages and audio between the clients are relayed with separate Photon Unity Network and Photon Voice plugins. The CityCompass VR

application architecture can be seen in Figure 9. The video content used in CityCompass VR is 360 x 180 degrees, and the viewport has a FOV of 60 degrees. In addition to the HMD, users wear a headset for communicating.

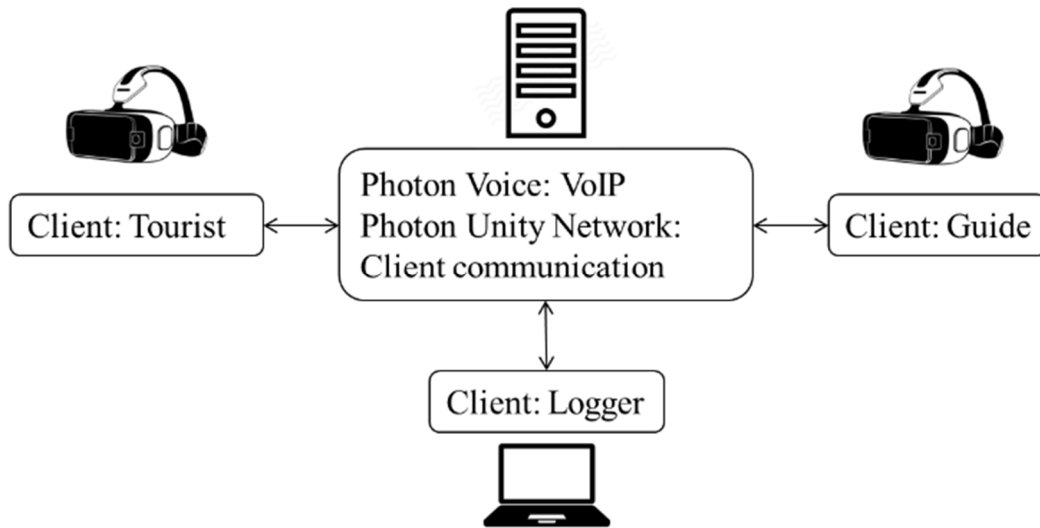


Figure 9. CityCompass's VR architecture (Publication VII).

Wayfinding Scenario

CityCompass VR currently has only one route. Its starting location is at the Tampere railway station, from which users try to find their way to the Finlayson business district. Unlike previous versions, users can choose either one of the two sub-routes while performing the task. Depending on which sub-route is selected, the route consists of eight or nine scenes. (In CityCompass VR, intersections are called *scenes* instead of *panoramas*, because they consist of video content.) This route with its intersections can be seen in Figure 10.

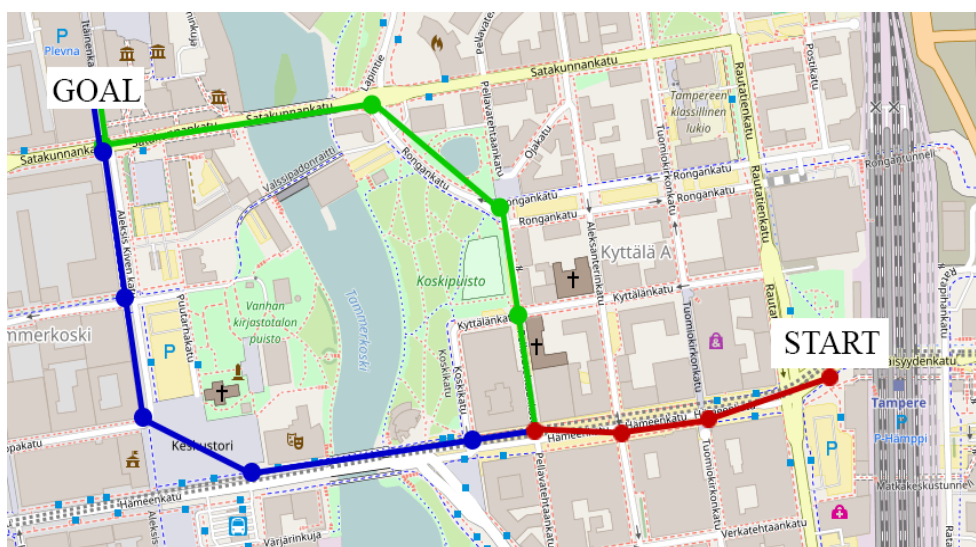


Figure 10. The route used in CityCompass VR. This route has two different sub-routes presented in blue and green. (Publication VII)

The dead-end scenarios in CityCompass VR are a bit different from the previous versions. When the tourist activates a dead end, both users are transitioned to a dead-end scene. These scenes are indicated with a red lock at the center of both users' viewport. Once in one of these scenes, the users' roles are flipped: now the guide needs to find the correct route from several exits, and the tourist has to guide them (See Figure 11). After activating the correct exit, they are sent back to the previous scene. This role-switching was done so that each user can experience both roles at least to some degree, and was considered to be useful based on the insights from evaluating the previous versions of the application.

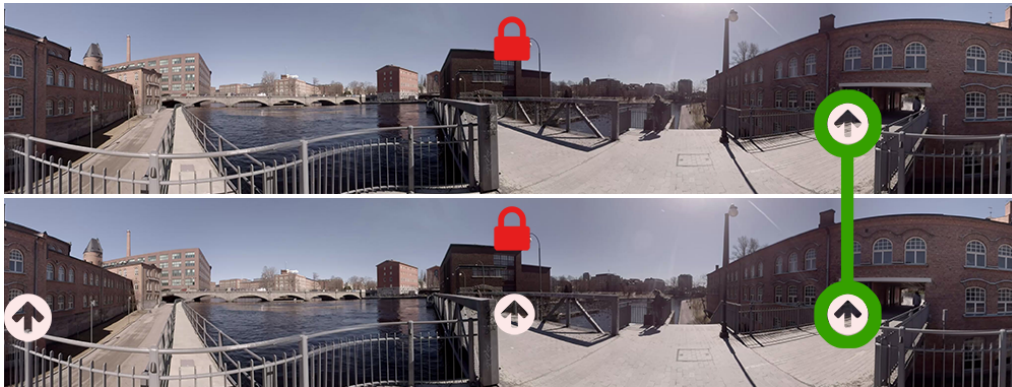


Figure 11. Example of a dead-end scene with the tourist's view above and the guide's below. The lock icon at the top of the screen indicates a dead-end scene, and the correct route is marked with the green line at the right side of the image. The other two arrows are incorrect. Activating them only keeps both users in the same panorama (Publication VII).

2.4 SUMMARY

This chapter introduced the three evolutionary stages of the CityCompass application. All three applications contain similar characteristics and were developed as an iterative process. The main similarities and differences between these applications can be seen in Table 1.

	Content type	FOV (degrees)	Interaction method	Architecture
Berlin Kompass	360-degree panoramic images	Depends on projection, up to 180	Embodied interaction (Kinect)	Client-client
CityCompass	360-degree panoramic images	Depends on projection, up to 180	Mouse/ Touchscreen	Client-server
CityCompass VR	Omnidirectional videos	60	Head position-based dwell timer	Client-server

Table 1. CityCompass applications and their features.

Although originally developed for language learning, these applications provide a platform for several research purposes. They have been mainly used for collaborative wayfinding studies, and each application offers its own approach to this topic. High-quality images and videos of real landscapes provide a sufficient level of immersion, and different interaction methods can also offer various research possibilities. The content creation process for each application is planned so it can be crowdsourced. For Berlin Kompass and CityCompass, content can be created with a smartphone that has panorama image capabilities. By changing the content, these applications can be flexibly used in assorted contexts. For example, they could be used for educational purposes in biology or science training, or for industrial showroom purposes. Each application still has several possibilities for further development. CityCompass VR especially offers numerous new research topics.

.....



3 Wayfinding and Landmarks

To understand the process of pedestrian wayfinding, spatial cognition, and cognitive mapping, we must embark on interdisciplinary research in cognitive psychology.

3.1 HUMAN WAYFINDING

Human movement is often divided into two categories: *navigation* and *wayfinding*. *Navigation* is described as the “processing of spatial information regarding position and rate of travel between identifiable origins and destinations summarized as a course to be followed” (Golledge, 1999), and *wayfinding* is the process of “selecting path segments from an existing network and linking them as one travels along a specific path” (Golledge, 1999). The selected path can vary based on the purpose of the trip and its requirements such as travel speed and efficiency. Wayfinding as a process is manifold, requiring them to know the origin and seek a possibly unknown destination. In addition, it requires the person to estimate turn angles in the correct sequence, remember how long route segments are, determine the direction of one’s movement along a segment, maintain one’s orientation, estimate one’s location based on landmarks, and differentiate between cues along or off the route (Golledge, 2000).

Allen (1999) introduced a taxonomy for wayfinding tasks (and the means for accomplishing them) with the following main categories:

- a) Traveling, where the goal is to reach a familiar destination,
- b) Exploratory traveling with no goal, where the traveler eventually returns to a familiar point of origin, and
- c) Traveling with the goal of reaching a novel destination.

The most used method of wayfinding is travel between common locations, for example, commuting from home to work and vice versa. Another common task is explorative traveling, which happens especially in scenarios where the person has moved to another location or when one visits a new environment, for example, on vacation. Wayfinding to novel destinations is often supported by symbolic spatial information that is then communicated to the wayfinder via different media (paper maps, verbal directions, wayfinding applications such as Google Maps, etc.). This type of wayfinding has also been observed in nonhumans, for example, in honey bees, who provide spatial information (e.g., in migration scenarios) via a specific dance.

Wayfinding is an activity that can be “observed and recorded as a trace of sensory motor actions through an environment. This trace is called the route” (Golledge, 1999). The selected route results from a travel plan, which comprises route segments and turns that lead the wayfinder to his or her destination (Golledge, 2000). This travel plan is determined by the criteria of the path selection (i.e., by the motivation of the traveler), such as the shortest distance, the shortest time, or the scenic nature of the path (see Table 2 for route selection criteria). These travel plans can also be organized by their legibility or the ease with which the route can become known to the person.

Wayfinding takes place in large-scale environments (Montello, 1993), such as cities and buildings. This means that the traveler cannot perceive the route from a single viewpoint and therefore must travel through the space to experience them (Nothegger, Winter and Raubal, 2004). To navigate these landscapes, people must utilize their spatial and cognitive abilities. This includes the person’s capability to process perceptions and information, previous knowledge, and motor functions (Allen, 1999). The cognitive requirements of wayfinding also depend on the task, meaning that wayfinding through a cityscape uses a different set of cognitive abilities than wayfinding inside a building (Nothegger, Winter and Raubal, 2004).

Longest leg first	Maximizing aesthetics
Shortest leg first	Minimizing effort
Fewest turns	Minimizing actual or perceived cost
Fewest lights or stop signs	Minimizing the number of intermodal transfers
Fewest obstacles or obstructions	Fastest route
Variety of seeking behaviors	Least hazardous in terms of known accidents
Minimizing negative externalities	Less likely to be patrolled by authorities
Avoiding congestion	Minimizing the number of segments in a chosen route
Avoiding detours	Minimizing the number of curved segments

Table 2. Types of Route Selection Criteria (Golledge, 1999)

There are various wayfinding strategies used by humans (and other animals), including:

- Oriented search
- Following a marked trail
- Piloting (moving from landmark to landmark)
- Habitual locomotion
- Path integration
- Referring to a cognitive map (Allen, 1999)

An *oriented search* is a simple way of reaching a destination in which the wayfinder first orients himself or herself according to a source of information and then searches until the destination is reached. This wayfinding method is utilized by many species. Even though some species rely on distal visual (sonar, lunar, and stellar), tactile (wind and water currents), geomagnetic, and olfactory information, humans rely most heavily on visual, vestibular, and proprioceptive information (Allen, 1999). An oriented search is most useful in the exploratory travel of short distances where the wayfinder finally returns to a familiar point of origin. *Following a marked trail* is a rather commonly used method of wayfinding and it is often found, for example, in hospitals or hiking trails. Marked trails are designed to minimize uncertainty and, therefore, to reduce the cognitive demands of the wayfinder. The problem with marked trails is that, when multiple instances are located in one segment (e.g., highway interchanges), the cognitive demands of the wayfinder increase (Allen, 1999). They are also relatively expensive to construct.

Piloting from landmark to landmark is a common method of wayfinding for many species. In landmark-based piloting, the wayfinder relies solely on sequential knowledge, meaning that a landmark is associated with only two types of information—the direction and the distance to the next landmark on the route. This type of wayfinding is an efficient way of traveling to familiar or novel destinations when in a well-known environment, and it is usually the standard method of wayfinding in an unfamiliar environment. Wayfinding instructions based on piloting consist of condition-action lists. The success with this method relies heavily on the recognition of landmarks. Piloting is also a common technique in explorative wayfinding (Allen, 1999). *Habitual locomotion* is a wayfinding method that is only utilized with familiar locations. After repetition, the wayfinder gets increasingly experienced with specific routes, which can lead to automatized locomotion on these routes. In time, the attention to the environment required for traveling the route diminishes. For example, many people returning from work pay little to no attention to the trip that brought them home. *Path integration* is “orienting by means of external and internal sources of information regarding direction and speed of movement” (Loomis et al., 1999). Path integration depends on the monitoring of one’s own self-movement. Path integration is utilized by other species, including small rodents (Alyan and McNaughton, 1999) and ants, who are extremely adept at it (Graham and Cheng, 2009). The most sophisticated model of wayfinding involves the use of an internal representation of relationships between places referred to as a *cognitive map*. The following section will explain this concept and expand on the cognitive aspects of wayfinding.

The possible utility of wayfinding methods for divergent wayfinding tasks can be seen in Table 3. Multiple methods can be used for the same wayfinding task, and most means can be used for multiple tasks. Finally, there are more methods for traveling to familiar destinations than exploratory travel, which in turn has more methods than traveling to novel destinations. To put it another way, there is flexibility in solving each type of wayfinding task, but this flexibility is greater when traveling to familiar destinations than in exploratory travel and more in exploratory travel than when traveling to novel destinations. This type of categorization is important when addressing individual differences in wayfinding performance, as one should also consider the nature of the wayfinding task. Travelers may differ in their wayfinding abilities because they use distinct methods (e.g., path integration versus piloting when returning home from exploratory travel). In addition, they may differ in their ability to assess these methods in their wayfinding (e.g., poor ability to identify landmarks).

Wayfinding experiments are often divided into two categories: those done in closed spaces such as buildings or rooms (Shanon, 1984) and those conducted in open, often large-scale environments such as cityscapes or campuses (publications I and II).

Wayfinding tasks

Wayfinding means	Travel to familiar destination	Exploratory travel	Travel to novel destinations
Oriented search	x	x	x
Following a trail	x	x	x
Piloting	x	x	x
Path integration	x	x	
Habitual locomotion	x		
Referring to a cognitive map	x	x	x

Table 3. The possible utility of proposed wayfinding methods for various wayfinding tasks (Allen, 1999).

3.2 COGNITIVE ASPECTS OF WAYFINDING

Cognitive Maps

People make wayfinding decisions based on a previously acquired spatial understanding of their world; this spatial representation of the environment is called a *cognitive map*. This term was originally introduced by Tolman (1948), and it refers to the mental representation of spatial relationships between essential objects (landmarks, locations, etc.) in human environments and the possible connections between these objects (Golledge et al., 2000). Humans develop these maps to answer questions such as: Where am I located? Where is my home? Where is my destination? Which route do I select to reach my destination? How will I know when I am lost (Boswell, 2001)? These questions are the basis of wayfinding, and they are also the reason we form cognitive maps (Golledge, 1999). In optimal situations, a cognitive map offers the possibility of locating the position of a specific destination and enables the wayfinder to find (or plan a route to) this destination (Ellard, 2009). Therefore, cognitive maps are “the internal representation of experienced external environments, including the spatial relations among features and objects” (Golledge, 1999), and cognitive mapping is the process of “encoding, storing and manipulating experienced geo-referenced information” (Golledge, 1999). It is still unknown how humans conduct this mapping, and it is an active topic of research within the field of neuropsychology and related fields. For example, Kitchen and Blades (2002) have integrated cognitive theories from geography and

psychology to enable a better understanding of environment-behavior interactions and cognitive maps.

For humans to travel, two active processes are required to facilitate spatial knowledge acquisition:

- a) *Person-to-object relations*, or the so-called *egocentric referencing* that changes as movement takes place, and
- b) *Object-to-object relations*, or the so-called *anchoring structure* of a cognitive map, which remains stable during a person's movement (Sholl, 1996).

In real-life scenarios, a traveler can become disoriented because of poor person-to-object comprehension. In these situations, the traveler can still understand the basic structure of the environment in which he or she is moving. Errors in the encoding of object-to-object relations may lead to scenarios in which the wayfinder misspecifies the anchor point's geometry. These scenarios often produce the distortions and fragmentations found in spatial products like maps (Golledge et al., 2000).

A knowledge of human wayfinding can be divided into three general categories (Golledge et al., 2000):

- a) *Route learning*, in which the traveler navigates a novel environment and tries to find his or her way around,
- b) *Route knowledge* acquisition, in which travelers understand their location along the route in a larger frame of reference, and
- c) *Survey knowledge* acquisition, which is the highest level of spatial knowledge, including spatial layouts and information such as locations, orientations, and distances between objects along the route. This information can then be linked to a network that can act as a frame of reference for environmental knowledge (Golledge, 1999).

Humans usually rely heavily on their visual, sensory-motor, and proprioceptive senses instead of using instruments or mapped representations when building a representation of their surroundings. Therefore, humans' environmental knowledge is mostly obtained during their movement through the environment (MacEachren, 1992). However, human senses are not reliable, and spatial representations are often incomplete. This can produce distortions or fragmentations in spatial awareness and lead to errors in wayfinding tasks.

Spatial Abilities

Imagine a scenario in which two of friends visit another for a week and take several trips around town. After their journey, one friend might have acquired a detailed knowledge of the town, while another friend may only remember the name of their hotel. Montello (1998) has pointed out that even

with equal levels of exposure, the spatial knowledge of two individuals may differ greatly. The ability to remember and recall environmental knowledge varies between individuals (see, e.g., Ishikawa and Montello, 2006), and the nature of this knowledge also varies. Evidence also supports the presence of individual changes in the development of the ability to learn route and survey knowledge (Piaget and Inhelder, 1967). This ability differs between age groups. For example, Pellegrino et al. (1990) observed large differences in spatial learning between pre-teen, teenaged, and adult participants. Part of this can be explained by the better understanding of spatial layouts and configurational structures in adults (Bell, 2000).

Spatial abilities can be grouped based on their function, that is, the situations in which they are used or based on their purpose. Allen (1999) stated that the most used and recognized of these abilities are the following:

- a) *Visualization*, or “the ability to imagine or anticipate the appearance of complex figures or objects after a prescribed transformation” (Lohman, 1988),
- b) *Spurred rotation*, or the ability to determine whether one object is a rotated version of another, and
- c) *Spatial orientation*, the ability of “an observer to anticipate the appearance of an object or object array from a prescribed perspective” (Allen, 1999).

There are several methods for evaluating these spatial abilities. More traditional samples can be found in Ekstrom et al. (1976); Ekstrom, French, and Harman (1979) provided information about the development of these samples.

Allen (1999) and Golledge et al. (2000) placed spatial abilities into three distinct categories:

- a) A stationary individual and manipulable objects,
- b) A stationary or mobile individual and moving objects, and
- c) A mobile individual and stationary objects.

Out of these three categories, the last is most related to the process of wayfinding, that is, a traveler is moving in large-scale environments consisting of both mobile and stationary objects. Thus, spatial abilities play a critical role in human wayfinding, including the construction and use of cognitive maps.

Spatial Knowledge

When the human mind is constructing a spatial representation of the surrounding environment, it contains a collection of geographic features. Lynch (1960) divided these features into four distinct categories: *paths*, *districts*, *edges*, *landmarks*, and *nodes*. All these features have coordinates,

distances among them, and all the other knowledge required for orientating oneself in the environment. This spatial representation aids travelers in locating and moving themselves within an environment and prevents them from getting lost (Siegel and White, 1975). Spatial knowledge is usually gained through the exploration of an environment, but it can also be gained from indirect sources, such as spoken instructions and maps (Burnett and Lee, 2005).

Thorndyke (1981) divided spatial knowledge into three categories:

- a) *Landmark knowledge*: knowledge of salient features or objects in the environment
- b) *Procedural knowledge*: knowledge of route representation, that is, the sequences that connect locations or segments in the environment
- c) *Survey knowledge*: knowledge about the global organization of features and the relationship between routes

It has been suggested by several studies that a traveler's knowledge increases sequentially, meaning that spatial knowledge progresses first from landmark knowledge to procedural and finally to survey knowledge with increased familiarity with the environment (Thorndyke, 1981). Based on Siegel and White (1975), landmarks and routes are necessary and sufficient elements for wayfinding to occur.

3.3 ROUTE DIRECTIONS

By dividing a travel plan into segments, it can be transformed into route directions. *Route directions* are a "set of instructions that prescribe the actions required in order to execute that course, step by step, in an appropriate manner" (Allen, 2000; Denis, 1997; Denis et al., 1999; Fontaine and Denis, 1999; Golding, Graesser and Hauselt, 1996; Lovelace, Hegarty and Montello, 1999). Their basic function is to describe sequential, ordered actions that take the wayfinder from his or her origin to a goal. These actions often include reorienting the traveler along the route.

While moving along a route, the wayfinder perceives his or her surroundings, which is why route directions rely on the perceptive nature of their users. Therefore, the comprehension and following of route directions are outcomes of "a collaborative, goal-directed communication process" (Golding, Graesser and Hauselt, 1996). For example, a route direction, "turn left after the church," requires the user to locate the church and then reorient himself or herself after passing that specific landmark. This means that the objective of route directions is to "deliver a combined set of procedures and descriptions that allow someone using them to build an advanced model of the environment to be traversed" (Michon and Denis, 2001). After the route has been followed several times, the wayfinder might

start remembering path components for later use. Michon and Denis (2001) referred to this process as *route learning*.⁴ If following the route becomes a daily habit, the traveler might enter a state of habitual locomotion where little to no attention is paid to the environment along the route.

Michon and Denis (2001) categorized the descriptive components of route directions into three categories. The first category contains the travel nodes on which the wayfinder is moving. Examples of these travel nodes are streets, paths, and alleys. Travel nodes are required for a two-dimensional extension, and they should have both a length and a width (Michon and Denis, 2001). Travel nodes are usually presented in vectors, which can be categorized by their type (e.g., “street” or “road”) or by their proper name (e.g., “Oxford Street”).

The second set of entities are the specific points along the travel nodes. These points refer to a location where reorientation should be performed. These entities are often referred to by linguistic expressions, for example, “from the intersection” or “at the edge of a forest.” They are to be distinguished conceptually from objects, such as landmarks, that may be located at these points. That is, they are coordinate points that have a metric value.

The third set contains the objects, for example, landmarks, which are located along the travel nodes. These objects refer to points or areas that have limited size (Michon and Denis, 2001). These entities serve a variety of functions in route directions. Most often, they are used to signal locations where reorientations should take place. Their second function is to support the locating of other landmarks (piloting). The third function of these entities is to confirm their location to the wayfinder. These objects, which are often landmarks, have a crucial role in route directions. They are generally considered one of the most important components of wayfinding and for constructing the cognitive maps and spatial representations used during wayfinding (Michon and Denis, 2001).

Route Strategy and Survey Strategy

Taylor and Tversky (1996) separated wayfinding strategies into two distinct categories: route strategy and survey strategy. With route strategy, the point of view originates from the wayfinder, meaning that locations are typically described using descriptors (left, right, front, back). For example, “to get to the square, take a left at the intersection and go straight. The square will be on your right.” Route strategy proceeds segment by segment, while adopting the wayfinder’s point of view that is updated after each segment. With survey strategy, a fixed reference frame using the surrounding environment is adopted (Allen, 2000; Taylor and Tversky,

⁴ It should be noted that Golledge et al. (2000) used the term *route learning* to refer to the process by which a person travels through a novel environment.

1996). According to this strategy, locations are usually described with cardinal directions (north, south, east, west) and in distance units. For example, “to find the square, turn west at the intersection. Travel west for 50 meters; the square will be facing north.”

Dabbs et al. (1998) stated that culture and evolution may determine the adoption of either strategy. They theorized that, in hunter-gatherer communities, males were predominantly hunters and females gatherers (Silverman et al., 2000). Hunters often navigated by using cardinal directions and global landmarks (for instance, the sun), that is, using the survey strategy, which provided them a higher level of space constancy (Bisiach et al., 1997). As gatherers moved around in smaller environments, the use of local landmarks and features, meaning route strategies, benefited them more. The distinction between route and survey strategies has also been supported by research from the field of neurobiology. Goldman-Rakic (1995) reported that those individuals who preferred route strategies showed increased activity in their right parietal and prefrontal areas, which are considered responsible for handling information regarding landmarks. Additionally, those travelers who preferred survey strategies had more activation in the left hippocampal areas. These areas are commonly connected to the use of more “bird’s eye views” (Walkowiak, Foulsham and Eardley, 2015). Lawton and Kallai (2002) developed the International Wayfinding Scale for measuring one’s preference for wayfinding strategies. Route directions should be designed so they are easy and quick to comprehend and understand (Lovelace et al., 1999). In optimal situations, these directions should characterize an external representation of a route that supports the wayfinder’s spatial cognitive processes and knowledge representations (Klippel, 2003). This is also the reason human spatial cognition needs to be studied.

What constitutes an effective wayfinding direction, and which cues, survey, or route is more effective? Lovelace, Hegarty, and Montello (1999) suggested two possible methods for assessing this issue. First, an effectiveness rating can be measured by asking participants how effective a route would be in wayfinding to a specific destination. Generally, directions that contain landmarks receive higher effectiveness ratings than those without them (Denis et al., 1999; Lovelace, Hegarty and Montello, 1999). The second method for calculating the effectiveness of wayfinding directions is to measure certain behavioral indices of wayfinding (Lovelace, Hegarty and Montello, 1999). These indices may contain the duration of wayfinding task completion, number of errors, and time spent on reorientations at decision points. There is no consensus on what constitutes “good route directions,” but many researchers (e.g., Allen, 1997; Denis et al., 1999; Waller, 1985) have made suggestions about the most important features in them, for instance: a) priming the wayfinder for upcoming choice points, b) mentioning landmarks at the choice points, c) informing

the wayfinder if he or she has made an error in the task, d) providing landmark information instead of street names, e) giving distances between the choice points, f) informing the wayfinder which way to proceed from a choice point, g) providing sufficient information for error recovery, and h) providing minimal redundant information.

Allen (2000) has stated that adults commit fewer errors when they use directions containing route cues rather than survey cues. When asked about their preferences, participants often state that route cues, such as landmarks, are one of the most useful features in effective wayfinding directions (Hölscher, Tenbrink and Wiener, 2011; Padgitt and Hund, 2012). Many experiments regarding route knowledge take place in urban environments or inside buildings, where route alternatives are rather limited. Hurlebaus et al. (2008) conducted an experiment in an open environment lacking any road networks, predefined locations, and unique landmarks. Their results stated that, at least in these environments, humans tend to rely on a combination of both route and survey knowledge, and that these two strategies actually complement each other.

There are still some discrepancies with the results. For instance, in one study (Chai and Jacobs, 2009), the survey strategy was reported to correlate with better wayfinding performance. Further comparisons between survey descriptors and route descriptors have been conducted with model towns (e.g., Hund and Minarik, 2006; Hund, Haney and Seanor, 2008). The results from these experiments suggest that survey descriptors are more efficient (i.e., result in faster task completion times and fewer errors), but the problem with this type of experiment is that it removes the essential factors such as motor functions and the exploration of space from the process of wayfinding. Taube, Valerio, and Yoder (2013) stated that spatial orientation relies heavily on locomotion and motor, vestibular, and proprioceptive systems. They suggest that the absence of these motion-based systems should be considered when interpreting results from wayfinding tasks in which the participant is stationary to achieve a more accurate understanding of the underlying mechanisms regarding wayfinding.

It is also possible that these discrepancies are due to the nature of the wayfinding task and its working memory demands. In a study by Denis et al. (1999), participants found their destination more efficiently with route cues once they were given more time (a total of two minutes) to memorize and learn them. As this study suggests, the discrepancies may also be due to the way the participants memorize these wayfinding instructions, as following directions from memory differs greatly from following directions segment by segment. For example, modern route guidance applications such as Google Maps provide segment-by-segment instructions, but also allow users to see their overall progression along the route. The role of working memory in wayfinding tasks was studied by Meilinger, Knauff,

and Bühlhoff (2008) with a dual task paradigm. In this study, the participants learned two routes through the VE of a city while conducting visual, spatial, or verbal tasks. Their performance was hindered while they were performing verbal and spatial secondary tasks, but not the visual task. These results suggest that utilizing verbal and spatial working memory resources are necessary for wayfinding tasks (see also Wen, Ishikawa and Sato, 2011). Therefore, it is important to make the distinction between following wayfinding directions from memory or from segment to segment. Allen (2000) has also suggested that, while performing wayfinding tasks, the wayfinder's memory is more taxed during the latter portions of the task, meaning that effective route directions should emphasize descriptives during these segments.

Sense of direction, or the confidence in one's ability to keep track of one's location within an environment (Kozlowski and Bryant, 1977), also plays a part in wayfinding performance. Sense of direction is usually measured by pointing accuracy and complemented with self-reported measures (e.g., Hund and Nazarczuk, 2009). It has been reported that, when one's sense of direction improves, so does one's performance at wayfinding tasks (Hund and Nazarczuk, 2009; Kato and Takeuchi, 2003). Those individuals who have a good sense of direction often adopt optimal strategies for wayfinding tasks (Kato and Takeuchi, 2003), but also suggested further studies on the subject for different wayfinding contexts.

One factor that has also been suggested to affect wayfinding effectiveness is *mental rotation*, the ability to "process spatial details by mentally rotating objects or environmental features" (Hund and Gill, 2014). This ability can be measured by using the Mental Rotation Test (MRT) developed by Vandenberg and Kuse (1978). In MRT, the participants must match rotated three-dimensional objects to a target object. This ability was connected to spatial abilities and map learning by De Beni, Pazzaglia, and Gardini (2006), who stated that individuals who scored higher on the MRT were better at learning maps. Moreover, Padgitt and Hund (2012) remarked that participants with high scores on the MRT made fewer errors with wayfinding tasks while using survey cues. This research suggests that mental rotation ability affects wayfinding when survey strategies are utilized. One final factor that appears to affect wayfinding is *spatial anxiety*. Hund and Minarik (2006) examined this phenomenon in a study. They found that people who self-reported more spatial anxiety made more errors during wayfinding tasks, suggesting that spatial anxiety affects wayfinding effectiveness. In addition, Lawton and Kallai (2002) discovered that females report higher levels of spatial anxiety than males.

Hund, Schmettow, and Noordzij (2012) investigated cultural differences in wayfinding. They studied subjects in the United States and in the Netherlands. These participants provided wayfinding instructions for

fictional recipients from both route and survey perspectives. Individuals from the United States referred more to street names than the Dutch, whereas the Dutch relied more on landmark information. In addition, US participants used more cardinal descriptors, whereas the Dutch ignored them almost completely. This research suggests that people from different cultures adopt a large variety of wayfinding strategies. It also is worth noting that both are Western cultures, which are generally considered to be quite similar. Cultural differences in wayfinding strategies definitely require more study in the future.

Gender Differences in Wayfinding

Gender differences in wayfinding have been a topic of thorough research throughout the years. Voyer et al. (1995) conducted a meta-analysis of 286 studies regarding these differences and reported significant distinctions between males and females in this regard. Males were more adept at tasks that required mental rotation skills (78 studies reported a male advantage), spatial perception (92 studies reported a male advantage), and spatial visualization (116 studies reported a male advantage).

Hund and Gill (2014) studied wayfinding tasks involving route and survey cues. They noticed that wayfinding task completion time between these two varied significantly with females (who were more effective with route cues), but not with males. This suggests that females prefer route cues over survey cues. Similar results were also reported in other studies (e.g., Galea and Kimura, 1993; Ward et al., 1986). Differences between males and females have also been reported regarding pointing accuracy in both indoors and outdoors environments (Holding and Holding, 1989; Lawton, 1996), suggesting differences in sense of direction. Kim et al. (2007) stated that females may be more efficient at two-dimensional matrix tasks when landmark instructions are provided. Additionally, the research presented in this dissertation suggests that gender differences diminish while performing collaborative wayfinding tasks (Publication IV).

3.4 COLLABORATIVE WAYFINDING

Collaborative wayfinding has been studied relatively infrequently. Dickinson and McIntyre (1997) have created a general model for teamwork that can also be applied to collaborative wayfinding tasks. This model consists of the seven core components of teamwork found in Figure 12.

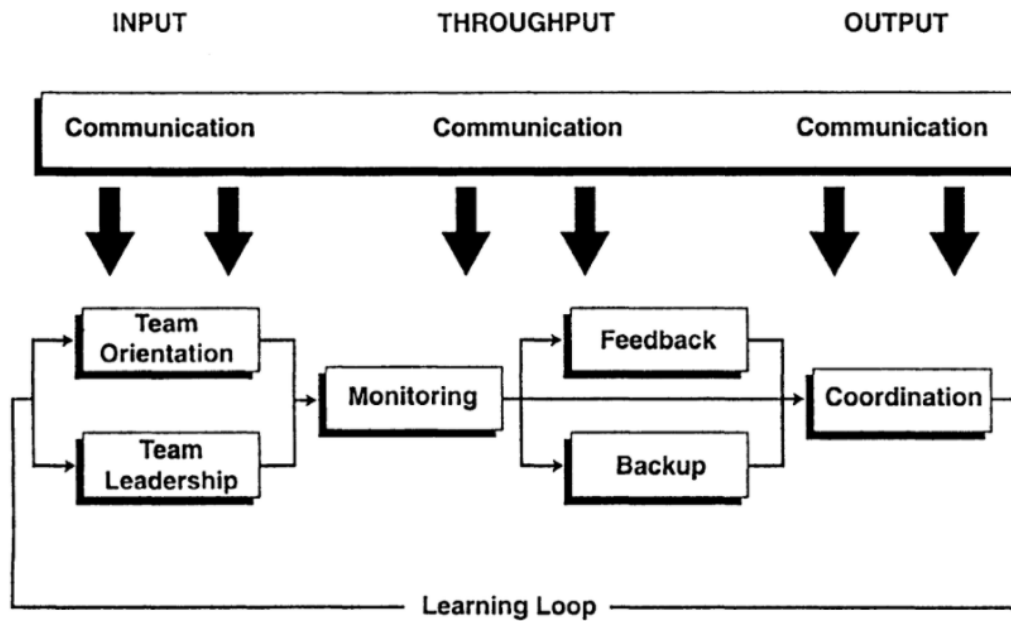


Figure 12. Model for collaboration (Dickinson and McIntyre, 1997).

Dickinson and McIntyre (1997) defined *communication* as “the exchange of information between two or more team members.” The function of communication in collaboration is to transfer, clarify, or acknowledge information. It is considered the link between other aspects of teamwork, and it is commonly believed to be a critical aspect of a functioning team (Cooke, Salas and Cannon-Bowers, 2000). Kraiger and Wenzel (1997) stated a team performs more proficiently overall when team members understand their individual responsibilities in communication and the team’s communication is well coordinated with concise statements, questioning, and confirmation. *Team orientation* includes team members’ attitudes towards one another and the task (Boswell, 2001). Stout et al. (1999) referred to team orientation in their research on shared mental models (SMMs) and their importance to successful collaboration. SMMs provide the team with a “common understanding of who is responsible for what task and what the information requirements are for each team member” (Boswell, 2001).

Team leadership refers to the “direction and coordination of activities of the team members” (Boswell, 2001). Leadership is not restricted to just one team member, but can be spread throughout the whole team. Prince et al. (1997) commented that consistent, identifiable behavior is key to successful leadership, and it improves the team’s performance as a whole and on an individual level. *Monitoring* happens when a team member is aware of other team members’ activities. It is a crucial element in the adjustment and adaptation of team strategies. *Feedback* is the “critical discussion of performance among team members” (Boswell, 2001). Feedback is an honest evaluation and critique of both individual performances and the team’s performance as a whole, as conducted by team members. Feedback is a critical component in teamwork; it generally improves coordination and

generates trust among team members (Tannenbaum, Smith-Jentsch and Behson, 1998). Particularly leaders who recognize their own faults and flaws often inspire similar behavior among their team members. Constructive peer criticism and its acceptance may also increase team performance.

Backup happens when team members help other team members with their current assignments. This activity requires cross-training among the team's members (Cannon-Bowers, Salas and Converse, 1993). When team members understand their own responsibilities and the responsibilities of other members, they are more likely to contribute to the team under stressful conditions (Boswell, 2001). *Coordination* ties together the rest of the components in the model. Boswell (2001) defined *coordination* as the synchronization of the team's efforts and abilities to achieve a common goal. *Coordination* can also be defined as the team's achievement of a higher "degree of shared mental model" (Boswell, 2001). Higher-level SMM makes critical components such as feedback and backup automatic (Stout et al., 1999).

In conclusion, teamwork and collaboration require efficient communication, but also a combination of planning, leadership, and team and individual goals. Successful backup and feedback components lend the team greater coordination and efficiency (Boswell, 2001). Coordination happens when the team communicates successfully and when the other components of the collaboration model are successfully adopted into the teamwork. This general model by Dickinson and McIntyre is also suitable for collaborative wayfinding tasks, and its components are applicable to the wayfinding experiments introduced in this dissertation (publications II, III, IV, and VII). All the components discussed here are present in these tasks, communication being the most important aspect. In addition, leadership is evident in these use scenarios even when there are only two collaborators. Boswell (2001) suggested that, by combining generalized models of wayfinding with the model of teamwork provided by Dickinson and McIntyre (1997), one could have a model that supports the behavioral requirements of collaborative wayfinding in disparate contexts, including VEs. Boswell introduced a model for collaborative wayfinding (Figure 13) that included a *story generation* pattern, which functions as a "connection between wayfinding and collaboration" (Boswell, 2001). This model was modified from Chen and Stanney's (1999) model for collaborative wayfinding, in which the communication part needed defined in more detail. In Boswell's story generation phase, the team develops a list of goals, expectations, and actions that then guide the team through the wayfinding task. In this phase, the team reviews the available materials, decides on the following actions, and identifies the objectives to reach a goal. Stout et al. (1999) called this a *team experience*, and it is tied to the shared experience between team members.

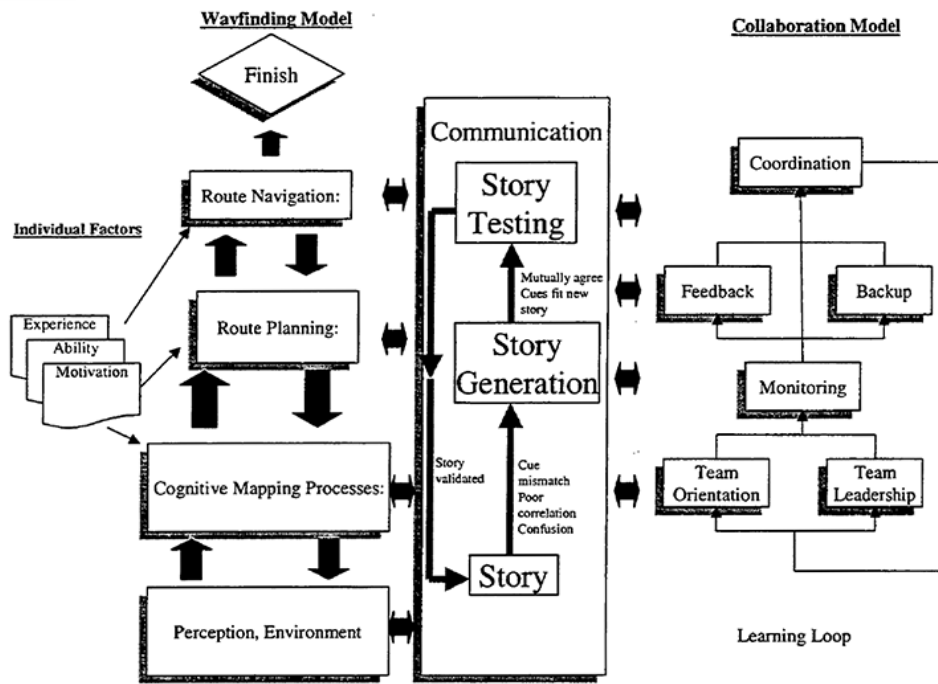


Figure 13. Model for collaborative wayfinding (Boswell, 2001).

3.5 LANDMARKS IN WAYFINDING

The Merriam-Webster dictionary defines the term *landmark* as “an object or structure that marks a locality and is used as a point of reference.” This definition has also been used in scientific literature (e.g., Cornell, Heth and Broda, 1989). The prominence of a landmark is not only dependent on its individual properties, but also on its contrast to the surrounding environment (e.g., a modern building on a block with only old buildings). Landmarks are often used as “mental representations of space” (Siegel and White, 1975; Hirtle and Heidorn, 1993), and they are often employed to communicate route directions (Denis et al., 1999; Lovelace, Hegarty and Montello, 1999).

People often rely on route directions from others to facilitate wayfinding. As mentioned earlier, these directions may contain cues, such as left-right turns, landmarks, and surveys, including cardinal directions and distances (Lawton, 1994; Taylor and Tversky, 1996). Padgitt and Hund (2010; 2012) have stated that route cues are the most effective method in terms of preference ratings and success in finding a destination. However, the effectiveness of these cues depends heavily on the situation in which they are being used (Chai and Jacobs, 2009). Some studies suggest that even though ratings indicate preferences for route cues, survey cues facilitate efficient wayfinding in indoor environments and model towns (Hund and Minarik, 2006; Hund and Nazarczuk, 2009). The purpose of route directions is to provide a “set of procedures and descriptions that allow someone using them to build an advance model of the environment to be traversed”

(Michon and Denis, 2001, p. 293). Landmarks can provide much support for building this model. The inclusion of landmarks in route directions also raises user confidence consistently and reduces wayfinding errors significantly (Ross, May and Thompson, 2004).

Landmarks are often located at decision points (locations where re-orientation is required) or potential decision points (locations where re-orientation is possible) (Lovelace et al., 1999). They can also be used to confirm that the wayfinder is on the correct path. In addition, they can be located at a distance. The first three types of landmarks are often called *local landmarks*, and the last type is referred to as *global landmarks*. Hansen et al. (2006) stated that, on a conceptual level, landmarks can be used either in a point-like (e.g., buildings), line-like (e.g., bridges), or area-like (e.g., squares and plazas) manner. This categorization depends on the landmark's spatial relationship with the route—a factor that diverges from traditional top-down maps in which all landmarks are considered areas. For example, in *turn right at the church*, the church acts as a point-like reference, whereas in *walk alongside the church*, it can be considered a line-like conceptualization. For area-like conceptualizations, route directions such as *walk around the church* are often used.

Several studies have shown that landmarks are often used in route directions at decision points (e.g., Habel, 1988; Michon and Denis, 2001). However, Lovelace, Hegarty, and Montello (1999) said, “More than 50% of the landmarks on unfamiliar routes and more than 40% of landmarks on familiar routes are mentioned at places other than decision points.” When comparing wayfinding in underground and open urban environments, people use landmarks as a reference point more often in the latter. Signs often dominate underground locations. For instance, in subway stations wayfinding and orientation is often solely based on signs that guide the user to a destination (Fontaine and Denis, 1999).

It has been suggested on many occasions (Deakin, 1996; Denis et al., 1999; Michon and Denis, 2001; Tom and Denis, 2003) that using landmarks and survey knowledge in route directions increases their effectiveness. Survey knowledge produces a more comprehensive understanding of a large-scale environment, as it offers a more absolute reference frame. A study by Burnett and Lee (2005) actually stated that modern wayfinding applications contribute “much less to the development of cognitive spatial models” than traditional maps. The lack of these models makes situations where users are lost more challenging, as they might not have a clear image of the environment they have navigated. This also makes it more difficult for them to consider and evaluate alternative routes, for example, in the case of road construction, roadblocks, etc. (Hipp et al., 2010). Modern wayfinding applications such as Google Maps already consider roadblocks and construction.

Identifying Landmarks

Lynch (1960) defined landmarks as “external points of reference.” According to this definition, landmarks are not part of a route itself. Lynch stated that landmark’s saliency is tied to its attributes, including a) a clear form, b) a contrast to its background, and c) a prominent location. The main contributing factor to a landmark’s saliency is its contrast to the environment (Figure 14). This contrast can be due to any of its attributes (or combination of attributes) that makes it unique in form or function when compared to its surroundings. There exist several categorizations for landmark attributes. For example, Sorrows and Hirtle (1999) categorized them by visual (visual contrast), structural (prominence of location), and cognitive properties (use or meaning). The effects of these properties can be cumulative. Hence, a visually interesting, culturally important landmark that is prominently located tends to attract a traveler’s attention more easily.

Raubal and Winter (2002) replaced cognitive properties with semantic properties. This classification was also used as the basis for our work in Publication I and Publication II. Winter (2003) expanded this model by utilizing advanced visibility for salient landmarks at decision points. In this model, the visibility of the landmark from the wayfinder’s point of view was also a factor when deciding its usability in route directions. The concept of advanced visibility was also evaluated in the VEs in Publication I of this dissertation.



Figure 14. National Fisheries Development Board in Hyderabad, India. The building has a clear form and stands in high contrast to its background. In addition, it is in a prominent location, making it a potential landmark for route guidance. Photo by Ra Chandroo.

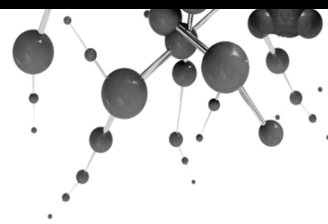
Data mining techniques for automatically retrieving landmark information have also been attempted in many studies. Elias (2003) used map and laser scanning data to retrieve height and layout data for landmarks. They gathered the data of visually prominent objects and the area sizes in which these objects could be seen by using common data mining techniques (ID3 and clustering with Cobweb). Tezuka and Tanaka (2005) modified spatial information with traditional text mining methods to obtain landmark information from the web. This approach has been used extensively in linguistic studies, for example, by Nicholson and Baldwin (2006), who employed Google to investigate the use of compound nominals on the web. These techniques have not been utilized much for landmark-based wayfinding studies recently, even though the concept of big data has been the focus of academics and the mainstream for several years now. Li et al. (2016) revisited current methods for retrieving geographical data to test if they are still capable of handling huge amounts of data. They also synthesized problems, major issues, and challenges in current developments of big data analysis regarding geographical data. Sester and Dalyot (2015) also introduced a concept for enriching route directions with landmark data and attributes, but no experiments for evaluating this model were conducted. Like Publication I, they suggested the use of crowdsourced geographic datasets, such as Wikipedia and Foursquare.

3.6 SUMMARY

In this chapter, I introduced the basic cognitive process of human wayfinding, with detailed introduction on how we navigate through space, form wayfinding related knowledge, and adopt different wayfinding strategies to reach our destination. I also introduced the concepts of cognitive mapping and spatial abilities, and their effects on human wayfinding. By using these abilities, humans can form route directions and route knowledge. For this, one of two different wayfinding strategies, route

strategy or survey strategy, is adopted. One of the most studied subject in wayfinding is gender differences. I briefly introduced the research and results in this field.

As the context of this dissertation is collaborative wayfinding, I explained the basic concepts of collaboration and collaborative wayfinding. Humans often rely on landmark information while performing wayfinding tasks, thus the use of these landmarks was presented. Landmark saliency and identification is relevant in the context of publications I and II, thus the basic procedure of identifying salient landmarks in the scenery were introduced and explained.



4 Virtual Environments

Virtual environments (VEs) are generally described as “three-dimensional, computer-generated environments which the user can explore and interact with” (Virtual Reality Society, 2017). The user is immersed in the environment and can manipulate objects or perform a range of actions in it. Wann and Mon-Williams (1996) defined *virtual environments* as a representation that “capitalizes upon natural aspects of human perception by extending visual information in three spatial dimensions.” Mikropoulos and Bellou (2006) made a clear distinction between Virtual Reality (VR) and VEs. By their definition, *virtual reality* refers to the technology or the building blocks for VEs, whereas *virtual environments* are considered three-dimensional spatial representations built with said technology. They also defined immersion and multimodal and intuitive interaction as other important characteristics of VR. These technologies are the basis for creating three-dimensional VEs that may represent both real (e.g., military training) or fictional scenarios (e.g., games taking place in a fantasy world).

One of the first attempts to use technology for creating the illusion humans are present somewhere they actually are not were Charles Wheatstone’s use of stereoscopic images in 1838 (Figure 15), as seen through a stereoscope. Wheatstone’s experiments showed that human brains perceive two different two-dimensional images from each eye and process them into one single three-dimensional object. Watching these images through the machine provided the user a sense of depth and, thus, the feeling of immersion—a technique called *stereoscopy*. Wheatstone’s apparatus was further developed by David Webster into the lenticular stereoscope in 1849 and by William Gruber into the View-Master in 1939. These devices can be seen in Figure 16. The same design principles used in these devices are still

utilized today with Google's Cardboard and other low-budget HMDs used with smartphones.



Figure 15. An early (c. 1860) stereoscopic image card of a park in Boston.

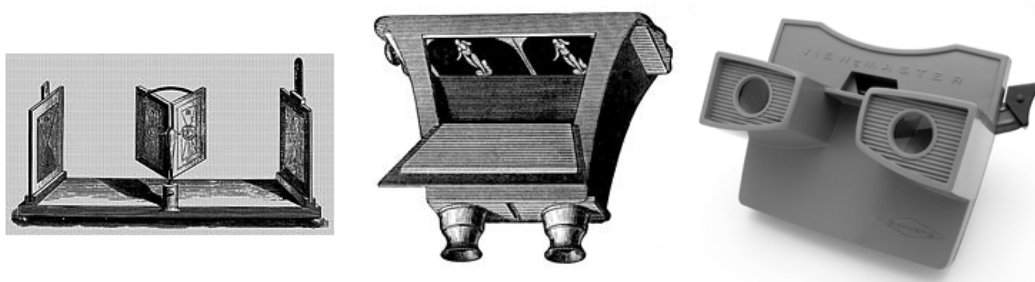


Figure 16. From left: Charles Wheatstone's Stereoscope (1838), David Brewster's Lenticular Stereoscope (1849), and William Gruber's View-Master (1939).

Cinematographer Morton Heilig developed several devices for experiencing immersive VEs. His first prototype, the Sensorama, was first described in a paper entitled "The Cinema of the Future," published in 1955. This vision was finally built in 1962. It featured stereo speakers, a stereoscopic display, a vibrating chair, and smell generators, thus allowing the user to be immersed in a truly multisensory experience. Heilig created a total of six short films for his invention. The Sensorama can be seen in Figure 17. Heilig also developed the first prototype for an HMD, the Telesphere Mask. It played back non-interactive recordings without any kind of motion tracking, but provided stereoscopic 3D with stereo sound.

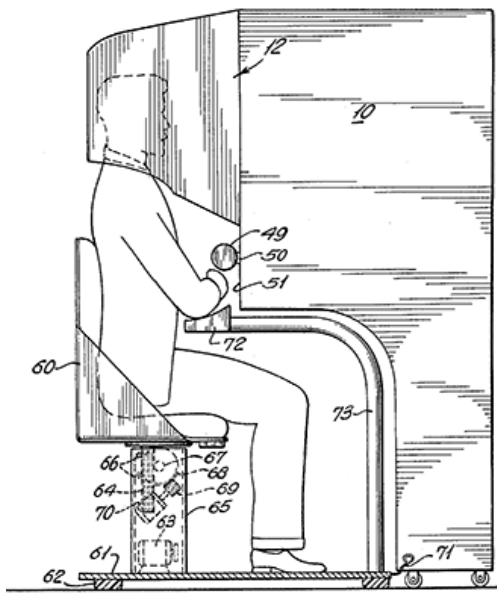


Figure 17. Left: The Sensorama, as presented in Heilig’s patent. Right: The physical setup of the Sensorama.

The first HMD with motion tracking, Headsight, was developed by Comeau and Bryan from the Philco Corporation in 1961. Headsight incorporated separate video screens for each eye and a magnetic motion tracking device attached to a camera. It was developed for the military for remote viewing of dangerous scenarios. The user’s head movements would refresh the viewports, allowing the user to naturally explore the scenery. It was the first step toward the development of modern HMDs. Four years later, a computer scientist, Ivan Sutherland, introduced his paper, “The Ultimate Display” (1965), in which he introduced the idea of an apparatus that would offer the experience of simulated reality so the user could not tell the difference between this experience and actual reality. This experience would be a computer-generated virtual world that would be seen through a head-mounted display with 3D sound and tactile feedback. In this environment, the user could interact with objects located inside the virtual world realistically. This publication later became a blueprint for many future concepts regarding VEs and VR technology. Subsequently, Sutherland developed the Sword of Damocles, the first VR HMD that was connected to a computer (Sutherland, 1968). The computer graphics provided by the system consisted of wireframe rooms and objects.

The term *virtual reality* was coined as late as 1987 by Jaron Lanier. Lanier’s company, VPL, released a range of VR products such as Dataglove and the EyePhone HMD. Dataglove was one of the first integrations of haptics into VEs. During the early ‘90s, several video game companies, including Sega and Nintendo, released their own VR headsets for gaming purposes, but all of these were commercial failures. After this juncture, VR and VEs mostly disappeared from the commercial market, but remained a prominent subject of research, for example, in terms of immersion (see, for instance,

Baños et al., 2000; Barfield, Baird and Bjorneseth, 1998) and the transfer of spatial ability (see, for example, Astur et al., 1998; Lawton and Morris, 1999). The validity of this research was confirmed by further research on the transfer of knowledge between the VE and real world, as Witmer et al. (1996) suggested that once sufficient fidelity and immersion are accomplished, the knowledge transfer between these two media is good. For this reason, VEs have been used extensively for therapy and real-world training purposes.

The development of mobile technologies during the first 15 years of the 21st century has brought VR and VEs again to the mainstream. The availability of powerful smartphones has enabled a new generation of practical devices for VR implementations. Many large companies, including Facebook, Samsung, and Google, have their own development projects for HMDs. Many of these devices use the same basic principles as Wheatstone's stereoscope, which was developed almost 200 years ago.

Virtual Learning Environments

Virtual learning environments, or VLEs (also called *educational virtual environments*, or EVEs) are “virtual environments that are based on a certain pedagogical model” (Mikropoulos and Natsis, 2011). They also incorporate didactic objectives and often offer experiences that would be impossible in the real world. They should also have carefully planned and defined learning outcomes. A meta-analysis of VLEs showed that most of these applications refer to science, technology, mathematics, and language learning. Bricken (1990) defined *cognitive presence* as the main features of supporting learning in VEs. VLEs have been studied in assorted educational settings, including elementary schools (e.g., Adamo-Villani and Wilbur, 2008), high schools (e.g., Schrader and Bastiaens, 2012), and higher education institutions (e.g., van der Land et al., 2013). Research on this topic has focused on different aspects of these applications, including comparing the use of VLEs with traditional teaching and learning methods (Codd and Choudhury, 2011) and comparing various media representations (van der Land et al., 2013). A recent meta-analysis by Merchant et al. (2014) suggested that mixed approaches with combinations of simulations, games, and VEs resulted in the most effective learning results. This analysis also showed that most of the research in this field is comparative studies between the use of VEs and traditional teaching instead of studying the characteristics of these environments.

Helsel (1992) described VR as “a process that enables users to become participants in abstract spaces where the physical machine and physical viewer do not exist,” reinforcing the importance of the immersive sensation in VEs. Pantelidis (1993) reported more active participation and higher interactivity as the main features of VR applications benefiting learning outcomes. Winn suggested the sense of immersion users experience with VEs is of main importance for their learning process (2000). The concept of

learning is a very complex process. Any activities performed in VLEs should not be segregated into isolated entities; nevertheless, these activities play a role in the learning outcome (Salzman et al., 1999). For this reason, it is important to study and define the basic features and outcomes of VLEs. Trindade, Fiolhais, and Almeida (2002) studied science learning in VEs, and their results concluded that the main strengths of using VLEs is the ability to visualize situations that could not be present in the real world. They also stated that the feeling of immersion is a crucial element within these environments. The results from this study also suggest that learning scientific information with VLEs increases the student's motivation to learn.

VLEs have also been employed in contexts other than education. They have been studied extensively, for example, in different military settings (Boswell, 2001; DeBrine and Morrow, 2000). These VE applications offer a range of scenarios for theater planning, training, and mission rehearsal. The problem with using VEs for teaching is that they often lack a well-defined goal (Berns et al., 2011). These researchers' solution for this problem was to design a collaborative task with common goals and limit the user's options and mobility in the VE.

In those VLEs that contain navigable elements, it is crucial to design for effective wayfinding. Minocha and Hardy (2011) stated that the following design features in VLEs can have negative impacts on wayfinding in three-dimensional learning spaces:

- VLE does not resemble real-world physical spaces.
- Functional areas of the environment can be difficult to find or reach.
- The VLE does not provide sufficient navigational assistance.
- Any navigational aids are difficult to understand and/or use.
- Sufficient help for use is not provided.

Difficult and poorly defined wayfinding in VLEs also affects the student's learning experience. Minocha and Hardy (2011) reported the following effects as the result of poor wayfinding design:

- Students may abandon an activity altogether.
- Learning in a VLE may take longer than necessary.
- Students may become frustrated.
- Students may wander aimlessly around the VLE without a coherent goal.
- Students may start guessing or making incorrect assumptions.

4.1 WAYFINDING IN VIRTUAL ENVIRONMENTS

Before computer-generated environments, spatial ability in humans was investigated with two methods. In the first method, researchers from the

field of psychometrics evaluated these abilities with pen-and-paper experiments. As part of these experiments, participants imagined the rotations and movements of small objects, such as blocks, cards, and flags (for example, Carroll, 1993; Pellegrino and Kail, 1982). The second method for evaluating spatial abilities was to investigate people's wayfinding process in large-scale environments, including building interiors and city centers (for example, McNamara, Rump and Werner, 2003). Both methods involve spatial cognition, but there is not much evidence that connects spatial cognition and results from psychometric experiments. A meta-analysis of spatial ability by Hegarty and Waller (2005) revealed that performance on pen-and-paper experiments regarding spatial abilities usually "accounts for about 5% of the variance in their ability to learn and find their way in a large-scale environment." Self-reported measures of spatial ability (for example, the Santa Barbara Sense-of-Direction scale) can have correlations of around .4 or .5 with spatial abilities (Hegarty et al., 2002). The problem with self-reported measures is that even though they are better at predicting wayfinding human behavior than psychometric tests, they still offer "very little insight about the psychological processes or mechanisms that underlie people's spatial ability" (Waller, 2005).

There is still a void in the understanding of the process regarding how we acquire spatial information about our environment. VEs provide a more flexible and diverse alternative to traditional pen-and-paper experiments. With VEs, we can simulate actual movement within large-scale environments. This activity is often connected to our ability to understand space, as it allows the user to reason about the results of one's own movement in these environments (Waller, 2005).

The biggest limitation of VEs that run on desktop computers is the lack of embodied interaction, thus removing body-based modalities from the wayfinding process. The importance of these modalities, including vestibular and kinesthetic senses, in wayfinding is still a matter of debate (see, for example, Riecke, van Veen and Bühlhoff, 2002; Waller, Loomis and Steck 2003); however, research with humans has repeatedly shown the important role that active movement plays in wayfinding. Humans' (and rodents' and monkeys') sense of spatial orientation depends on proprioceptive feedback and vestibular signals. Research has also shown that active exploration usually results in greater spatial knowledge of large-scale environments. For example, when an array of objects rotates relative to a stationary observer, the observer's scene recognition was impaired. This same effect was not detected when the observer moved relative to a stationary display (Simons and Wang, 1998). Further studies suggested that information obtained through self-motion made the scene recognition easier from novel, diverse viewpoints compared to when the observer passively viewed scenery (Wang and Simon, 1999). A study by Witmer and Kline (1998) stated that participants often experienced difficulties with

estimating distances, often underestimating them, in VEs compared to walking in the real world.

Many VEs use embodied interaction and gestures as an interaction method. We also wanted to address this issue by integrating this interaction method into some of our applications (publications II, III, and IV). This type of interaction is not equivalent to real-world wayfinding, but it still incorporates the body-based sensory modalities used in these scenarios.

Ultimately, choosing varying environments for evaluating spatial ability comes down to the setup and premise of the experiment: computer-generated VEs provide researchers an “ecologically relevant environment” in which to examine human behavior with “high control over the environment’s properties” (Waller, 2005). For this reason, VEs are an exceptionally useful tool, even with their limitations, for assessing individual distinctions in spatial cognition.

Richardson et al. (1999) reported that users commonly experience greater difficulty in forming spatial knowledge about VEs than the real world. This may lead to poorer performance in wayfinding tasks in VEs. Some factors may improve this performance, for example, increasing the FOV (McCreary and Williges, 1998), embodied interaction (Zanbaka, 2004; publications II and IV), and heightening the visual information available to the wayfinder (Gillner and Mallot, 1998).

4.2 LANDMARKS IN VIRTUAL ENVIRONMENTS

Landmarks provide information that helps individuals to identify their location and orientation, and often serve as the main component in route planning through virtual and real-world large-scale environments. Vinson (1999) suggested various guidelines for landmark design in VEs. Some of these guidelines are introduced and discussed briefly in the following chapters:

VE should contain several landmarks.

Once the traveler gains experience with a particular route, he or she increases the representational precision of distances and positions of landmarks (Evans, 1981), which in turn might change the spatial representation from route knowledge to survey knowledge. This allows the traveler to adopt the most suitable perspective of the environment for a wayfinding task (Thorndyke and Hayes-Roth, 1982). Vinson (1999) suggested that the types of landmarks to be included in VEs should follow Lynch’s (1960) categorization. These include paths, edges, districts, nodes, and landmarks. Each of these has a specific function, but each individual object can also have more functions than just one (Table 4). For example,

landmarks can also be used as focal points and reference points while traveling.

Type	Examples	Function
Paths	Street, canal, transit	Channel for traveler movement line
Edges	Fence, river	Indicate district limits
Districts	Neighborhood	Reference point
Nodes	Town square, public building	Focal point for travel
Landmarks	Statue	Reference point into which one does not enter

Table 4. Landmark types and functions (Vinson, 1999)

All five types (paths, edges, districts, nodes, and landmarks) should be included in a VE.

Because most individuals are used to navigating large-scale environments in the real world by wielding these types as reference points, it is important to include all of them in VEs. The designer of a VE usually has the benefit of choosing the type of landmark and its location in the environment. As stated before, the transfer of knowledge between VEs and the real world is relatively good if the level of fidelity and immersion is sufficient (Witmer et al., 1996). The relationship between landmark features and recalling them has been examined by Evans et al. (1984). Consequently, they presented a set of features that make landmarks and buildings more memorable or easier to locate. Some features can also contribute to both memorability (marked with ^m in the list below) and location recall (marked with ^l in the list below). These features (from Evans et al., 1984) are as follows:

- Significant height ^m
- Complex shape ^m
- Bright exterior ^l
- Large, visible signs ^m
- Expensive building materials and good maintenance ^l
- Freestanding (visible) ^{lm}
- Surrounded by landscaping ^m
- Unique exterior color, texture ^l

^m Improves memorability

^l Improves location recall

In addition, Vinson (1999) suggested the use of landmarks from natural environments, including fabricated items such as roads, sheds, and fences; land contours such as hills, slopes, and cliff faces; and water features such as lakes, streams, and rivers. From these elements, we can provide the third guideline:

Make the landmarks distinctive by using the features presented by Evans et al. (1984) for urban environments and Vinson (1999) for natural environments.

As a fourth guideline, Vinson (1999) suggests that designers:

Use concrete, non-abstract objects as landmarks.

Ruddle, Payne, and Jones (1997) concluded that memorable landmarks increase effective wayfinding. In their study, they used familiar three-dimensional objects such as cars. In the same study, they suggested that abstract objects such as complex paintings did not help the traveler in his or her task. In natural environments, any manufactured constructs stand out from the rest of the environment (Whitaker, 1996). Expert orienteers most often rely on land contours and water features in addition to synthetic constructs when traveling in natural environments (Whitaker and Cuglock-Knopp, 1992).

A landmark must be distinguishable from its environment.

Landmarks presented in VEs should be distinctive compared to other nearby landmarks, as those objects that contrast with their surroundings stand out from their environment (Evans et al., 1984). Confusing one landmark with another is a very common mistake in wayfinding tasks, and in natural environments, this error has been named the *parallel error* (Darken and Banker, 1998). In addition, landmarks should have distinct sides that have enough differences so travelers can tell from which direction they are looking at it.

The saliency of a landmark can be increased by placing other objects nearby.

In some situations, landmarks can complement each other. For example, placing a colorful landmark among many monochromatic ones makes it prominent in a landscape, rendering it a good landmark for wayfinding purposes. For example, consider a landmark that is symmetrical on every side. It is very challenging for a traveler to discern the direction he or she is viewing it from. Inserting another, distinct object next to it makes the orientation easier.

Place landmarks along travel nodes and at decision points.

The memorability of a landmark is also affected by its location in the environment (Evans et al., 1984), especially if it is located on a major travel node or at an intersection. The most convenient place for a landmark is at the decision points, that is, those locations where travelers need to reorient themselves. Placing landmarks relatively close to each other also supports those travelers who rely on piloting (traveling from landmark to landmark). For example, the following positions for a building contribute to their memorability (marked with ^m in the list below) and location recall (marked with ^l in the list below):

- Located on a major path ^m
- Visible from a major road ^{lm}
- Direct access from a street ^{lm}
- Located at an important choice point ^m

^m Improves memorability

^l Improves location recall

These general guidelines can be useful in designing landmarks and navigable content for VEs, but do not really consider realistic representations of urban or natural environments. For example, when using photographic images or videos for VE content, one can only choose the locations (e.g., a plaza or a street corner) that are presented. Each application presented in this dissertation provides this kind of VE content. To design these environments, one should consider locations with prominent landmarks that are in high contrast with their surroundings. Another model for landmark presentation was presented in Publication I. This model highlights landmarks based on their saliency, which is calculated with three distinct properties: semantic, cultural, and structural. This model was then evaluated in a collaborative VE (see Publication II). Because many wayfinding experiments were conducted a long time ago, the technology for implementing more sophisticated three-dimensional VEs did not exist. Wayfinding evaluations are often performed in three-dimensional mazes where the participant navigates from the start point to the goal. In these mazes, landmarks are often presented with either three-dimensional models (see Figure 18a) or with two-dimensional icons (see Figure 18b). These models are not very good representations of the real world because of their simplifications. Interactive omnidirectional videos of real-world environments would be a better representation for wayfinding studies because of their realism. These representations were introduced in publications VI and VII.

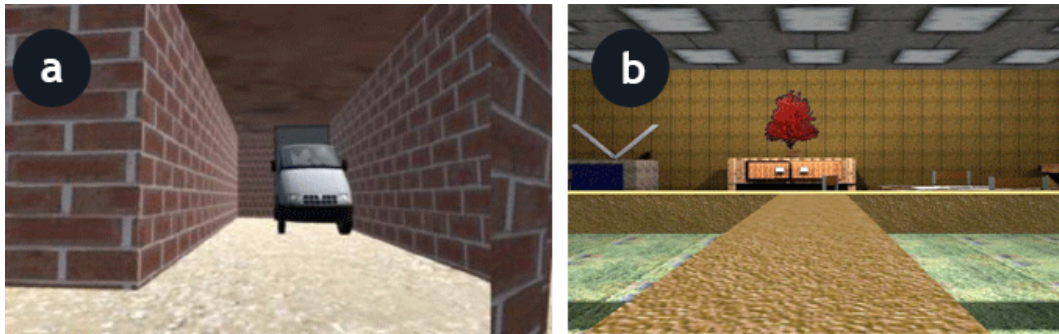


Figure 18. a) A three-dimensional maze where landmarks are presented as icons (Sharma et al., 2017). b) A three-dimensional maze where landmarks are presented as three-dimensional objects (Astur et al., 2016).

Gender Differences in Wayfinding in Virtual Environments

Most research on individual differences in wayfinding in VEs has focused on gender (Walkowiak, Foulsham and Eardley, 2015). Results from these studies generally suggest that males outperform females in spatial tasks, and that these differences are even larger in VEs than in real-world scenarios (Astur, Ortiz and Sutherland, 1998). Castelli, Corazzini, and Geminiani (2008) found that males perform more efficiently (i.e., complete a task faster) and make fewer errors than females while completing a wayfinding task where they must utilize a survey strategy. Moffat et al. (1998) and Waller (2000) reported similar results in their studies. Lin et al. (2012) have provided possible explanations for these distinctions. They suggested that males are more explorative in their wayfinding. Males also traveled large distances even when they were still not familiar with the environment. This was not evident in females, who adopted more conservative strategies during the wayfinding task. This exploratory nature of wayfinding among males was also reported by Coluccia et al. (2007).

Wayfinding experiments are commonly conducted with either the virtual Morris Water Task (vMWT) (Morris, 1984) or the multiple T-maze (Tolman, 1948). In the vMWT, participants try to find their way through a virtual water maze by using various navigational cues. The avatar in this maze is usually controlled by a mouse or a keyboard. Time and distance to locate the goal across trials are then used to describe the user's wayfinding performance. Virtual corridor mazes contain a start location, interconnecting corridors, and a goal. Task completion time, errors made, and number of trials to the criterion are usually used for measuring performance (e.g., Moffat, Hampson and Hatzipantelis, 1998). The same variables apply to virtual wayfinding measures as to real-world wayfinding tasks, and more traditional tasks such as a mental rotation task and paper map tasks are used to complement the results from these wayfinding tasks performed in VEs (e.g., Astur et al., 2016).

Walkowiak, Foulsham, and Eardley (2015) suggested that one variable affecting wayfinding in VEs is computer experience. In their study, females

who reported more computer and video game experience completed the wayfinding maze task faster and made fewer errors during its completion. Similar results were reported by Lin et al. (2012) and Head and Isom (2010). Because less computer experience is detrimental to the user's feeling of computer self-efficacy (Compeau and Higgins, 1995), some of these effects may be due to a lack of familiarity with this type of task.

In general, wayfinding experiments in VEs follow the same procedures as in real life. For this reason, the same differences between genders have been reported. For example, males recall non-vivid descriptions, which are generally harder to memorize, more effectively than females (Tom and Tversky, 2012). This difference diminished once the descriptions were made more vivid, thus benefiting both genders. Males have also consistently outperformed females in locating information, but there are also studies that did not find gender variations in information location (see, e.g., Tom and Tversky, 2012; Halpern, 2000; Wolbers and Hegarty, 2010). The problem with these evaluations is that they are simplified versions of real-life scenarios, and they often do not utilize the individual's motor, vestibular, and proprioceptive systems. Taube, Valerio, and Yoder (2013) suggested that ruling out these factors needs to be considered when reporting any wayfinding measurements. More sophisticated VEs could provide these functions in addition to more traditional spatial tasks, thus complementing the results gained from these experiments.

4.3 IMMERSION/PRESENCE

The terms *immersion* and *presence* have several definitions within the research community. In this chapter, I will describe the most common definitions for both of these and how they were defined in the scope of this dissertation.

Presence

The term *presence* was introduced in a work by Akin et al. (1983), who defined it as an experience that happens when:

“at the worksite, the manipulators have the dexterity to allow the operator to perform normal human functions. At the control station, the operator receives sufficient quantity and quality of sensory feedback to provide a feeling of actual presence at the worksite.”

Presence as a phenomenon has many definitions, depending on the field and the context of research, but it is most commonly referred to as the feeling of “being there.” Witmer and Singer (1998) described presence as “the subjective experience of being in one place or environment, even when one is physically situated in another.” Skarbez, Brooks, and Whitton (2017) stated, “presence has the advantage of being a metric that is applicable to

any VE,” meaning that, for example, presence-related questionnaires between two completely opposite applications can be compared. They continue by defining presence as a *quale* (plural *qualia*), “a subjective and internal feeling of elicited by sense perceptions.” By their definition, this internal and subjective experience is extremely difficult to measure.

Immersion

The term *immersion* is still a subject of some debate, and there are discrepancies between research fields on the meaning of this term. Slater (1999) defined it as an objective characteristic of a VE application. Witmer and Singer (1998) considered it to be a “psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.” Lombard et al. (2000) defined these two conceptions respectively as *perceptual immersion* and *psychological immersion*. Basically, Slater’s immersion is required to experience Witmer and Singer’s immersion, but the use of these terms interchangeably has caused some discrepancy in the terminology. *Immersion* has also been used synonymously with *presence* (for example, publications III, IV, VI, and VII; Psootka, Davison and Lewis, 1993; Wikipedia, 2017). Slater (2009) has also stated, “Immersion provides the boundaries within which [presence] can occur.” The problem with this statement is that immersion in VE applications has other utilities in addition to enabling presence (Bowman and McMahan, 2007).

Social Presence and Copresence

One limitation of the term *presence* is that there are no aspects of social interaction involved. Current interpretations of this term include only single user applications and their interaction with the VE (Skarbez, Brooks and Whitton, 2017). For VEs that involve other characters or avatars, which can be controlled by either humans or computers, the definitions *copresence* and *social presence* have been developed. *Copresence* was first introduced by Goffman (1963), who described it as “exist[ing] when people sensed that they were able to perceive others and that others were able to actively perceive them...render[ing] persons uniquely accessible, available, and subject to one another.” Another definition by Zhao (2003) is as follows: “a condition in which instant two-way human interactions can take place,” and even more briefly, as “being there together” (Schroeder, 2002). Seemingly similar and clearly related, these definitions have important differences: Goffman’s and Zhao’s definitions “refer to properties of a communication medium,” thus being “objective, immersive characteristics of a system,” whereas Schroeder’s definition refers to the *feeling* of being together in a place (Skarbez, Brooks and Whitton, 2017). The first definition for *social presence* comes from Short et al. (1976), who defined as “the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships.” Bull (1983) defined *social presence* as “when one person feels another person is ‘there.’” All the definitions

mentioned overlap and share similar characteristics. This has also caused them to be used interchangeably (e.g., Blascovich, 2002). Skarbez, Brooks, and Whitton (2017) defined *copresence* as “the sense of being together with another or others.” For *social presence*, they used a description as follows: “the moment-by-moment awareness of the copresence of another sentient being accompanied by a sense of engagement with them.” These two definitions differ to “the extent to which one’s experience depends on the other or others.” Thus, *copresence* is already present when another being exists in the same space and the observer is aware of this. For *social presence*, some kind of interaction between the users should be present, and this interaction should affect the other user’s behavior and/or vice versa. To prevent the definitions from being confusing and contradictory, instead of *social presence*, Skarbez, Brooks, and Whitton (2017) coined a new term, *social presence illusion*, referring to the “illusory feeling of being together with and engaging with a real, sentient being.” With the same terms, they suggested the term *copresence illusion* be used instead of *copresence* to refer to the “feeling of ‘being together’ in a virtual space.”

Other Concepts Related to *Presence* and *Immersion*

Another issue regarding presence being used as a general measure is that it does not consider the level of realism that the VE provides. This aspect can be very important when VEs are being used for practicing real-life scenarios, such as training in military settings. For measuring VEs’ realism, Alexander et al. (2005) defined the term *fidelity*, which refers to “the extent to which the virtual environment emulates the real world.” Again, there is an overlap with the terms *fidelity* and *immersion*, but it is feasible, for example, to build a high level of immersion in an environment with low fidelity, and vice versa.

To avoid confusion concerning the presence construct, Slater (2009) divided it into two categories: *place illusion* (PI) and *plausibility illusion* (Psi). They defined *place illusion* as “the...illusion of being in a place in spite of the sure knowledge that you are not there,” and *plausibility illusion* as “the illusion that what is apparently happening is really happening (even though you know for sure that it is not).” With these definitions, PI replaces *presence* as the traditional definition as the feeling of “being there,” and Psi refers more to the experience of believing what you are seeing, albeit knowing that said experience is illusory. This theoretical framework reduces the overlap and interchangeable use of the terms regarding presence.

Coherence, a definition coined by Skarbez (2016), is “the set of reasonable circumstances that can be demonstrated by the scenario without introducing unreasonable circumstances.” For instance, if the user has been led to believe he or she will be part of a VE with fantastic elements, then the presence of fantastic creatures or people performing superhuman acts would be coherent activities. However, if users are told they will participate

in a realistic training simulation scenario, such exotic events would be incoherent, thus decreasing their feeling of presence/PI.

Another term that is commonly used within the context of VEs is *embodiment*. It generally refers to a representation, commonly known as an *avatar*, of the user in a VE. Gabbard (1997) commented, “Representing the user within a VE is known as *user embodiment*,” whereas Benford et al. (1995) remarked, “User embodiment concerns the provision of users with appropriate body images to represent them to others (and to themselves) in collaborative situations.” *Ownership* is “the sense that a body (or body part) is one’s own” (Skarbez, Brooks and Whitton, 2017). Embodiment is a prerequisite for ownership, but they are not the same phenomena. The feeling of embodiment can be achieved with technology and/or the help of tools, but does not automatically result in the user’s feeling of ownership. The illusion of body ownership with virtual bodies has been studied by Slater et al. (2010). Another definition that is related to both the feeling of embodiment and ownership is *self-presence*. Biocca (1997) defined this as “the users’ mental model of themselves inside the virtual world.” It refers to “the effect of embodiment in [a] virtual environment on mental models of the self.”

Involvement and *engagement* are often used interchangeably in VE literature. Witmer and Singer (1998) have described *involvement* as a state of focus and/or interest. *Involvement* and *engagement* are included as subscales in the Witmer-Singer Presence Questionnaire (Witmer and Singer, 1998), the ITC Sense of Presence Inventory (ITC-SOPI) (Lessiter et al., 2001), and the Igroup Presence Questionnaire (IPQ) (Schubert, Friedmann and Regenbrecht, 2001). Skarbez, Brooks, and Whitton (2017) have argued that because presence is logically orthogonal to involvement/engagement, it should not be included in presence questionnaires.

Flow is a concept that has been studied in many contexts and commonly refers to “an optimal state of focus and concentration” or a “state in which individuals are so involved in an activity that nothing else seems to matter” (Csikszentmihalyi, 1990). Takatalo (2002) found that more presence may provide more of a feeling of flow. *Absorption* is a term coined by Baños et al. (1999), who defined it as “the ability to get lost in the task at hand whether it is watching a movie, reading a book, or experiencing VR.” Brockmyer et al. (2009) suggested this was a stronger state of engagement than flow. *Transportation* is more commonly used in narrative worlds and is related to the feeling of presence in VEs (Gerrig, 1993). When in a state of transportation, the individual “loses access to some real-world facts in favor of accepting the narrative world that the author has created” (Green and Brock, 2000). *Transportability* is a term that refers to an individual’s “ability to be transported by a narrative” (Skarbez, Brooks and Whitton, 2017).

As this summary of presence-related terms shows, there are several factors that may affect the user's experience in VEs. In addition, there may be VEs that are not a specific representation of a space and thus would not benefit from the feeling of "being there," but may benefit from other features of the application, for example, immersion and fidelity. In those VEs that attempt to represent the real world as realistically as possible, a high level of fidelity and immersion are more important (Skarbez, Brooks and Whitton, 2017). In VEs where the user is performing real work, the level of involvement and flow may increase effectiveness and/or productivity. When the VE is used for training for real-world events with a sufficient level of knowledge transfer, an appropriate amount of immersion and fidelity should be provided.

4.4 INTERACTIVE OMNIDIRECTIONAL VIDEOS (IODVS)

Omnidirectional videos (ODVs, or 360° videos) have been studied extensively in recent years. There is a vast number of algorithms and devices to capture, construct, and display omnidirectional video content, and large enterprises including Vimeo and YouTube offer their own platforms for viewing these videos. Omnidirectional videos have been used, for instance, in remote operations and telepresence applications (see, for example, publications VI and VII; Onoe et al. 1998; Boulton, 1998). They have also been employed to supply immersive experiences to users in cultural context, say, in museums (Kwiatkiewicz and Woolner, 2010) and theaters (Decock et al., 2014). Other domains where ODVs have been utilized include education, for example, in teaching secondary languages (Publication VII) and sign language (Järvinen and Ekola, 2014).

One interesting field of research in which ODVs have been used recently is health care. VEs have been useful tools for studying and treating patients, for which the term *virtual reality exposure therapy*, or VRET (Riva, Botella, Légeron and Optale, 2004) has been adopted. There is some evidence that people with higher levels of anxiety report higher levels in presence (Alsina-Jurnet, Gutiérrez-Maldonado, & Rangel-Gómez, 2011). This also makes them experience greater anxiety when they are exposed to phobia- or fear-inducing stimuli within the VE. These findings support the notion that VEs, including those that use ODVs, are useful tools for studying and treating phobias and fears. Fassbender and Heiden (2014) implemented an application, Atmosphaeres, for stress and pain management, and Rizzo et al. theorized about practical uses for ODVs in therapy.

One aspect of applications that utilize ODVs is their interactivity. In addition to viewing their video content, one can also add other interactive elements to them. For example, the user can navigate video content or gather contextual information about the environment presented in the content. For these applications, we invented the term *interactive ODVs*, or

iODVs for short (Publication VI). Guidelines for designing these applications have been suggested by Saarinen et al. (2017) and Argyriou et al. (2016). Multimodality and interaction have also been studied in the context of ODVs and, for example, gesture-based interaction (Rovelo-Ruiz, 2014) and second screen interfaces (Zoric et al., 2013) have been used for interacting with ODV content. Benko and Wilson (2010) implemented an application that utilizes ODVs, has a multi-user support, and supports mid-air gestures. In publications VI and VII, we presented an application that utilizes position-based interaction with a dwell timer for HMD applications, and for CAVE systems, we developed an interaction method that employs a rotating chair with a built-in rotation sensor.

Collaboration within VEs that utilize ODV content have also been studied in recent years. Singhal and Neustaedter (2017) developed an application, *BeWithMe*, that allows long-distance couples to collaborate and communicate. In this application, the users can share ODVs about their daily life and experiences. The use of ODVs in a shared guided tour was evaluated by Tang and Fakourfar (2017). Participants in their study had difficulties in building “a shared understanding of what was being looked at and discussed,” which might be due to the low interactivity of the application. Ramalho and Chambel (2013) simulated wind to enhance the subject’s experience with ODVs. This type of multisensory augmentation is another potential research subject related to ODVs.

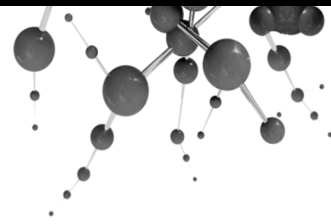
4.5 SUMMARY

This chapter explained the history and basic concepts of VEs and VR systems. The first virtual reality experiments date back to the 1800s, when Wheatstone experimented with the use of stereoscopic images. This again led to the development of the lenticular stereoscope in 1849 and the *View-Master* in 1939. These devices applied the same principles that modern HMD devices use. The first VEs connected to a computer were developed during the 1960s. In addition, a short introduction on collaborative VEs and VLEs was presented.

Subsequently, this researcher summarized the background of and work related to wayfinding in VEs. The basic problems were stated concerning the use of pen-and-paper evaluations and self-reported measures for wayfinding studies. In addition, it was suggested that using traditional interaction methods such as a keyboard and mouse may not provide researchers with comprehensive results on wayfinding, as the motor, vestibular, and proprioceptive systems are not utilized like they usually are when traveling in real large-scale environments.

Guidelines were provided for designing prominent landmarks that support wayfinding for VEs for both urban and natural environments, and it was

suggested that VEs using photorealistic environments may supply a more realistic experience for wayfinding than three-dimensional models. Gender differences were also discussed regarding wayfinding in VEs, including the possible reasons for differences in wayfinding performance between females and males. Then *immersion* and *presence* were described, alongside concepts related to these phenomena. I will discuss this terminology further in the Discussion chapter of this thesis. Finally, research was mentioned concerning ODVs and iODVs, including potential research topics for wayfinding studies with these content types.



5 Introduction to the Publications

The research for this dissertation consists of designing and evaluating a model for highlighting landmarks for pedestrian wayfinding, as well as implementing applications for collaborative wayfinding in VEs and the user studies conducted with them. The research articles presented in this dissertation target the following topics:

- This paper presented a model for highlighting landmarks in pedestrian wayfinding applications (Publication I).
- This work assessed the model developed in the previous publication (Publication II).
- These articles explored the user experience with and immersion in collaborative VEs (publications III, IV, and V).
- This work evaluated the sensation of immersion and user experience with iODVs between CAVE systems and HMDs (Publication VI).
- Interactive ODVs (Publication VII): this model was used in the design and implementation of contextual information for the applications.

The switch of focus from landmark-based wayfinding to collaborative wayfinding in VEs came naturally because the model for landmark highlighting (Publication I) that was developed required an application for evaluating it. After reading the related work for pedestrian wayfinding, it was evident that there was still a large gap in research in collaborative wayfinding and that the model developed earlier could be utilized in this research. When sufficient realism and fidelity are provided, the transfer knowledge between VEs and real-world situations are comparable (Witmer

et al., 1996). Waller et al. (1998) suggested that, because of the variability of VEs, some training scenarios can be even superior to real-world setups.

The first application that utilized the model was Berlin Kompass (Publication II), which created contextual information that supported the collaborative wayfinding task.

In Publication III, this concept of collaborative wayfinding was utilized in the context of language learning. The results suggested that this concept has pedagogical potential, but also stated its clunky, complex installation might be a hindrance for its actual use in pedagogical settings.

After this, the same application was utilized for evaluating the influence of gender and game experience on user behavior (Publication IV).

Subsequently, the clunky setup of Berlin Kompass was developed further into a web version that employed modern web technologies and allowed users to improve their wayfinding and language skills with just a web browser. This application, CityCompass (Publication V) was much easier to set up and use in various environments.

Publication VI compared the feeling of immersion and user experience between two VEs, HMD and CAVE. This aim of the study was to detect any differences between the two media and give guidance on which platform would be more suitable for future applications.

For Publication VII, a CityCompass VR application utilizing iODVs and HMD was developed. This paper studied variations in the sensation of immersion between genders while interacting in collaborative VEs. The following sections will explore and explain each publication in greater detail, starting with Publication 1.

5.1 MODEL FOR LANDMARK HIGHLIGHTING IN MOBILE WEB SERVICES

Reference

Kallioniemi, P., and Turunen, M. (2012). Model for landmark highlighting in mobile web services. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*. ACM, New York, NY, USA, Article 25, 10 pages. doi:10.1145/2406367.2406398

Objective

The objective of the study was to develop a model that highlights prominent landmarks to support pedestrian wayfinding in route guidance applications. The model calculates a saliency score for each landmark in the scenery based on its properties (for example, cultural significance or size). The landmarks that obtain the highest scores in the model are then

highlighted to support the pedestrian's wayfinding process. The total saliency of a landmark is based on a model by Raubal and Winter (2002) that calculates visual, structural, and semantic properties and adds these for a total saliency score. The researchers' new model changed this model so it is more suitable for mobile services, for example, to be used with Google Maps. For the calculations, data from external databases such as Panoramio and the Helsinki Real Estate Department database were used. The paper also introduced a model that calculates the visibility of a landmark from a specific viewpoint. This model was not integrated into the existing model because there was no sufficient three-dimensional material available.

For this paper, an experiment was conducted with 20 participants (13 of which were male and seven female). They selected the most salient landmarks from panoramic images in two locations. Their selections were then compared with the ones made by the model to discern if it could make selections like the participants'.

Results and Discussion

In Intersect 1 (top of Figure 19), the landmarks with the highest ratings from the participants were also ranked as the most salient by the model. The model ranked two museums as the most salient landmarks on the route, after which it ranked a hotel. The model provides relatively high scores for museums, as they are often located in old buildings or have abnormal architecture (e.g., Reina Sofia in Madrid, Spain). These landmarks were also ranked as the most salient by the participants, but in a slightly different order – the hotel was ranked as more salient than the second museum. The model assigned the hotel a lower score because of its distance from the viewpoint. All three landmarks were also located along the route node. The first intersect was interesting because it contained some quite famous landmarks, including the library of the Finnish Parliament, but they were not ranked highly by either the users or the model since they were too far from the actual route node or the façades were facing the wrong way. These results suggest that landmarks located along the travel node are the most suitable for use in route directions.

Intersect 2 (bottom of Figure 19) had only five landmarks from which to choose. Two were along the route but a bit too far away for good recognition. Consequently, they received a much lower score among the participants than the model. In this situation, the calculation for landmark visibility would have been useful, as neither of these two landmarks would have been visible from the viewpoint. In this intersect, both the model and the participants provided the highest score (but in a different order) for a shopping mall and a movie theater.

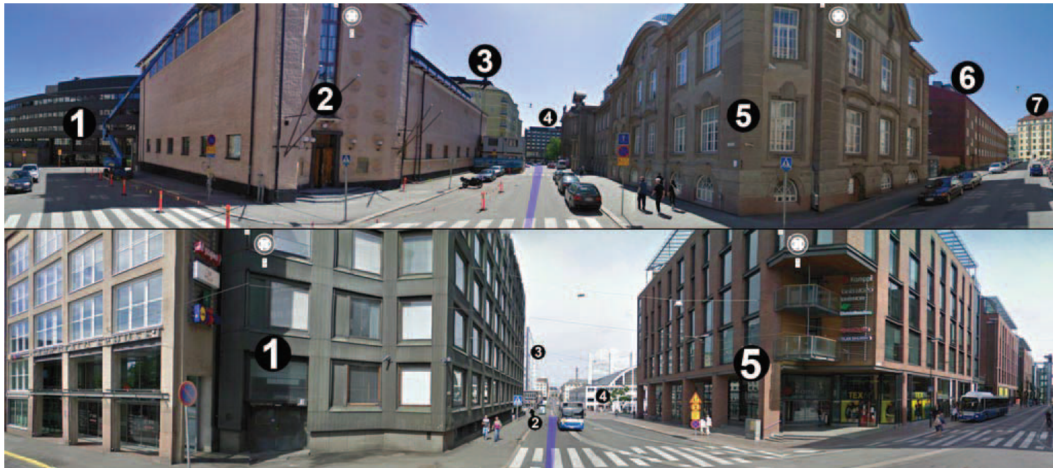


Figure 19. Images of the intersects with landmarks. Intersect 1 above and Intersect 2 below. (from Publication I)

Based on these results, two statements can be made: a) the most salient landmarks are usually placed along the route node, 2) and their façades face the user. Based on questionnaire answers, participants ranked landmark saliency higher in cases in which they recognize the landmark (recognition versus recall). In many cases, selecting a landmark was only based on semantic value and knowability. Accordingly, researchers should study situations in which the user does not know a location (see Publication II).

In conclusion, this evaluation suggests that landmarks ranked highest by the model corresponded well with landmarks selected by participants in both intersects. For future work, evaluations with different day cycles were planned but never conducted. The reason was that future studies concentrated more on collaborative aspects of wayfinding taking place in VEs. The model was again evaluated in Publication II, but due to a lack of external geographical databases to provide more metadata to support wayfinding, it was very challenging to integrate the model into modern wayfinding applications such as Google Maps.

5.2 EVALUATING LANDMARK ATTRACTION MODEL IN COLLABORATIVE WAYFINDING IN VIRTUAL LEARNING ENVIRONMENTS

Reference

Kallioniemi, P., Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., Uusi-Mäkelä, M., Hietala, P., Okkonen, J., and Raisamo, R. (2013). Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 33, 10 pages. doi:10.1145/2541831.2541849

Objective

This study's objective was to evaluate the landmark attraction model from previous publication in the context of collaborative wayfinding in VLEs. It also provides insights into creating suitable content for language learning scenarios in collaborative VLEs. For this evaluation, the collaborative VLE application, Berlin Kompass, was used. This research was a multidisciplinary project conducted in collaboration with education experts. The users were collaborating and communicating in a foreign language to reach a tourist attraction. The application was utilized by two simultaneous users in remote locations. Their goal was to reach a tourist attraction. The model for highlighting landmarks presented in Publication I was applied in each panorama to quantify the relative difficulty of each panorama.

Results and Discussion

Our analysis showed that the panoramas with prominent landmarks would be solved faster by the participants. The participants also required less help from the application in these panoramas. This indicates that this model can be used as a tool for designing content for VLEs that have navigational aspects. The model could be a basis for route planning, for example, making those panoramas that have little or no prominent landmarks part of more challenging routes. In addition, panoramas with many salient landmarks could be used in routes for beginners.

We suggested that the model could be integrated into the content creation process, where it could determine the difficulty level of the route based on the landmark's contents and their saliency. This would also require assistance from the subject matter's experts, for example, language teachers who would then provide contextual information for the route. The participants experienced many difficulties in solving one of the locations. This panorama, DDR4, can be seen in Figure 20. In this location, the users had to resort to referring to moving landmarks, such as cars along the road and pedestrians walking on the streets, to find their way to the next panorama. In this location, some of the more creative solutions were also seen, such as using degrees to communicate directions.

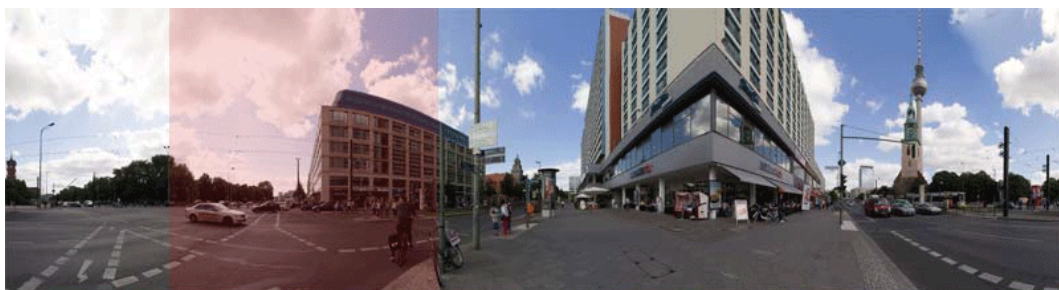


Figure 20. The DDR4 panorama with which the participants had the most difficulties. The correct route is highlighted in the image (Publication IV).

The biggest problem with the model is the lack of datasets from which the metadata could be retrieved. Some geographic data can be retrieved through APIs such as Google Maps, Wikipedia, and Google Album Archive (formerly Panoramio). As can be seen in the case of Panoramio, many of these datasets tend to change substantially over time, often making retrieval solutions obsolete.

The hotspot content in the application was used relatively little. The researchers discerned two possible reasons for this: 1) the content's representation was not suitable for the intended users, and 2) their language skills were insufficient for understanding the content (Pihkala-Posti et al., 2014; publication III). These could be fixed with more careful planning and creating a more dynamic application that would offer contextual information based on the user's skill level in the language used. There were also some technical issues. For example, speakers were used instead of headsets. In later experiments, these were replaced with Bluetooth headsets; a wireless solution was required since the users moved around considerably for the embodied interaction.

Finally, the paper suggested many changes for future experiments. A detailed analysis of the speech's content was performed in Publication IV, and target selection mechanisms were improved for publications III and IV. Hotspot content for dead ends was also added for these experiments. A head tracking system was also planned, but was not implemented because the application was to be migrated into a browser environment (Publication V) and later into head-mounted displays (Publication VII).

5.3 BERLIN KOMPASS: MULTIMODAL GAMEFUL EMPOWERMENT FOR FOREIGN LANGUAGE LEARNING

Reference

Kallioniemi, P. , Pihkala-Posti, L., Hakulinen, J., Turunen, M., Keskinen, T., and Raisamo, R. (2015). Berlin Kompass: Multimodal gameful empowerment for foreign language learning. *Journal of Educational Technology Systems*. 43: pp. 429-450. doi:10.1177/0047239515588166

Objective

In this experiment, the researchers studied the Berlin Kompass learning environment with 99 Finnish upper secondary school students (aged 16 to 19). The study was conducted as an interdisciplinary experiment with people from the field of interactive technology and education. The goal of the study was to answer several research questions:

1. "Does embodied collaborative learning in virtual environments provide any added value when compared with other approaches?"
2. "How do the participants experience this learning approach?"

3. "Does prior experience with video games have any influence on this experience?"

In addition, we collected feedback about the application and its aspects (immersion, collaboration, and embodied interaction) with a questionnaire with open-ended questions.

Results and Discussion

The results of this experiment suggest that video game experience affects how well the participants adopted the gesture-based interaction method of the application and how immersed they felt during task completion. This effect was not detected between those participants who played video games occasionally and those who never played video games. A significant effect was detected between gaming experience and an understanding of the speech synthesis output provided by the hotspots. In addition, those participants who had more video game experience perceived the audio outputs as more pleasant. One possible explanation for this could be prior experiences with video games that use speech synthesis (instead of voice actors) as output.

Regarding subjective feedback, all the negative feedback was targeted towards the technical qualities of the application, such as audio quality and gestural interaction. Participants reported positive outcomes with the authenticity, collaborative aspects, and embodied interactions of the application. Overall, the questionnaire's results indicate that the application can be effectively used for collaborative language learning tasks that are related to wayfinding and the description of complex visual surroundings. The results also suggest that the level of immersion in the application is sufficient, especially for participants who acted as tourists during the task. This difference in the sensation of immersion between the roles (tourist and guide) might be explained with the more active role of the tourist.

Our observations during the task revealed that there were many strategies to approach the task and all of them resulted in eventually finding the goal. Hotspot contents were used for support by those whose language skills were not as advanced. Participants who were less proficient with the target language also gave more precise, clear descriptions of their surroundings and often managed to complete the task without making errors. (Pihkala-Posti et al., 2014)

The challenge in evaluating the results of this study is that it involves several independent cognitive processes, for example, spatial cognition, language skills, and interaction and collaboration. Nevertheless, this experiment provided interesting results about the relationship of video game experience with the application's immersion and interaction. For future work, a more mobile application was suggested, and during the publication of this article, it was already being developed. Using alternative speech synthesis software for providing the audio output was also

considered for future work and was changed for the next evolution of the application. However, no comparative experiment was conducted.

5.4 COLLABORATIVE NAVIGATION IN VIRTUAL WORLDS: HOW GENDER AND GAME EXPERIENCE INFLUENCE USER BEHAVIOR

Reference

Kallioniemi, P., Heimonen, T., Turunen, M., Hakulinen, J., Keskinen, T., Pihkala-Posti, L., Okkonen, J., and Raisamo, R. (2015). Collaborative navigation in virtual worlds: How gender and game experience influence user behavior. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST '15)*, Stephen N. Spencer (Ed.). ACM, New York, NY, USA, pp. 173-182. doi:10.1145/2821592.2821610

Objective

The objective of this research was to study the difference between females and males in collaborative wayfinding in VEs. For this study, the Berlin Kompass application was utilized. Data for analysis from various sources were collected. This data included application logs, audio and video recordings, and questionnaire data, and the experiment had more than 200 participants. Like publications II and III, the context of the use was foreign language learning. Participants were also asked how immersed in the task they felt during the experiment. The research question in this study was as follows:

- “What are the main differences between genders in collaborative wayfinding, and more specifically, are there gender differences in interaction patterns during wayfinding task completion?”

Participants self-reported their user experience by filling out a questionnaire based on a survey designed for evaluating multimodal systems called SUXES. They responded to nine user experience-related claims on a seven-point Likert scale. This questionnaire was adjusted so it also included the immersion-related item. Regarding video game experience, the participants stated how often (never, less than every month, every month, weekly, or daily) they used a computer to play video games. Audio content analysis was performed on a subset of 20 participants.

Results and Discussion

Findings from this study suggest a multitude of differences between females and males in collaborative wayfinding in VEs. Males spent more time playing video games than females, which also correlated with several questionnaire statements, for example, with “Using the system is pleasant” and “Using the system is entertaining.” Male participants also reported a higher desire to use the application again.

No significant variations were detected in interaction between the two genders. Males finished the task faster in general, but this effect was within the margin of error. No time limitations or goals were provided at the introductory stage of the experiment, and the time spent on the task was not indicative of the participant's language learning experience in any way. In this sense, comparing the completion time between genders does not provide meaningful results regarding spatial abilities. More interesting findings lie in the wayfinding strategies that both genders applied during the task. As Lin et al. (2012) stated, males tend to engage in an exploratory mode of wayfinding that, with this application and context can result in quicker decision making (i.e., moves) but also in less optimal routes. In addition, they suggest that females adopt more conservative strategies during wayfinding, leading to slower decision making but also to fewer detours. These strategies were confirmed in our audio analysis and then verified with the recorded videos.

Analysis of the audio content revealed that males spoke in longer sentences than their female counterparts. Immersion could be one explaining factor here, as those who are more immersed in the collaborative task might tend to communicate more actively to reach the goal faster. It was observed that males started experimenting and interacting with the application soon after or even before the introduction. This was not observed among females, who were generally more patient and waited to finish the introduction before they started using the application. A study by Thompson (1975) came to a similar conclusion and stated that males focus more on competition, status, and independence, and females concentrate on intimacy and consensus.

Respecting the design of collaborative VEs, these communication-related differences should be considered. Gender plays a big role in communication and interaction, and in ideal situations, these applications should cater to both females and males. In summary, the following guidelines are suggested for designing collaborative VEs:

- The application should emphasize varied styles of communication.
- The application should contain a tutorial that introduces the basic interaction. The users should be able to try these interactions during the tutorial phase.
- When used for measuring spatial and/or wayfinding abilities, the use of dynamic landmarks in still images should be avoided. Another solution for this issue would be to use video content in panoramas.

For future studies, it was suggested to use videos instead of images, and this was implemented in the HMD version of the application (Publication VII). Further studies on types of gaming experience were not conducted due to a lack of expertise in the field.

To answer the research question raised in the beginning, “What are the main differences between genders in collaborative wayfinding, and are there gender differences in interaction patterns during wayfinding task completion?” the results revealed that males communicated in longer sentences than their female counterparts. This could be explained by the greater sensation of immersion (i.e., presence) and/or by the more goal-oriented performance resulting from playing video games.

5.5 CITYCOMPASS: A COLLABORATIVE ONLINE LANGUAGE LEARNING APPLICATION

Reference

Kallioniemi, P., Sharma, S., and Turunen, M. (2016). CityCompass: A collaborative online language learning application. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, pp. 94-97. doi:10.1145/2818052.2874334

Objective

In this paper, the researchers introduced CityCompass, a collaborative VE for language learning. This version of the application supports a traditional mouse and keyboard method for interaction. The researchers also introduced Amaze360, a framework for creating collaborative VEs that utilize omnidirectional videos. This framework was later used for developing CityCompass VR for smartphones and head-mounted displays.

Results and Discussion

The browser version of CityCompass was developed with the easier organization of cross-cultural collaboration in mind. The non-mobile setup of the previous Berlin Kompass application was not suitable for extensive studies between countries and thus was unsuitable for the experiments. CityCompass uses modern web technologies, such as three.js for the graphic interface, WebRTC for audio and video transmission, and NodeJS for server-side functionality. The goal for this implementation was to avoid all external plugins or applications, thus running the whole application with a modern browser.

Simultaneously, the researchers were developing Amaze360, a framework for developing VR applications utilizing omnidirectional videos. CityCompass provided the basis for this development.

5.6 USER EXPERIENCE AND IMMERSION OF INTERACTIVE OMNIDIRECTIONAL VIDEOS IN CAVE SYSTEMS AND HEAD-MOUNTED DISPLAYS

Reference

Kallioniemi, P., Mäkelä, V., Saarinen, S., Turunen, M., Winter, Y., and Istudor, A. (2017). User experience and immersion of interactive omnidirectional videos in CAVE systems and head-mounted displays. In: Bernhaupt, R., Dalvi, G., Joshi, A. K., Balkrishan, D., O'Neill, J., and Winckler, M. (Eds.) *Human-Computer Interaction – INTERACT 2017. INTERACT 2017*. Lecture Notes in Computer Science, Vol. 10516. Springer, Cham. doi:10.1007/978-3-319-68059-0_20

Objective

The objective of this study was to compare the user experience and the feeling of immersion between two iODV applications: a CAVE system called cCAVE and an HMD application called Amaze360. Both applications utilized omnidirectional videos. cCAVE was operated from a chair with a rotational sensor, and UI elements were activated with a dwell timer. In the Amaze360 application, the interaction was accomplished while standing, and it was based on the head/device position. Like cCAVE, Amaze360 also used a dwell timer for activation. The participants filled out a questionnaire regarding user experience in which they stated both their expectations and their experiences. In addition, they reported their level of immersion while using the application. The researchers then compared the results from these two applications. The video content of these applications was from both indoors and outdoors environments. Immersion has been studied before in both CAVE and HMD applications, but this sort of comparison was the first of its kind.

The main research questions of this study were the following:

RQ1: What are the differences in the user experience between CAVE and HMD applications?

RQ2: How immersive are interactive CAVE and HMD applications utilizing omnidirectional videos, and are there differences in the level of immersion between these two media?

This experiment was also conducted with the future of the CityCompass application in mind. The researchers had already conducted evaluations with BerlinKompass inside CAVE systems, but also wanted to find the main differences between CAVEs and HMDs regarding user experience and immersion. In this study, the researchers also introduced a new term, *interactive ODVs*, or *iODVs*. iODVs are applications that utilize ODV with additional interactions in addition to looking around the scene, for example, in the form of activating UI elements for more information on objects in the scene or transitioning from one ODV scene to another.

Results and Discussion

The results of this study suggest that the participants' experiences exceeded their expectations greatly, especially with the HMD application. The UX results were positive in general, and both the CAVE system and the HMD application received high scores on the seven-point Likert scale, especially regarding pleasantness, clarity, and performance. Participants expected the CAVE system to be more pleasant to use, perhaps because of the interaction method—using the application while seated can be considered more comfortable than while standing by many. Moreover, it might be difficult to make any estimations about the user experience aspects of the application on the black box-type of device like the HMD, as there are no cues on the method of interaction. Amaze360 was also considered faster than the cCAVE, which can be explained by the interaction method: natural head movements are much faster than spinning on a rotating chair. Both applications were also considered easy to learn, but HMD had a significantly higher score in this metric.

Regarding immersion, the general result is that iODV is a very immersive medium. Amaze360 was also considered more immersive than the cCAVE, for which there are three explanations: 1) HMD obscures the surrounding world completely, thus helping the user to focus more on the content, 2) the sense of depth provided by the stereoscopy of the HMD, and 3) the viewport in the application is based on the head/device orientation, making the exploration of the scenes more natural.

These results indicate that both CAVE systems and HMD applications utilizing iODVs are regarded as useful, easy to learn, and very immersive. The applications had very simple user interfaces where all interactions were based on dwell timers. This type of interaction worked well in both applications, but more sophisticated methods (controllers, embodied interaction, etc.) should be considered if the interaction becomes more complex.

5.7 EFFECT OF GENDER ON IMMERSION IN COLLABORATIVE IODV APPLICATIONS

Reference

Kallioniemi, K., Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., and Ronkainen, K. (2017). Effect of gender on immersion in collaborative iODV applications. In *Proceedings of the 16th International Conference on Mobile and*

Objective

The aim of this experiment was to study the difference in immersion between males and females with collaborative applications that utilize iODVs. As was stated in Publication VI, iODVs are a very immersive medium, and in this work, the researchers studied this phenomenon further by comparing this feeling of “being there” with males and females. The researchers also expanded the concept of immersion into six subscales: *spatial immersion, interaction, involvement, realness, auditory, and physical*. Gender has been repeatedly suggested as a big factor in immersion, and distinctions between genders have been detected in many topics, including watching TV shows, playing video games, and interacting with VEs. For this experiment, the participants used the latest version of CityCompass, CityCompass VR.

The questions we wanted to answer in this study were the following:

RQ1: Are there differences in immersion between the genders while performing collaborative tasks in iODV applications?

RQ2: Are there any gender differences in the task performance (task completion time, navigational mistakes)?

We also wanted to create an immersion questionnaire that would be suitable for a wider age group. For example, in experiments regarding language learning, many of the participants were in elementary school, and many of the commonly used questionnaires were too complex for them. In the researchers’ suggested questionnaire, the aim was to make the statements and questions simple and suitable for most age groups.

Results and Discussion

There was a significant difference in both spatial immersion and involvement subscales, as males reported higher scores than females in both cases. There are three possible explanatory factors for these variations: self-efficacy with technology, computer experience, and video game experience. Felnhofer et al. (2012) claimed that self-efficacy with technology may be one reason for this greater feeling of immersion among males. Computer experience was suggested to be a factor in a study by Waller, Hunt, and Knapp (1998). Video game experience was suggested to be a factor for the gender difference in immersion by Lachlan and Krcmar (2011), and this was supported by these researchers’ results in Publication IV. For increasing the sensation of immersion among females, the researchers suggest the addition of television-type, dramatized content such as the Bollywood method for tasks performed in VEs.

Another finding in this study was that “performing interactive, collaborative tasks in iODV applications helps build a shared understanding between the users” (Publication VII). Collaborative tasks in which the users share a task and aim to reach a common goal keep them more focused on the task and enhance their communication. They “also have to consider the other user’s viewpoint and situation” while performing these tasks.

In addition, the setup’s technical and physical limitations might affect immersion. Blurred lenses, errors in videos’ looping sequences, and stitching errors in ODV content all can harm the user’s experience. These problems can be overcome by following iODV content design guidelines (e.g., Saarinen et al., 2017; Argyriou et al., 2016).

There were no distinctions between males and females in the task completion data. These results are like the ones presented in Publication IV. Even though spatial ability was not the focus of this experiment, it is still an interesting finding, as the differences that are usually detected between genders in wayfinding tasks are not present when this collaborative aspect is added (Publication IV).

Another contribution made in this publication is the immersion questionnaire. This custom questionnaire with six subscales was planned to be simpler than currently used questionnaires and therefore should be suitable for people of various ages. For this questionnaire, the researchers added subscales that are not commonly present: interaction, physical, and auditory. For future studies, items related to physical and auditory subscales could be added to the questionnaire. In addition, the researchers are interested in adding and observing dramatized content to determine how it affects the user’s feeling of immersion. Video game experience, self-efficacy, and technology acceptance are also metrics that should be considered in future studies.



6 Discussion

The two main research questions of this dissertation were as follows:

RQ1: What strategies do people use to find their way in collaborative virtual environments?

RQ2: What aspects affect collaborative wayfinding tasks?

These questions were answered throughout this dissertation, and these findings can be useful in developing new, better collaborative VEs with wayfinding aspects. Publication I lays the groundwork for landmark-based wayfinding in VEs, which was then evaluated extensively in Publication II. Previous research has shown that landmark-based wayfinding is an effective method for pedestrian wayfinding (e.g., Siegel and White, 1975; Hirtle and Heidorn, 1993; Denis et al., 1999; Lovelace, Hegarty and Montello, 1999), and this was confirmed to be the case in collaborative VEs as well (publications II and IV). The model for highlighting prominent landmarks in scenery was successful in selecting the same landmarks in a wayfinding situation as humans in a laboratory setting, which makes it a good tool for designing large-scale VEs that utilize landmarks.

For evaluating the model and collaborative wayfinding in general, the researchers developed a collaborative VE application, CityCompass, with three evolutionary stages. These applications were also used in the context of language learning. All these applications presented 360-degree sceneries with either still images or videos, and they were used simultaneously by two subjects. Our first application, Berlin Kompass, utilized embodied interaction and a large projection screen. It was evaluated extensively, and the results were reported in publications II, III, and IV. The next application, CityCompass, was developed for the desktop environment. The interaction

with this application was accomplished with either a keyboard and mouse or with a touchscreen. This application was introduced in Publication V. This stage was utilized by companies that provide language learning laboratories for schools. Thus, it was not evaluated in the context of this thesis, but it laid the groundwork for the next application, CityCompass VR (Publication VII).

When all three applications and their implementations are examined more closely, two main observations can be made: First, the mobility of these installations increased with each evolutionary stage. Berlin Kompass was a large installation with its wall projections and required much space because of the embodied interaction. CityCompass could be used on a regular desktop computer or even on a tablet or a smartphone. CityCompass VR used only a smartphone and an HMD device, making it very easy to move around and utilize in many environments. Of course, this brings with it new challenges, such as overheating and battery life issues. For CityCompass VR, the researchers also suggested a new term, *interactive omnidirectional video*, or *iODV*, for interactive applications that utilize omnidirectional videos. For more research on this topic see, for example, Saarinen et al. (2017).

Concerning collaborative wayfinding strategies, the researchers observed several techniques in navigating the cityscapes. In Publication II, the researchers stated that, because of the lack of prominent landmarks, users attempted a range of solutions in solving the wayfinding scenario. For example, they communicated in degrees (especially males) and resorted to using their native language (instead of the target language in the language learning scenario). Differences between males and females were evident in wayfinding strategies and communication during the wayfinding tasks. Some of these are like those found in individual wayfinding (for example, Astur et al., 1998; Voyer et al., 1995; Coluccia and Louse, 2004), but the researchers also detected that the distinctions between males and females in wayfinding task performance diminish when they are conducted collaboratively (publications IV and VII). Lin et al. (2012) suggested that males tend to engage more in an exploratory mode of wayfinding than females, which in our studies resulted in quicker decisions but not necessarily correct or optimal routes. Females adopted more conservative strategies that resulted in slower wayfinding but also fewer detours.

Regarding communication, some interesting observations were made. Our analysis of voice recordings during the collaborative wayfinding task suggested that males spoke longer on average than females (Publication IV). This finding could be related to the feeling of immersion. Thompson (1975) claims that females and males communicate differently, and based on their results, males concentrate more on competition and independence, whereas females tend to focus on intimacy and consensus. This was also evident in the studies of these researchers, in which males often started exploring the

VE even before the introductory part of the experiment, whereas females tried to maintain a consensus and discuss strategy with the other user before even attempting to interact with the application.

One limitation regarding wayfinding that the first two applications shared was the prominence of dynamic landmarks. Because the panoramic content in these applications consisted of still images, they also contained dynamic objects, such as cars and pedestrians, in the scenery. These were sometimes referred to by the users (males more than females) during the wayfinding task. This was changed in the latest application, CityCompass VR, which utilized omnidirectional videos as content, making it a more realistic depiction of real-world situations. As some of these observations are more related to collaborative language learning, they are outside the scope of this thesis. A summary of these strategies is reported in some detail in Publication III.

The researchers also studied the effects of immersion (publications IV and VII) and video game experience (publications III and IV) on collaborative wayfinding in VEs. In Publication IV, the questions regarding the feeling of immersion and video game experience were integrated into the SUXES questionnaire to avoid placing too many items in one questionnaire. In Publication III, video game experience was measured with two items (experience and frequency of playing video games). These somewhat simplified questionnaires revealed that video game experience does indeed affect the user's feeling of immersion and how well he or she adopts gesture-based interaction techniques. It also had a significant effect on how users perceived the speech synthesis output of the application. Females with less video game experience also had more negative user experiences with the application, whereas this was not detected between males.

The experiment in Publication VII concentrated wholly on the feeling of immersion and the distinctions between females and males respecting this phenomenon. With a customized questionnaire, the researchers measured immersion with six subscales: *spatial immersion*, *interaction*, *involvement*, *realness*, *physical*, and *auditory*. The goal was to implement a questionnaire that is simple and suitable for all age groups, including elementary school students. This questionnaire was filled out by the participants after they completed a collaborative wayfinding task with the CityCompass VR application. The results showed that males "reported significantly higher scores in *spatial immersion* and *involvement* subscales" (Publication VII). Again, no single explanatory factor for this can be stated, but related work has suggested that this could be due to video game experience (Lachlan and Krcmer, 2011; publications III and IV). It has also been suggested that greater self-efficacy with computers among males could be an explanatory factor in this phenomenon (Felnhofer et al., 2012). In addition, the researchers studied the difference in immersion between two VR systems

utilizing iODVs, HMD, and CAVE (Publication VI). The results indicate that HMD is more immersive than CAVE. This difference can be attributed to three factors: a) the head mount obscures outside stimuli from the user, b) the stereoscopic view creates a sense of depth, and c) the viewport on the HMD is based on the orientation of the head, allowing users to naturally look around at their surroundings. There are still possibilities for future studies with these two systems, as the CAVE did not have a stereoscope or any kind of user and/or head tracking device.

The current mainstream hype surrounding VR and its applications is very similar to the one the researchers experienced during the '90s. Regardless of its success, it is still a very useful tool for science and education. Its applications range from exposure theory (Price and Anderson, 2007; Riva, Botella, Légeron and Optale, 2004) to education (publications II and III; Kelton, 2007). Many of these applications have navigational aspects, in which the user moves around in VEs, thus utilizing their spatial abilities. This process requires cognitive resources, including attention from the user, making these resources unavailable for the actual task at hand. In this dissertation, the researchers provide guidance on how to plan and implement collaborative VEs to make them easier to navigate. These guidelines are supported by subjective and objective data gathered from the experiments provided by this thesis.

In summary, this dissertation provides results and guidance for developing collaborative VEs that have wayfinding aspects. If these applications are well designed regarding interactions and UI, the user needs to pay less attention to them and may concentrate solely on the actual purpose (e.g., language learning) of the application. In addition, if provided an immersive experience within these VEs, the user may feel more presence and thus have a stronger experience within the environment. For example, in the case of educational applications, this may lead to a better learning experience. Finally, this work resulted in three collaborative VEs that can be adjusted for many fields of research, including education and language learning.



7 Conclusion

In summary, this dissertation reported research on collaborative wayfinding in VEs. This research was disseminated in seven individual publications that focused on various aspects of this phenomenon. This dissertation contributes to these issues in the following ways:

- It provides a model for landmark-based wayfinding for VEs (which is also usable in real-world situations). This model highlights the most salient landmarks in the landscape.
- It introduces an evolutionary cycle of collaborative VE applications that utilize wayfinding tasks. All these applications have their own unique aspects and features, and can be used in a range of contexts (e.g., language learning and other educational contexts).
- It introduces an array of collaborative wayfinding strategies used while navigating VEs collaboratively. These strategies were observed by the researchers during task completion and then confirmed by system log data and audio and video recordings.
- It also provides information about how gender affects collaborative wayfinding in terms of communication and interaction.
- The researchers also introduced a new concept, interactive omnidirectional video, or iODV, for interactive VE applications that utilize omnidirectional videos. The researchers evaluated iODVs with both HMD and CAVE applications.

- The researchers studied the feeling of immersion in collaborative VEs and iODV applications, and discovered several factors: 1) video game experience affects immersion with collaborative VEs, 2) HMD is a more immersive medium than CAVE, 3) when comparing immersion between males and females, males tend to be more immersed while using collaborative VEs. The researchers also provided possible explanations for these phenomena.

These findings are useful for the design and implementation of collaborative VEs with wayfinding aspects. The results suggest that factors such as gender, video game experience, and the feeling of immersion affect the user's experience with such applications and that people use different wayfinding strategies when completing collaborative navigational tasks. A knowledge of these strategies will also benefit the design process with such applications.

7.1 FUTURE WORK

The work introduced in this dissertation provides a basis for future work studying collaborative wayfinding in VEs. The collaborative aspects of wayfinding have been studied relatively little, and modern technology provides the means for developing the understanding of this phenomena with further experiments. The added fidelity and immersion in VEs supports these studies by making the environment more realistic, thus improving the user's experience with the application. By using iODVs as content and HMDs as devices in VEs, the researchers can provide the user with a highly immersive system that is relatively cheap to produce. HMDs also allow better mobility. By adding walkable VEs, for example, the researchers can add the use of motor and vestibular systems to these experiments.

CityCompass VR is still under development and has a large potential for conducting studies in different contexts. All the applications presented in this dissertation have been studied in the educational context, and the results reported are encouraging for further researchers. The applications still have much room for improvement, and spatial audio, for instance, could be added to all versions. One could also implement other sensory stimulations in them. For example, one might study the effects of olfactory stimuli or the effects of artificial wind on wayfinding performance.

All these applications could also be used in several domains of research, including education and language learning. CityCompass has already been integrated into a learning laboratory as a language learning module. As the content in all three applications is flexible, the context in these applications can be shifted easily, for example, for learning biology, history, or art.



8 References

Adamo-Villani, N., and Wilbur, R. (2008). Two novel technologies for accessible math and science education. *IEEE Multimedia–Special Issue on Accessibility*, Vol. 15, No. 4. pp. 38-46. doi:10.1109/MMUL.2008.97

Aginsky, V., Harris, C., Rensink, R., and Beusmans, J. (1997). Two strategies for learning a route in a driving simulator. *Journal of Environmental Psychology*, Vol. 17, pp. 317–331. doi:10.1006/jevp.1997.0070

Akin, D.L., Minsky, M.L., Thiel, E.D., and Kurtzman, C.R. (1983). Space applications of automation, robotics, and machine intelligence systems (ARAMIS), Phase II, Vol. 3: Executive Summary. Technical Report. Massachusetts Institute of Technology, Cambridge, MA. Retrieved on 29.12.2017 from: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19840002515.pdf>

Alexander, A.L., Brunyé, T., Sidman, J., and Weil, S.A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC'05)*. NTSA.

Allen, G. L. (1999). Spatial abilities, cognitive maps, and wayfinding: Bases for individual differences in spatial cognition and behavior. In *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*, R.G. Golledge (Ed.) Baltimore: The Johns Hopkins University Press, pp. 46-80.

Allen, G.L. (2000). Principles and practices for communicating route knowledge, *Applied Cognitive Psychology*, Vol. 14, pp. 333–359. doi:10.1002/1099-0720(200007/08)14:4<333::AID-ACP655>3.0.CO;2-C

Alsina-Jurnet, I., Gutiérrez-Maldonado, J., and Rangel-Gómez, M.-V. (2011). The role of presence in the level of anxiety experienced in clinical virtual environments. *Computers in Human Behavior*, Vol. 27, pp. 504–512. doi:10.1016/j.chb.2010.09.018

Alyan, S., and McNaughton, B.L. (1999). Hippocampectomized rats are capable of homing by path integration. *Behavioral Neuroscience*, Vol. 113, No. 1, pp. 19–31. doi:10.1037/0735-7044.113.1.19

Argyriou, L., Economou, D., Bouki, V., and Doumanis, I. (2016). Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications. In *15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)*, Granada, 2016, pp. 145–152. doi:10.1109/IUCC-CSS.2016.028

Astur, R.S., Ortiz, M.L., and Sutherland, R.J. (1998). A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference. *Behavioural Brain Research*, Vol. 93, pp. 185–90. doi:10.1016/S0166-4328(98)00019-9

Astur, R.S., Purton, A.J., Zaniwski, M.J., Cimadevilla, J., and Markus, E.J. (2016). Human sex differences in solving a virtual navigation problem. In *Behavioural Brain Research*, Vol. 308, pp. 236–243, doi:10.1016/j.bbr.2016.04.037

Baños, R.M., Botella, C., García-Palacios, A., Villa, H., Perpiña, C., and Gallardo, M. (1999). Psychological variables and reality judgment in virtual environments: The roles of absorption and dissociation. *CyberPsychology & Behavior*. Vol. 2, No. 2 (April 1999), pp. 143–148. doi:10.1089/cpb.1999.2.143

Benford, S., Bowers, J., Fahlén, L.E., Greenhalgh, C., and Snowdon, D. (1995). User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press/Addison-Wesley Publishing Co., pp. 242–249. doi:10.1145/223904.223935

Benko, H., and Wilson, A.D. (2010). Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, pp. 19–28. doi:10.1145/1936652.1936657

Berns, A., Gonzalez-Pardo, A., and Camacho, D. (2011). Implementing the use of virtual worlds in the teaching of foreign languages. In S.

Czepielewski (Ed.), *Learning a Language in Virtual Worlds: A Review of Innovation and ICT in Language Teaching Methodology*. Warsaw Academy of Computer Science, Warsaw, pp. 33–40.

Biocca, F. (1997). The Cyborg's dilemma: Progressive embodiment in virtual environments, *Journal of Computer-Mediated Communication*, Vol. 3, No. 2. doi:10.1111/j.1083-6101.1997.tb00070.x

Bisiach, E., Pattini, P., Rusconi, M.L., Ricci, R., and Bernardini, B. (1997). Unilateral neglect and space constancy during passive locomotion, *Cortex*, Vol. 33, pp. 313–322. doi:10.1016/S0010-9452(08)70007-8

Blascovich, J. (2002). A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments*. ACM, pp. 25–30. doi:10.1145/571878.571883

Boswell, J.E. (2001). User-centered iterative design of a collaborative virtual environment. Master's thesis, Naval Postgraduate School. <http://hdl.handle.net/10945/10874>

Boult, T.E. (1998). Remote reality via omnidirectional imaging. In *ACM SIGGRAPH 98 Conference Abstracts and Applications (SIGGRAPH '98)*. ACM, New York, NY, USA. p. 253. doi:10.1145/280953.282215

Bowman, D.A., and McMahan, R.P. (2007). Virtual reality: How much immersion is enough? *Computer*, Vol. 40, No. 7 (July 2007), pp. 36–43. doi:10.1109/MC.2007.257

Bricken, W. (1990). Learning in virtual reality. Technical Report No. HITL-M-90-5. Retrieved on 8.8.2017 from: <http://www.wbricken.com/pdfs/03words/03education/02vr-education/01learn-in-VR.pdf>

Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., and Pidruzny, J.N. (2009). The development of the game engagement questionnaire: A measure of engagement in video game-playing. In *Journal of Experimental Social Psychology*, Vol. 45, No. 4, pp. 624–634. doi:10.1016/j.jesp.2009.02.016

Bull, P. (1983). *Body Movement and Interpersonal Communication*. New York: John Wiley & Sons. ISBN:0471900699

Burnett, G.E., and Lee, K. (2005). The effect of vehicle navigation systems on the formation of cognitive maps. In: Underwood, G. (Ed.) *Traffic and Transport Psychology: Theory and Application*. Amsterdam: Elsevier, pp. 407–418. doi:10.1016/B978-008044379-9/50188-6

- Cannon-Bowers, J.A., Salas, E., and Converse, S. (1993). Shared mental models in expert team decision making. In N.J. Castellan, Jr. (Ed.), *Individual and Group Decision Making: Current Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 221–246.
- Carroll, J.B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press. ISBN: 0-521-38712-4.
- Castelli, L., Corazzini, L.L., and Geminiani, G.C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human Behaviour*, Vol. 24, pp. 1643–1667. doi:10.1016/j.chb.2007.06.005
- Chai, X.J., and Jacobs, L.F. (2009). Sex differences in directional cue use in a virtual landscape. *Behavioral Neuroscience*, Vol. 123, No. 2, pp. 276–283. doi: 10.1037/a0014722
- Chen, J.L., and Stanney, K.M. (1999). A theoretical model of wayfinding in virtual environments: Proposed strategies for navigational aiding. In *Presence*, Vol. 8, No. 6, pp. 671–685. doi:10.1162/105474699566558
- Codd, A.M., and Choudhury, B. (2011), Virtual reality anatomy: Is it comparable with traditional methods in the teaching of human forearm musculoskeletal anatomy? *Anatomical Sciences Education*, Vol. 4, pp. 119–125. doi:10.1002/ase.214
- Coluccia, E., and Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, Vol. 24, No. 3, pp. 329–340. doi:10.1016/j.jenvp.2004.08.006
- Coluccia, E., Iosue, G., and Brandimonte, M.A. (2007). The relationship between map drawing and spatial orientation abilities: A study of gender differences. *Journal of Environmental Psychology*, Vol. 27, No. 2, pp. 135–144. doi:10.1016/j.jenvp.2006.12.005
- Compeau, D.R., and Higgins, C.A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, Vol. 19, pp. 189–211. doi:10.2307/249688
- Cooke, N.J., Salas, E., and Cannon-Bowers, J.A. (2000). Measuring team knowledge. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 42, No. 1, pp. 151–173. 10.1518/001872000779656561
- Cornell, E.H., Heth, C.D., and Broda, L.S. (1989). Children's way finding: Response to instructions to use environmental landmarks. *Developmental Psychology*, Vol. 25, pp. 755–764. doi:10.1037/0012-1649.25.5.755

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row. ISBN:0061339202

Dabbs, J.M.J., Chang, E.-L., Strong, R.A., and Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior*, Vol. 19, pp. 89-98. doi:10.1016/S1090-5138(97)00107-4

Darken, R.P., and Banker, W.P. (1998). Navigating in natural environments: A virtual environment training transfer study, in *Proceedings of IEEE 1998 Virtual Reality Annual International Symposium*, IEEE Computer Society: Los Alamitos, CA. pp. 12-19.

De Beni, R., Pazzaglia, F., and Gardini, S. (2006). The role of mental rotation and age in spatial perspective-taking tasks: When age does not impair perspective-taking performance. *Applied Cognitive Psychology*, Vol. 20, pp. 807-821. doi:10.1002/acp.1229

Deakin, A. (1996). Landmarks as navigational aids on street maps. *Cartography and Geographic Information Systems*, Vol. 23, No. 1, pp. 21-36. doi:10.1559/152304096782512159

DeBrine, J.D., and Morrow, D.E. (2000). Re-Purposing Commercial Entertainment Software for Military Use. Master's Thesis. Naval Postgraduate School, Monterey, California. Retrieved on 30.12.2017 from: <https://calhoun.nps.edu/handle/10945/26726>

Decock, J., Van Looy, J., Bleumers, L., and Bekaert, P. (2011). The pleasure of being (there?). An explorative study into the effects of presence and identification on the enjoyment of an interactive theatrical performance using omnidirectional video. In *Proceedings of the International Society for Presence Research Annual Conference. ISPR 2011*, Vol. 29, No. 4, pp. 449-459. doi:10.1007/s00146-013-0487-6

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, Vol. 16. pp. 409-458.

Denis, M., Pazzaglia, F., Cornoldi, C., and Bertolo, L. (1999). Spatial discourse and navigation: An analysis of route directions in the city of Venice. *Applied Cognitive Psychology*. Vol. 13, pp. 145-174. doi:10.1002/(SICI)1099-0720(199904)13:2<145::AID-ACP550>3.0.CO;2-4

Dickinson, T.L., and McIntyre, R.M. (1997). A conceptual framework for teamwork measurement. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Series in Applied Psychology. Team Performance Assessment and Measurement:*

Theory, Methods, and Applications. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 19–43.

Ekstrom, R.B., French, J.W., and Harman, H.H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, Vol. 79, p. 84.

Ekstrom, R.B., French, J.W., Harman, H.H., and Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Elias, B. (2003). Extracting landmarks with data mining methods. In: Kuhn, W., Worboys, M.F., Timpf, S. (Eds.) *COSIT 2003*. LNCS, Vol. 2825, pp. 375–389. Heidelberg: Springer. doi:10.1007/978-3-540-39923-0_25

Ellard, C. (2009). *You Are Here: Why We Can Find Our Way to the Moon, But Get Lost in the Mall*. Knopf Doubleday Publishing Group.

Evans, G.W., Marrero, D.G., and Butler, P.A. (1981). Environmental learning and cognitive mapping. *Environment and Behavior*, Vol. 13, pp. 83–104. doi:10.1177/0013916581131005

Evans, G.W., Skorpanich, M.A., Gärling, T., Bryant, K.J., and Bresolin, B. (1984). The effects of pathway configuration, landmarks, and stress on environmental cognition. *Journal of Environmental Psychology*, Vol. 4, pp. 323–335. doi:10.1016/S0272-4944(84)80003-1

Felnhofer, A., Kothgassner, O.D., Beutl, L., Hlavacs, H., and Kryspin-Exner, I. (2012). Is virtual reality made for men only? Exploring gender differences in the sense of presence. In *Proceedings of the International Society on Presence Research*, pp. 103–112.

Fontaine, S., and Denis, M. (1999). The production of route instructions in underground and urban environments. In Freksa, C., and Mark, D.M. (Eds.), *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, COSIT '99*, Vol. 1661 Lecture Notes in Computer Science. Berlin: Springer-Verlag, pp. 83–94. doi:10.1007/3-540-48384-5_6

Gabbard, J.L. (1997). *A Taxonomy of Usability Characteristics in Virtual Environments*. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. Retrieved on 1.1.2018 from: <http://www.sdml.cs.kent.edu/library/Gabbard97.pdf>

Galea, L.A., and Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences*, Vol. 14, No. 1, pp. 53–65. doi:10.1016/0191-8869(93)90174-2

Gerrig, R.J. (1993). *Experiencing Narrative Worlds*. New Haven, CT: Yale University Press.

Goffman, E. (1963). *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: The Free Press. ISBN:0029119405

Golding, J.M., Graesser, A.C., and Hauselt, J. (1996). The process of answering direction-giving questions when someone is lost on a university campus: The role of pragmatics. *Applied Cognitive Psychology*, Vol. 10, No. 23-39. doi:10.1002/(SICI)1099-0720(199602)10:1<23::AID-ACP357>3.0.CO;2-H

Goldman-Rakic, P.S. (1995). Cellular basis of working memory. *Neuron*, Vol. 14, No. 3, pp. 477-85. doi:10.1016/0896-6273(95)90304-6

Golledge, R.G. (1999). *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*. JHU Press. doi:10.1046/j.1468-1331.2000.t01-1-00082.x

Golledge, R.G., Jacobson, R.D., Kitchin, R., and Blades, M. (2000). Cognitive maps, spatial abilities, and human wayfinding. *Geographical Review of Japan, Series B* 73(2), pp. 93-104. doi:10.4157/grj1984b.73.93

Graham, P., and Cheng, K. (2009). Ants use the panoramic skyline as a visual cue during navigation. *Current Biology*, Vol. 19, No. 20, R935-R937.

Green, M.C., and Brock, T.C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, Vol. 79, No. 5 (November 2000), pp. 701-721. doi:10.1037/0022-3514.79.5.701

Habel, C. (1988). Prozedurale Aspekte der Wegplanung und Wegbeschreibung (Procedural aspects of route planning and directions). In *Sprache in mensch und computer*. pp. 107-133. VS Verlag für Sozialwissenschaften, Wiesbaden. doi:10.1007/978-3-322-86251-8_5

Halpern, D. F. (2000). *Sex Differences in Cognitive Abilities: 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN:0805827927

Hansen, S., Richter, K.F., and Klippel, A. (2006). Landmarks in OpenLS—a data structure for cognitive ergonomic route directions. In *International Conference on Geographic Information Science*, Vol. 4197, pp. 128-144. Heidelberg: Springer. doi:10.1007/11863939_9

Head, D., and Isom, M. (2010). Age effects on wayfinding and route learning skills. *Behavioural Brain Research*, Vol. 209, pp. 49-58. doi:doi.org/10.1016/j.bbr.2010.01.012

Hegarty, M., and Waller, D. (2005). Individual differences in spatial abilities. *The Cambridge Handbook of Visuospatial Thinking*, pp. 121–169. ISBN: 9780521807104

Hegarty, M., Richardson, A.E., Montello, D.R., Lovelace, K., and Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, Vol. 30, pp. 425–447. doi:10.1016/S0160-2896(02)00116-2

Helsel, S. (1992). Virtual reality and education. *Educational Technology*. XXXII, 5 (May 1992), pp. 38–42.

Hipp, M., Schaub, F., Kargl, F., and Weber, M. (2010). Interaction weaknesses of personal navigation devices. In *Proceeding of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 129–136. doi:10.1145/1969773.1969796

Hirtle, S.C., and Heidorn, P.B. (1993). The structure of cognitive maps: Representations and processes. In T. Gärling and R.G. Golledge (Eds.), *Advances in Psychology*, Vol. 96. *Behavior and Environment: Psychological and Geographical Approaches*. pp. 170–192. doi:10.1016/S0166-4115(08)60043-6

Holding, C.S., and Holding, D.H. (1989). Acquisition of route network knowledge by males and females. *Journal of General Psychology*, Vol. 116, pp. 29–41. doi:10.1080/00221309.1989.9711108

Hund, A.M., and Gill, D.M. (2014). What constitutes effective wayfinding directions: The interactive role of descriptive cues and memory demands, In *Journal of Environmental Psychology*, Vol. 38, pp. 217–224. doi:10.1016/j.jenvp.2014.02.006

Hund, A.M., and Minarik, J.L. (2006). Getting from here to there: Spatial anxiety, wayfinding strategies, direction type, and wayfinding efficiency. *Spatial Cognition and Computation*, Vol. 6, pp. 179–201. doi:10.1207/s15427633scc0603_1

Hund, A.M., and Nazarczuk, S.N. (2009). The effects of sense of direction and training experience on wayfinding efficiency. *Journal of Environmental Psychology*, Vol. 29, pp. 151–159. doi:10.1016/j.jenvp.2008.05.009

Hund, A.M., and Padgitt, A.J. (2010). Direction giving and following in the service of wayfinding in a complex indoor environment. *Journal of Environmental Psychology*, Vol. 30, pp. 553–564. doi:10.1016/j.jenvp.2010.01.002

Hund, A.M., Haney, K.H., and Seanor, B.D. (2008). The role of recipient perspective in giving and following wayfinding directions. *Applied Cognitive Psychology*, Vol. 22, pp. 896–916. doi:10.1002/acp.1400

Hurlebaus, R., Basten, K., Mallot, H.A., and Wiener, J.M. (2008). Route learning strategies in a virtual cluttered environment. *Spatial Cognition VI*, pp. 104–120. doi:10.1007/978-3-540-87601-4_10

Hölscher, C., Tenbrink, T., and Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, Vol. 121, pp. 228–247. doi:10.1016/j.cognition.2011.06.005

Ishikawa, T., and Montello, D. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, Vol. 52, pp. 93–129. doi:10.1016/j.cogpsych.2005.08.003

Janzen, G., and van Turenout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, Vol. 7, No. 6, pp. 572–574. doi:10.1038/nn1257

Järvinen, A., and Ekola, L. (2014). Turning point – A Practical Assessment of Using 360 Video in Sign Language Interpreting Studies. Diaconia University of Applied Sciences. Degree Programme in Sign Language Interpretation. Bachelor Thesis.

Kato, Y., and Takeuchi, Y. (2003). Individual differences in wayfinding strategies. *Journal of Environmental Psychology*, Vol. 23, pp. 171–188. doi:10.1016/S0272-4944(03)00011-2

Kelton, A. (2007). Second Life: Reaching into the virtual world for real-world learning. *Educause Center for Applied Research: Research Bulletin*, Issue 17.

Kim, B., Lee, S., and Lee, J. (2007). Gender differences in spatial navigation. *World Academy of Science, Engineering and Technology*, Vol. 31, pp. 297–300. doi:10.1999/1307-6892/5947

Kitchin, R., and Blades, M. (2002). *The cognition of geographic space*. London: I.B. Tauris. ISBN: 978-1860647048

Klatzky, R.L., Beall, A.C., Loomis, J.M., Golledge, R.G., and Philbeck, J.W. (1999). Human navigation ability: Tests of the encoding-error model of path integration. *Spatial Cognition and Computation*, Vol. 1, pp. 31–65. doi:10.1023/A:1010061313300

Klippel, A. (2003). Wayfinding choremes: Conceptualizing wayfinding and route direction elements. SFB/TR 8 *Spatial Cognition*. ISBN:3887225902

Kohler, M., and Wehner, R. (2005). Idiosyncratic route-based memories in desert ants, *Melophorus bagoti*: How do they interact with path-integration

vectors? *Neurobiology of Learning and Memory* 83, pp. 1–12. doi:10.1016/j.nlm.2004.05.011

Kozlowski, L.T., and Bryant, K.J. (1977). Sense of direction spatial orientation, and cognitive maps. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 3, pp. 590–598. doi:10.1037/0096-1523.3.4.590

Kraiger, K., and Wenzel, L.H. (1997). Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In Michael T. Brannick, Eduardo Salas, and Carolyn Prince, *Team Performance Assessment and Measurement*, Lawrence Erlbaum Associates, pp. 63–84.

Kwiatek, K., and Woolner, M. (2010). Transporting the viewer into a 360° heritage story: Panoramic interactive narrative presented on a wrap-around screen. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference*, pp. 234–241. doi:10.1109/VSM.2010.5665980

Lachlan, K.A., and Krcmar, M. (2011). Experiencing presence in video games: The role of presence tendencies, game experience, gender, and time spent in play. *Communication Research Reports*, Vol. 28, No. 1, pp. 27–31. doi:10.1080/08824096.2010.518924

Lawton, C.A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, Vol. 30, pp. 765–779. doi:10.1007/BF01544230

Lawton, C.A., and Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles*, Vol. 47, pp. 389–401. doi:10.1023/A:1021668724970

Lawton, C.A., and Morrin, K.A. (1999). Gender differences in pointing accuracy in computer simulated 3D mazes. *Sex Roles*, Vol. 40 No. 1/2, pp. 73–92. doi:10.1023/A:1018830401088

Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, Vol. 10, No. 3 (June 2001), pp. 282–297. doi:10.1162/105474601300343612

Li, S., Dragicevic, S., Castro, F.A., Sester, M., Winter, S., Coltekin, A., Pettit, C., Jiang, B., Haworth, J., Stein, A., and Cheng, T. (2016) Geospatial big data handling theory and methods: A review and research challenges. In *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 115, pp. 119–133. doi: 10.1016/j.isprsjprs.2015.10.012

Lin, C-T., Huang, T-Y., Lin, W-J., Chang, S-Y., Lin, Y-H., Ko, L-W., Hung, D.L., and Chang, E.C. (2012). Gender differences in wayfinding in virtual environments with global or local landmarks. *Journal of Environmental Psychology*, Vol. 32, No. 2, pp. 89-96. doi:10.1016/j.jenvp.2011.12.004

Lohman, F.D. (1979). *Spatial Ability: A Review and Reanalysis of the Correlational Literature*. Technical Report. Stanford University, California, CA. 226 pages. Retrieved on 22.9.2017 from: <http://www.dtic.mil/docs/citations/ADA075972>

Lombard, M., Ditton, T.B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., Rossman, J., and Park, S. (2000). Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument. Presented at the *Third International Workshop on Presence*, Delft, The Netherlands.

Loomis, J.M., Klatzky, R.L., Golledge, R.G., and Philbeck, J.W. (1999). Human navigation by path integration. *Wayfinding behavior: Cognitive mapping and other spatial processes*. pp. 125-151.

Lovelace, K.L., Hegarty, M., and Montello, D.R. (1999). Elements of good route directions in familiar and unfamiliar environments. In C. Freksa and D.M. Mark (Eds.), *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*. Berlin: Springer-Verlag. pp. 65-82. doi:10.1007/3-540-48384-5_5

Lynch, K. (1960). *The Image of the City*. Boston: The MIT Press. ISBN:0-262-62001-4.

MacEachren, A. M. (1992). Application of environmental learning theory to spatial knowledge acquisition from maps. *Annals of the Association of American Geographers*, Vol. 82, No. 2, pp. 245-274. doi:10.1111/j.1467-8306.1992.tb01907.x

Mallot, H., and Gillner, S. (2000). Route navigation without place recognition. What is recognized in recognition triggered responses? *Perception*, Vol. 29, pp. 43-55. doi:10.1068/p2865

McNamara, T.P., Rump, B., and Werner, S. (2003). Egocentric and geocentric frames of reference in memory of large-scale space. *Psychonomic Bulletin and Review*, Vol. 10, pp. 589-595. doi:10.3758/BF03196519

Meilinger, T., Knauff, M., and Bühlhoff, H.H. (2008). Working memory in wayfinding – A dual task experiment in a virtual city. *Cognitive Science*, Vol. 32, No. 4, pp. 755-770. doi:10.1080/03640210802067004

Merchant, Z., Goetz, E.T., Cifuentes, L., Keeney-Kennicutt, W., and Davis, T.J. (2014). Effectiveness of virtual reality-based instruction on students'

learning outcomes in K-12 and higher education: A meta-analysis, In *Computers and Education*, Vol. 70, pp. 29-40. doi:10.1016/j.compedu.2013.07.033

Michon, P., and Denis, M. (2001). When and why are visual landmarks used in giving directions? In: *Montello, D. (Ed.) Proceedings of the International Conference COSIT 2001, Spatial Information Theory*. Springer, Verlag, pp. 292-305. doi:10.1007/3-540-45424-1_20

Mikropoulos, T.A., and Bellou, J. (2006). The unique features of educational virtual environments. In P. Isaias, M. McPherson, and F. Banister (Eds.), *Proceedings of E-society 2006*, Vol. 1. IADIS, pp. 122-128. doi:10.1007/s10055-006-0039-1

Mikropoulos, T.A., and Natsis, A. (2010). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers and Education*, Vol. 56, No. 3, pp. 769-780. doi:10.1016/j.compedu.2010.10.020

Minocha, S., and Hardy, C.L. (2011). Designing navigation and wayfinding in 3D virtual learning spaces. In: *OzCHI 2011 Design, Culture and Interaction*, 28 Nov.-02 Dec. 2011, Australian National University, Canberra, Australia, pp. 211-220. doi: 10.1145/2071536.2071570

Moffat, S.D., Hampson, E., and Hatzipantelis, M. (1998). Navigation in a "virtual" maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior*, Vol. 19, pp. 73-87. doi:10.1016/S1090-5138(97)00104-9

Montello, D.R., Hegarty, M., Richardson, A.E., and Waller, D. (2004). Spatial memory of real environments, virtual environments, and maps. In G. Allen (Ed.), *Human Spatial Memory: Remembering Where*. Mahwah: Lawrence Erlbaum Associates. pp. 251-285.

Morris, R.G.M. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *Journal of Neuroscience Methods*, 11, pp. 47-60. doi:10.1016/0165-0270(84)90007-4

Mäkelä, V., Heimonen, T., Luhtala, M., and Turunen, M. (2014). Information wall: Evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14)*. ACM, New York, NY, USA, pp. 228-231. doi:10.1145/2677972.2677998

Nicholson, J., and Baldwin, T. (2006). Interpretation of compound nominalisations using corpus and web statistics. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties* (pp. 54-61). Association for Computational Linguistics. ISBN:1-932432-84-1

Nothegger, C., Winter, S., and Raubal, M. (2004). Selection of salient features for route directions. *Spatial Cognition and Computation*, Vol. 4, No. 2, pp. 113–136. doi:10.1207/s15427633scc0402_1

Onoe, Y., Yamazawa, K., Takemura, H., and Yokoya, N. (1998). Telepresence by real-time view-dependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, Vol. 71, No. 2, pp. 154–165.

Padgitt, A.J., and Hund, A.M. (2012). How good are these directions? Determining direction quality and wayfinding efficiency. *Journal of Environmental Psychology*, Vol. 32, pp. 164–172.

Panitz, T. (1996). A Definition of Collaborative vs Cooperative Learning. *Deliberations*, London Metropolitan University; UK, Retrieved on 4.1.2018 from: <http://www.londonmet.ac.uk/deliberations/collaborative-learning/panitz-paper.cfm>

Pantelidis, V.S. (1993). Virtual reality in the classroom. *Educational Technology*, April, pp. 23–27.

Panula, H.S., and Elina Kouki (Hrsg.) (2014). *Ainedidaktisia tutkimuksia 7: Opettaminen valinkauhassa*. Ainedidaktinen symposiumi Turussa 15.3.2013. Turku: Suomen ainedidaktinen tutkimusseura. Available at: <https://helda.helsinki.fi/handle/10138/42530>

Pellegrino, J.W., and Kail, R. (1982). Process analyses of spatial aptitude. *Advances in the Psychology of Human Intelligence*, Vol. 1, pp. 311–365. doi:10.3758/BF03198314

Pellegrino, J.W., Golledge, R.G., Gale, N.D., and Ruggles, A. (1987). Effects of route complexity, age, and gender on route learning processes. Unpublished manuscript, Department of Geography, University of California, Santa Barbara.

Piaget, J., and Inhelder, B. (1967). *The Child's Conception of Space*. New York: Norton. ISBN:978-0393004083

Pihkala-Posti L. & Uusi-Mäkelä M. (2013). Kielenopetuksen tilat muutoksessa. (Die Sprachlernumgebungen in Veränderung) In: Eija Yli-Panula, Harry Silfverberg & Elina Kouki (Eds.) (2013). *Ainedidaktisia tutkimuksia 7. Opettaminen valinkauhassa*. Ainedidaktinen symposiumi Turussa 15.3.2013. Turku: Suomen ainedidaktinen tutkimusseura. <https://helda.helsinki.fi/handle/10138/42530>

Pihkala-Posti, L. (2014). Innovative training of oral communication: Berlin Kompass. In S. Jager, L. Bradley, E. J. Meima, and S. Thouésny (Eds),

CALL Design: Principles and Practice; *Proceedings of the 2014 EUROCALL Conference*, Groningen, The Netherlands. Dublin: Research-publishing.net. pp. 312-317. doi:10.14705/rpnet.2014.000237

Pihkala-Posti, L., Kallioniemi, P., Uusi-Mäkelä, M., Hietala, P., Hakulinen, J., Turunen, M., Okkonen, J., Kangas, S., Raisamo, R., and Keskinen, T. (2014). Collaborative learner autonomy and immersion in embodied virtual language learning environment. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014*. Chesapeake, VA: AACE. pp. 1313-1322. Available at: <https://www.learntechlib.org/p/147664/>

Price, M., and Anderson, P. (2007) The role of presence in virtual reality exposure therapy. *Journal of Anxiety Disorders*, Vol. 21, pp. 742-751. doi:10.1016/j.janxdis.2006.11.002

Prince, A., Brannick, M.T., Prince, C., and Salas, E. (1997). The measurement of team process behaviors in the cockpit: Lessons learned. In M.T. Brannick, E. Salas, and C. Prince (Eds.), *Series in Applied Psychology. Team Performance Assessment and Measurement: Theory, Methods, and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates. pp. 289-310.

Psotka, J., Davison, S., and Lewis, S.A. (1993). Exploring immersion in virtual space. *VR Systems*, Vols. 1, 2 (Fall 1993), pp. 70-82.

Ramalho, J., and Chambel, T. (2013). Windy sight surfers: Sensing and awareness of 360° immersive videos on the move. In *Proceedings of the 11th European Conference on Interactive TV and Video, EuroITV '13*, ACM (New York, NY, USA, 2013), pp. 107-116. doi:10.1145/2465958.2465969

Raubal, M., and Winter, S. (2002). Enriching wayfinding instructions with local landmarks. In: Egenhofer, M.J., Mark, D.M. (Eds.) *Geographic Information Science. Lecture Notes in Computer Science*, Vol. 2478. Springer, Berlin, pp. 243-259. doi:10.1007/3-540-45799-2_17

Richardson, A.E., Montello, D.R., and Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition*, Vol. 27, pp. 741-750. doi: 10.3758/BF03211566

Riecke, B.E., Van Veen, H.A.H.C., and Bühlhoff, H.H. (2002). Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence: Teleoperators and Virtual Environments*, Vol. 11, No. 5, pp. 443-473. doi:10.1162/105474602320935810

Riva, G., Botella, C., Légeron, P., and Optale, G. (2004). *Cybertherapy: Internet*

and Virtual Reality as Assessment and Rehabilitation Tools for Clinical Psychology and Neuroscience. Amsterdam: IOS Press. ISBN:1586034111

Rizzo, A.A., Ghahremani, K., Pryor, L., and Gardner, S. (2003). Immersive 360-degree panoramic video environments: Research on creating useful and usable applications. In *Human-Computer Interaction: Theory and Practice*, Vol. 1, J. Jacko and C. Stephanidis, Eds. Lawrence Erlbaum Associates: Mahwah, NJ, USA, pp. 1233–1237.

Ross, T., May, A., and Thompson, S. (2004) The use of landmarks in pedestrian navigation instructions and the effects of context. In: Brewster, S., and Dunlop M. (Eds.) *Mobile Human-Computer Interaction – MobileHCI 2004*. Lecture Notes in Computer Science, Vol. 3160. Springer, Berlin, Heidelberg. pp. 300–304. doi:10.1007/978-3-540-28637-0_26

Rovelo Ruiz, G.A., Vanacken, D., Luyten, K., Abad, F., and Camahort, E. (2014). Multi-viewer gesture-based interaction for omnidirectional video. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*, ACM (New York, NY, USA, 2014), pp. 4077–4086. doi:10.1145/2556288.2557113

Ruddle, R.P., Payne, S.J., and Jones, D.M. (1997). Navigating buildings in “desk-top” virtual environments: Experimental investigations using extended navigational experience. *Journal of Experimental Psychology: Applied*, Vol. 3, No. 2. pp. 143–159. doi:10.1037/1076-898X.3.2.143

Saarinen, S., Mäkelä, V., Kallioniemi, P., Hakulinen, J., and Turunen, M. (2017). guidelines for designing interactive omnidirectional video applications In: Bernhaupt, R., Dalvi, G., Joshi, A. K., Balkrishan, D., O’Neill, J., Winckler, M. (Eds.) *Human-Computer Interaction – INTERACT 2017*. INTERACT 2017. Champagne: Springer, pp. 263–272. doi:10.1007/978-3-319-68059-0_17

Schrader, C., and Bastiaens, T. (2012). Learning in educational computer games for novices: The impact of support provision types on virtual presence, cognitive load, and learning outcomes. *The International Review of Research in Open and Distributed Learning*, Vol. 13, No. 3, pp. 206–227. doi:10.19173/irrodl.v13i3.1166

Schroeder, R. (2002). Copresence and interaction in virtual environments: An overview of the range of issues. In *Proceedings of the 5th Annual International Workshop on Presence (PRESENCE’02)*, pp. 274–295. doi:10.1007/978-1-4471-0277-9_1

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual*

Environments, Vol. 10, No. 3 (June 2001), pp. 266–281.
doi:10.1162/105474601300343603

Sester, M., and Dalyot, S. (2015). Enriching navigation instructions to support the formation of mental maps. In: Harvey F., Leung Y. (Eds.) *Advances in Spatial Data Handling and Analysis. Advances in Geographic Information Science*, pp. 15–33. doi:10.1007/978-3-319-19950-4_2

Shanon, B. (1984). Room descriptions. *Discourse Processes*, Vol. 7, pp. 225–255. doi:10.1080/01638538409544591

Sharma, G., Kaushal, Y., Chandra, S., Singh, V., Mittal, A.P., and Dutt, V. (2017). Influence of landmarks on wayfinding and brain connectivity in immersive virtual reality environment. In *Front Psychology*, Vol. 8, p. 1220. doi:10.3389/fpsyg.2017.01220

Sholl M.J. (1996) From visual information to cognitive maps. In: Portugali J. (Eds.) *The construction of cognitive maps. GeoJournal Library*, Vol. 32. Springer, Dordrecht, pp. 157–186. doi:10.1007/978-0-585-33485-1_8

Short, J., Williams, E., and Christie, B. (1976). *The Social Psychology of Telecommunications*. John Wiley & Sons, Ltd., Hoboken, NJ. ISBN:0608176761

Siegel, A., and White, S. (1975). The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior* Vol. 10, pp. 9–55. doi:10.1016/S0065-2407(08)60007-5

Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., and Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex differences. In *Evolution and Human Behavior*, Vol. 21, No. 3, pp. 201–213. doi:10.1016/S1090-5138(00)00036-2

Simons, D.J., and Wang, R.F., (1998). Perceiving real-world viewpoint changes. *Psychological Science*, Vol. 9, No. 315–320. doi:10.1111/1467-9280.00062

Singhal, S., and Neustaedter, C. (2017). BeWithMe: An immersive telepresence system for distance separated couples. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17 Companion)*. ACM, New York, NY, USA, pp. 307–310. doi:10.1145/3022198.3026310

Skarbez, R. (2016). Plausibility Illusion in Virtual Environments. Ph.D. Dissertation. The University of North Carolina at Chapel Hill. ProQuest:10145960

Skarbez, R., Brooks, F.P. Jr., and Whitton, M.C. (2017). A survey of presence and related concepts. *ACM Computing Surveys (CSUR)*, Vol. 50, No. 6, Article 96 (November 2017), 39 pages. doi:10.1145/3134301

Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 364 (2009), pp. 3549–3557. doi:10.1098/rstb.2009.0138

Slater, M., Alberto, C., and Usoh, M. (1995). In the building or through the window? An experimental comparison of immersive and non-immersive walkthroughs, *Virtual Reality Environments in Architecture*, Leeds 2–3rd November 1994, Computer Graphics Society.

Slater, M., Spanlang, B., Sanchez-Vives, M.V., and Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLOS ONE* Vol. 5, No. 5 (May 2010), pp. 1–9. doi:10.1371/journal.pone.0010564

Sorrows, M.E., and Hirtle, S.C. (1999). The nature of landmarks for real and electronic spaces. In: Freksa, C., Mark, D.M. (eds) *Spatial information theory. Cognitive and Computational Foundations of Geographic Information Science. COSIT 1999. Lecture Notes in Computer Science*, Vol. 1 1661: pp. 37–50. Springer: Berlin, Heidelberg. doi:10.1007/3-540-48384-5_3

Stout, R.J., Cannon-Bowers, J., Salas, E.M., and Milanovich, D. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors*, Vol. 41, pp. 61–71. doi:10.1518/001872099779577273

Sutherland, I.E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, part I (AFIPS '68 (Fall, part I))*. ACM, New York, NY, USA, pp. 757–764. doi:10.1145/1476589.1476686

Takatalo, J. (2002). Presence and Flow in Virtual Environments: An Explorative Study. Master's thesis. University of Helsinki.

Tang, A., and Fakourfar, O. (2017). Watching 360° Videos Together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4501–4506. doi:10.1145/3025453.3025519

Tannenbaum, S.I., Smith-Jentsch, K.A., and Behson, S.J. (1998). Training team leaders to facilitate team learning and performance. In J.A. Cannon-Bowers and E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training*. Washington, DC: American Psychological Association. pp. 247–270.

- Taylor, H.A., and Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language*, Vol. 35, No. 3, pp. 371-391. doi:10.1006/jmla.1996.0021
- Thompson, S.K. (1975). Gender labels and early sex role development. *Child Development*, Vol. 46, No. 2 (June, 1975), pp. 339-347. doi:10.2307/1128126
- Thorndyke, P.W. (1981). Distance estimation from cognitive maps. *Cognitive Psychology*, Vol. 13, No. 4, pp. 526-550. doi:10.1016/0010-0285(81)90019-0
- Thorndyke, P.W., and Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, Vol. 14, pp. 560-589. doi:10.1016/0010-0285(82)90019-6
- Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychological Review*, Vol. 55, No. 4, pp. 189-208. doi:10.1037/h0061626
- Tom, A., and Denis, M. (2003). Referring to landmark or street information in route directions: What difference does it make? In Kuhn, W., Worboys, M., Timpf, S. (Eds.) *COSIT 2003*, LNCS 2825. Springer, Berlin, pp. 362-374. doi:10.1007/978-3-540-39923-0_24
- Tom, A.C., and Tversky, B. (2012). Remembering routes: Streets and landmarks. *Applied Cognitive Psychology*, Vol. 26, pp. 182-193. doi:10.1002/acp.1805
- Trindade, J., Fiolhais, C., and Almeida, L. (2002). Science learning in virtual environments: A descriptive study. *British Journal of Educational Technology*, Vol. 33, pp. 471-488. doi:10.1111/1467-8535.00283
- Voyer, D., Voyer, S., and Bryden, P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis consideration of critical variables. *Psychological Bulletin*, Vol. 117, pp. 250-270. doi:10.1037/0033-2909.117.2.250
- Vandenberg, S.G., and Kuse, A.R. (1978). Mental rotations, a group test of threedimensional spatial visualization. *Perceptual and Motor Skills*, Vol. 47, pp. 599-604. doi:10.2466/pms.1978.47.2.599
- Vinson, N.G. (1999). Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, pp. 278-285. doi:10.1145/302979.303062
- Virtual Reality Society. (2017). What is Virtual Reality? Retrieved on 10.10.2017 from <https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html>

- Walkowiak, S., Foulsham, T., and Eardley, A.F. (2015). Individual differences and personality correlates of navigational performance in the virtual route learning task. *Computers in Human Behavior*, Vol. 45, pp. 402-410. doi:10.1016/j.chb.2014.12.041
- Waller, D. (2000). Individual differences in spatial learning from computersimulated environments. *Journal of Experimental Applied Psychology*, Vol. 6, No. 4, pp. 307-321. doi:10.1037/1076-898X.6.4.307
- Waller, G. (1985). Linear organization of spatial instructions: Development of comprehension and production. *First Language*, Vol. 6, No. 16: pp. 53-67. doi:10.1177/014272378600601605
- Waller, D. (2005). The WALKABOUT: Using virtual environments to assess large-scale spatial abilities, In *Computers in Human Behavior*, Vol. 21, No. 2, pp. 243-253, ISSN 0747-5632, doi:10.1016/j.chb.2004.02.022
- Waller, D., Hunt, E., and Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperations and Virtual Environments*, Vol. 7, No. 2, pp. 129-143.
- Waller, D., Loomis, J.M., and Haun D.B.M. (2004). Body-based senses enhance knowledge of directions in large-scale environments. *D.B.M. Psychonomic Bulletin & Review* (2004) Vol. 11. pp. 157-163. doi:10.3758/BF03206476
- Wang, R.F., and Simons, D.J. (1999). Active and passive scene recognition across views. *Cognition*, Vol. 70, pp. 191-210.
- Wann, J., and Mon-Williams, M. (1996). What does virtual reality NEED?: Human factors issues in the design of three-dimensional computer environments. *International Journal of Human-Computer Studies*, Vol. 44, pp. 829-847. 10.1006/ijhc.1996.0035
- Ward, S.L., Newcombe, N., and Overton, W.F. (1986). Turn left at the church, or three miles North. A study of direction giving and sex differences. *Environment and Behavior*, Vol. 18, No. 2, pp. 192-213. doi:10.1177/0013916586182003
- Welch, R.B. (1999). How can we determine if the sense of presence affects task performance? *Presence: Teleoperators and Virtual Environments*. Vol. 8, No. 5, pp. 574-577. doi:10.1162/105474699566387
- Wen, W., Ishikawa, T., and Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. *Applied Cognitive Psychology*, Vol. 25, pp. 654-662. doi:10.1002/acp.1737

Whitaker, L.A. (1996). Getting around in the natural world. *Ergonomics in Design*, Vol. 4, No. 3, pp. 11-15.

Whitaker, L.A., and Cuqlock-Knopp, G. (1992). Navigation in off-road environments: Orienteering interviews. *Scientific Journal of Orienteering*, Vol. 8, pp. 55-71.

Wikipedia. (2017). Immersion (Virtual Reality)—Wikipedia, The Free Encyclopedia. Retrieved on 29.12.2017 from [https://en.wikipedia.org/w/index.php?title=Immersion_\(virtual_reality\)&oldid=769910362](https://en.wikipedia.org/w/index.php?title=Immersion_(virtual_reality)&oldid=769910362)

Winn, W., and Windschitl, M. (2000). Learning Science in Virtual Environments: The Interplay of Theory and Experience. *Themes in Education*, Vol. 1, No. 4, pp. 373-389.

Winter, S. (2003). Route adaptive selection of salient features. In *International Conference on Spatial Information Theory*, Vol. 2825, pp. 349-361. Springer: Berlin, Heidelberg. doi:10.1007/978-3-540-39923-0_23

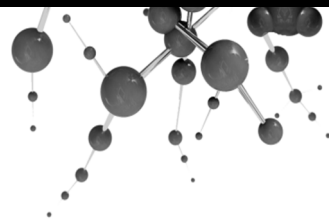
Witmer, B.G., and Kline, P.B. (1998). Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators and Virtual Environments*, Vol. 7, pp. 144-167. doi:10.1162/105474698565640

Witmer, B.G., and Singer, M.J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, Vol. 7, NO. 3 (June 1998), pp. 225-240. doi:10.1162/105474698565686

Witmer, B.G., Bailey, J.H., Knerr, B.W., and Parsons, K.C. (1996). Virtual spaces and real world places: Transfer of route knowledge. *International Journal of Human-Computer Studies*, Vol. 45, No. 4 (October 1996), pp. 413-428.

Zhao, S. (2003). Toward a taxonomy of copresence. *Presence: Teleoperators and Virtual Environments*, Vol. 12, No. 5 (October 2003), pp. 445-455. doi:10.1162/105474603322761261

Zoric, G., Barkhuus, L., Engström, A., and Önnvall, E. (2013). Panoramic video: Design challenges and implications for content interaction. *Proceedings of the 11th European Conference on Interactive TV and Video*, ACM, pp. 153-162. doi:10.1145/2465958.2465959



Publication I

Kallioniemi, P., and Markku Turunen. (2012). Model for landmark highlighting in mobile web services. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*. ACM, New York, NY, USA, Article 25, 10 pages. doi:10.1145/2406367.2406398

© ACM, 2012. Reprinted with permission.

Model for Landmark Highlighting in Mobile Web Services

Pekka Kallioniemi
University of Tampere
Kalevantie 4
33014 Tampere
pekka.kallioniemi@uta.fi

Markku Turunen
University of Tampere
Kalevantie 4
33014 Tampere
markku.turunen@sis.uta.fi

ABSTRACT

We introduce a model for landmark highlighting for pedestrian route guidance services for mobile devices. The model determines which landmarks are the most attractive based on their properties in the current context of user's orientation and the location on the route and highlights these landmarks on the mobile map. The attractiveness of a landmark is based on its visual, structural and semantic properties which are used for calculating the total attractiveness of a single landmark.

This model was evaluated with voluntary users conducted in laboratory environment. Test subjects were shown images of street intersections from where they selected the most attractive and prominent landmarks in the route's context. We then compared these results with the landmarks selected by the model. The results show that landmarks highlighted by the model were the same ones that were selected by the participants as most salient landmarks.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: *web-based services*

General Terms

Design

Keywords

Mobile web services, landmarks, pedestrian route guidance

1. INTRODUCTION

Imagine that you have arrived to a foreign city and need quickly navigate to an important meeting that you are having pretty soon. This is your first time visiting the city, so you do not know the way and need good route instructions. Naturally, you start to use your new pedestrian route guidance service on your mobile device, but you quickly notice that you need to translate its map-based information to the actual physical environment. You can ask directions from locals passing by, and they usually will give instructions which contain information such as names of buildings and other important landmarks. You realize that your new fancy mobile application does not really support landmark-based navigation, which is natural and efficient in real world.

One of the challenges is that existing route guidance services are usually designed for car navigation and pedestrian route guidance services use the same information, which is usually just street names and distance measures. But because pedestrians are less restricted and constrained in their movements while navigating,

they should be offered richer and more meaningful route guidance. Many studies [1][2][3] have shown that landmark-based navigation is an intuitive and effective form of direction giving and works particularly well in pedestrian navigation.

In contrast, modern web mapping services rarely make any reference to landmarks. One reason for this is that there are no data available about landmarks or even agreement on what is considered a landmark. Many countries and cities have their own data for buildings but these are not in any standardized form and the actual information of buildings can be limited.

This paper presents a model for incorporating landmarks and their visibility into pedestrian route guidance systems. The model is based on earlier model by Raubal and Winter [4] and Winter [5]. Their work introduced a grading system for landmarks based on their visual, semantic and structural properties. From these properties one can determine total saliency score of a landmark. This score can be used for highlighting landmarks on route guidance systems such as Google Maps.

To evaluate the solution, we conducted a user study with 20 participants. In this experiment we displayed images of street crossings which were decision points in a route. From these images the users selected the landmarks that they thought were the most salient in the route context. Then we compared these results to the ones that the model had selected. Our results show that the landmarks highlighted by the model were the same ones that were selected by the participants as the most salient landmarks. These results indicate that there is need for further studies about benefits of highlighting salient landmarks in the route guidance context. Current model can be used as a basis for these studies and it can also be developed further to support different data sources as properties for the landmark grading.

1.1 Previous and Related Work

1.1.1 Landmarks

Landmarks are defined as prominent features in the environment that are unique or contrast with their neighboring objects. They characterize a geographic location and structure routes by forming points to move to or away from. An informal but influential characterization of the term 'landmark' was done by Lynch in his book *The Image of the City* [7]. In this book landmarks are defined as "points of reference which are external to the observer, and sees them as simple physical elements which may vary in scale". Lynch also notes, that the key characteristic of a landmark is its singularity, in other words some aspect of the landmark is unique and/or memorable.

As was mentioned before, Raubal and Winter created a grading model for landmark saliency [4]. The model is based on landmark characteristics according to Sorrow and Hirtle [6]. They identified three basic types of landmarks: visual, semantic and structural landmarks. These types are not mutually exclusive, and one landmark can be, for example, eye-catching (visual landmark) and culturally important (semantic landmark). Raubal and Winter suggested that visual properties of landmark could be measured

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MUM '12, December 4-6, 2012, Ulm, Germany.

Copyright 2010 ACM 1-58113-000-0/00/0010...\$10.00.

by façade area, shape, colour and visibility, semantic properties could be measured by cultural and historical importance, and structural attraction could be measured by surrounding nodes (roads, highways, etc.) and boundaries of the landmark.

1.1.2 Landmarks in route instructions

Previous studies [8, 9] have shown that routes enriched with landmarks lead to a better guidance and higher confidence than routes primarily based on street names. Study by Tom and Denis [9] compared these two routing methods, and the results showed that landmark-based route guidance took less time to process than street names and that the participants could recall landmarks more easily. Ross et al. [10] conducted a study which showed that the addition of landmarks to textual directions improved pedestrian navigation and increased the user's confidence. May et al. [11] even argued that landmarks should be used as the primary means of providing directions.

Lovelace [12] conducted a study about location of landmarks as navigational aids and found out that landmarks located near to decision points and route marks are most commonly used when navigating through unfamiliar environments, and that distant landmarks were used more sparsely. It has been observed that landmarks within urban environment are not evenly distributed along the route, but instead they are often in locations where re-orientation or decision making is required. They also tend to be in locations where re-orientation is not mandatory but could reoccur, e.g. places where several possible directions could be followed [13]. The inclusion of landmarks within route descriptions improves the changes of success in navigation and reduces the likelihood of getting lost.

Recent research has also been concentrating on how to detect landmarks from available datasets and how these landmarks could be integrated into current route guidance instructions [17, 18]. Elias [19] took a different approach by investigating whether it is possible to extract the most salient landmarks from spatial databases using data mining methods. This research focused on an approach to select landmarks automatically from a building database with attributes for each building including size, height and distance to road. Most salient landmarks were then selected using hierarchical clustering. Clustering results similar objects being grouped together and the unique and salient landmarks to stand out, therefore becoming potential landmarks.

Another popular approach that has been taken by many authors has been to study the ways in which the Internet could be used as a data source for landmarks. Tezuka and Tanaka [20] and Furlan et al. [21] have studied web mining as a possible method for acquiring knowledge about landmarks with good results.

2. A MODEL FOR LANDMARK HIGHLIGHTING

2.1.1 Original model

The existing work by Winter and Raubal was selected as a basis for the landmark highlighting model since it is the most complement work in the area, being developed and tested over time [14, 15, 16]. Most importantly, it contains several elements that can be integrated into existing mobile map services such as Google Maps and Bing Maps. However, the original model was created for textual directions, and based on the assumption that there are databases, which provide data about buildings. In order to implement a real world service to be used with applications such as Google Maps the model required numerous modifications.

The original model divides the total attraction of landmark into three different categories: visual, semantic and structural attraction. There are formulas for calculating attraction value for each of these categories and from these values we can determine the total attractiveness value of each landmark.

Visual attraction - Landmarks are visually attractive if they have some certain characteristics, such as sharp contrast with their surroundings or a prominent location [4]. The original model used four measures for calculating the visual attractiveness of a landmark: façade area, shape, color and visibility. For individual properties and measurements of visual attraction, see Table 1.

Table 1. Properties for calculating visual attraction of a landmark.

Property	Example	Measurement
Façade area	$\alpha = 25m * 15m = 375m^2$	$\alpha = \int x x \in \text{façade}$
Shape	<ul style="list-style-type: none"> Shape factor $\beta_1 = 15m / 25m = 0.6$ Shape deviation from rectangle $\beta_2 = (375m^2 - 295m^2) / 375m^2 = 21\%$ 	$\beta_1 = \text{height} / \text{width}$ $\beta_2 = (\text{area of minimum bounding rectangle} - \alpha) / \text{area of minimum bounding rectangle}$
Color	$\gamma = [255, 0, 0] = \text{red}$	$\gamma = [R, G, B]$

Semantic attraction – Semantic measures for the original model of landmark saliency comprise cultural and historical importance of the landmark. It also takes into account any explicit marks, such as signs or boards on the building façade. In this model, cultural and historical importance is determined with Boolean value and again refined with predefined scale from 1 to 5. Explicit marks are only assigned with Boolean values. See Table 2 for individual properties and measurements of semantic attraction.

Table 2. Properties for calculating visual attraction of a landmark.

Property	Example	Measurement
Cultural and historical importance	$\epsilon = T$ $\epsilon = 1$ (building very famous for its architecture)	$\epsilon \in \{T, F\}$ $\epsilon \in \{1,2,3,4,5\}$ Scale of importance: 1 (high) – 5 (low)
Explicit marks	$\zeta = T$ Sign in front of the building	$\zeta = \{T, F\}$

Structural landmarks – A landmark is structurally attractive if it has a major role or prominent location in the structure of the spatial environment [4]. Examples of a structurally attractive landmark could be a big plaza or a road intersection. Therefore, if a landmark is located in an intersection with many nodes in the vicinity, it is considered to be structurally attractive. Winter and Raubal hypothesized that a boundary is the more prominent the larger its resistance is. For example, canals and railroads usually have only few crossovers (bridges, tunnels, etc.), so such barriers form significant shapes in city maps. For individual properties and measurements of structural landmarks, see Table 3.

Table 4. Calculating the total attraction value for landmarks

Measure	Property	Value	Significance (property)	Significance (Measure)	Weight	Weighted significance	Total
Visual attraction	α	...	S_α	$S_{vis} = (S_\alpha + S_{\beta1} + S_{\beta2} + S_\gamma + S_\delta) / 5$	W_{vis}	$S_{vis} * W_{vis}$	$S_{vis} * W_{vis} + S_{sem} * W_{sem} + S_{str} * W_{str}$
	$\beta1$...	$S_{\beta1}$				
	$\beta2$...	$S_{\beta2}$				
	γ	...	S_γ				
Semantic attraction	δ	...	S_δ	$S_{sem} = (S_\epsilon + S_\zeta) / 2$	W_{sem}	$S_{sem} * W_{sem}$	
	ϵ	...	S_ϵ				
Structural attraction	ζ	...	S_ζ	$S_{str} = (S_\eta + S_\theta) / 2$	W_{str}	$S_{str} * W_{str}$	
	η	...	S_η				
	θ	...	S_θ				

Table 3. Properties for calculating visual attraction of a landmark.

Property	Example	Measurement
Nodes	$\eta = (4*2+4*2)$	$\eta = (i+o)$ Weighted incoming (i) and outgoing (o) to and from a node
Boundaries	$\theta = 2500$	$\theta = \text{cell size} * \text{form factor}$ Form factor = long side / short side

Once calculations for visual, semantic and structural attraction have been calculated, these can be combined for total attraction value of a single landmark. These measures can also be given a predefined weight value, which allows for an adaptation of the context (e.g. mode of travel) or individual user preferences. See Table 4. for the measurement of total attraction for a single landmark.

Later on Winter also introduced a formal definition for advanced visibility [14] for the model. In this additional feature, it was acknowledged that that a façade that is hardly visible is not useful for navigational purposes. Advanced visibility requires an external, 2-dimensional dataset for streets and building models.

2.1.2 New model

The original model is not feasible for mobile map services context for two reasons: First, there is no global dataset with all properties required for measurements available. For example, it is very difficult to find a dataset which contains façade areas, cultural or historical importance values or explicit marks for landmarks. And second, the original model was intended to be used manually and lacks the automatic processes for attractiveness calculation. Therefore the original model needs to be adjusted for this context,

so that it can be implemented as a service for pedestrian navigation.

To calculate any landmark attractiveness scores, we would need a dataset with properties and measurements. Currently there is no any single dataset with all required data. Instead, we need to use multiple data sources and combine them together. Many countries have their own datasets where they keep track of building locations, information about when they were built, number of storeys, etc. Often these datasets also contain street plan which has all roads, highways, railways, etc. This data can be used for measuring visual and structural attractiveness of a landmark.

Using the street plan data we can extract the surrounding nodes of a landmark. Most of the time there is no distinction between small roads and highways in the dataset, so we just have to rate all nodes with same weighting score. Instead of measuring boundaries and resistance of landmarks, we measured if the landmark has exceptional architecture. For this, we calculated the coordination points of landmark boundaries – more points the landmark has, more complex its architecture is. For example, a building with four points (rectangle) is considered a normal landmark with no exceptional architecture, but building with more points is more complex and stands out from other buildings architecture-wise. One exception in this measurement is buildings with three coordination points – they are always considered having an exceptional architecture.



Figure 1. Geotagged images of Notre Dame in Paris.

For semantic attraction, we can use another external data source for geotagged images. For example, we can safely assume that Notre Dame in Paris has much more geotagged images (see Figure 1) around it than any normal building of residence in the suburbs of the same city. We can take the area of the building (and some of the surroundings, since the images are often tagged outside the actual landmark) and count the geotagged images inside it for rough measurement of cultural and historical importance of a landmark (Table 5).

Table 5. Cultural and historical importance of a landmark is determined by the amount of geotagged images in its area

# of images	Cultural and historical importance
0	1
1-10	2
11-30	3
31-60	4
over 60	5

Kolbe [15, 16] computed the salience of building facades using information theoretic measures. This approach determines the peculiarity of surprise in visual characteristics of facades. The entropy of the façade is 0 if its probability is 1 (thus, if it looks like all the other facades in the neighborhood), and it becomes larger the smaller the probability of its occurrence in the surrounding area is. The façade with the largest entropy in the area is then identified as a landmark. For example, old wooden church would be visually striking if it were surrounded by concrete office buildings. We can use this approach in addition to the previous measurements for determining if the landmark stands out from its surroundings.

We also created a model for detecting visibility in 3D space for pedestrian navigation, but this model was not implemented in the final application because of lack of 3D material required for the algorithm. In this model we divide visibility into two properties: visibility area and visibility from the viewpoint. Visibility area is the percentage of the area that is visible to the navigator. Visibility can be partly or fully concealed by another landmark. If landmark is only partially visible, there is no reason to highlight it on the map as a salient landmark. Visibility from the viewpoint is a Boolean value and is true if the visibility area is greater than 20 % and false in other cases.

Visibility area can be calculated using sweeping algorithm by Goldman [16] where we determine if two line segments, s_1 and s_2 intersect, and if they do, we also have to determine the point of intersection. Line segments are determined in the following way:

$s_1 = p + t \cdot r$ where $0 \leq t \leq 1$ and p, r are 2-dimensional vectors and

$s_2 = q + u \cdot s$ where $0 \leq u \leq 1$ and q, s are 2-dimensional vectors.

Next we compute the cross products to determine intersections. The 2-dimensional cross product of $p_1 \times p_2$ is given by:

$$\det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = x_1 y_2 - x_2 y_1 = -p_2 \times p_1$$

If the cross product is zero, the two vectors are collinear and are pointing either in the same or the opposite direction. The lines intersect when $p + t \cdot r = q + u \cdot s$ and by crossing both sides with s we get:

$$(p + t \cdot r) \times s = (q + u \cdot s) \times s = q \times s + u \cdot s \times s = q \times s.$$

From this we solve for t obtaining

$$t = \frac{(q - p) \times s}{r \times s}$$

and u :

$$u = \frac{(q - p) \times r}{r \times s}$$

If both t and u get a value between 0 and 1, the two line segments intersect and the intersection point is given by $p + t \cdot r$. If either t or u is not between 0 and 1, the line segments do not intersect. In cases where $r \times s = 0$ we cannot solve t or u which means the two line segments are parallel. If $(p - q) \times r \neq 0$, the segments are collinear and we have to project them onto the x-axis and determine if their projections intersect.

2.1.3 Implementation of the model

For the evaluation of the new model, JavaScript based application was created. This application uses Google Maps JavaScript API V3. Google Maps offers good tools for creating custom maps, and it is also least vulnerable to runtime errors and exceptions [17]. For our user experiment, the application utilized landmark data fetched from Helsinki Real Estate Department and it contains about 1 square kilometer area of downtown Helsinki with a center point at the crossing of Aurorankatu and Nervanderinkatu. Data contains 2-dimensional building models and 2-dimensional street model from this area. It also contains some basic information (number of stories, building material, etc.) about the buildings in MapInfo format.

This information was transformed into KML (Keyhole Markup Language) files which are supported by the Google Maps API. KML is XML notation for geographic annotation and it can also be used for visualization of data in Google Maps. KML data is fetched using server-side PHP which shortens the loading times inside the application. Panoramio API was used for fetching data about surrounding geotagged images, which were then used for calculating the historical and cultural attraction of each landmark.

Application itself was implemented for Android 2.2 platform, so it works in every Android device released during and after 2010. The overall architecture of the application is illustrated in Figure 2.

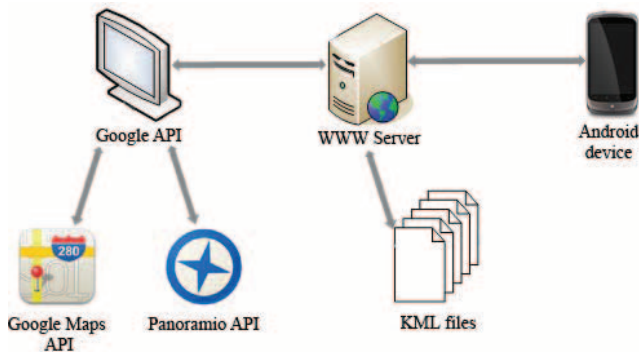


Figure 2. Application architecture.

2.1.4 Application functionality

The application is initialized with predefined parameters and routes. After this it calculates the orientation of the user and selects all landmarks in 200 meter semicircle in front of the user (see Figure 3.). After we have calculated the coordinates for the semicircle, we fetch the landmark data from KML file and detect which landmarks have coordinates inside it. Orientation for the semicircle is calculated using the location of the user and starting and ending point of the current leg on the route. This way we do not have to make any connection with the device's compass.

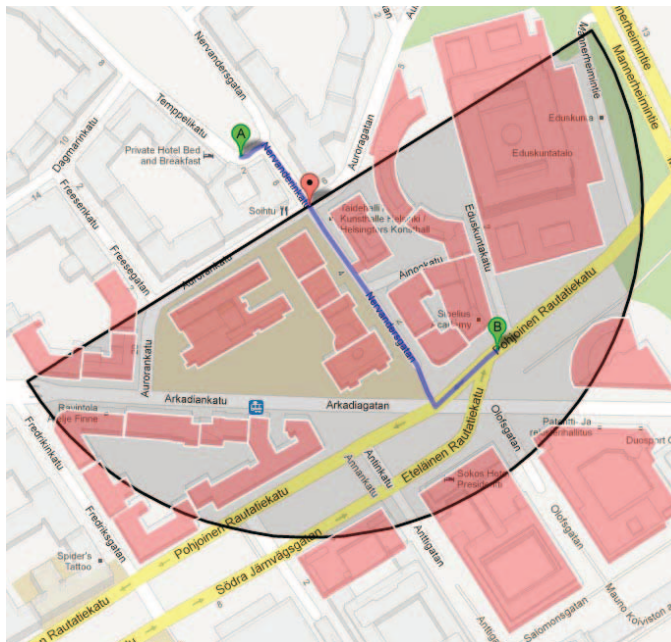


Figure 3. Landmarks in the area of sight radius are selected. Sight radius can be seen as a black semicircle.

After all data for all landmarks in sight radius have been retrieved, the application calculates which landmarks are actually visible to the user from the viewpoint (Figure 4.). At the present stage the application calculates the visibility only in 2-dimensional space and the model does not take into account the altitude of the landmarks or landscape.

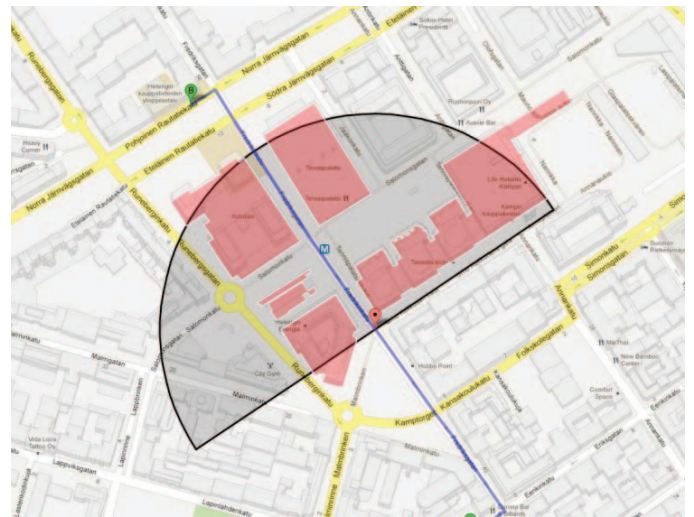


Figure 4. Only the landmarks that are visible to user are highlighted.

When all visible landmarks have been filtered, the application performs saliency comparison and selects the most attractive landmarks which are then highlighted on the map. Landmark selection is carried out by using data from two different datasets. For the data we have for each landmark, see Table 6. After the comparison of landmark properties has been done, we can select which landmarks we highlight on the map (see Figure 5.).

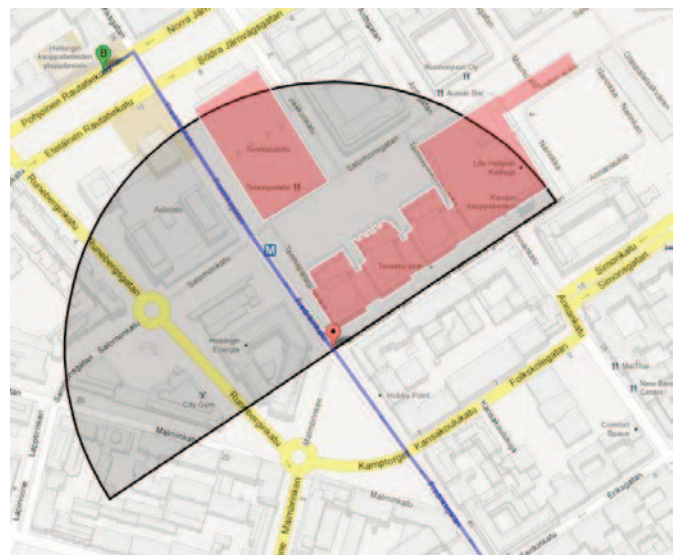


Figure 5. Highlighted landmarks along the route. Current route is presented on the map as a blue line and visibility area as a darkened semicircle.

Table 6. KML Data retrieved for a movie theater in Helsinki. Currently this data is combined from two different datasets.

Property	Example	Dataset
Coordinates for the building polygon	24.930612,60.169059,0 24.929658,60.16982,0 24.930525,60.170089,0 24.931482,60.169329,0 24.930612,60.169059,0	Helsinki Real Estate Department
Building year	1937	Helsinki Real Estate Department
Façade material	Stone	Helsinki Real Estate Department
# of storeys	8	Helsinki Real Estate Department
Geotagged images	4	Panoramio API
Use	Movie theater	Helsinki Real Estate Department
Nodes	4	Helsinki Real Estate Department
Distance from user	120 meters	Google Maps API

3. USER EXPERIMENT

To investigate whether the model selects landmarks similar to humans, we conducted an evaluation that focused on people’s landmark selection in decision points on the route. We displayed participants 180° images of two intersections from downtown Helsinki in laboratory environment. Current route was presented on the images as a transparent blue line. This was done to emphasize the route context in the selection of the landmarks. Participants then chose 3-5 landmarks that they considered to be the most salient. They were also handed a mobile device with a route and current location of the user. After this, they scored these landmarks in the scale of 1-5 from most attractive (5) to least attractive (1) in the context of the route. It was emphasized that temporary structures and buildings (such as construction sites and vehicles) could not be selected as landmarks.

In addition, we performed some questions about why the participants selected these particular landmarks. After this we compared the study results to the landmarks selected by the model to see if the model corresponds with human-selected landmarks. This evaluation method was selected because it gives us information about how humans select landmarks in decision points. The model can be developed further based on this information to correspond the human landmark selections i.e. make the model’s landmark selection more humanlike.

Participation in the evaluation was voluntary, anonymous and unpaid. Participants filled out a digital form during the test to avoid rejection of results due to bad handwriting. The evaluation was conducted in a darkened laboratory where the images of the intersections were projected on the wall. To avoid the use of

complex street names, intersects were named simply Intersect 1 and Intersect 2. A blue line was also added to the images to depict the route shown in the mobile device.

3.1 Intersects

Next we introduce the intersections that were used for this evaluation. For this study only two intersections in downtown Helsinki were used.

Intersect 1 – Intersect 1 was located on the route Temppliaukion Kirkko – Sibelius Academy (Figure 6.). It is located on very central location in the city center. This intersect contains buildings from different centuries and architectural styles. One very famous building, the Finnish parliament, is very close to the intersection but because it is not located on the route or any of the surrounding nodes and it is not clearly visible, it is not included in the calculation. This intersect can be seen in Figure 6.

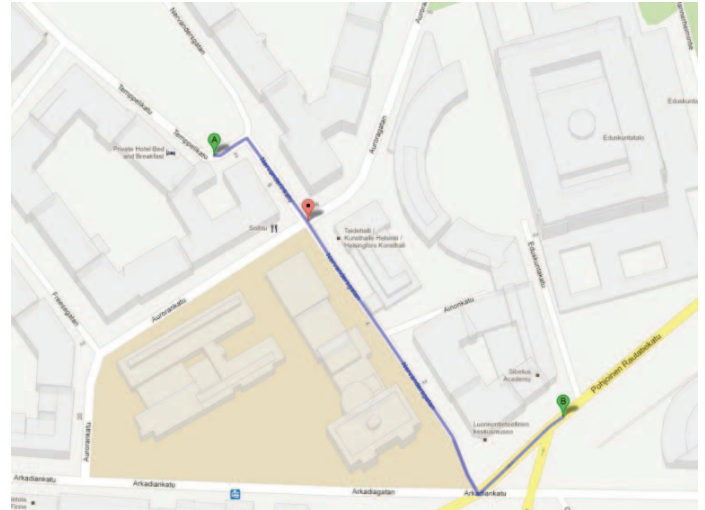


Figure 6. Route on Intersect 1.



Figure 7. 180 degree image of Intersect 1.

Intersect 2 - This intersect is located on the route movie theater Orion – Guild house of Helsinki School of Economics. In this intersection there are fewer landmarks to be used and they are more generic. This location was selected because there is no single landmark that is distinctive and prominent compared to other landmarks in the vicinity. These kind of intersects are the kind where people usually take wrong directions, get lost, etc. and they are often found on the smaller side streets in big cities. To see the route on the map, see Figure 8. This intersect can be seen in Figure 9.

Table 7. Results of the evaluation and comparison with human subjects and results from the model

Intersect 1

Landmark	Attractiveness					Score from evaluation	Score from model	Grade in evaluation	Grade in model
	1	2	3	4	5				
1	5	4	1	2	4	2,6	2,5	4	6
2	5	5	6	2	1	3,4	4,33	1	1
3	0	1	4	3	2	1,2	2,33	5	7
4	7	3	5	1	1	3,2	4	2	3
5	3	7	2	4	1	2,9	4,33	3	1
6	0	0	2	2	3	0,65	3,33	6	4
7	0	0	0	0	1	0,05	2,67	7	5

Intersect 2

Landmark	Attractiveness					Score from evaluation	Score from model	Grade in evaluation	Grade in model
	1	2	3	4	5				
1	1	5	8	4	0	3,25	4,17	3	3
2	0	3	1	3	8	1,45	2,33	5	5
3	2	0	3	5	2	1,55	4	4	4
4	13	3	2	2	0	4,35	4,5	1	2
5	4	9	6	0	1	3,75	4,83	2	1

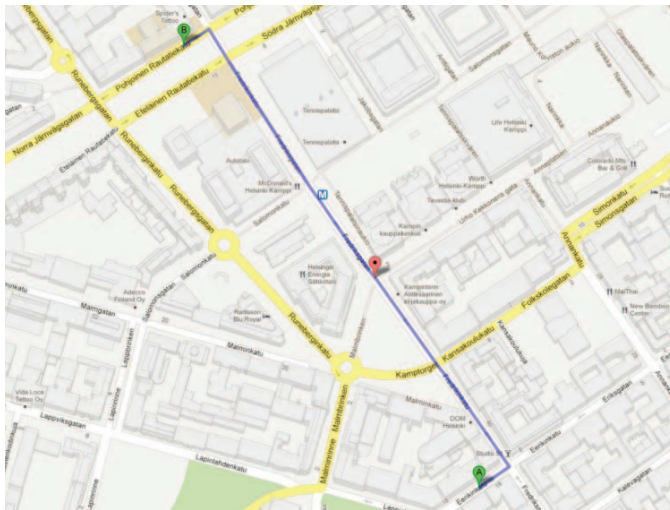


Figure 8. Route on Intersect 2.



Figure 9. 180 degree image of Intersect 2.

3.2 Results

The main question of interest in this evaluation was if the model can make similar selections of landmarks with humans. In this section we show the results of the evaluation and compare the differences between landmarks selected by humans and the ones selected by the model. In total we had 20 participants (13 male, 7 female).

We had a lot of additional information from the freeform questions of the evaluation. The most distinctive features of landmarks were its shape, location on the route (landmarks next to the route were selected more often) and semantic meaning. Intersect 1 was fairly unknown for the participants, whereas Intersect 2 was more commonly known because of some of the quite famous buildings on it. 3 of the participants recognized Intersect 1 and 15 recognized Intersect 2. This can be seen in the results – many of the participants named individual landmarks in the second intersection.

The final goal of the study was to compare the landmarks selected by human subjects and the model defined in this paper. In the model every landmark has an attraction value based on its properties. Some weighting was done for distance and architectural complexity properties. Usually landmarks that are in 200 meters radius are seen quite well so the distance does not play a major role when scoring attractiveness. Total attractiveness of a landmark is sum of following properties: number of storeys, building year, use, amount of geotagged images in vicinity, surrounding nodes, distance from the user (weighted down) and architectural complexity (weighted down). Each property was valued between 0-5 and the final attraction value was the average of these values.

Participants valued landmarks based on their attractiveness grading them from 1 (most attractive) to 5 (least attractive). Each attractiveness grade of 1 was worth 5 points, grade 2 worth 4 points, etc. If the landmark had no grade, it was worth 0 points. From this we got average for each landmark between 0 (minimum) and 100 (maximum) which was then divided by the number of participants for a score between 0 (minimum score) and 5 (maximum score). See Table 7 for results of the evaluation and Figures 10 and 11 for routes with numbered landmarks and



Figure 12. Images of the intersects with landmarks. Intersect 1 above and Intersect 2 below.

Figure 12 for images of the intersects with the same numbering.

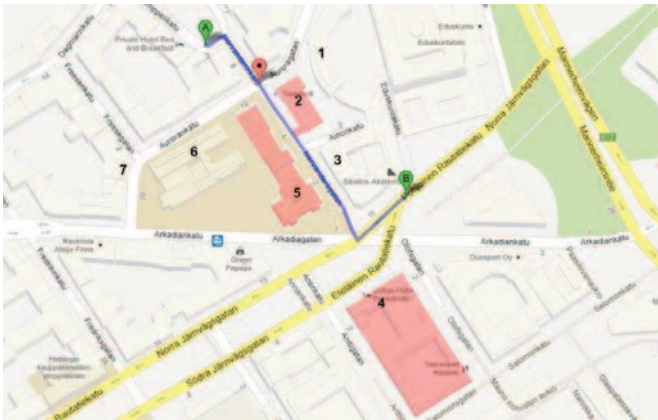


Figure 10. Intersect 1 with highlighted landmarks. Numbering of landmarks is based on Table 7.



Figure 11. Intersect 2 with highlighted landmarks. Numbering of landmarks is based on Table 7.

As we can see from the results, in the case if Intersect 1, all three landmarks with highest scores in the user tests had also the highest scores in the model. Landmark number 2 had the highest

score both in evaluation and in the model and in the freeform answers it was recognized fairly well, which gave it higher semantic value. This was also taken into consideration in the model where museums are generally ranked higher than for example shops and residential buildings.

Second highest ranking in the user experiment was for a hotel that was located directly ahead on the route. It is large and quite well known hotel in a very central location which might explain its high score in the user experiment. This building is also built in different age than the rest of the buildings in this area, which gave it better scoring in the model. In the model this landmark was ranked at third place and this was mostly due to its distance from the user.

Landmark number 5 was ranked third in the user experiment but was in shared first place in the model. This building is Helsinki Museum of Natural Sciences which is quite famous building but it was not recognized by many of the participants, maybe because this image was from behind the building and the front façade is easily recognizable with its famous statue. In this case the distance to the user was the main reason for the high ranking in the model. It is close to the route and to the user which gives it a higher rating.

Landmark number 1 is the library of the Finnish parliament, also quite famous building in the city center. Still it was graded quite poorly (4th in the evaluation, 6th in the model). This was mainly because it is not close to the route and does not really help the navigator along the route. Same happened with landmarks 6 and 7 – they were simply too far from the route to be helpful in the guidance context. Landmark number 3 is on the route but did not get very high scores in the model or the evaluation.

All three landmarks highlighted in the Intersect 1 were along the route which gave them better rating than the other buildings in the image. From this we can safely assume that unless there are some especially prominent landmarks in the vicinity, landmarks along the route are the best landmarks for supporting the guidance.

In Intersect 2 there were only 5 landmarks to choose from which two were bit too far and not visible enough for recognition. These landmarks (numbers 2 and 3) got the lowest ratings from the participants but landmark number 3 got quite high score from the

model. Both of these situations are ones that would've really benefited from the visibility calculation and both of these landmarks would probably have been removed from the attraction calculations because of their low visibility. Currently the model does not make distinction between fully visible and partially visible landmarks and cases where we have barely visible building gets very high score in the model.

In Intersect 2 the highlighted landmarks were number 4 and 5. Number 5 is a big and well known shopping mall in the city center and it was widely recognized in the freeform answers. Landmark number 4 is a movie theater and also very well known. Both of these landmarks were also the ones with the highest scores from the participants, but in different order. This is probably because of the semantic value of the theater – almost everyone who has been in downtown Helsinki knows this building. Landmark number 3 got relatively high scores both in the model and the evaluation, but didn't stick out from other buildings enough.

If we look at the results we notice that many of the landmarks with highest scores have some similarities: for example, they are very close to the route and their facades are facing the navigator. We learned from the freeform answers that many of the participants recognized and ranked these landmarks higher if they recognized them. In many cases semantic value was the only criteria for selecting a landmark as salient in the route's context. This is usually not the case when people are navigating in an unknown environment, which is usually the case when navigational aid is required. Therefore there could be room for further studies where the environment is completely unknown to the participant.

Our evaluation shows that landmarks selected by the participants correspond quite well with the landmarks selected by the model. The most attractive landmarks were very similar in both intersections and highlighted landmarks were among the highest scores, only in slightly different order. There's still a lot of room for further studies in different landscapes and scenarios, but these preliminary results are quite promising.

4. CONCLUSION AND FUTURE WORK

In this study we have introduced a new model for highlighting landmarks on web-based route guidance services. This model grades landmarks based on their properties that are retrieved from external datasets. Base for this model was done by Raubal and Winter [4] which was then modified to be more suitable for context of mobile map services. We presented a proof-of-concept implementation of the model in Android mobile devices.

The new model was evaluated by conducting a controlled experiment where images of intersects were shown to study participants, and they were asked to grade landmarks in intersects from most prominent to least prominent. The results were compared to the landmarks selected by the model. The results show, as the landmarks selected by the model and the landmarks selected by participants were almost exactly the same in all cases.

There is a lot of room for future research. For example, the model does not take into account the time of the day (landmarks look different during the night and the day). The major problem in integrating this kind of model for real world applications is the need of external datasets – it is impossible to find an up-to-date, worldwide and free external dataset of landmarks so the properties for the landmarks need to be retrieved from different datasets. In our proof-of-concept implementation this information was

gathered from the Helsinki Real Estate Department, Google Maps API and Panoramio API.

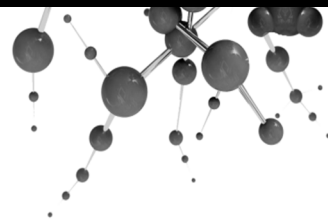
For extending the semantic attraction of landmarks, we plan to use Google Places API to add information about shops and restaurants to determine the historical and cultural attractiveness of a landmark. This could be developed even further by adding the search for Wikipedia articles on the landmarks. This way we could set higher semantic attractiveness value for landmarks with Wikipedia article. In the future it would also be very important to implement a visibility calculation in three-dimensional space. This way we could calculate which landmarks are visible to the user and which are not. This would require a very large three-dimensional dataset which is not available to this date. Model for this is already done so the implementation should be quite trivial once the data is widely available.

In conclusion, we believe that this model offers a good method for calculating landmark saliency and can be developed further for highlighting landmarks in decision points for pedestrian navigation.

5. REFERENCES

- [1] May A.J., Ross T., Bayer S.H. & Tarkiainen M.J. (2003). *Pedestrian Navigation Aids: Information Requirements and Design Implications*. Personal Ubiquitous Computing 7. S. 331-338.
- [2] Peebles, D., Davies, C., and Mora, R. (2007). Effects of geometry, landmarks and orientation strategies in the 'drop-off' orientation task. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds). *Spatial Information Theory*. Springer.
- [3] Gunzelmann, G., Anderson, J. R., & Douglass, S. (2004). Orientation Tasks Involving Multiple Views of Space: Strategies and Performance. *Spatial Cognition and Computation*, 4(3), 209-256.
- [4] Raubal M, & Winter S. (2002). *Enriching Wayfinding Instructions with Local Landmarks*. Geographic Information Science. Springer Verlag. S. 243-259.
- [5] Winter S. (2003). *Route Adaptive Selection Of Salient Features*. Technical University of Vienna.
- [6] Nothegger C., Winter S. & Raubal M. (2004). *Computation of the salience of features*. Technical report, Institute for Geoinformation, Technical University Vienna.
- [7] Lynch K. (1960). *The Image of the City*. Cambridge, MA, MIT Press.
- [8] Allen G. (2000). *Principles and practices for communicating route knowledge*. *Applied Cognitive Psychology* 14: 333–59
- [9] Tom A. & Denis M. (2003). *Referring to landmark or street information in route directions: What difference does it make?* In Kuhn W, Worboys M F, and Timpf S (eds) COSIT 2003. Berlin, Springer-Verlag: 362–74.
- [10] Ross T., May A., & Thompson S. (2004). *The use of landmarks in pedestrian navigation instructions and the effects of context*. In Brewster S and Dunlop M (eds) *Mobile HCI 2004*. Berlin, Springer-Verlag: 300–04
- [11] May A.J., Ross T., Bayer S.H. & Tarkiainen M.J. (2003). *Pedestrian Navigation Aids: Information Requirements and Design Implications*. Personal Ubiquitous Computing 7: 331-338.
- [12] Lovelace K., Hegarty M., & Montello D. (1999). *Elements of good route directions in familiar and unfamiliar*

- environments*. In Freksa C and Mark DM(eds) *Spatial Information Theory*. Berlin, Springer: 65–82.
- [13] Michon P.E. & Denis M. (2001). *When and why are visual landmarks used in giving directions?* In Montello D (ed) *Spatial Information Theory*. Berlin, Springer: 292–305
- [14] Kolbe T. H. (2002). *Fugängernavigation und Routenplanung in Innenstädten und Gebäuden mit Videos und Panoramen*. Institut für Geoinformatik, Universität Münster.
- [15] Kolbe T. H. (2004). *Augmented videos and panoramas for pedestrian navigation*. Second Symposium on Location-Based Services and TeleCartography 2004. Technical University Vienna.
- [16] Glassner A.S. (1998). *Graphics Gems*. San Diego, USA. Academic Press.
- [17] Caduff D. & Timpf S.(2005). *The landmark spider: Representing landmark knowledge for wayfinding tasks*. In Barkwosky T, Freksa C, Hegarty M, and Lowe R (eds) *Reasoning with Mental and External Diagrams: Computation Modeling and Spatial Assistance*. Stanford, CA, AAAI Press: 30–35
- [18] Sadeghian P. & Kantardzic M. (2008). *The new generation of automatic landmark detection systems: Challenges and guidelines*. *Spatial Cognition and Computation* 8: 252–87
- [19] Elias B. (2003). *Extracting landmarks with data mining methods*. In Kuhn W, Worboys M.F. & Timpf S. (eds) *COSIT 2003*. Berlin, Springer-Verlag: 375-89
- [20] Tezuka T. & Tanaka K. (2005). *Landmark extraction: A web mining approach*. In Mark D.M. (ed) *Spatial Information Theory*. Berlin, Springer: 379-98
- [21] Furlan A., Baldwin T. & Klippel A. (2007). *Landmark classification for route directions*. In *Proceedings of the fourth ACL-SIGSEM Workshop on Prepositions*. Prague, Czech Republic: 9-16



Publication II

Kallioniemi, P., Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., Uusi-Mäkelä, M., Hietala, P., Okkonen, J, and Raisamo, R. (2013). Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 33, 10 pages. doi:10.1145/2541831.2541849

© ACM, 2013. Reprinted with permission.

Evaluating Landmark Attraction Model in Collaborative Wayfinding in Virtual Learning Environments

Pekka Kallioniemi¹, Jaakko Hakulinen¹, Tuuli Keskinen¹, Markku Turunen¹, Tomi Heimonen¹, Laura Pihkala-Posti², Mikael Uusi-Mäkelä¹, Pentti Hietala¹, Jussi Okkonen¹, Roope Raisamo¹

School of Information Sciences¹, School of Modern Languages and Translation Studies²
33014 University of Tampere, Finland
firstname.lastname@uta.fi

ABSTRACT

In Virtual Learning Environments efficient navigation is a major issue, especially when it is used as a component in the learning process. This paper addresses the challenges in creating meaningful navigation routes from language learning perspective. The work is grounded on findings from a specific case on German language learning, wherein two remotely located users communicated in a wayfinding guidance scenario. The users navigated through 360-degree virtual panoramic images using body gestures and could receive communication help via spoken hints by pointing at objects in the scenery. An important design consideration is how to choose these objects, as they have both navigational importance and pedagogical significance in terms of learning the desired language. Wayfinding interactions from 21 participants were compared to the values provided by a landmark attraction model applied on the landmarks along the routes. The results show that there was a clear connection between prominence of landmarks and time spent on each panorama. This indicates that together with pedagogical planning, the model can aid in selecting the interactive content for language learning applications in virtual environments.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: *Artificial, augmented, and virtual realities*

General Terms

Design, Experimentation, Human Factors

Keywords

Wayfinding, virtual environments, gesture-based interfaces, embodied interaction, second language learning.

1. INTRODUCTION

In the last years we have seen huge advances in web-based technologies and interactive learning media, which have provided a range of options for educational purposes. Technology is becoming a part of classrooms and it is speculated to become an integral part of the educational setting in the near future [22].

In classroom use, simulated and realistic virtual environments can offer a viable alternative to visiting real places and have the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

MUM '13, December 02 - 05 2013, Luleå, Sweden
Copyright 2013 ACM 978-1-4503-2648-3/13/12...\$15.00.
<http://dx.doi.org/10.1145/2541831.2541849>

potential of providing a greater sense of presence compared with two-dimensional environments. Virtual Learning Environments (VLEs) have been used successfully in teaching of science, mathematics, and languages [4]. In a multidisciplinary research project focusing on active learning spaces, we have created an application that facilitates second language learning by tasking the users in route guidance dialogue in a virtual environment. Route guidance is a relevant task for language learning and virtual environments provide a context where this kind of communication can be practiced in a simulated environment that immerses the students better than traditional tools, such as a two dimensional map. Since movement in the virtual is natural part of such task, it can also provide intrinsic motivation for the learning.

In navigation and route guidance, landmarks are one of the fundamental items used in communicating directions. Availability of prominent landmarks affects how easy or hard it is to provide guidance in certain location. In language learning context, the challenge level of the language learning task could be adjusted by manipulating the amount of landmarks, and thus communication cues, available to the users. The main research question in this work is how to systematically analyze and select salient features of the scenery to act as placeholders for communicative cues and how they affect the navigation and guidance in the virtual environment.

We conducted an exploratory lab-based user study with 21 participants, collecting log data from the wayfinding interactions and observations of the use of a multimodal language learning application. During the wayfinding task, two remotely located students communicated with each other in German language to reach the destination. One of the users was a tourist going around the Berlin, while the other was a local who gave directions to the tourist during the navigation task. Using 360-degree photographic panoramas, the students could explore some of the most famous sights in Berlin, such as the DDR museum and Hackescher Höfe. Interaction with the environment took place in front of a large display screen using embodied gestures, which consisted of forward movement (to advance on the route), upper body rotation (to pan the panorama scene), and pointing to activate landmarks in the panorama that provided communication cues via audio.

In order to study the navigational effects of landmarks on the wayfinding task, we applied a landmark attraction model on the different landmarks along the route to quantify the relative difficulty of the panoramas. Results show that within the panoramas with prominent landmarks users would find the correct route faster and needed less help from the system. This finding indicates that the landmark attraction model can be a helpful tool when content for second language learning tasks is being prepared. For example, advanced students could be challenged with routes that contain less attractive landmarks whereas beginners would benefit from routes that contain prominent

environmental features. However, in a second language learning scenario the pedagogical value of the landmarks, i.e., contribution towards the expected learning outcomes and the students' knowledge and understanding of the topical matter, should also be carefully considered.

Next, we cover previous research, introduce the virtual learning environment used in our user study, and describe our evaluation setting and its results. Discussion of the key findings and their implications concludes the paper.

2. RELATED WORK

Our review of relevant previous research includes studies on virtual learning environments, how gesture-based and embodied interaction has been used previously in such environments, and landmark based navigation in real and virtual environments. Finally, we describe the landmark attraction model used as the basis for our analysis.

2.1 Virtual Reality and Virtual Learning Environments

Jonassen [1] and Smeets [2] consider information and communication technologies to be one of the most powerful tools for supporting the learning process. Virtual Reality (VR) technologies are computer-simulated environments that can be used to simulate locations in either real or virtual worlds. Mikropoulos and Bellou [3] described their technological characteristics as follows:

- creation of 3D spatial representations
- multisensory channels for user interface and interaction
- immersion of the user using the virtual environment
- intuitive interaction in real time

Virtual Learning Environments are virtual environments that are based on a certain pedagogical model and incorporate didactic objectives. According to a ten-year review of empirical research on educational applications of virtual reality [4], most of the VLEs conducted refer to science, technology and mathematics, but for example Second Life has been successfully used in language learning in over 200 universities [5]. Berns et al. [6] stated that sometimes it might be difficult to use virtual environments in teaching as they often lack a well-defined goal. This can be fixed by limiting users' movement in the environment so that the users do not get lost in the virtual world and lose interest in it.

2.2 Gesture and Embodied Interfaces in Virtual Environments

Embodied interfaces in virtual worlds have a long history and the first work was done in the 1970's by Krueger [18] who was the first one to combine an image of the user in a virtual world with two-dimensional gestural interaction. The ALIVE system created in MIT Media Lab was the first to allow three-dimensional interaction within a virtual environment [24]. Later the SURVIVE system with support for a number of full body interfaces that used stereo camera for tracking the human form was created in MIT [19].

In 2010, Microsoft unveiled the Kinect system that uses depth-sensing camera to detect humans and objects, which can be used to track body movements and gestures. It also has a microphone array for detecting audio and the captured audio and gesture data can serve as commands to interact with digital content presented in games or other applications. The benefits of Kinect in learning

applications can be divided to two major aspects. First, it is a stimulating tool, and if the interaction and pedagogical tasks are carefully designed, the classroom with Kinect devices should have the potential to create enjoyable and interesting interaction types to motivate the students. Second, Kinect can be used with specifically tailored software to enhance its role as a learning tool. A lot of Kinect-based teaching applications have been made available via KinectEDucation, which is "an educator-driven community resource for developers, teachers, students, enthusiasts, and any other education stakeholders to promote the use of Kinect applications in classrooms." [7]. Kinect-based games in virtual environments have also been used for rehabilitating patients with neurological impairments and older adults at risk of falls [8] and for teaching the sign language [9].

2.3 Wayfinding in Virtual Environments

Wayfinding is ones "ability to find a way to a particular location in an expedient manner and to recognize the destination when reached." [10]. It is based on consistent use and organization of sensory cues from the environment. Wayfinding instructions often consist of environmental features such as landmarks, pathways and choice points. Landmarks serve as sub-goals that keep the traveler connected to the point of origin and the destination along a specified path of movement. Pathways are nominals which refer to actual channels of movement, such as streets, highways or sidewalks. Choice points are nominal that refer to places affording options for the navigator with intersections being the most typical example [25].

Several studies have examined wayfinding in virtual environments. Darken and Sibert [11] focused on wayfinding in large virtual worlds. To summarize their work:

- Users need adequate source of directional cues, for example landmarks
- A large world without explicit structure is difficult to search exhaustively
- Path following is a natural spatial behavior (for example, following a coastlines as if it was a path)
- A map allows for optimizations in search strategies

Witmer et al. [16] studied the transfer of knowledge between virtual environment and the real world, and they found out that if a sufficient level of realism is provided, the performance in virtual environment is comparable to the actual environment. This finding was confirmed in study made by Waller et al. [17], who even suggested that training in virtual environments could be superior to real-world training. Unlike traditional maps, VEs represent spatial information by iconic simulation. This offers "a more naturalistic medium in which to acquire spatial information, and potentially allow users to devote less cognitive resources" [30] for wayfinding tasks when compared to maps.

The gender differences in navigation have been studied a lot, partly because these differences may be the most sizable of all differences in male and female cognitive abilities. [27] Lin et al. [28] studied gender differences in wayfinding in virtual environments, and found out that "males moved faster than females but did not necessarily navigate the spatial surroundings more efficiently. Each gender showed different strengths related to wayfinding; these differences require the application of both overall and fine-grained measures for accurate assessment."

2.4 Landmarks in Navigation

Landmarks are defined as prominent features in the environment that are in contrast with their neighboring objects [12]. This definition is commonly used in studies on navigation and wayfinding. Landmarks are a relevant part of navigation both in the real world and in virtual environments. Many previous studies have shown that landmark-based route navigation is common strategy in human wayfinding [13, 14]. Steck and Mallot [15] studied the role of global and local landmarks in virtual environment navigation. Distant landmarks such as towers or mountain peaks are considered global where as local landmarks are visible only from a small distance. The results from this study showed that in virtual environments both local and global landmarks are used in wayfinding tasks, but different participants relied on different strategies – some of the participants used only local landmarks while others relied on global landmarks. These two types are not mutually exclusive, as global landmarks in some phase of the task may serve as local landmarks. In spoken guidance, like in our case, the landmarks are central part of the communication and most of the communication is likely to be related to landmarks.

Vinson [26] defined design guidelines for landmarks to increase the accuracy of a navigator’s spatial knowledge, thus supporting navigation in virtual environments. Some of these guidelines are very essential for this study:

- It is essential for VE to contain several landmarks
- Include all five types (paths, edges, districts, nodes and landmarks) in VEs
- Use concrete objects instead of abstract ones; and
- place landmarks on major paths and at path junctions

Since we use photographic panoramas in our system, we cannot effect on how the virtual environment is represented, but we can use these guidelines when we want to design new panoramas (e.g. choose intersections with several distinctive landmarks) or if we want to highlight some of the landmarks on the route.

2.5 A Model for Landmark Highlighting

More recent studies have also investigated how to automatically identify landmarks from datasets and how these landmarks could be integrated into route guidance instructions. Elias [20] used a data mining method for finding the most salient landmarks from spatial databases. This approach focused on selecting the most salient landmarks automatically using hierarchical clustering from an external building database based on attributes such as size, height and distance to roads. This method emphasized landmarks that stand out from other surrounding objects. Raubal and Winter [21] introduced a grading system for landmarks based on their visual, semantic and structural properties. By using these properties, a total attraction value for each landmark can be calculated.

The model for landmark highlighting used in this evaluation is based on the model introduced by Kallioniemi and Turunen [23]. Similarly to Raubal and Winter [21], the model calculates attraction values for landmarks based on their visual, semantic and structural properties. The model contains three property categories: visual attraction, semantic attraction, and structural landmarks. All of these properties get a value on scale 1 to 5. Once calculations for visual, semantic and structural attraction have been calculated, these can be combined for total attraction value of a single landmark. These measures can also be given a predefined weight value, which allows for an adaptation of the context (e.g. mode of travel) or

individual user preferences. Whereas the model by Raubal and Winter was created for textual directions, the model presented by Kallioniemi and Turunen can be used in graphical route guidance systems such as Google Maps. The model uses data sources such as geotagged photos for determining the attraction value for each landmark. The property categories of the model proposed in [23] include:

Visual attraction – “Landmarks are visually attractive if they have some certain characteristics, such as sharp contrast with their surroundings or a prominent location.” [23] For determining visual attraction, we can combine the properties from both of these models because we can detect them from the panoramas (unlike in the study done in mobile context). Visual attraction of a landmark is based on the following measures: façade area, number of storeys, usage, temporality and visibility.

Examining the façade area of an object is very important property for determining its salience compared its surroundings. If the landmark’s façade exceeds or falls below the average size of the façades in the vicinity, these differences are often noticed by people observing them. Number of storeys can also tell about the saliency of a landmark and was therefore included in the model of this study. Use of the building is divided into different categories with different weight values. For example, churches and synagogues have a higher attraction value than residential buildings due to their cultural and historical value and distinctiveness from their surroundings.

Temporality was a property that was added to this evaluation, and it determines how temporary the landmark is – e.g. street vendors or ice cream kiosks are temporary landmarks. Temporality is only assigned with Boolean values. The last property determining the visual attraction of a landmark, the visibility, can be observed directly from the panoramas.

Semantic attraction – “Semantic measures for the original model of landmark saliency comprise cultural and historical importance of the landmark. It also takes into account explicit marks, such as signs or boards on the building façade. In this model, cultural and historical importance is determined with Boolean value and again refined with predefined scale from 1 to 5.” [23] Cultural and historical importance is determined by the amount of geotagged images in the area of the building (and some of its surroundings, since the images are often geotagged outside the actual landmark). Explicit marks can be observed from the panoramas.

Structural landmarks – “A landmark is structurally attractive if it has a major role or prominent location in the structure of the spatial environment. Examples of a structurally attractive landmark could be a big plaza or a road intersection. Therefore, if a landmark is located in an intersection with many nodes in the vicinity, it is considered to be structurally attractive.” [23] Raubal and Winter [21] determined the prominence of a landmark based on its resistance – as the resistance becomes larger, the more prominent the landmark is. For example, a big plaza, bridge or a road intersection is usually structurally attractive. Their model also takes into account nodes. “Nodes in a travel network are its intersections” and “the central structural characteristic of a node is its grade connectivity.” [21] Nodes were not included in the landmark attraction calculations of this study, as the virtual environment sets the user already in a very prominent location on the route and all the available exits were predetermined. Distance from the user is also one of the properties for the structural attraction – landmarks that are closer to the user are valued higher than those further away.

3. THE CITY COMPASS LANGUAGE LEARNING ENVIRONMENT

Our study was organized within a collaborative language learning environment called City Compass, where two remotely located users communicate using a foreign language in the context of a wayfinding task, one providing guidance to the other. The route in the virtual environment consists of a sequence of panorama pictures with multiple exits that must be taken to progress onward. Each user has his/her own panorama view, which he/she can pan freely. Both users' views can be seen in Figure 1. The physical setup of the environment consists of two separate locations, which are connected via a Voice over IP (VoIP) connection to enable two-way spoken communication.

3.1 Interaction Design

The panoramas contain hotspots that contain dialogue support for the users in form of spoken German language. A user can find hotspots by pointing at them on the screen with his/her hand (see Figure 2). Once a hotspot has been activated, a spoken utterance is played back to the user. The utterance can be a single noun or adjective about the target, or a longer description of the object and provides the user with words and phrases to use when communicating with the other user.

Interaction with the virtual environment is carried out using embodied gestures. The users can pan the view by turning their shoulder line left or right. This gesture emulates the turning of the upper body while person is looking around his or her surroundings. To move to the next panorama, the guided user walks towards the screen. A user can also point at screen where a cursor is displayed. When the user points at a hotspot, the related highlight-object is displayed.

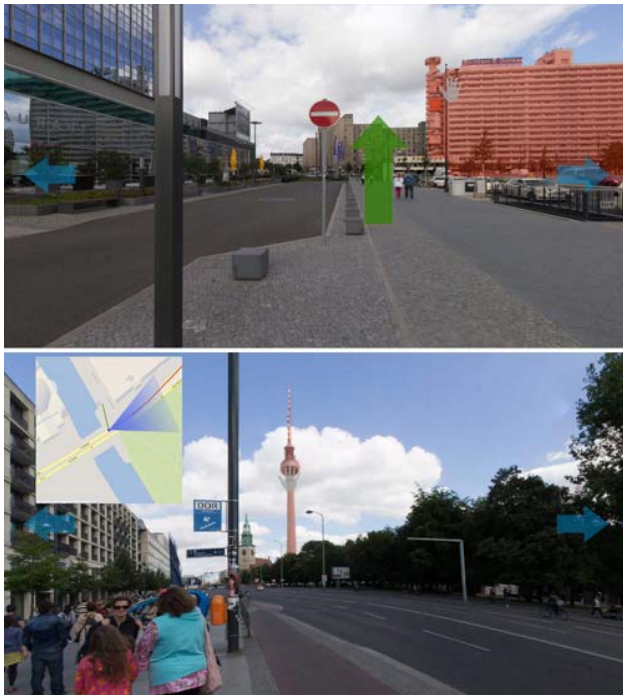


Figure 1. Above image is the interface for the tourist with an activated hotspot (highlighted with red) and an active exit. Below is the guide view with map and visibility sector cone in the upper left corner of the screen.



Figure 2. Participant is browsing hotspots in the panorama by pointing at the display.

The environment aims at providing a reasonably realistic environment and task to motivate and force the students to communicate in rich manner. Photographs of real environment provide plenty of details for the students to describe and discuss and the successful communication is required to advance in the route. The real location and advancement in the course provides intrinsic motivation for the students. Furthermore, the system is used by physically moving to activate and immerse users into the task.

3.2 Wayfinding Scenario

In the wayfinding scenario one of the users is a guide who gives directions to the other user who acts as a tourist. Starting from a given location, the goal is to reach a given destination within the city. Only the tourist can activate transitions, which take both users to the next panorama. The panoramas are in most cases one city block away from each other. There are on average four possible exits in each panorama, of which only one is the correct exit moving forward on the route. The exits in a panorama are presented as green arrows to the tourist while the guide can see the correct route on his/her panorama. The incorrect exits take both users either to the previous panorama or to a dead end. In the dead end the tourist supposedly gets lost and the guide needs to find out where she is so they can go back on the route. The tourist has a single image representing a location where she is lost. The guide has four different images of which one is the same image as the tourist's. The tourist needs to give a description of the location and the guide needs to find the correct image by this description. The guide then unlocks the exit by walking towards the screen. Once unlocked, the tourist can activate the exit to return to the route. Both dead end views are presented in Figure 3.

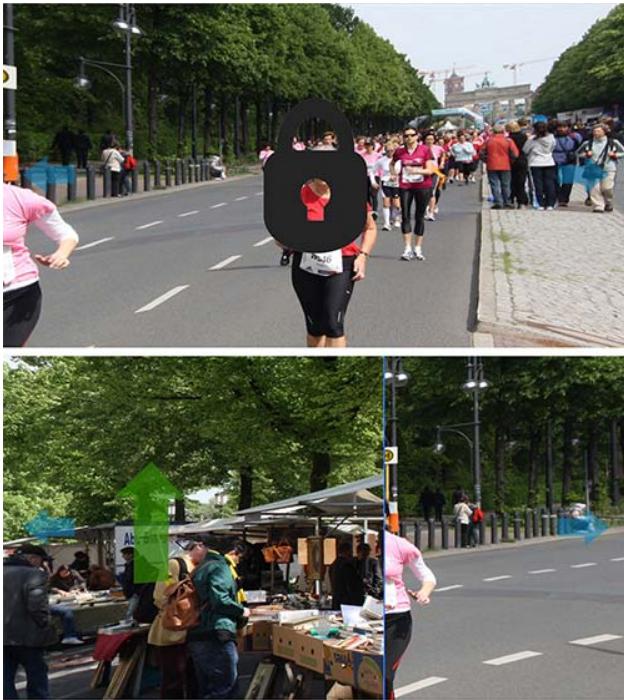


Figure 3. Above is the view for the tourist. Lock indicates that the guide has not yet found the right location. Below is the view for the guide – the user is between two images and the green arrow indicating exit node can be seen on the left.

3.3 System Description

The City Compass system consists of four components: the Central Logic, Graphics and Voice Service, Kinect Service, and Audio Transmission Service (Figure 4). The central logic handles the overall program logic and communicates with all the services and the other user's instance of the system, sending messages whenever exits are activated.

The virtual environment supports different languages and locations. A route consists of a set of panoramic images, the configuration of the exit nodes and hotspots. The textual content associated with hotspots can be any language as long as there is a speech synthesis available for the said language. In this case, system utterances were spoken using ScanSoft Steffi, a SAPI based speech synthesizers, which has been utilized for example in providing access to the online information system of the German

Ministry of Health¹, and was evaluated to be suitable for language learning by a German language teacher.

The Kinect Service built on top of Microsoft Kinect SDK² provided the locations of detected users, and joint data for users who are being tracked. The pointing location on the screen was calculated using a ray casting method that used the user's shoulder and wrist coordinates as reference points. The Audio Transmission Service is based on H.323 Plus open source package³, whose VoIP protocol implementation is used to set up a call between the two computers running the system.

The Graphics and Voice Service, built on top of Panda3D⁴ graphics engine, displays cylindrical panoramic images with 90 degree field of view. This is overlaid with the pointing cursor, exit arrows, and other user interface elements.

4. EVALUATION

We investigated the panorama-based collaborative wayfinding approach by conducting a study with high school students who are studying German as their second language (starting from the third grade). The aim of the evaluation was to study how the participants' wayfinding performance and interactions is influenced by the pedagogical and navigational qualities of the panoramas and their associated landmarks.

4.1 Participants

We had 21 participants (10 male, 11 female), who were eleventh grade high school students and thus 17–18 years old. A native German speaking research assistant filled in as a pair for the odd participant and is not included in the results. Each user excluding the native German speaker were native Finnish speakers. Experience with gesture-based games, such as Kinect or Wii games, was rare as only three participants reported to play such games, out of whom the most active player did so in general a few times a week. Playing mobile and computer or console games was more common as they all were individually played by between eight and nine participants each. Two of the participants had prior experience on the route that we used in the application of which one was the native speaker mentioned earlier.

4.2 Procedure

The users were in separate rooms and communicated with each other via VoIP using wireless clip-on microphones and speakers. In the beginning of the evaluation, the application was introduced to the participants by a researcher. During the introduction the participants were allowed to test the functionality of the application and ask questions. They could practice the use of the

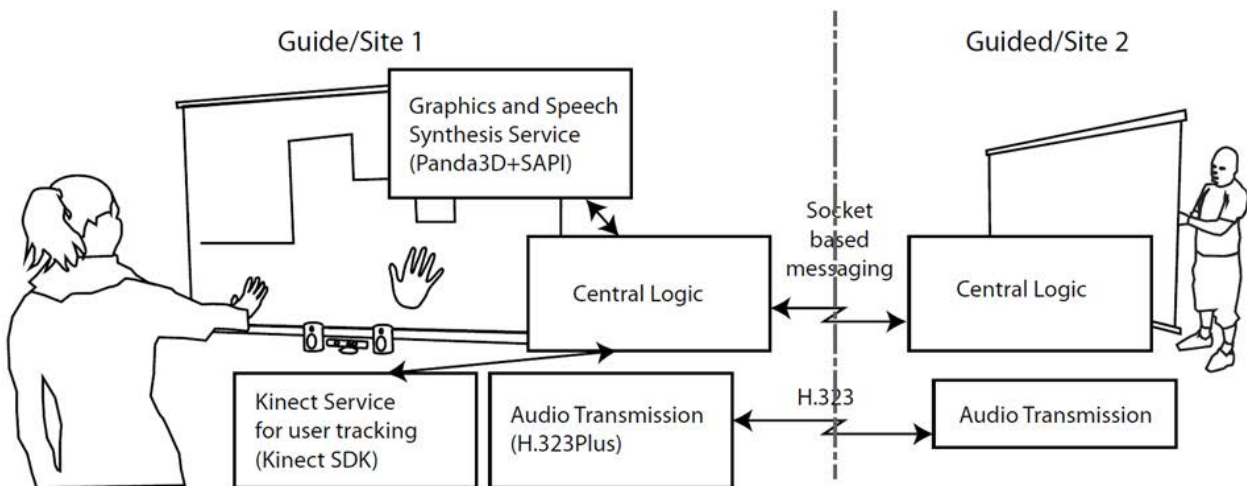


Figure 4. Virtual learning environment architecture

gesture interface, e.g., how to pan the screen and how to point at the hotspots on the screen. This was done in the start screen, i.e., in the first panorama of the route. After the training part, the participating pair performed a navigation task consisting of six panoramas.

4.2.1 Navigation Task

The navigation task followed a route in downtown Berlin from the ParkInn hotel near Alexanderplatz to the DDR museum along the Karl-Liebknecht-Straße (Figure 5). This route has a lot of prominent and famous landmarks such as Fernsehturm Berlin and Berliner Dom. However, these landmarks were not visible in every panoramic image, and thus could not be used as a global landmark for every panorama. Because of this, the participants had to use a good mixture of local and global landmarks in describing their surroundings. The average attraction score for landmarks and highest attraction score for a single landmark on correct route can be seen in Table 1.

	Highest attraction score for a landmark	Average attraction score for landmarks
DDR1	2,8	2,4
DDR2	3,2	2,8
DDR3	3	2,6
DDR4	2,6	2,3
DDR5	4,7	3,5

Table 1. Highest attraction score for a single landmark and average attraction score for landmarks in each intersection. ParkInn was excluded from this since it was used mostly for practicing the interaction with the system.

The participants were instructed to use German language for communication throughout the navigation task. One participant of each pair acted as the guide, i.e., giving instructions where to go, while the other was the tourist, i.e., moving according to the instructions. The introduction and completing the navigation task took a maximum of 15 minutes, after which the participants filled in questionnaires about their experiences. The participants were also interviewed verbally as pairs.



Figure 5. The route and its intersections. Each intersection had their own panoramic image. 1. ParkInn, 2. DDR1, 3. DDR2, 4. DDR3, 5. DDR4, 6. DDR5.

4.3 Data Collection

The application logged all data on the actions of each user during the evaluation. This data includes the time spent on the route and in each panorama, amount of pans the user executed per panorama and the total time of pans. The system also logged the amount of hotspots activated during the task.

In addition, we calculated attraction value for each landmark in the field of view of the next route leg to see if these values correlate with how much time it took for the users to find the correct exit in each panorama. We used the following properties to calculate the total attractiveness of each landmark: number of stories, amount of geotagged photos in the vicinity of the landmark, distance from the user, temporality and use. The number of geotagged images of landmarks was calculated manually from the data available in Panoramio⁵. Google Maps⁶ was used for calculating the distance by using the distance functions of Google Maps API⁷. Number of stories, use and temporality properties were gathered from the panorama pictures.

Subjective data from the participants was gathered with questionnaires and interviews after the navigation task. The questionnaires included 31 statements that were answered on a five-point Likert-type scale (1=Totally disagree – 5=Totally agree). The statements concerned more general user experience aspects of the application (e.g., pleasantness, learnability, naturalness, usefulness), interaction and modality related issues (e.g., easiness to control using gestures and movement, pleasantness of speech output), and aspects related to the wayfinding setting. In addition, some pedagogical aspects were covered (e.g., how well the environment promoted language learning). The interviews focused on establishing the rationale for questionnaire answers, any outstanding positive and negative properties of the environment, and development ideas.

5. RESULTS

The results of the study clearly show that landmark attraction model is a useful tool in designing routes for collaborative virtual learning environments. Subjective data also shows that using the system was easy to learn and that the underlying idea of the application is good.

5.1 Wayfinding Performance

Each pair completed the navigation task successfully, with an average task time of 650 seconds (SD = 144, min = 342, max = 914) spent on the route. The differences in wayfinding interactions between tourists and guides are not statistically significant, which suggests that the role taken by the participant does not markedly affect the use of the system. On average, the tourists panned the view 83 times (SD = 34) and the guides 63 times (SD = 22) during the route completion. The average total panning time of the tourists (mean = 158 seconds, SD = 42) similar to that of the guides (mean = 155 seconds, SD = 47). Similarly, the average total number of hotspot activations did not vary much between tourists (mean = 19.0, SD = 7.6) and guides (mean 15.4, SD = 5.9).

When comparing the wayfinding performance in different panoramas along the route one panorama, DDR4, appears more challenging than the others. A repeated measures MANOVA

⁵ <http://www.panoramio.com/>

⁶ <http://maps.google.com>

⁷ <https://developers.google.com/maps/>

revealed significant effect of panorama on time, $F(2.1, 39.3) = 11.79$, $p < 0.001$, number of pans, $F(2.1, 40.3) = 7.75$, $p < 0.05$, and hotspot activation, $F(2.7, 51.2) = 3.45$, $p < 0.05$. Mauchly's test indicated violations in the assumptions of sphericity and therefore Greenhouse-Geisser corrections were used. Bonferroni-corrected pairwise comparisons indicated that DDR4 required more time than the other panoramas ($p = 0.05$ or smaller), and that users panned the view more in DDR4 than in DDR2 ($p < 0.01$) and DDR5 ($p < 0.05$). Also, seven out of nine of dead ends occurred in this particular panorama. Similar effects were also observed between DDR2, the fastest of the panoramas, and DDR1, the second slowest, in time ($p < 0.01$) and pans ($p < 0.01$). The first panorama (ParkInn) was removed from the analysis

because most users used it for getting used to the user interface and gestures. For summary of the results, see Table 2.

	Avg. time (s)	# dead- ends	Avg. # hotspots activated	Avg. # pans
ParkInn	161 (SD = 89)	1	4.2 (SD = 5.0)	17.7 (SD = 10.0)
DDR1	97 (SD = 43)	1	1.2 (SD = 2.0)	11.6 (SD = 6.5)
DDR2	57 (SD = 21)	0	1.2 (SD = 1.4)	6.5 (SD = 4.0)
DDR3	85 (SD = 72)	0	1.5 (SD = 1.4)	9.3 (SD = 6.4)
DDR4	177 (SD = 92)	7	2.8 (SD = 2.4)	18.5 (SD = 13.9)
DDR5	71 (SD = 21)	0	2.0 (SD = 1.7)	9.7 (SD = 6.3)

Table 2. Log data from the system for all panoramas.

A possible explanation to the difficulties experienced in DDR4 is revealed when investigating the relationship between landmark attractiveness and panorama use. We restricted this analysis on the correct route (i.e., field of view containing the exit that leads to the next panorama), as this is the view the guide orients towards to assist the tourist. DDR4 had the lowest average attraction score (2.3) for landmarks among the panoramas (with others ranging between 2.8 and 3.5). Spearman correlation coefficients show that panoramas with a higher average attractiveness of landmarks resulted in shorter time spent on navigating through the scene ($r_s = -.90$, $p = .037$, $n = 5$). Given the low number of observations these findings are not conclusive, but strongly suggest that landmark attractiveness could be a viable predictor for task performance in this context.

5.2 Subjective Feedback

Although this study was focused on the wayfinding performance, it is also needful to investigate the user experience of system use. Establishing the participants' motivation and perceptions of system functionality form the basis for understanding the overall impact and use of the system. If a system such as this does not lead to an improved sense of language learning, or is extremely difficult to use, it is clear that any conclusions drawn of the study are of questionable validity.

5.2.1 Overall User Experience

The overall user experience of the embodied learning environment was very positive. Every participant was in full agreement that using the application is easy to learn and they really liked the underlying idea (median = 5). Using the learning environment was also reported to be rather pleasant, clear, effortless, natural, useful, entertaining, and interesting (median = 4). Participants felt that

controlling the application with gestures and movement was both easy and comfortable, and that the application functioned as it should (as a consequence of their actions) (median = 4).

Unsurprisingly, previous experience with games influenced the students' perceptions of the system. For example, computer game players liked the idea of the application more than non-computer game players, whereas console game players felt it was easier to learn and were overall more satisfied than non-console game players. What this suggests is that non-gamers might require additional support and motivation to use such a system, particularly in terms of introducing the interaction methods that may be unfamiliar to users with no prior experience from gesture-based interaction and virtual environments.

5.2.2 Wayfinding and Language Learning

Considering the wayfinding experiences, the participants reported that the navigation felt relatively realistic (median = 4). Interestingly, the tourists felt more like "being on the street" (median = 4, $n = 9$) than the guides (median = 3, $n = 7$). From a pedagogical point of view, the participants reported that they believe the application promotes German language learning (median = 4). According to the results, the hints provided by the application were of suitable extent and they were also useful (median = 4). However, the participants also considered that they had some difficulties in remembering German words (median = 4). Because users were communicating using a foreign language, limitations in vocabulary likely affected the wayfinding instructions, communication between the users, and also led to the appropriation and circumvention of system features in an effort to successfully complete the wayfinding task.

6. DISCUSSION

The primary goal of our study was to observe how to systemically analyze and select salient features of the scenery and observe how wayfinding tasks in multimodal virtual learning environments are affected by communicative cues attached to prominent landmarks. If this is the case, the landmark attraction model could be integrated into the route planning of wayfinding tasks in virtual learning environments.

In the following chapters we present four key design implications inspired our findings. First, we discuss how the landmark attraction model could be used to address different levels of language skill and learning outcomes. Second, we describe how our participants' utilized transient landmarks in wayfinding and how such landmarks could be utilized as a component of route planning. Third, we describe how the process of extracting landmark features should be improved with an eye towards better fit for use in virtual learning environments. Fourth, we propose ways to improve the use of the landmarks in navigation by manipulating the associated content.

6.1 Addressing Different Levels of Language Skill with Landmark Selection

As our results showed, the landmark attraction model scores correlated with the participants' navigation times. This indicates that the model could be used in planning the routes for the system. This could be done by analyzing each panorama and determining attraction values for each landmark on the route. For example, if we wanted to integrate a new route in the application, we would first analyze the panoramic images of said route and calculate attraction values for each landmark in the panoramas. These values could then be used for determining the overall difficulty of the route. The difficulty level could then be adjusted by the subject matter experts who are responsible selecting the hotspots



Figure 6. Two panoramas from the application. Above is the DDR4 panorama in which the participants had most difficulties with. Below is the DDR2 in which the participants solved with relative ease.

and their content. This way one could create routes and hotspots for users with different skill levels – the routes with landmarks with high attraction values could be suitable for users with low level language skills, whereas users with better skills could navigate through routes with less hotspots and with landmarks of low attraction values. In some cases the difficulty of the route can be changed simply by changing the hotspots and their content (see below in section 6.3).

6.2 Appropriation of Transient Landmarks in Wayfinding

As was seen from the results, one of the panoramas (DDR4, Figure 6) was particularly difficult for many users. Part of this can be explained by the lack of attractive landmarks. Some of the participants had to refer to moving objects such as taxis and people wearing clothing of specific color to find their way to the next leg of the route. We also observed that some of the users had difficulties finding correct words in panoramas that lacked prominent landmarks. In these situations, users attempted creative solutions such as communicating directions in degrees, or even using their native language for communication, which goes against the learning objectives of the scenario.

From wayfinding perspective, this appropriation of transient objects such as cars and people was particularly interesting, since these would be impossible to use as direction cues in normal wayfinding. Although it is not realistic in terms of real world navigation, adding such objects could be used from a language learning perspective as a way to decrease the difficulty of panoramas that lack natural, attractive landmarks.

6.3 Extracting Landmark Features

Although the landmark attraction model can potentially predict possible problem spots, the model would probably give better results after some changes to the properties or their weighting. The key problem with the current model is that it relies on properties that need to be extracted from external datasets, which are often not available for public use. Fortunately some of these

datasets are available for everyone through APIs (e.g. Google Maps, Wikipedia, and Panoramio), and the rest the property data can be gathered from the panoramic images themselves. Also, new properties such as searching for Wikipedia entries for the landmarks or using explicit signs and logos for determining cultural and historical attraction could make the model better for determining landmark saliency.

6.4 Hotspot Content Presentation

Despite the overall success of the wayfinding tasks, hotspots were relatively rarely used during the navigation task. For the landmark based wayfinding concept to be effective, the content and representation of the hotspots needs to be appropriate for the intended users. Based on our observations there were issues with speech quality and the language skills of the users were not sufficient to understand most of the content of the hotspots.

Speech quality issues apply both to the technical setup of the virtual environment and the quality of the speech output. Although the directional audio speakers used in our setup free the users from wearing a wireless headset, using one might be more appropriate in this setting. Not only would this eliminate any distractions caused by the environment, it would also have the benefit of masking unwanted audio cues in environments where total separation between the users is not possible (e.g., due to logistical reasons or lack of specialized equipment). In our setting this was not entirely possible and at times resulted in undesirable behavior whereby the participants utilized hotspot synthesis as audio cues for the orientation of the other user. After this, they just had to find the same hotspot and give directions from that orientation. With respect to the quality of the generated speech output, word-level adjustment of the speech synthesis prosody (e.g., rate and emphasis) can help make the content easier to understand. For example, one could add emphasis on the key terms or phrases in the content to make them stand out better in the speech output.

One of the challenges in designing language learning content is the variable skill level of users. In a virtual environment where the system can be adapted on the basis of users' interactions, this problem can be tackled by providing progressively easier content when comprehension difficulties are encountered, for example if the user activates the same hotspot repeatedly, the users are stuck on a particular panorama for a lengthy duration, or if the guidance leads to dead ends repeatedly. Depending on the particular context, this could be realized either by adding emphasis on the critical words, or by altering the content to improve comprehension (e.g., by using easier to understand terminology) and wayfinding support (e.g., the content is expanded to include features of the landmarks that can assist in describing the direction to take).

6.5 Limitations and Future Work

Overall, the findings presented in this paper form the starting point for our research into utilizing landmark salience to enhance the creation of language learning scenarios for virtual environments. Particularly, a detailed analysis of the communication between the users is currently underway. We are also planning to evaluate an improved version of the City Compass system with more participants in the context of English language learning. Also, Santa Barbara sense-of-direction scale [29] should be implemented to all future studies.

Although the user experience of the panorama-based virtual learning environment was rated to be of good quality, there are always ways to improve the interface, interaction and task flow. The ray casting based pointing was not particularly precise and pointing at small objects can be tedious. Adding target selection mechanisms such as "magnetism" to the hotspots, so that they pull the pointer in when it gets close by, is one solution we are investigating.

In the current implementation, the user's field of view was fixed and the only way to reveal more content was to pan the view. We are looking to implement head tracking to adjust the field of view of the virtual camera to study if the further immersion this can provide is beneficial.

In the tasks used in the user study, dead end views did not include any hotspots, which may have made it unduly difficult to return from the dead end to the wayfinding flow. One promising solution is to add hotspots also to the dead end screens, which should help the users in returning to the navigational task and also broaden their vocabularies.

7. CONCLUSION

This paper addressed the challenges in creating meaningful navigation routes from language learning perspective. An important consideration in the design of wayfinding tasks for virtual environments is how to choose the scenes presented to the users, as the prominent landmarks visible to the users have both navigational importance and pedagogical significance in terms of learning the desired language.

Our findings are based on a specific case on German language learning show that there was a clear connection between prominence of landmarks and time spent on each panorama. This indicates that together with pedagogical planning, the model can aid in selecting the interactive content for learning applications in virtual environments.

8. ACKNOWLEDGEMENTS

This work is part of the "Active Learning Spaces" project, funded by the Finnish Funding Agency for Technology and Innovation.

REFERENCES

- [1] Jonassen, D. H. 1994. Thinking technology: toward a constructivist design model. *Educational Technology* 34, 4, 34–37.
- [2] Smeets, E. 2005. Does ICT contribute to powerful learning environments in primary education? *Computers & Education*, 44, 3, 343–355.
- [3] Mikropoulos, T. A. and Bellou, J. 2006. The unique features of educational virtual environments. In P. Isaias, M. McPherson, & F. Banister (Eds.), *Proceedings of e-society 2006, Vol. 1*. IADIS, 122–128.
- [4] Mikropoulos, T.A. and Natsis, A. 2010. Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education* 56, 3, 769–780.
- [5] Kelton, A. 2007. Second Life: Reaching into the virtual world for real-world learning. *Educause Center for Applied Research: Research Bulletin*, 2007, 17.
- [6] Berns, A., Gonzalez-Pardo, A., and Camacho, D. 2011. Implementing the use of virtual worlds in the teaching of foreign languages. In S. Czepielewski (Ed.), *Learning a Language in Virtual Worlds: A Review of Innovation and ICT in Language Teaching Methodology*. Warsaw Academy of Computer Science, Warsaw, 33–40.
- [7] KinectEducation. 2013. What is KinectEducation? Retrieved from <http://www.kinecteducation.com/blog/kinect-in-education>, referenced on 15.4.2013.
- [8] Lange, B., Koenig, S., McConnell, E., Chang, C-Y., Juang, R., Suma, E., Bolas, M., and Rizzo, A. 2012. Interactive Game-Based Rehabilitation Using the Microsoft Kinect. In *IEEE Virtual Reality Short Papers and Posters (VRW)*. IEEE Computer Society, 171-172.
- [9] Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., and Presti, P. 2011. American sign language recognition with the kinect. In *Proceedings of the 13th international conference on multimodal interfaces*. ICMI '11. ACM, New York, NY, USA, 279-286. DOI=<http://doi.acm.org/10.1145/2070481.207053>
- [10] Peponis, J., Zimring, C., and Choi, Y.K. 1990. Finding the Building in Wayfinding. *Environment and Behavior* 22, 5, 555-590.
- [11] Darken, R. P. and Sibert, J. L. 1996. Wayfinding strategies and behaviors in large virtual worlds. In Michael J. Tauber (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '96. ACM, New York, NY, USA, 142-149. DOI=<http://doi.acm.org/10.1145/238386.238459>.
- [12] Lynch, K. 1960. *The Image of the City*. MIT Press, Cambridge, Mass.
- [13] Thorndyke, P. W. and Hayes-Roth, B. 1982. Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology* 14, 4, 560–589
- [14] Gale, N., Golledge, R. G., Pellegrino, J. W., and Doherty, S. 1990. The acquisition and integration of route knowledge in an unfamiliar neighborhood. *Journal of Environmental Psychology* 10, 1, 3–25.
- [15] Steck, S. D. and Mallot, H. A. 2000. The Role of Global and Local Landmarks in Virtual Environment Navigation.

- Presence: Teleoper. Virtual Environ.* 9, 1 (February 2000), 69-83.
- [16] Witmer, B. G., Bailey, J. H., Knerr, B. W. and Parsons, K. C. 1996. Virtual spaces and real world places: Transfer of route knowledge. *Int. J. Hum.-Comput. Stud.* 45, 4 (October 1996), 413-428.
- [17] Waller, D., Hunt, E. and Knapp, D. 1998. The transfer of spatial knowledge in virtual environment training *Presence: Teleoper. Virtual Environ.* 7, 2, 129-143.
- [18] Krueger, M., *Artificial Reality*. Addison-Wesley, 1983.
- [19] Tollmar, K., Demirdjian, D., and Darrell, T. 2004. Navigating in virtual environments using a vision-based interface. In *Proceedings of the third Nordic conference on Human-computer interaction*. NordiCHI '04. ACM, New York, NY, USA, 113-120. DOI=<http://doi.acm.org/10.1145/1028014.1028033>
- [20] Elias, B. 2003. Extracting landmarks with data mining methods. In W. Kuhn, M. F. Worboys, & S. Timpf (Eds.) *COSIT 2003, LNCS 2825*. Springer-Verlag, Berlin Heidelberg, 375-389.
- [21] Raubal M. and Winter, S. 2002. Enriching Wayfinding Instructions with Local Landmarks. In M. J. Egenhofer and D. M. Mark (Eds.), *Proceedings of the Second International Conference on Geographic Information Science*. GIScience '02. Springer-Verlag, London, UK, 243-259.
- [22] U.S. Department of Commerce. 2002. Visions 2020: Transforming Education and Training Through Advanced Technologies. Washington, DC. Retrieved from <http://purl.access.gpo.gov/GPO/LPS34165>.
- [23] Kallioniemi, P. and Turunen, M. 2012. Model for landmark highlighting in mobile web services. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. MUM '12. ACM, New York, NY, USA, Article 25. DOI=<http://doi.acm.org/10.1145/2406367.2406398>.
- [24] Maes P., Darrell T., Blumberg B., and Pentland, A. 1997. The ALIVE system: wireless, full-body interaction with autonomous agents. *Multimedia Syst.* 5, 2 (March 1997), 105-112.
- [25] Allen, G. L. 2000. Principles and Practices for Communicating Route Knowledge. *Applied Cognitive Psychology* 14, 4 (July/August 2000), 333-359.
- [26] Vinson, N.G. 1999. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. CHI '99. ACM, New York, NY, USA, 278-285.
- [27] Lawton, C. A., & Morrin, K. A. 1999. Gender differences in pointing accuracy in computer-simulated 3D mazes. *Sex Roles*, 40 (1-2), 73-92.
- [28] Chin-Teng Lin, Teng-Yi Huang, Wen-Jing Lin, Shu-Yen Chang, Yin-Hung Lin, Li-Wei Ko, Daisy L. Hung, Erik C. Chang. 2012. Gender differences in wayfinding in virtual environments with global or local landmarks, *Journal of Environmental Psychology*, Volume 32 (2), 89-96.
- [29] Hegarty M., Richardson A.E., Montello D.R., Lovelace K., Subbiah I. 2002. Development of a self-report measure of environmental spatial ability. *Intelligence*, Volume 30 (5), 425-447.
- [30] Montello D.R., Hegarty M., Richardson A.E. 2004. Spatial Memory of Real Environments, Virtual Environments, and Maps. in Allen G.L. ed. *Human Spatial Memory: Remembering where*, Lawrence Erlbaum Associates, 251-285.



Publication III

Kallioniemi, P., Pihkala-Posti, L., Hakulinen, J., Turunen, M., Keskinen, T., and Raisamo, R. (2015). Berlin Kompass: Multimodal gameful empowerment for foreign language learning. *Journal of Educational Technology Systems*, Vol. 43, pp. 429–450. doi:10.1177/0047239515588166

© SAGE Publishing, 2015. Reprinted with permission.

Berlin Kompass: Multimodal Gameful Empowerment for Foreign Language Learning

Pekka Kallioniemi¹, Laura-Pihkala Posti², Jaakko Hakulinen¹, Markku Turunen¹, Tuuli Keskinen¹, Jussi Okkonen¹, Roope Raisamo¹

¹School of Information Sciences, ²School of Language, Translation and Literary Studies
University of Tampere, Finland
{firstname.lastname}@staff.uta.fi

Abstract

This paper presents an innovative, gameful, multimodal and authentic learning environment for training of oral communication in a foreign language – a virtual adventure called *Berlin Kompass*. After a brief presentation of the pedagogical and technological backgrounds the system is described. Central results of a series of pilots in autumn 2013 with around 100 upper secondary pupils are described and further steps discussed. The results are highly promising, the concept was highly appreciated by the pupils regardless some technical problems that the prototype had during the pilots. The researcher observations and the questionnaire results show that the concept manages to create a new motivating collaborative learning context with clear added value compared with equal tasks realized in a typical classroom approach.

Introduction

Advances in modern web-based technologies and interactive learning solutions have provided a range of new options for the field of education. It has been speculated that virtual classrooms will become an integral part of the future educational environment. As haptic and embodied interfaces become less prohibitive and computational power in our devices increases, virtual environments that provide the users with immersive and natural interaction and a new type of learning experience can become a real possibility.

In this paper we present our innovative gameful approach for learning oral communication in a foreign language. The state-of-the-art technologies are used in order to create learning possibilities that go beyond the traditional approaches with school books or individual computer supported learning. *Berlin Kompass* is an adventure in a virtual but authentic city environment using an embodied collaborative approach. In this application, two remotely located users communicate with each other by speaking foreign language in the context of wayfinding. *Berlin Kompass* is based on embodied interaction. One of the users takes a role a tourist in a new, unknown city and the other user acts as a guide who helps the tourist along the route. This route consists of realistic 360 degree panoramic images. These panoramas have multiple exits and the correct one must be chosen in order to progress further along the route. The system has a novel approach to language learning by combining embodied interaction and spoken communication together with challenging language learning-related tasks. These tasks contain a clear starting point and a goal, as well as problem solving on the way to this goal. The aim of this approach is to shift the focus of the user from thinking of being in a foreign language learning situation into immersive gaming experience where the users collaborate in order to reach their destination.

The pedagogical concept of the *Berlin Kompass* application has been previously introduced in Pihkala-Posti & Uusi-Mäkelä (2013), Pihkala-Posti et al (2014), Pihkala-Posti (2014 a, b, c) and the technological part in Kallioniemi et al (2013) and Kallioniemi et al (2014). This paper goes further with deepening the earlier insights and discussing new aspects. Our main research questions that are discussed in this paper are as follows:

- Are there indications for any added value for learning compared with other approaches used before? This concerns especially offering feelings of success and motivation for different types of learners and communicators.
- How do the users experience the new approach?
- Does an earlier experience in playing computer games have an influence on the acceptance and use of the application?
- How the innovative aspects of the application are brought up by the open ended answers of the questionnaire, this means authenticity / immersion, collaboration and the embodied interaction?

The computer scientific expert/author of this paper is Pekka Kallioniemi and pedagogical expertise is presented by the co-author Pihkala-Posti. Both quantitative and qualitative approaches were used in this study as we collected feedback from the participants on the *Berlin Kompass*. Interesting results of our user questionnaires and the researcher observations from the pilots in the end of the year 2013 are presented and discussed.

Related Work

The following chapter describes the research that is related to the topic of this study.

Learning Theoretical Aspects: Multimodality, Action-orientation, Authenticity, Gamification and Collaboration

Through the last decades there have been different learning theories emphasizing the importance of different kinds of approaches in order to support diverse learner styles, types and kinds of learners, e.g. Gardner's theory of Multiple Intelligences. Many of these theories have been criticized because no systematical evidence could be found for them in the tests that have been carried out (e.g. Coffield et al. 2004). Yet it is still quite popular among the pedagogics to divide the learners into three main categories: visual, aural and kinesthetic learners. The teacher education uses these terms and the learners might even categorize themselves according to these categories as our research material shows (cf. Pihkala-Posti & Uusi-Mäkelä 2014). Regardless of the problems connected to these kinds of categorizations the brain research seems to support the idea of utilizing different modalities in order to support the learning. This is because the use of different modalities activates the brain in more areas simultaneously which has positive effects on learning (e.g. Grein 2013, Gee 2007, Pihkala-Posti 2014 b,c). The modern web-technologies enable the combination of modalities in new ways which is meaningful when designing learning environments. The combination of visual and aural modalities is already commonly used.

There are approaches at least since 1940's that focus on the importance of the corporeity of the perception (Merleau-Ponty 2013, Liimakka 2011). In this case embodied experience must be essential for learning too. E.g. Gee (2004) stresses the situatedness of the meanings of signs in embodied experiences of the learner. The language use should not be separated from the learner as a whole to be just a written or spoken text. In this way approaches that take the body and the embodied (kinesthetic) experience in account should be of special interest and relevance. Although the role of the embodiment for learning has been recognized for a quite long time at least theoretically, its practical implications / application have not been as easy.

Learning concepts emphasizing aspects as *learning by doing* and *experiential learning* (cf. Dewey 1915, Kolb 1984, Gee 2004, Aldrich 2005) have not managed to change the schooling mainstream although they are being occasionally used. New technology-supported action-oriented approaches might be able to change the situation. *Berlin Kompass* that will be closer described in the next chapters is one promising example of such approaches.

James Paul Gee has been a leading promoter for use of games in language learning (Gee 2004, 2007, Gee & Hayes 2011) as well as for the meaning of multimodality. His principles include many factors that are of interest for our context (Gee 2007). Gamification as well as using real games in supporting informal or formal (language) learning is an actual trend that has shown promising results in developing learner attitudes, motivation and skills e.g. in EFL classes (Sundqvist & Sylven 2014, Chik 2014, Uuskoski 2011, Oksanen 2014). A clear difference should in our opinion be made between playing real games in order to learn language (cf. e.g. Minecraft; Pihkala-Posti & Uusi-Mäkelä 2014, Pihkala-Posti 2014c) and using a concept designed for language pedagogical purposes that includes gameful elements – as for example our *Berlin Kompass*. Then the question is about gamification, that is defined as the use of game elements and game thinking in non-game environments in primarily non-entertaining contexts (Deterding et al. 2011, Betts 2013). The goal can among others be to enhance the level of engagement for learning. Or as in our case, to diminish the fear of speaking a foreign language by distracting the attention to other matters.

A central pedagogical issue is that of authenticity. The most essential factor for our definition of authenticity is the experience of the learner of the situation (or of the learning material) (cf. Pihkala-Posti in press 2015). Authenticity belongs according to our understanding closely together with the learner agency, as van Lier (1996) and Kaikkonen (2002b) reflect. "Authentic foreign language learning is enhanced through interactive and reflective encounters with the use of the foreign language, whereby the the perception and meaning validation play an important role in real language cultural situations" (Kaikkonen 2002b, S.40). [Translation from German original „Authentisches Fremdsprachenlernen wird durch interaktive und reflektierte Erfahrungen über den fremden Sprachgebrauch gefördert, wobei Wahrnehmung und Bedeutungsüberprüfung in wirklichen sprachkulturellen Situationen eine wichtige Rolle spielen“ Pihkala-Posti]

In our context the authenticity has the meaning of experiencing things and communicating in contexts outside of the language classroom in order to reach something real (cf. Pihkala-Posti in press 2015). We are not dealing here with purely instructed exercises anymore, or with artificial constructed ones, but approaching experiences of authentic communication and situations (cf. Kaikkonen 2002a; 2002b). Further, the feelings of immersion are related to those of authenticity in our multimodal virtual context (cf e.g. sensory immersion in Ermi & Mäyrä 2005).

Dillenbourg (1999) describes collaborative learning as a construction of shared knowledge through activities with others, where these participants are committed in shared goals and problem solving. Usually this collaboration includes social interaction between group members and is a prerequisite for the completion of any given task (Arvaja, Häkkinen & Kankaanranta (2008). Collaborative learning is based on the idea that participants share the task-solving, and achieve something that any one individual could not achieve on their own (Oksanen, 2014). Serious games (video games for learning) are one way to support this collaboration through games. A study by Sung & Hwang (2013) indicates that collaborative serious games can improve learning motivation and attitudes. When designing these games, one should take into account both the theoretical grounding and the game-design perspective to harness the full potential of collaborative games (Oksanen, 2014).

Virtual Learning Environments

Based on studies by Jonassen (1994) and Smeets (2005), information and communication technologies are one of the most powerful tools for supporting the learning process in the near future. Among these, Virtual Reality (VR) technologies have become more and more common in the field of education. These computer-simulated environments can be used to simulate locations in either real or virtual worlds. Mikropoulos and Bellou (2006) described the characteristics of VR in following manner:

- “creation of 3D spatial representations, namely virtual environments
- multisensory channels for user interface and interaction
- immersion of the user using the virtual environment
- intuitive interaction in real time”

Virtual Learning Environments (VLEs) are virtual environments that are “based on a certain pedagogical model, incorporates or implies one or more didactic objectives, provides users with experiences they would otherwise not be able to experience in the physical world and redounds specific learning outcomes” (Mikropoulos & Natsis, 2011). A ten-year review of empirical research on educational applications of virtual reality by Mikropoulos & Natsis (2011) revealed that most of the VLEs conducted refer to science, technology and mathematics, but to a lesser extent also in language learning. For example, Second Life has been used in language learning in over 200 universities (Kelton, 2007).

As stated by Steuer (1992), virtual reality is usually described as “a collection of technological hardware”, thus having a very technology-concentrated focus. Because of this approach, there was no conceptual framework for the educational uses of virtual reality. Bricken (1990) specified natural semantics and cognitive presence as the cornerstones of virtual reality and claimed constructivism as the theoretical model that supports VLEs. Helsel’s (1992) conceptual orientation to VLEs (and VR in general) described it as “a process that enables users to become participants in abstract spaces where the physical machine and physical viewer do not exist”.

Pantelidis (1993) provided several reasons for the use VLEs in classroom environment, including active participation, high interactivity and individualization. Winn and Windschitl (2000) reported that with VLEs the sense of presence is a major factor in the learning process because it enhances the “first hand” experiences and the psychological activity when people are interacting directly with real or virtual worlds. Learning is a complex process and features in VLEs should not be segregated as an isolate entity from this process, but they play a role in the learning outcome (Salzman, Dede, Loftin, & Chen, 1999). Therefore, it is important to study and define these features so that we can understand further the contribution of VLEs to learning process and outcomes.

Embodied Interaction

Recent years have brought us new interaction styles which aim at leveraging human-computer interaction. Leaps in modern technology and the expanding context of where we use it suggest that there is need for a new ways of interacting with our devices. Mouse and keyboard have dominated our interaction with computers for long time, and modern HCI research and especially research on *embodied interaction* has looked into methods that could replace this metaphor. Dourish used this term to describe an approach to interaction design where the emphasis is on understanding and incorporating our relationship with the surrounding world into designing and use of interactive systems (Dourish 2001). Antle et al. (2009) described it as "an approach to understanding human-computer interaction that seeks to investigate and support the complex interplay of mind, body and environment in interaction."

Antle et al. (2009) also emphasizes the importance of intuitiveness in designing embodied interaction. In intuitive interaction the user can immediately use the interface without problems and the interface works in a way how people

expect it to work. Natural body movements provide more intuitive form of interaction than the more traditional interaction methods such as mouse and keyboard (Djajadiningrat et. al., 2004). These findings have led the interaction design to support natural body movements, which are recognized with gesture-recognition technologies and are then translated into commands and interactions in applications.

Gesture-recognition technologies such as Microsoft Kinect and Nintendo Wii were recognized as very promising future trends in the field of educational applications. They have already been used extensively for research, including the following topics:

- Exergames - using embodied interaction in physically activating games (Kiili et al. 2012),
- Groups with special needs (Abirached et al 2012),
- Mathematics and natural sciences (Johnson et al. 2013),
- Foreign language learning (Searson et al. 2012; Cooper et al. 2013; Edge et al. 2013).

In foreign language learning, these gesture-based user interfaces typically emphasize the ability to use body movements in addition to speech. For example, in *SpatialEase* game by Edge et al. (2013), the users respond by moving their bodies for executing simple audio movement commands issued in the foreign language, and in the case of incorrect body movement patterns, they receive corrective feedback from the system. Cooper et al. (2013) developed *The Virtual Interviewing System* for helping students to develop both their verbal and non-verbal communicational skills in English and there are plans to include coordination of the gestures with speech.

A project similar to ours was suggested by al-Issa (2013). This application combines Google Earth Street View with Kinect and by using gestures and body movements, the users can move around the landscapes, as well as zoom in or out in the environment while standing in front of a projection. Al-Issa's application uses a quite similar approach than our system, although it is a more of an explorative system without having a clear task, end point or a goal. The biggest difference is that our approach is social constructionistic (cf eg. Montola 2012, Oksanen 2014): two learners collaborate and communicate together and the result of their adventure depends on their interaction. Regardless, we agree with al-Issa (2013) that this kind of a "project is relevant because it brings part of that immersion experience to the student, where he can experience foreign countries in an authentic life-like way that engages him visually, audibly and kinesthetically in the environment. Moreover, this project is relevant because it changes the classroom environment from a teacher based to a learner based one that challenges the innovation of the learner and his ability using technologies that simulate game environments."

Berlin Kompass Application

A collaborative virtual environment called *Berlin Kompass* was developed for the evaluation. In this application, two remotely located users are communicating and trying to collaborate to find their way through sequential 360 degree panorama images. One of the users takes the role of a tourist and the other acts as a guide and the guide instructs the tourist along the route until they find the goal. Each user has their own view in the application and they can look around freely but only the tourist can move along on the route based on the guide's instructions.

Interaction Design

Berlin Kompass supports embodied interaction – the users can pan the view by turning their shoulders to left or right to look around the panorama. This emulates the turning of the upper body when looking around their surroundings. The tourist can move along on the route once he/she finds a green arrow which indicates an exit. This movement is done by actually walking towards the screen and it takes both users to the next panorama. The user interaction is detected with the Kinect device.

The user can also use pointing gesture at different hotspot objects that can be found in the panoramas. These objects give out dialogue support and vocabulary and they are always overlaid on landmarks. Once a hotspot has been activated, a spoken utterance is played and this information is also shown as text to support the adoption of the term. These utterances are output via speech synthesis. The hotspot information can be a single noun or adjective about the target or a longer description that can be then used to describe the surroundings to the other user.



Figure 1. Berlin Kompass application used in laboratory environment.

The application view is shown as a projection in the front of the user (see Figure 1). The default field of view for the panorama images was 90 degrees but this could be extended by utilizing multiple displays. Based on Polys et al (2005), a larger field of view is more efficient with search-based tasks and in this study it was also perceived as more satisfactory than interfaces with lower field of view. In Figure 1, a field of view of 160 degrees was used.

Berlin Kompass provides a reasonably realistic virtual environment by providing photographic panorama images of real world geographic locations. Combined together with motivating tasks that encourages the users communicate and collaborate with each other it provides fairly effective way for language learning and wayfinding studies, especially when combined with embodied interaction and a large projection or screen size. Large projection was selected as it is proven to improve performance in spatial tasks (Tan et al., 2006).

Wayfinding Scenario

In the evaluation scenario one of the users takes a role of a tourist who has just arrived to a new city and is interested in finding a local tourist attraction. To find their way, they need guidance from the second user who acts as a guide and knows which way the tourist needs to go. Only the tourist can move along the panoramas and moving takes both users to the next panorama. There are on average four exits in each panorama of which only one is the correct one (takes the users closer to the goal). The incorrect exits take both users to a dead end panorama, in which the tourist needs to describe their location and the guide needs to find a correct location out from four possibilities (See Figure 2). After finding the correct location, the guide activates it by walking towards the screen after which the tourist can activate the exit on their end which takes the both users back to the previous panorama.

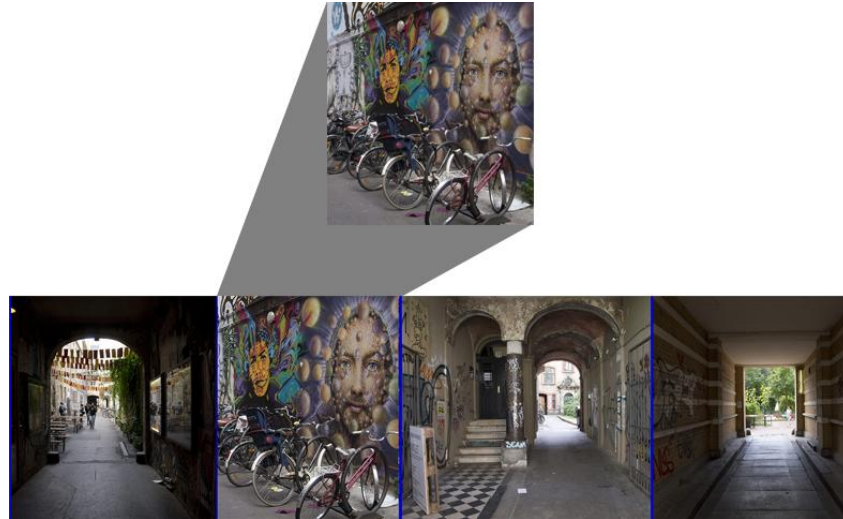


Figure 2. In the dead end situations the tourist describes their location so that the guide can select the correct image.

Method

Participants

We recruited a total of 99 participants (43 females, 56 males, $Age = 16.50$, $SD = 1.02$) who were Finnish students studying German at the upper secondary school level (16-19 year olds). Most of the evaluations were done during school hours, but those students who participated after lessons were compensated with a movie ticket (worth about 12 €). All participants were naïve as to the purpose of the evaluation.

Procedure

In the evaluation scenario the participants work in pairs, as one of the users takes the role of a tourist while the other one acts as a guide. It was up to the participants to decide which role they want to take. The users communicated to each other via audio connection and the guide gave out instructions for the tourist on how to go forward along the route. The goal of the task was to find one of the tourist attractions (DDR Museum, Hackesche Höfe or Pergamon Museum). There was no time limit for the task. The route selection was done by the guide at the beginning of the task after agreeing on it with the user who acts as a tourist. Both users were given instructions on how to interact with the system in the first panorama. Researchers were actively observing the participants during the task.

Data collection and analysis

All participants filled out a web questionnaire after they had completed the task. This questionnaire consisted of questions with Likert scale, containing seven ordered levels (strongly disagree to strongly agree) with 4 (neither agree nor disagree) in the middle, and open-ended questions about the positive and negative features of the application. The participants also answered to questions about their video game experience and frequency of playing video games. This questionnaire data was then analyzed using the repeated measures MANOVA. We observed the use of the system and recorded audio and video recordings of the participants which were then observed later again for further findings. Through this data analysis, we identified several main result themes.

Results

The reported results are combined from the questionnaire data and observations done during the evaluations. In some interesting cases the observations and findings were supported by the audio and video recordings. For the questionnaire feedback, the results concentrate on general feedback and comments on the system features and user experiences, and statistical significance between video game experience and other affecting factors.

Questionnaire Feedback

A repeated measures MANOVA revealed significant effect of video game experience on how well the users adopted the gesture-based interaction within the application, $F(4, 78) = 2.5$, $p < 0.05$, and on how immersed the users felt during the task, $F(4, 78) = 6.23$, $p < 0.01$. Although, there was no statistical significance on the gesture adoption ($F(1, 96) = 1.34$, $p > 0.05$) or with the feeling of immersion ($F(1, 96) = 1.34$, $p > 0.05$) between those who played video games

occasionally and those who never played video games. Another repeated measures ANOVA was used to analyze the effects of video game experience and audio intelligibility in both speech synthesis output and communication between the two users. There was statistical significance between video game experience and how well the participant understood the speech synthesis output from hotspots, $F(4, 78) = 3.42, p < 0.05$ and also how pleasant these outputs were perceived, $F(4, 78) = 3.01, p < 0.05$.

77 % of the participants chose an option that they believe that the application promoted their language learning ($M = 6, SD = 1.22$). 14 % neither agreed nor disagreed with this statement. 88 % agreed with the statement that the application is pleasant to use ($M = 6, SD = 1.29$), and 82 % agreed with the statement that the functional principle of the application is clear ($M = 5, SD = 1.30$), 73 % agreed with the statement that the application is easy to use ($M = 6, SD = 1.57$) and 73 % agreed with the statement that the application is easy to learn ($M = 6, SD = 1.1$).

93 % agreed with the statement that the application is entertaining ($M = 6, SD = 1.1$) and 83 % agreed with the statement that using the application is useful for learning ($M = 6, SD = 1.35$). 83 % of the participants would like to use the application again ($M = 6, SD = 1.44$). In addition, 93 % of the users liked the idea of the application ($M = 6, SD = 0.98$), and additionally 50 % of the participants strongly agreed with this statement. 75 % of the participants who completed the task agreed with the statement "When I was guided, I felt like I was actually moving on the streets". For the guides, this statement was "When I was guiding the other user on the route, I felt like they were actually moving on the streets". 63 % of the guides agreed with this statement.

Subjective Feedback

Of subjective feedback, we report some of the more frequent replies about positive and negative feedback found in the answers from open-ended questions. All of the negative feedback was targeted towards the technical qualities of the application (audio transmission, gesture interaction). Translations in the following examples from Finnish to English were done by the researchers.

Authenticity

Several participants stated that the photographic panoramas made the language use and interaction scenario more realistic. Authenticity was mentioned often times in the subjective feedback and it was connected to both panoramic images based on real world and the wayfinding task in general:

- "Using language in authentic situation when the other user is 'present in the same image' and the cityscape is realistic" ("Positiivista on, että oppii käyttämään kieltä todentuntuisessa tilanteessa kun toinen henkilö on oikeasti "läsnä samassa kuvassa" ja maisema on aito.")
- "Nice, authentic cityscape with enough details [easy to describe the environment]" ("Mukavat, todentuntuiset maisemat, joissa tarpeeksi yksityiskohtia [helppo kuvailla ympäristöä].")
- "Graphics were high quality and realistic." ("Grafiikka oli hyvää ja todentuntuista.")
- "Nice, authentic landscapes with enough details (easy to describe your surroundings)." ("Mukavat, todentuntuiset maisemat, joissa tarpeeksi yksityiskohtia [helppo kuvailla ympäristöä].")

Collaboration

Other participants commented on the collaboration and the need to use the foreign language and interact with the other user to complete the task. Some users also mentioned the need to improvise with language use during the task:

- "One must interact in foreign language in order to guide the other user. Necessity motivates to interact or else the game does not progress." ("Kieltä on pakko käyttää opastaakseen kaveria. Pakko motivoi puhuman sillä ilman sitä peli ei etene.")
- "Communication worked well. Different hints in the application were supporting the use. I liked the idea of the application and I feel that it would be pleasant to study languages more with this method." ("Kommunikointi toimi hyvin. Erilaiset vihjeet sovelluksessa toimivat hyvänä tukena. Pidän sovelluksen ideasta ja minusta olisi mielekästä opiskella kieliä enemmänkin tällä tavalla.")
- "These kind of guiding situations are not practiced enough, even though they are useful. And you also have to adapt, if you do not remember the correct words." ("Tuollaisia opastustilanteita ei harjoitella kunnolla, vaikka ne ovat hyödyllisiä. Ja joutuu myös soveltamaan, jos ei muista tarkalleen jotain sanaa.")

Embodied interaction

Several participants also commented on the embodied interaction. This interaction, and especially the pointing, was also criticized for being inaccurate:

- “Movement as a modality” (“liikkuminen ohjauksen välineenä”)
- ”Best thing was that you do something besides sitting still.” (“Parasta että tulee tehtyä jotain eikä vain istuttua paikallaan.”)
- “It is fun to interact with the application with your body and the positive thing about it is that you also learn German at the same time.” (“On hauska ohjata sovellusta vartalolla ja positiivista siinä on se, että oppii saksaa samalla.”)
- ”At first it was difficult to perceive the gestures used in the application, but it was successful after practice.” (“Aluksi oli vaikea hahmottaa ohjelman liiketoimintoja, onnistui kuitenkin harjoituksen myötä.”)
- ”I couldn’t lift my arm because it would activate the hotspots in the application.” (“Kättä ei pysty nostamaan, tai se antaa sanavihjeitä sovelluksesta.”)
- “Before you get used to the application it takes attention away from the language.” (“Ennen kuin ohjelmaan tottuu sen käyttö vie huomiota kielestä.”)

Observations

One main finding is the learner-centeredness and flexibility of our concept. The observations revealed how different strategies of approaching the task all resulted in eventually finding the goal. In some cases, the lack of language skills was compensated with more elaborative wayfinding methods, for example using degrees as directional cues. This sort of activity was observed especially in boys. Some users explored the whole panorama and hotspots before they started conversing with their counterpart. Using longer sentences and more complex vocabulary did not necessarily enhance the performance (time-wise or by activating less dead ends), but led to some very interesting conversations between the participants. In those cases where the task was fulfilled quickly with relevant expressions it became more gamelike. Then the language learning procedures are not necessarily less effective but more unconscious. A conscious language learning aspect might have been focused more in those cases, where a longer time was used for exploring the panoramas and studying the content of the hotspots or where longer discussions were needed in order to bring the tourist to the goal. These sort of remarkable variations with task completion strategies do not mostly occur as much in more traditional instructed learning situations (cf Pihkala-Posti et al 2014).

There was also a lot of variance in the interaction with the application. Hotspot activation was more evident with those participants who were not as adept with the language used, but some more advanced language users also still checked the hotspot contents, possibly out of interest or in order to expand their vocabulary. One of the key observation on the task completion was that in addition to language skills, it also required spatial cognition skills, with precision being the key factor. Those language users who were less proficient with the target language but gave precise, clear descriptions of their surroundings often managed to complete the task without making any errors (e.g. ending up in a dead end or going back on the route). Also, immersion-related scenarios were observed during the evaluations. One of the most interesting observations was when one of the participants was looking towards a restaurant which name was not fully visible. She was trying to spell the name of this restaurant to the other user, but the name was covered by a pillar. Trying to read the name, the student turned her/his head aside in order to be able to see behind the said pillar, which of course was not possible in the static image (cf Pihkala-Posti et al 2014).

According to our observations many pupils used the application as if it would have been a game. Some aspects in the interface also led towards this kind of thinking, e.g. a congratulation screen after the users had reached their destination. The meta-discussions for example in situations where a tourist ended up in a dead end show how seriously pupils took the situation and how eager they were to correct the misunderstandings and get along in the route. Even the weaker language users came out of their normal social pressure in the classroom and completed the task. The multimodal approach seemed to create clear advantages enhancing and deepening the learning. This happened through combining oral communication and visual surroundings with embodied action, in this case going in the right or wrong direction and experiencing thereby a satisfaction of coming further to the next panorama or a disappointment of ending up in a dead end. These experiences were quite often verbalized by the users, e.g. through expressions of happiness or frustration. Further meta-discussions about how to correct the communication in order to find a problem solution took place. This kind of behavior is a sign of gameful thinking about the task and the learning environment. It is not usual to do this kind of things in a normal language class e.g. after have finished a way finding dialogue.

Discussion

The results from this study indicate that those with more video game experience felt that they adopted the interaction paradigm more easily than their counterparts with less gaming experience and that they also felt more immersed during the execution of the wayfinding task. One interesting finding was the statistical significance between video game experience and how well the participants understood the speech synthesis audio output by the application and also how pleasantly these output felt to the user. This could be explained by previous experience with video games that use speech synthesis as audio output (instead of using for example actual voice actors). This issue needs further investigation and for future studies for example experience with other speech synthesis software (e.g. the ones used in smartphones) need to be taken into account. As the video game experience was a significant factor in how the participants adopted the interaction techniques of the application and also how immersed they felt during the task, we can imply that many of the participants experienced *Berlin Kompass* as a game instead of research apparatus.

Based on the questionnaire results, the application was well received. These results indicate that the application can be effectively used for language learning tasks which are related to description of complex visual surroundings, problem solving, wayfinding and location-related conversations which are practiced in classroom situations. The questionnaire feedback also indicates that the level of immersion in the application is sufficient. Especially those participants who were acting as the tourist and actually moved along the route felt immersed during the task. One explanation for the difference in the level of immersion between these two roles could be in the gesture and interaction design applied in the application: the tourist uses more gestures during the task and has more active role in general, whereas the guide acts more as an observer. This finding is also something that requires further research. Something that needs to be taken into account with the questionnaire results is the novelty of the application: with an exception of a group of 10 students who participated in an early pilot, all participants were naïve to *Berlin Kompass* before the evaluation.

The pedagogical concept worked well: the pupils mentioned in their open-ended feedback answers central aspects of the pedagogical concept without knowing about them in advance (cf. Pihkala-Posti et al 2014). As previously mentioned in the results, the biggest challenges we had during the evaluation were technical. One of the biggest occurrences with gesture-based interaction we experienced was with the hand gestures. Several users were making movements with their arms while conversing which led into hotspots being activated unintentionally, thus activating the audio output. The audio output was also reported as being disruptive with the conversation, which could also be the result of accidental hotspot activations.

Overall, the collaboration with other participants got the most positive feedback in the subjective, open-ended questions. As the application requires the users to converse in order to complete the task, it is crucial that this conversation and its necessity are considered as positive traits by the participants. The results show that this was the case in this study, which encourages the further development of the application.

Conclusion and future work

The impression of the majority of the participants who took part in this evaluation is that *Berlin Kompass* offers a fresh approach for training oral communication in a foreign language, despite of their earlier computer game experience. This result is very good compared with the satisfaction with other computer applications we have tested with the same pupils including applications for both oral and written communication (cf. Pihkala-Posti Uusi-Mäkelä 2013). This evaluation has shown the pedagogical potential of the application, and our future studies will concentrate on the further development of the concept based on findings and problems that occurred.

One of the biggest challenges with the future development of *Berlin Kompass* is with the mobility: the current system requires a lot of technical apparatuses. This problem will be solved with web-based application which enables the collaborative wayfinding between students in remote locations. This work is already well under way: early pilots for non-embodied version were conducted in November of 2014 and extensive pilots for this new version will be done during 2015. The goal of this new version is to remove any dependencies on external applications (Kinect implementation, being the exception) besides the browser. This client-server based architecture also enables the analysis of the speech and giving the users a real-time feedback based on this analysis. As the subjective feedback and our observations show, the embodied interaction increases the level of immersion during the task, and therefore will also be one of the interaction methods in the upcoming version of the application.

The next pedagogically demanding part is to build up different customized hotspot offers for language learners on different skill levels of the target language. The *Berlin Kompass* interface will be adjusted accordingly this integration by analyzing the task progression and dynamically offering hotspot information based on this analysis. This and the planned use of speech analysis and customized learner feedback will support the learners to develop their skill

optimally in sense of Vygotskian zone of proximal development (cf e.g. Lantolf & Thorne 2006). The gamefulness of the application could also be further developed and researched.

Acknowledgements

We would like to thank the rest of our *Berlin Kompass* development team *Tampere Unit for Computer Human Interaction* with the researchers Pentti Hietala and Sanna Kangas as well as Mikael Uusi-Mäkelä from *Tampere Research Center for Information and Media*. Thanks also go to *Tekes* for financing our project *Active Learning Spaces*.

References

- Abirached, B., Zhang, Y. & Park, J.H. (2012). *Understanding User Needs for Serious Games for Teaching Children with Autism Spectrum Disorders Emotions*. In T. Amiel & B. Wilson (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2012* (pp. 1054-1063).
- Aldrich C. (2005). *Learning by Doing: A Comprehensive Guide to Simulations, Computer Games, and Pedagogy in e-Learning and Other Educational Experiences*. San Fransisco: John Wiley and Sons.
- Al-Issa, M. (2013). *Virtual Language Immersion*. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 3577-3579). Chesapeake, VA: AACE.
- Antle, A.N., Corness, G. and Droumeva, M. (2009) 'Human-computer-intuition? Exploring the cognitive basis for intuition in embodied interaction', *Int. J. Arts and Technology*, Vol. 2, No. 3, pp.235–254.
- Arvaja, M., Häkkinen, P., & Kankaanranta, M. (2008). *Collaborative learning and computer supported collaborative learning environments*. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 267–279). New York: Springer.
- Betts, B. (2013). *Gamification, Meet Gamefulness*. Retrieved January 18, 2015 from <https://www.td.org/Publications/Blogs/Learning-Technologies-Blog/2013/03/Gamification-Meet-Gamefulness>
- Bricken, W. (1990). *Learning in virtual reality*. Technical Report No. HITL-M-90-5. Retrieved 17 December 2014, from <http://www.hitl.washington.edu/publications/m-90-5>
- Chik, A. (2014). *Digital gaming and language learning: Autonomy and community*. *Language Learning & Technology*. 18/2.
- Coffield, F., & Moseley, D., Hall, E. & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning*. A systematic and critical review. London: Learning and Skills Research Centre.
- Cooper, T., Tsukada, A., Naruse, Y. & Yamaguchi, A. (2013). *What about Non-verbal Communication? Teaching Total Communication in EFL Lessons*. In R. McBride & M. Searson (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2013* (pp. 2054-2058).
- Djajadiningrat, T., Wensveen, S., Frens, J. and Overbeeke, K. (2004) 'Tangible products: redressing the balance between appearance and action'. in *Personal and Ubiquitous Computing*, Vol.8, No.5, pp.294—309.
- Dewey, J. (1915). *The School and Society*. Chicago: University of Chicago Press.
- Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011). *From Game Design Elements to Gamefulness: Defining "Gamification"* MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments Pages 9-15 , ACM New York, NY, USA
- Dillenbourg, P. (1999). *Introduction: What do we mean by collaborative learning?* In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1–19). Heerlen: Open Universiteit Nederland.
- Edge, D., Kai-Yin C., K-Y., Whitne, M. (2013). *SpatialEase: learning language through body motion*. *Proceedings of CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, Pages 469-472.

- Engelbert, B., Greweling, C. & Morisse, K. (2013). *The Use and Benefit of a Xbox Kinect based Tracking System in a Lecture Recording Service*. In: Jan Herrington et al. (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2013* (pp. 179-184).
- Ermi, L. & Mäyrä, F. (2005). *Fundamental components of gameplay experience: analysing immersion*. *Changing Views: Worlds in Play*. *Proceedings of DiGRA 2005 Conference*. Vancouver, DiGRA, 15–27.
- Gee, J. P. (2007). *What video games have to teach us about learning and literacy*. Revised and updated edition. New York: Palgrave Macmillan.
- Gee, J.P. (2004). *Situated Language and Learning: A Critique of Traditional Schooling*. New York: Routledge
- Gee, J.P. & Hayes, E.R. (2011). *Language and Learning in the Digital Age*. New York, USA: Routledge
- Grein, M. (2013). *Neurodidaktik*. Grundlagen für Sprachenlehrende. Ismaning, Hueber.
- Helsel, S. (1992). *Virtual reality and education*. *Educational Technology*, 32(5), 38–42.
- Johnson, K., Pavleas, J., Chang, J. (2013). *Kinecting to Mathematics through Embodied Interactions*. *IEEE Computer*, 46, 10 (October 2013), 101-104.
- Jonassen, D. H. (1994). *Thinking technology: toward a constructivist design model*. *Educational Technology* 34, 4, 34–37.
- Kaikkonen, P. (2002a). *Authentizität und authentische Erfahrung in einem interkulturellen Fremdsprachenunterricht*. In: *Info DaF*, 29 (1), S. 3–12.
- Kaikkonen, P. (2002b) *Identitätsbildung als Zielvorstellung im interkulturellen Fremdsprachenunterricht*. In: Kohonen, Viljo/Kaikkonen, Pauli (Hg.): *Quo Vadis Foreign Language Education? Reports from the Department of Teacher Education in Tampere University*, Bd. A27, S. 33–44.
- Kallioniemi, P., Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., ... & Raisamo, R. (2013). *Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments*. In *Proc. 12th International Conference on Mobile and Ubiquitous Multimedia* (p. 33). ACM.
- Kallioniemi, P., Pihkala-Posti, L., Hakulinen, J., Uusi-Mäkelä, M., Turunen, M., Hietala, P., Okkonen, J., Kangas, S. & Raisamo, R. (2014). *The Berlin Kompass Language Learning Environment*. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014*, (pp. 1240 -1245). Chesapeake, VA: AACE.
- Kelton, A. (2007). *Second Life: Reaching into the virtual world for real-world learning*. *Educause Center for Applied Research: Research Bulletin*, 2007, 17.
- Kiili, K., Tuomi, P., Perttula, A. (2013). *Exerbraining for schools: Combining body and brain training*. *Procedia Computer Science* 15 (2012) 163 – 173.
- Kolb D.A. (1984). *Experiential Learning – Experience as The Source Learning and Development*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Lantolf, J.P. & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.
- van Lier, L. (1996) *Interaction in the language Curriculum*. Awareness, Autonomy and Authenticity. London: Longman.
- Liimakka, S. (2011). *Ruumis, kokemus, muutos: Ruumiinkokemus Merleau-Pontyn ja Bourdieun avulla tulkittuna*. In Latomaa, Timo & Suorsa, Teemu (eds.): *Kokemuksen tutkimus II: Ymmärtävän psykologian syntyhistoriaa ja kehityslinjoja*. Lapin yliopisto: Lapin yliopistokustannus, pp. 142–168
- Merleau-Ponty, M. (2013). *Phenomenology of Perception*. Routledge: London.

- Mikropoulos, T. A. and Bellou, J. (2006). *The unique features of educational virtual environments*. In P. Isaias, M. McPherson, & F. Banister (Eds.), *Proceedings of e-society 2006*, Vol. 1. IADIS, 122–128.
- Mikropoulos T.A. & Natsis A.(2011). *Educational virtual environments: A ten-year review of empirical research (1999-2009)*. *Comput. Educ.* 56, 3 (April 2011), 769-780.
- Oksanen, K. (2014). *Serious game design: supporting collaborative learning and investigating learners' experiences*. Finnish Institute for Educational Research 31, Jyväskylä.
- Pantelidis, V. (1993). *Virtual reality in the classroom*. *Educational Technology*, 33(4), 23–27.
- Pihkala-Posti L. & Uusi-Mäkelä M. (2013). *Kielenopetuksen tilat muutoksessa. (Language learning environments in transition)*. In: Eija Yli-Panula, Harry Silfverberg & Elina Kouki (Hrsg.) (2014). *Ainedidaktisia tutkimuksia 7 Opettaminen valinkauhassa Ainedidaktinen symposiumi Turussa 15.3.2013*. Turku: Suomen ainedidaktinen tutkimusseura. Available at: <https://helda.helsinki.fi/handle/10138/42530>
- Pihkala-Posti, L., Kallioniemi, P., Uusi-Mäkelä, M., Hietala, P., Hakulinen, J., Turunen, M., Okkonen, J., Kangas, S., Raisamo, R. & Keskinen, T. (2014). *Collaborative Learner Autonomy and Immersion in Embodied Virtual Language Learning Environment*. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014* (pp. 1313-1322). Chesapeake, VA: AACE.
- Pihkala-Posti, L. (2014a). *Innovative training of oral communication: Berlin Kompass*. In S. Jager, L. Bradley, E. J. Meima, & S. Thouësny (Eds), *CALL Design: Principles and Practice; Proceedings of the 2014 EUROCALL Conference, Groningen, The Netherlands* (pp. 312-317). Dublin: Research-publishing.net. doi:10.14705/rpnet.2014.000237
- Pihkala-Posti, L. (2014b). *Kuva, sana, puhe ja liike multimodaalisen kielenoppimistilan elementteinä. (Image, Word, Speech and Movement as Elements of a Multimodal Language Learning Environment)*. In Höglund, Mikko; Kakko, Tommi; Miettinen, Mervi; Parviainen, Hanna; Rickman, Paul (Eds.) *Word and Image: Theoretical and Methodological Approaches*. Tampere Studies in Language, Translation and Literature : B2. Tampere: Tampereen yliopisto 2014, 195-213.
- Pihkala-Posti, L. (2014c). *Die Entwicklung multimodaler interaktiver E-Learning-Konzepte für den Unterricht Deutsch als Fremdsprache*. In: Thomas Tinnefeld (Hg.) unter Mitarbeit von Bürgel, Christoph/ Busch-Lauer, Ines-Andrea/ Kostrzewa, Frank/ Langner, Michael/ Lüger, Heinz-Helmut/ Siepmann, Dirk: *Fremdsprachenunterricht im Spannungsfeld zwischen Sprachwissen und Sprachkönnen*. SAARBRÜCKER SCHRIFTEN ZU LINGUISTIK UND FREMDSPRACHENDIDAKTIK (SSLF); B: Sammelbände; Bd. 2.Herausgegeben von Thomas Tinnefeld. Saarbrücken: htw saar 2014, S. 195-212.
- Pihkala-Posti, L. (in press 2015). *Lernerautonomie und Authentizität durch die Computerapplikation Berlin Kompass*. To be published in *Jahrbuch Deutsch als Fremdsprache*.
- Salzman, M. C., Dede, C., Loftin, R. B., & Chen, J. (1999). *A model for understanding how virtual reality aids complex conceptual learning*. *Presence: Teleoperators and Virtual Environments*, 8(3), 293–316.
- Searson, M., Yan, X. & Shepherd, G. (2012). *'Kinecting' with Student Learning: Applications of Game Consoles in World Language Programs*. In P. Resta (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2012* (pp. 2626-2628).
- Smeets, E. (2005). *Does ICT contribute to powerful learning environments in primary education?* *Computers & Education*, 44, 3, 343–355.
- Steuer, J. (1992). *Defining virtual reality: dimensions determining telepresence*. *Journal of Communication*, 42(4), 73–93.
- Sundqvist, P. & Sylvén, L.K. (2014) *Language-related computer use: Focus on young L2 English learners in Sweden*. *ReCALL / Volume 26 / Issue 01 / January 2014*, pp 3-20. European Association for Computer Assisted Language Learning 2014.
- Sung, H-Y., & Hwang, G-J. (2013). A collaborative game-based learning approach to improving students' learning

performance in science courses. *Computers & Education*, 63 (4), 43–51.

Tan D., Gergle D., Scupelli P. & Pausch R. (2006). *Physically large displays improve performance on spatial tasks*. *ACM Trans. Comput.-Hum. Interact.* 13, 1 (March 2006), 71-99.

Uusikoski, O. (2011). *Playing video games: A waste of time... or not? Exploring the connection between playing video games and English grades*. Department of Modern Languages, University of Helsinki.

Visions 2020: Transforming Education and Training Through Advanced Technologies, Retrieved December 17, 2014 from <http://www.visions2020.gov/>

Winn, W., & Windschitl, M. (2000). *Learning science in virtual environments: the interplay of theory and experience*. *Themes in Education*, 1(4), 373–389.



Publication IV

Kallioniemi, P., Heimonen, T., Turunen, M., Hakulinen, J., Keskinen, T., Pihkala-Posti, L., Okkonen, J., and Raisamo, R. (2015). Collaborative navigation in virtual worlds: How gender and game experience influence user behavior. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST '15)*, Stephen N. Spencer (Ed.). ACM, New York, NY, USA, pp. 173182. doi:10.1145/2821592.2821610

© ACM, 2015. Reprinted with permission.

Collaborative Navigation in Virtual Worlds: How Gender and Game Experience Influence User Behavior

Pekka Kallioniemi¹, Tomi Heimonen¹, Markku Turunen¹, Jaakko Hakulinen¹, Tuuli Keskinen¹, Laura Pihkala-Posti², Jussi Okkonen¹, Roope Raisamo¹

¹TAUCHI, School of Information Sciences, University of Tampere

²School of Language, Translation and Literary Studies, University of Tampere



Figure 1: 360 degree panorama image from collaborative language learning application, Berlin Kompass.

Abstract

There exists a large base of evidence for gender differences in human navigation. However, there is not much research on gender differences in collaborative aspects of navigation, including the interaction of individuals during collaborative wayfinding tasks in virtual environments. In light of this, we present a study of a collaborative virtual environment, Berlin Kompass. The goal of this study was to find out the main differences between genders in collaborative wayfinding. The application was evaluated in the context of foreign language learning in schools with over 200 students, where the users navigated through cityscapes while interacting verbally with each other. We collected and analyzed interaction logs, questionnaire data and audio and video recordings to gain insights into gender-related differences in wayfinding in virtual worlds. Our findings suggest that several differences that are evident in single user systems are not present when the collaborative aspect is added. Male users were more immersed during the task than females. One of the explaining factors for this might be video game experience. Genders also communicated differently – males spoke in longer utterances whereas females had more, shorter utterances. Males referred more to relative directions and dynamic landmarks such as cars and pedestrians while navigating. Males with more video game experience also provided more positive subjective user experience feedback on the application.

CR categories: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities

General Terms: Design, Experimentation, Human Factors

Keywords: Wayfinding, virtual environments, gender differences

1 Introduction

With the technological advances we have seen in the last years, virtual environments have become prominent in several domains, and these applications have been utilized especially for the educational purposes [Mikropoulos & Natsis 2010]. With the sufficient level of realism, the performance in these applications can be comparable to real world situations [Witmer et al. 1996]. Therefore they are useful tools for measuring real life phenomena such as wayfinding in a controlled environment.

It has been reported in several studies that males and females utilize different wayfinding strategies in real world scenarios [Lawton & Morrin 1999]. In spatial cognition, gender differences are generally considered among the largest differences in all cognitive abilities [Halpern 1992]. Men are more likely to refer to directions and distances but women, on the other hand, refer more to landmarks when finding their way in the environment [Miller & Santoni 1986; Ward et al. 1986]. These differences should be taken into consideration when designing both 2D and 3D virtual worlds, since these potentially subtle differences in wayfinding strategies can often be magnified, with males often outperforming females in virtual environment wayfinding [Waller et al. 1998]. Virtual environments have been used for wayfinding studies on several occasions, as they provide a good way to control the environment where the user, or users, are interacting. The representation of space in these applications varies, ranging from simple 3-dimensional models [Astur et al. 1998] to photorealistic panoramic images [Waller et al. 2004]. Interaction techniques in these virtual environments varies from traditional arrow keys on a QWERTY keyboard [Lin et al. 2013] and joysticks [Hurlebaus et al. 2008] to full embodied interaction utilizing the Kinect device [Kallioniemi et al. 2013].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Request permissions from Permissions@acm.org.

VRST '15, November 13 – 15, 2015, Beijing, China.

© 2015 ACM. ISBN 978-1-4503-3990-2/15/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2821592.2821610>

To date there has been very little research on collaborative wayfinding scenarios, and barely any that concentrates on gender differences in said scenarios. Working together with teachers and students of local elementary and high schools, we explored the gender-based differences in collaborative virtual navigation in the context of foreign language learning. This study was designed to look into the differences between males and females during collaborative navigation in virtual environments and whether they use different wayfinding strategies while navigating in the application. In our study the participants completed a collaborative wayfinding task with another student in a virtual learning environment called Berlin Kompass in which the users must collaborate and verbally interact in order to reach the goal by navigating through 360 degree panoramic images (see Figure 1). In this application one user takes the role of a tourist and the other acts as a guide, who helps the tourist along the route until they find the goal.

This study was conducted using mixed methods and for the results we analyzed the questionnaire and log data and transcribed, analyzed and categorized the contents of the audio recordings in order to gain insight about some of the gender-related differences related to wayfinding in virtual worlds. Our main research question for this study is as follows: “What are the main differences between genders in collaborative wayfinding, and more specifically, are there gender differences in interaction patterns during wayfinding task completion?”

Our findings suggest that there are several differences between genders in collaborative wayfinding in virtual environments. Males referred more to relative directions (e.g. “Turn right from the shopping mall”) and also referred more to dynamic objects such as cars and pedestrians during the performance of the task. The VAD (Voice Activity Detection) analysis of audio transcripts showed that males and females spoke close to same amount percentage-wise, but that males spoke in longer utterances on average. From this we can also predict that females had higher number (but shorter in length) of total utterances. The questionnaire data showed that males felt more immersed during the task and had more experience with video games than females. Video game experience also affected the subjective feedback, as those female users who acted as guides and played less video games tended to give more negative feedback on the user experience questionnaire than other user groups. Video game experience and the feeling of immersion could be some of the explaining factors for the differences between genders in collaborative wayfinding tasks in virtual environments.

In the following, we first summarize the related work on this topic, which is then followed by a comprehensive description of our prototype. Next, we introduce the methodology used in our evaluation and then report the both qualitative and quantitative results of the study. We conclude the paper by discussing these findings and how they could be used in designing future applications and prototypes that cater for both genders.

2 Background

Both the general principles of three-dimensional wayfinding and gender-related differences in spatial abilities have been studied extensively. In this chapter we describe the previous research in both of these fields.

2.1 Wayfinding in virtual environments

As testing people for spatial cognition at real locations (particularly with large-scale geographic spaces) can be prohibitively difficult for researchers, they have often relied on more conventional method of paper and pencil test that assessed knowledge of relative directions from the participant's viewpoint. This method has been mostly abandoned because of its inaccuracies and problems in relation to the real world scenarios and has been replaced by experiments done in virtual environments (VEs). Witmer et al. [1996] studied the transfer of knowledge between virtual environments and the real world, and concluded that when a sufficient level of realism is provided, performance in VEs is comparable to the real environment. Another study by Waller et al. [1998] came to the same conclusion and even suggested that training in VEs could be superior to training in real world situations.

It has been found convenient to utilize VEs, for example, in examining directional knowledge [Waller et al. 2004] and assessing spatial abilities [Waller 2005]. In the former study, Waller et al. concluded that computerized assessment allows the investigators to measure more dimensions about the participants' behavior and measure them more precisely and accurately than traditional paper and pencil methods. In addition to the greater fidelity in VEs, they might also be more effective than paper and pencil assessments because they are more interesting and immersive medium with which the participants can engage and interact. The standard for measuring spatial and place learning ability in mammals is the Morris water task, where the subject is required to use the spatial cues outside of a circular pool and swim to a hidden goal platform located in a fixed location. This setting was computerized by Astur et al. [1998] and has been since then used for evaluating different aspects of spatial cognition on humans, including gender differences [Astur et al. 1998], spatial memory impairments [Astur et al. 2002] and effects of exercise on spatial tasks [Herting & Nagel 2012].

As was mentioned earlier, 3-dimensional mazes are often used as virtual environments when evaluating spatial cognition. Lin et al. [2013] used a virtual maze for studying gender differences with local and global landmarks as did Lawton & Morrin [1999] in their research. Waller et al. [Waller et al. 2005], Kallioniemi et al. [2013] and Pihkala-Posti et al. [2014] used photorealistic panoramic environments that referenced real world locations. In addition, these systems used embodied interfaces for the interaction [Kallioniemi et al. 2013] and Head Mounted Displays (HDM) for presenting the visual information during the task [Waller et al. 2004].

Wayfinding in collaborative virtual environment is a rather novel research topic. One reason for this is that only through recent technological development have these collaborative elements become prominent in both research prototypes and computer games [Zagal et al. 2006]. Some earlier studies on the topic can be found - Bruckman [1998] found that game-like VEs enable community support and the development of social interaction and relationships. Kallioniemi et al. [2013] studied collaborative wayfinding in the context of landmark saliency and came to the conclusion that when one is designing and implementing meaningful collaborative wayfinding tasks for virtual environments, careful planning with both the context and contents of the application is required.

2.2 Gender and spatial ability

Kimura [1999] has summarized the known gender differences in spatial abilities and wayfinding strategies, and most of the results indicate male advantage in spatial tasks. A meta-analysis of 286 studies conducted on gender differences in spatial abilities [Voyer et al. 1995] showed that there are significant differences between sexes when evaluating mental rotation (the ability to rotate quickly and accurately 2- or 3-dimensional figures in imagination), spatial perception (the ability to determine spatial relations) and spatial visualization (the ability to manipulate complex spatial information in several stages). Out of these 286 studies, male advantage was reported in 78 studies of mental rotation, 92 studies of spatial perception and 116 studies of spatial visualization.

More recent studies suggest that these differences are even larger when the wayfinding task is evaluated in virtual environments [Astur et al. 1998]. Common method for evaluating spatial cognition in virtual environments is virtual mazes, but they favor males heavily because these mazes often rely on geometrical navigation, rather than landmark cues. Sandstrom et al. [1998] and Waller et al. [1998] have concluded that women rely heavily on landmark-based navigation, whereas men use both structural characteristics and landmarks as wayfinding cues. Gender differences are also evident when studying pointing accuracy in both indoors and outdoors – number of studies have shown that males are more accurate in pointing than females, with mean difference in error of pointing ranging from 4 to 40 degrees [Holding & Holding 1989; Lawton 1996]. Pointing accuracy is closely related to the orientation wayfinding strategies where the users rely more on directional relationships of different landmarks in the scenery. Orientation wayfinding strategies are often favored by men [Lawton 1996]. Pointing accuracy may also play a critical role in less restricted and unfamiliar environment, for example an unfamiliar city (or cityscape in virtual context) [Lawton & Morrin 1999].

These differences are reduced according to the task requirements. In a study by Bia et al. [1997], female participants responded faster than their male counterparts in a 2D matrix navigation task when landmark instructions were provided. Previous research by Levy et al. [2005] seems to show no gender differences in the use of different spatial strategies when navigating through water or



Figure 2: *Berlin Kompass application used by the researcher in laboratory environment. In this scenario the researcher is pointing at a hotspot which is highlighted in red on the screen.*

radial arm virtual mazes.

3 Berlin Kompass application

We implemented a collaborative virtual environment called Berlin Kompass for the evaluation. In Berlin Kompass, two remotely located users are communicating via audio connection and collaborating in order to find their way through sequential 360 degree panorama images. During the task, one of the users takes the role of a tourist while the other user acts as the guide, trying to instruct the tourist along the route until they reach the goal. Each user has their own view in the application and they can look around freely but only the tourist can move along on the route in accordance with the guide's instructions. The application view is shown as a projection in the front of the user (see Figure 2). In this chapter we describe the interaction design and system architecture of Berlin Kompass and finally describe the wayfinding scenario used in the application.

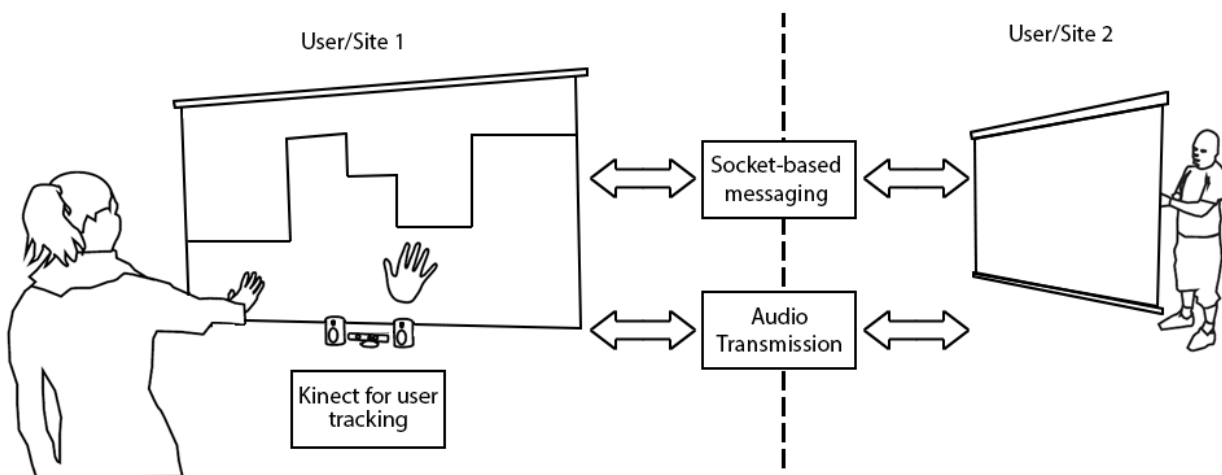


Figure 3: *Application architecture. Users are in remote locations and interact via audio connection. The application states are transferred via socket-based messaging.*



Figure 4: Participants using the application. The user on the left is acting as the guide (the correct route is indicated with the line) and the one on the right is going through the route in the role of a tourist. The participants are communicating via audio connection.

3.1 Interaction Design

A general overview of the Berlin Kompass application architecture can be seen in Figure 3. Interaction with the system is done with the user's body. The user can pan the panorama view by turning their shoulders to either left or right to look around the scenery. This gesture emulates the turning of the upper body when one is looking around their surroundings. The user who acts as a tourist can move to the next panorama once he/she finds a green arrow which indicates an exit. This movement is done by walking towards the screen. By activating one of these exits both users are moved to the next panorama. The user interaction and gestures are detected with the Kinect device which is positioned in the front of the user.

The application also detects pointing gestures. Once the user points at the screen, a hand icon is shown indicating the pointing location. The user can point at different hotspot objects that are scattered around the panoramas. These hotspots give out dialogue support and vocabulary for the users and they are always overlaid on landmarks such as store fronts, billboards and residential buildings. Once a hotspot is activated, a synthesized voice is played. This information is also shown as a text to support the adoption of the given term. The hotspot information is usually a description of the pointed landmark and it can be a single noun or adjective or a longer description. The default field of view used for the panorama view was 90 degrees but this could be extended by utilizing multiple screens or projections. Berlin Kompass application provides a reasonably realistic virtual environment by presenting photographic panorama images of real world geographic locations. When combined together with motivating tasks that encourages the users communicate and collaborate with each other it provides an effective method for language learning and wayfinding studies, especially when combined together with embodied interaction and a large projection or screen size.

3.2 System Architecture

The Berlin Kompass application is composed of four components: the Central Logic, Graphics and Voice Service, Kinect Service,

and Audio Transmission Service. The Central Logic controls the overall program logic and receives and sends messages from and to all the other services. It also handles the communication with the other user's instance of the application, sending messages whenever exits should be activated.

For the Kinect Service the Microsoft Kinect SDK is used. This service provides the physical locations and skeletal joint data of tracked users. The skeletal data is then used to control panning gestures while location of the user is used to active exits. For the pointing a relative method is used where the hand location of the user relative to the body is translated into screen coordinates. For this a physical interaction zone from the Kinect SDK was used [Vassigh et al., 2011] The Audio Transmission service can be used to record audio from any local microphone. This is then sent as UDP packages to the other site where it is played back. The Graphics and Voice Service is built on top of Panda3D graphics engine and it displays cylindrical panoramic images with 90 degree field of view [Kallioniemi et al., 2013].

3.3 Wayfinding Scenario

In the application one of the users takes a role of a tourist who has just arrived to a new and unknown city and wants to find a local tourist attraction. In the beginning the tourist can select one out of the three routes (DDR Museum, Hackescher Höfe or Pergamon museum). All three routes share same panoramas, but Hackescher Höfe and Pergamon museum routes have two extra panoramas. In order for the tourist to find their way to these attractions, they need to ask guidance from the second user who works as the guide and knows the way to the location where the tourist needs to go. Figure 4 shows both user roles in an actual use scenario. The users can communicate freely with each other using with a predetermined language via headsets. Only the user who plays as tourist can move along the panoramas and once they activate an exit, both users are taken to the next panorama. There are three to four exits in each panorama and only one of them is the correct one (takes the users closer to the goal).



Figure 5: In the dead end situations the tourist describes their location (above) so that the guide can select the correct image out of four different options (below).

If an incorrect exit is chosen, both users are moved to a dead end panorama, in which the tourist needs to describe their surroundings and the guide needs to find a correct location out of four different images (See Figure 5). After finding the correct location, the guide can activate it by walking towards the screen after which both users are taken back to the previous panorama.

Method

4.1 Participants

95 females (16 ± 8 years old) and 111 males (16 ± 3 years old) students participated in the evaluation after providing informed consent. The participants were from different levels of education, including elementary school, high school and university levels. The experiments were conducted as part of language learning curriculum and those students that attended outside school hours received a movie ticket as compensation. Of the total of 103 pairs, 36 pairs were female-female, 46 were male-male and 21 were mixed gender.

4.2 Procedure

In the evaluation scenario the participants work in pairs and one of the users takes the role of a tourist while the other one acts as a guide. It was up to the participants to decide whom they pair up with and which role they want to take. The users communicated to each other via audio connection and the guide gave out instructions for the tourist on how to go forward along the route. The goal of the task was to find one of the tourist attractions. For the application, three different routes were created: DDR Museum, Hackesche Höfe or Pergamon Museum. There was no time limit for the task. The route selection was done by the guide at the beginning of the task after agreeing on it with the user who acts as a tourist. Both users were given instructions on how to interact with the system in the first panorama and they could also practice the embodied interaction with the system before starting the actual task. The dead end scenarios were described in more detail to the participants only when they made an actual error on route progression and found themselves in a dead end. Evaluations were carried out in two normal, empty classrooms and only the participants and a researcher were present during the task.

4.3 Data collection and analysis

All participants filled out a web-based subjective feedback after they had completed the task. The content of the questionnaire was based on the SUXES [Turunen et al. 2009] which is a user experience evaluation method for collecting subjective user feedback of multimodal systems. The questionnaire consisted of 9

UX related claims to which the participants responded on a 7-point Likert scale, where 1 = “strongly disagree”, 4 = “neither agree nor disagree” and 7 = “strongly agree”. In addition, the questionnaire consisted of questions about video game experience with a custom scale (with question “How often do you use computer to play video games?” and with answers 1 = Daily, 2 = Weekly, 3 = Every month, 4 = Less than every month, 5 = I do not play video games) and the level of immersion they felt during the task using the same Likert scale as the user experience part of the questionnaire. These results were then analyzed using the two-way analysis of variance.

In addition to collecting subjective feedback, we observed the use of the system and collected audio and video recordings of the participants. The system also logged all interaction by the user during the task, including activating exits and dead ends, pans of the screen, activation of the hotspots and reaching the goal. After investigating the results for individual participants, we analyzed the results for each pair in order to detect any significant effects in male-male, female-female or mixed gender pair interaction during the task. As the routes were of different lengths, the route was considered as a factor in the data analysis.

We also analyzed the audio recordings of 20 randomly selected participants (5 female guides, 5 male guides, 5 female tourists and 5 male tourists) with a VAD (Voice Audio Detection) tool that detected the audio levels. By comparing them to the log data we could detect occurrences of speech and silence during the task and also in each individual panorama. The first panorama and first dead end were removed from the analysis of each task, since the participants used them for practicing the interaction and they also contained speech and assistance by the researchers. Observations were made during the evaluations and they are reported in the results. These findings were confirmed from the video recordings.

After this we also transcribed the actual dialogues for these participants for better understanding of the collaborative wayfinding process. Landmarks referred to were categorized into five different groups and their occurrences counted for measuring the gender differences in landmark-based wayfinding. One interrater was used for this categorization. The categorization was based on a landmark model by Kallioniemi & Turunen [2012]. This model ranks landmarks based on their saliency in three categories: visual, structural and semantic. As the original model was designed for mobile environments and the categorization was based on metadata from maps, and therefore some changes were made so it fits better for virtual environments. Visual landmarks were divided into two categories: residential and office/store front, semantic landmarks were changed into historical landmarks. In addition dynamic landmarks were added. Categories and their descriptions for the landmarks can be seen in Table 1.

Table 1: Landmark categorization

Category	Description
Residential	Residential buildings, private housing
Office/store front	Offices and store fronts which often have billboards and/or signs on their façade
Structural	Landmarks that have a major role in the structure of the spatial environment. E.g. roads, traffic signs and bridges.
Historical	Landmarks with cultural and historical importance. E.g. churches and castles.
Dynamic	Dynamic objects in the landscape. E.g. humans, cars and bicycles.

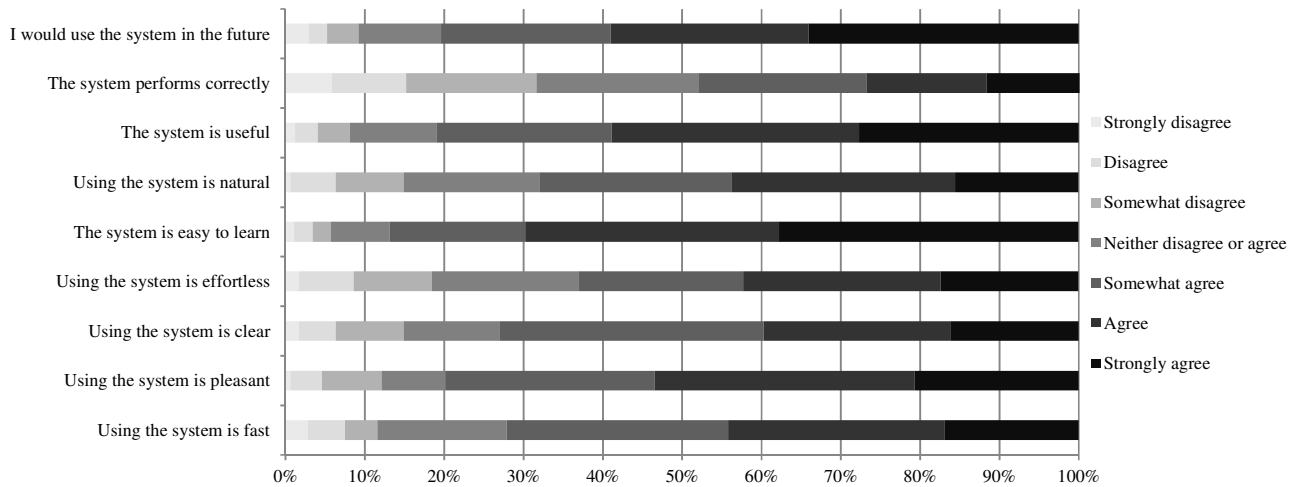


Figure 6: Distribution of users' experiences with Berlin Kompass.

In addition to the landmark information, we transcribed any remarks to directions (left/right), references to degrees (“Turn 180 degrees”) and also occurrences of hesitation/wayfinding anxiety. In this task, wayfinding anxiety was defined by a simultaneous silence of over 15 seconds of both users.

5 Results

The main research interests in this evaluation were the differences between the genders in collaborative virtual wayfinding. In this section we report both quantitative and qualitative results from the study. This includes data from the logs, our observations during the task, VAD analysis of the audio recordings and transcripts of said recordings.

5.1 Questionnaire

The questionnaire consisted of user experience related questions and questions about gaming experience and immersion. Regarding user experience, 88 % agreed with the statement that the application is pleasant to use ($M = 6$, $SD = 1.29$) and 73 % agreed with the statement that the application is easy to use ($M = 6$, $SD = 1.57$). 73 % agreed with the statement that the application is easy to learn ($M = 6$, $SD = 1.1$). See Figure 6 for distribution of users' experiences with the application. Role and gender had no significant effect on these results.

Regarding immersion, 75 % of the participants who completed the task agreed with the following statement: “When I was guided, I felt like I was actually moving on the streets”. For the participants who acted as guides, the statement was “When I was guiding the other user on the route, I felt like they were actually moving on the streets”. 63 % of the participants who acted as guides agreed with this statement. Based on the questionnaire results, males were more immersed than females during the task while they were acting as the tourist, $F(1, 95) = 7.257$, $p < 0.05$ but there was no statistically significant difference for the participants who acted as guides, $F(1, 86) = .015$, $p > 0.05$. As was the case with user experience questionnaire, role and gender had no significant effect on the immersion results.

There was also a statistically significant difference on video game experience, $F(1, 216) = 8.215$, $p < 0.05$, where males ($M = 2.24$,

$SD = 1.174$) spent more time playing games than females ($M = 1.78$, $SD = 0.894$). Video game experience correlated with several subjective feedback questionnaire statements. There was a negative correlation amongst females who acted as the guide between video game experience and the statement “Using the system is pleasant” ($r_s = -0.386$, $N = 37$, $p < 0.05$), the statement “Using the system is clear” ($r_s = -0.373$, $N = 37$, $p < 0.05$), the statement “using the system is natural” ($r_s = -0.334$, $N = 37$, $p < 0.05$) and “using the system is entertaining” ($r_s = -0.400$, $N = 37$, $p < 0.05$). These same correlations were not present with male or female tourists or male guides. With male guides, there was a positive correlation between video game experience and the statement “The system performs correctly” ($r_s = 0.291$, $N = 47$, $p < 0.05$) and with male tourists between video game experience and the statement “I would use the system in the future” ($r_s = 0.307$, $N = 46$, $p < 0.05$).

5.2 Interaction data

We compared the effect of three independent variables (gender, role and route) on the interaction action data collected from the log files. This data consists of total time spent on the task, number of dead ends, number of pans and activation of hotspots during the task. The participants were categorized into three different groups: male tourist with male guide, female tourist with female guide and mixed gender pairs.

There was no statistically significant difference on total time spent on the task between all male groups ($M = 432.52$, $SD = 48.269$), all female groups ($M = 552.95$, $SD = 57.516$) and mixed gender groups ($M = 593.11$, $SD = 70.299$). The average total time spent by each group can be seen in Figure 7. Also the selected route had no significant effect on these results. There was no statistically significant difference between the groups on the number of dead ends (male with male: $M = 0.78$, $SD = 0.964$, female with female: $M = 0.92$, $SD = 1.628$, mixed gender: $M = 1.52$, $SD = 1.750$), pans (male with male: $M = 63.69$, $SD = 33.306$, female with female: $M = 66.22$, $SD = 43.975$, mixed gender: $M = 71.05$, $SD = 26.150$), or activation of hotspots during the task (male with male: $M = 20.89$, $SD = 14.257$, female with female: $M = 23.06$, $SD = 19.613$, mixed gender: $M = 15.00$, $SD = 10.114$).

There was a statistically significant difference on how much the user panned the screen between the roles, $F(1, 216) = 7.221, p < 0.05$, where the tourist panned the image more than the guide. Another statistically significant difference was found on the activation of hotspots: the users acting as guides activated more hotspots than their tourist counterparts, $F(1, 216) = 6.650, p < 0.05$, regardless of gender or the route.

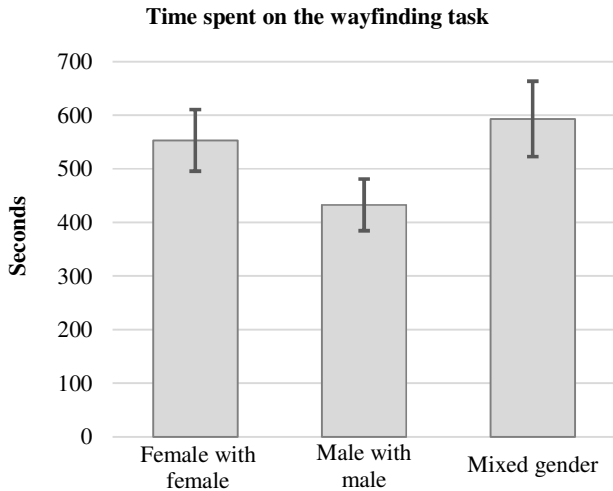


Figure 7: Total time spent on the wayfinding task by group.

5.3 Voice analysis

Voice activity detection was performed for a total of 20 audio files recorded during the task. There was a statistical significant difference on the average length of speech segments between the genders, $F(1, 20) = 4.541, p < 0.05$. Average length of speech segment was 2.9 ($SD = 1.178$) seconds for males and 1.7 ($SD = 0.719$) seconds for females for a total average of 2.3 ($SD = 1.34$) seconds (see Figure 8.). There was no statistically significant difference on total percentage of speech during the task between genders. On average, males spoke 35 ($SD = 12.3$) percent of the total time spent on task and females 33.1 ($SD = 15.99$) percent of the total time spent on the task for a total average of 34.1 ($SD = 13.92$) percent for both genders. Either age or role had no significant effect on the total percentage of speech.

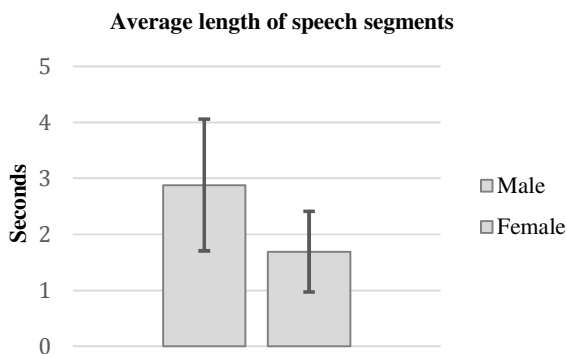


Figure 8: Average length of speech segments by gender.

5.4 Audio content analysis

The contents of the audio files were also transcribed and categorized. Males ($M = 5.7, SD = 3.860$) used more remarks to directions than females ($M = 1.5, SD = 2.121$), $F(1, 20) = 8.545, p < 0.05$, and males ($M = 4.2, SD = 0.787$) also referred more to dynamic landmarks than females ($M = 1.9, SD = 0.787$), $F(1, 20) = 5.395, p < 0.05$. There was no statistically significant difference in other landmark categories between genders. 30 percent of males used degrees as cues for wayfinding, whereas none of the females did this. Wayfinding anxiety was experienced by 20 percent of analyzed participants from each gender.

5.5 Observations

Based on our observations on the audio content, both males and females used mostly full sentences while describing the virtual environment or while asking for directions and cues. Male participants were active during the initial phase and some started interacting with the system and activating exits before the researcher was finished with the introductory part of the study. There were several ways of coping with wayfinding anxiety, for example asking help from the researcher or using the native language for communication. From total of 20 participants' whose audio content were analyzed, a total of 8 users resorted to using their native language during the task. In most cases, the guide was the more active participant, describing the environment surrounding the correct route to the user acting as tourist, but in 3 of these scenarios the tourist took more active role whereas the guide contented to single word utterances (often "Yes" or "No").

When moving to a new panorama, 4 out of 20 users panned through the whole panorama before continuing the conversation with their counterpart and all of these users were female. Some users used longer and more elaborate vocabulary while conversing which did not necessarily enhance the performance time-wise, but led to some interesting dialogue between the participants. Based on our observations, male participants considered the task more "game-like", and tried to accomplish it as quickly as possible, whereas females concentrated more on the conversation with the other user.

6 Discussion

The results of the study suggest that there exists differences between genders in collaborative virtual wayfinding. Some of these findings are similar to the ones observed in wayfinding tasks performed individually [Astur et al. 1998; Voyer et al. 1995]. Video game experience seems to be an affecting factor with gender differences on the user experience with collaborative virtual environments. This might also be the explaining factor for the higher feeling of immersion for males than females during the task. Those females who acted as guides and had less video game experience felt more negatively about the user experience factors such as "Using the system is pleasant" and "Using the system is entertaining". This indicates that the lack of video game experience affects the user experience among females but not among males. In addition, those male tourists (who usually are in more active role during the task) agreed more with the statement "I would use the system in the future".

6.1 Gender differences in interaction

There was no difference between gender groups in terms of interaction data, but as the context of the application was language learning, the time spent on the task was not indicative of the quality of the learning experience of the student. Our findings were quite different from previous studies with only one user. Based on a study by Coluccia & Louse [2004], males outperform females in virtual environment wayfinding when they have an opportunity to interact actively with the environment and that this is explained by the higher familiarity with virtual environments among males. Males performed faster on average, although these results were within the margin of error. Lin et al. [2012] stated that “males tended to engage in a more exploratory mode of wayfinding” which can be seen from quicker moves but not necessarily optimal routes. They also state that females adopt more conservative strategies where they make more stops and change their orientation and that this strategy leads to “slower moves but not few detours”. This was evident in our observation data with the subgroup of 20 participants whose performance was observed during the completion of the task and then verified from the video recordings.

The difference on number of pans between user roles can be explained with the design of the system. As the guide knows where she has to lead the tourist, they usually do not pan much after they have found the correct route. In most scenarios the guides described the environment to the users acting as tourists and looked out for vocabulary cues (e.g. “Can you find the souvenir stand?”) which explains the difference in the activation of hotspots.

6.2 Gender differences in communication

Voice activity analysis showed that males spoke in longer utterances on average than females. One of the explaining factors for this could be the immersion. Those immersed during the task might have tendency to communicate more in order to progress. When the results are analyzed in the context of language learning, some interesting observations can be made. A study by Thompson [1975] claimed that there are fundamental differences in the communication between males and females, and that males focus more on competition, status and independence whereas females concentrate more on intimacy and consensus. Our observations support this, as many male users started interacting with the application even before the initial introduction whereas females attempted to maintain a common consensus and strategy with the other user before starting the exploratory part of the task.

As the audio transcripts showed, males referred more to dynamic landmarks during the task. In most of the cases, these were driving or parked cars and in some cases people walking on the street. This is an interesting finding but does not apply to real world scenarios, as these objects are rarely observed twice in the same location and thus are not very good directional cues. Our aim is to create the virtual environment as realistic as possible and therefore these dynamic objects should be removed from the panoramas. Another option would be to create video-based panoramas where these objects were actually moving during the task. In this scenario the videos for both roles should be produced on different times because if the video material is same for both user roles, these objects could still be referred to even though they are moving on the screen.

As the interaction log data shows no evident differences between genders, creating gender-specific guidelines for the interaction is

not necessary. The communication-related differences are something that could be taken into account when designing virtual environments where collaboration and communication are key elements. Especially in the context of education, gender should be an affecting factor when the performance is being evaluated, as there are observed differences in how the groups interact with each other. In general, these applications should cater to both genders.

For example, wayfinding anxiety was prevalent in both genders which support the idea that these applications should have some kind of encouraging mechanism for users who remain quiet for a certain period of time. This feature could be implemented with an avatar or just a general text box that encourages the user to communicate more in order to progress with the task. In order to increase the feeling of immersion, some “game-like” elements like scoring system or story-based scenario could be added to the application. Another important addition would be support for different type of learners. Previous performances of users could be used to determine the type of support the application gives to the users, e.g. emphasizing elements that support spatial ability or the user’s language skills.

Summarization of this study in a form of gender-related design guidelines for collaborative virtual environments is as follows:

- These applications emphasize the type of communication used. For example, in the context of language learning it could encourage females to speak in longer utterances and males to speak more often
- In order to familiarize the users with the application, it should contain a tutorial where they can practice the interaction. This introduction might also increase the user’s level of immersion
- If the system is used for measuring spatial ability and the content consists of photorealistic images, the number of dynamic landmarks (such as cars and pedestrians) should be minimized

In contrast to our original question, “What are the main differences between genders in collaborative wayfinding and are there gender differences in interaction patterns during wayfinding task completion?” we can see some similar patterns to individually performed navigation tasks. Explaining factors for the abundance of male communication might be the higher feeling of immersion or goal-oriented performing from video game experience.

7 Conclusion

This paper studied the gender differences in collaborative wayfinding in virtual environments and then analyzed these results in the light of previous research. Berlin Kompass, a collaborative language learning application was evaluated in schools with over 200 students. We collected and analyzed interaction logs, questionnaire data and audio and video recordings to gain insights into gender-related differences in wayfinding in virtual worlds.

The main findings suggest male groups tend to communicate in longer utterances than their female counterparts in collaborative virtual environments. As males and females communicate close to same amount percentage-wise, this result also indicates that females tend to communicate in shorter utterances but in higher number than males. Males also had higher feeling of immersion while performing wayfinding. Interestingly, those females with

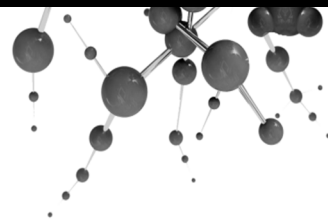
less video game experience who acted as the guide gave more negative user experience feedback on the application than the other user groups (male guides, male and female tourists). These findings also indicate that there is no need for gender-specific user interaction design for collaborative virtual environments, but the systems should be specifically designed to cater both genders by supporting their strengths in these scenarios.

For further studies, a type of gaming experience should be categorized in more detail. This should be done because different type of games might affect the performance and user experience differently (e.g. playing Candy Crush on mobile versus playing Minecraft on PC). Also, new methods of increasing the immersion of the users of both genders should be explored.

8 References

- Astur, R. S., Ortiz, M. L., & Sutherland, R., J. (1998). *A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference*. Behavioural Brain Research, 93, 185-90.
- Astur, R.S., Taylor, L.B., Mamelak, A.N., Philpott, L., Sutherland, R.J. (2002). *Humans with hippocampus damage display severe spatial memory impairments in a virtual Morris water task*. Behavioural Brain Research, Volume 132, Issue 1, 77-84.
- Bia, K., Sewon, L. & Jaesik, L. (2007). *Gender Differences in Spatial Navigation*. International Journal of Social, Education, Economics and Management Engineering Vol 1, No 7.
- Bruckman, A. (1998). *Community support for constructionist learning*. Computer Supported Cooperative Work: The Journal of Collaborative Computing, 7 (1-2), 47-86.
- Coluccia, E., & Louse, G. (2004). *Gender differences in spatial orientation: A review*. Journal of Environmental Psychology, 24(3), 329-340.
- Halpern, D. F. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum. 2nd edition.
- Herting, M., Nagel, B. (2012) *Aerobic fitness relates to learning on a virtual Morris Water Task and hippocampal volume in adolescents*. Behavioural Brain Research, Volume 233, Issue 2, 517-525.
- Holding, C. S., & Holding, D. H. (1989). *Acquisition of route network knowledge by males and females*. The Journal of General Psychology, 116, 29-41.
- Hurlebaus, R., Basten, K., Mallot, H. A., & Wiener, J. M. (2008). *Route learning strategies in a virtual cluttered environment*. Spatial Cognition VI: Learning, Reasoning, and Talking About Space, 5248, 104-120.
- Kallioniemi, P. & Turunen, M. (2012). *Model for Landmark Highlighting in Mobile Web Services*. In Proceedings of The 11th International Conference on Mobile and Ubiquitous Multimedia (MUM 2012). New York: ACM, 25.
- Kallioniemi, P., Hakulinen, J., Keskinen, T., Turunen, M., Heimonen, T., Pihkala-Posti, L., Uusi-Mäkelä, M., Hietala, P., Okkonen, J., Raisamo, R. (2013). *Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments*. In Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia, 33.
- Kimura, D. (1999). *Sex and Cognition*. MIT Press, Cambridge, Mass, pp. 1-66.
- Lawton, C. A. (1996). *Strategies for indoor wayfinding: The role of orientation*. Journal of Environmental Psychology, 16, 137-145.
- Lawton, C. A., & Morrin, K. A. (1999). *Gender differences in pointing accuracy in computer simulated 3D mazes*. Sex Roles, 40(1/2), 73-92.
- Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). *Individual- and gender-related differences in indoor wayfinding*. Environment and Behavior, 28, 204-219.
- Levy, L.J., Robert, S. A. & Karyn, M. F. (2005). *Men and Women Differ in Object Memory but Not Performance of a Virtual Radial Maze*. Behavioral Vol. 119, No. 4, 853-862.
- Lin, C-T., Huang, T-Y., Lin, W-J., Chang, S-Y., Lin, Y-H., Ko, L-W., Hung, D.L., Chang, E.C. (2012). *Gender differences in wayfinding in virtual environments with global or local landmarks*. Journal of Environmental Psychology (Impact Factor: 2.4). 32(2): 89-96.
- Miller, L. K., & Santoni, V. (1986) . *Sex differences in spatial abilities: Strategic and experiential correlates*. Acta Psychologica, 62, 225-235.
- Nicholas, F., Polys, S. K. & Bowman, D.A. (2005). *Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments*. In Proceedings of the ACM symposium on Virtual reality software and technology (VRST '05). ACM, New York, NY, USA, 46-55.
- Pihkala-Posti, L., Kallioniemi, P., Uusi-Mäkelä, M., Hietala, P., Hakulinen, J., Turunen, M., Okkonen, J., Kangas, S., Raisamo, R. & Keskinen, T. (2014). *Collaborative Learner Autonomy and Immersion in Embodied Virtual Language Learning Environment*. In Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2014 (pp. 1313-1322). Chesapeake, VA: AACE.

- Sandstrom, N.J., Kaufman, J. & Huettel, S.A. (1998). *Males and females use different distal cues in a virtual environment navigation task*. *Cognitive Brain Research*, 6, 351-360.
- Tan, D. S., Gergle, D., Scupelli, P. & Pausch, R. (2006). *Physically large displays improve performance on spatial tasks*. *ACM Trans. Comput.-Hum. Interact.* 13, 1 (March 2006), 71-99.
- Thompson, S.K. (1975). *Gender Labels and Early Sex Role Development*. *Child Development* Vol. 46, No. 2 (Jun., 1975), 339-347.
- Vassigh, A.M., Klein, C. & Pennington E.L. (2011). *Physical Interaction Zone for Gesture-based User Interfaces*. US Patent no. 20110193939 A1.
- Voyer, D., Voyer, S., & Bryden, P. (1995). *Magnitude of sex differences in spatial abilities: A meta-analysis consideration of critical variables*. *Psychological Bulletin*, 117, 250-270.
- Waller, D. (2005). *The WALKABOUT: Using Virtual Environments to Assess Large-scale Spatial Abilities*. *Computers in Human Behavior* 21(2), 243–253
- Waller, D., Beall, A. C. and Loomis, J. M. (2004). *Using Virtual Environments to Assess Directional Knowledge*. *Journal of Environmental Psychology* 24(1), 105–116.
- Waller, D., Hunt, E. and Knapp, D. (1998). *The transfer of spatial knowledge in virtual environment training*. Presence: Teleoper. Virtual Environ. 7, 2, 129–143.
- Waller, D., Hunt, E., & Knapp, D. (1998). *The transfer of spatial knowledge in virtual environment training*. Presence: Teleoperators and Virtual Environments, 7(2), 129-143.
- Ward, S. L., Newcombe, N., & Overton, W. F. (1986). *Turn left at the church, or three miles north*. *Environment and Behavior*, 18, 192-213.
- Witmer, B., Bailey, J., Knerr, B., & Parsons, K. (1996). *Virtual spaces and real world places: Transfer of route knowledge*. *International Journal of Human-Computer Studies*, 45, 413-428.
- Zagal, J., Rick, J., & His, I. (2006). *Collaborative games: Lessons learned from board games*. *Simulation & Gaming*, 37 (1), 24–40.



Publication V

Kallioniemi, P., Sharma, S., and Turunen, M. (2016). CityCompass: A collaborative online language learning application. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, pp. 94-97. doi:10.1145/2818052.2874334

© ACM, 2016. Reprinted with permission.

CityCompass: A Collaborative Online Language Learning Application

**Pekka Kallioniemi,
Sumita Sharma and
Markku Turunen**

University of Tampere
Tampere, 33100 FINLAND
Pekka.kallioniemi@sis.uta.fi
Sumita.s.sharma@sis.uta.fi
Markku.turunen@sis.uta.fi

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
CSCW '16 Companion, February 27 - March 02, 2016, San Francisco, CA, USA
ACM 978-1-4503-3950-6/16/02.
<http://dx.doi.org/10.1145/2818052.2874334>

Abstract

CityCompass supports conversational spoken language learning by means of way finding tasks. In CityCompass, two remotely located users, a tourist and a guide, collaboratively navigate in 360 degree panoramic views of a city, to reach a preassigned destination. Over one hundred and fifty students from schools in Germany, Finland and India have used the application for foreign language learning within their classroom activities. The project's goal is to establish a global network of schools to connect students from various cultures and backgrounds for conversational foreign language learning.

The application supports multiple interaction paradigms differing in their level of immersion. CityCompass supports a traditional mouse and keyboard interaction. A previous version of CityCompass designed for the city of Berlin employed embodied interaction using the Microsoft Kinect sensor. Going forward, an immersive virtual reality version, called Amaze360, is being developed. Amaze360 supports 360 degree video panoramas on a mobile phone placed inside virtual reality goggles.

Author Keywords

Collaborative Online Language Learning; Gesture-based Interaction

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction

Even as traditional classrooms evolve into collaborative online learning environments, the focus is on higher education with younger school children lacking exposure to cross-cultural interactions. CityCompass provides an easy to use collaborative interface to understand and study online cross-cultural collaborations with younger school students (aged between 10-15 years). Additionally, new content can be generated to suit several languages and learning needs.

We are already working with students from Germany, Finland and India. Our current work is focused on three key aspects (a) addressing the needs of underprivileged Indian children, who have access to computers but lack relevant educational material and also a global perspective and presence, (b) promoting online cross-cultural collaborations between younger school students and (c) extending the current interaction paradigms to immersive 360-degree virtual learning environments using interactive video panoramas.

CityCompass

CityCompass is a web-based language learning application. It supports remote collaboration for two users for exploring 360 degree panoramic views of a city. Being a web-based application, CityCompass has the potential to connect students from different countries for collaborative problem-solving, with

minimum resources. Thus, it allows easily to setup cross-cultural research.

CityCompass was initially developed as an embodied navigational application that supports full body interaction using the Microsoft Kinect sensor for gesture-recognition. Its pedagogical benefits [3] are well researched with over one hundred students from schools in Germany and Finland. It has also been successfully utilized for studying different wayfinding strategies based on an environment full of landmarks [1 and 2].

In CityCompass, two users, a tourist and a guide, work collaboratively to reach a preassigned destination in a city, using a traditional mouse interface for interaction. The guide sees a blue-marked route in the panoramas, in order to help the tourist navigate the new city, and reach the destination. The route in the application consists of a sequence of 360 degree panoramas of an actual city (e.g., Tampere, New Delhi or Berlin).

A panorama has multiple exits, represented by green arrows, which take both users to another panorama. Each panorama contains several informative audio-visual landmarks, called hotspots, played on mouse hovers. Each user has her own panoramic view with which she can interact freely. The tourist's view has a green arrow to select possible exists. An example hotspot (mouse on the trash can) and green arrow (possible way forward) for a tourist is shown in Figure 1. The guide's view has the actual route marked by a blue line, as shown in Figure 2.

Each panorama has about three to four exits, which are visible to the tourist as green arrows. Of these exits, only one is correct, visible to the guide by the blue line, and it takes the users closer to the goal. The incorrect exits take both users to a dead-end panorama in which the tourist needs to describe her location and the guide needs to find the correct location out of four possibilities.



Figure 1: A tourist's view of Tampere with a trashcan hotspot and one green arrow (exit).

After finding the correct exit and activating it, both users are taken back to the previous panorama. In this way, at each point, for dead-ends and panoramas, the tourist and guide have to communicate to reach the destination, encouraging the need for verbal communication between teammates. Additionally, panoramic views of actual cities further promote the experience and learning of cross-cultural artifacts and environments. A city route, from start to finish, consists of seven panoramas.



Figure 2: A guide's view of Tampere with a McDonald's hotspot and a blue line marking the route.

Technology

City Compass utilizes current web technologies for creating an immersive user experience. For example, `three.js`¹ is used for creating the graphic interfaces, WebRTC² is used for transmitting video and audio between the users, and Node.JS³, Express and MongoDB are used for server-side functionality and transferring data between the users. All components used by the application are executed in the browser stack, which enables better support for different environments and removes the need for external plugins or installations.

Amaze 360

To make the interaction with the system more immersive, we are developing a virtual reality version supporting omni-directional videos. For this, we have

¹ <http://www.threejs.org/>

² <http://www.webrtc.org/>

³ <http://nodejs.org/>

built a prototype application, Amaze 360, which utilizes fully interactive 360-videos for virtual reality headsets and mobile devices. With the 360° video panorama technology, one can record real world panoramic scenes as videos. These scenes can be then applied into an application used with modern mobile devices and virtual reality headsets and the user can explore the panoramas by embodied means. This novel solution has a lot of potential for virtual reality research and it offers a higher sense of presence than static panoramic images.

The benefit utilizing mobile devices and VR headsets compared to for example the Oculus Rift is the mobility and performance – as current smartphones have a large computational capacity, high resolution displays, built-in sensors and fast wireless communications, they have all the required elements for wholesome VR experience.

Conclusion

CityCompass provides a collaborative, immersive and engaging experience for conversational language learning. It currently supports German, Finnish and English, and is easily scalable to other languages. With the CityCompass application, we aim to create a network of schools where students can collaboratively achieve similar language learning goals. Thus, the project aims to provide a common platform for global

collaboration among school students from different cultures and backgrounds.

Acknowledgment

We would like to thank all the student participants and their teachers for their time and efforts in making this research possible.

Publications

1. Pekka Kallioniemi, Jaakko Hakulinen, Tuuli Keskinen, Markku Turunen, Tomi Heimonen, Laura Pihkala-Posti, Mikael Uusi-Mäkelä, Pentti Hietala, Jussi Okkonen, and Roope Raisamo. 2013. Evaluating landmark attraction model in collaborative wayfinding in virtual learning environments. *In Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. Article 33 <http://doi.acm.org/10.1145/2541831.2541849>
2. Pekka Kallioniemi, Pekka, Laura-Pihkala Posti, Jaakko Hakulinen, Markku Turunen, Tuuli Keskinen and Roope Raisamo. 2015. Berlin Kompass: Multimodal Gameful Empowerment for Foreign Language Learning. *Journal of Educational Technology Systems* 43, no. 4. 429-450.
3. Laura Pihkala-Posti, Pekka Kallioniemi, Mikael Uusi-Mäkelä, Pentti Hietala, Jaakko Hakulinen, Markku Turunen, Jussi Okkonen, Sanna Kangas, Roope Raisamo, and Tuuli Keskinen. 2014. Collaborative learner autonomy and immersion in embodied virtual language learning environment. *In World Conference on Educational Multimedia, Hypermedia and Telecommunications* 1:1381-1390.



Publication VI

Kallioniemi, P., Mäkelä, V., Saarinen, S., Turunen, M., Winter, Y., and Istudor, A. (2017). User experience and immersion of interactive omnidirectional videos in CAVE systems and head-mounted displays. In: Bernhaupt, R., Dalvi, G., Joshi, A. K., Balkrishan, D., O'Neill, J., Winckler, M. (Eds.) *Human-Computer Interaction – INTERACT 2017*. Lecture Notes in Computer Science, Vol 10516. Springer: Champagne. doi:10.1007/978-3-319-68059-0_20

© SpringerLink, 2017. Reprinted with permission.

User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays

Pekka Kallioniemi¹, Ville Mäkelä¹, Santeri Saarinen¹, Markku Turunen¹, York Winter², Andrei Istudor²

¹ Tampere Unit for Computer-Human Interaction, University of Tampere, Finland
E-mail: {firstname.lastname@uta.fi}

² Department of Neurobiology, Humboldt-University Berlin, Berlin, Germany
E-mail: {firstname.lastname@charite.de}

Abstract. Omnidirectional video (ODV) is a medium that offers the viewer a 360-degree panoramic video view of the recorded setting. In recent years, various novel platforms for presenting such content have emerged. Many of these applications aim to offer an immersive and interactive experience for the user, but there has been little research on how immersive these solutions actually are. For this study, two interactive ODV (iODV) applications were evaluated: a CAVE system and a head-mounted display (HMD) application. We compared the users' expectations and experience and the level of immersion between these systems. Both indoor and outdoor recorded environments were included. First, the results indicate that the user's experiences with these applications exceed their expectations greatly. Second, the HMD application was found to be more immersive than the CAVE system. Based on the findings of this study, both systems seem to have a great potential for presenting ODV content, thus offering the user an immersive experience for both indoor and outdoor content.

Keywords: Immersion, User Experience, Omnidirectional Video, CAVE, Head-Mounted Displays

1 Introduction

Omnidirectional videos (ODV) have been making their way into the mainstream in the last years. These videos are typically recorded with a set of cameras that cover 360 degrees of the recorded scenery. ODV content has been utilized in several interactive applications, including capturing events such as mountain climbing¹ and musical concerts². As the full contents of these videos cannot be viewed as-is due to the limitations

¹ Mammut #project360. Home page: <https://play.google.com/store/apps/details?id=ch.mammut.project360&hl=en> (Retrieved on 29.7.2015)

² Concert - 360-degree video from ZuZuVideo.
https://www.youtube.com/watch?v=1Kp1_icG328 (Retrieved on 29.7.2015)

in the human field of view, they pose two main design challenges: presentation of the content and interacting with it.

There are several different methods for ODV playback. Often these mediums are some kind of Virtual Reality (VR) applications, ranging from CAVEs (Cave Automatic Virtual Environment) [24] to HMDs [18], but ODV content can also be played with web-based applications (Youtube and Facebook 360 video support) and tablets [33]. In addition to the growing consumer markets, VR applications are used in many domains. For example, they have been found to be a promising tool for treating different kind of phobias, such as acrophobia [6] and agoraphobia [21]. ODV's also have potential in industry use, where they could replace for example 3-dimensional models or content recorded with a single camera, which are often used for demonstrating or training purposes. While numerous interesting solutions and applications exist, thorough understanding of omnidirectional video as a medium and its possibilities in different application domains is yet to be achieved. Our study focuses on iODVs, application that utilize ODV with additional interaction in addition to looking around the scene. This interaction could be, for example, in the form of activating UI elements for more information on different objects in the scene, or transitioning from one ODV scene to another.

One of the most important features of virtual reality applications, also the one's that utilize ODVs, is immersion. For example, in a study by Slater, Alberto and Usoh [27] results indicated that those individuals with a higher sense of immersion achieved better performance overall. The term itself has many definitions in the scientific community, but it is commonly referred to as the feeling of "being there". Our study looked into the differences in the feeling of immersion in two different interactive applications displaying omnidirectional video content – a CAVE system and a HMD application. Both mediums have been studied thoroughly in different contexts but in our study, we wanted to explore these applications further in the context of user experience and immersion. As they are both used extensively, e.g. in industrial use, the results from our study can help in designing future applications. Comparing two different methods of displaying interactive content can be very useful for future designs in this domain. In the two applications we implemented, the user could interact with the environment by activating either exits that took the user to another video or hotspots that offered the user contextual information about the environment. In addition to measuring the sense of immersion, we evaluated the user experience on both applications in order to validate them and to measure the differences in both expectations and experiences between the two systems. The user experience metrics measured the participant's opinion for example on usefulness, pleasantness and clarity of the application. In addition, we compared the different video content types to see if there are any differences between them in the user experience or in the feeling of immersion.

Our main research questions for this study were:

- What are the differences in the user experience between CAVE and HMD applications?

- How immersive are interactive CAVE and HMD applications utilizing omnidirectional videos and are there differences in the level of immersion between these two mediums?

Our findings suggest that the users' experiences exceeded their expectations greatly, especially with the HMD application. The user experience results were very positive in general, and both applications received high scores on the 7-point Likert scale on pleasantness, clearness and performance. One explanation for the contrast between expectations and experiences with the HMD application can be in its "black box" nature, which offers barely any cues on the method of interaction or the overall experience to the user. In the case of CAVE systems, their large size and futuristic look might increase the users' expectations. The positive feedback the HMD application received is also interesting when considering its technical limitations in the presentation of the content: our HMD application offered relatively limited field of view of 60 degrees, which is much more limited than that of the human eye, whereas the CAVE system had no physical limitations on its field of view. Interestingly, none of the users reported this as a limitation.

Regarding immersion, our results indicate that ODV is a very immersive medium. Overall, the HMD application was considered more immersive than the CAVE system with both indoor and outdoor video content. For this difference, we have three explanations: a) HMD obscures the outside world completely from the user, thus allowing them to better focus on the content, b) the sense of depth created by the stereoscopic effect (separate viewports for both eyes), and c) the viewport on the display is based on head orientation, allowing the user to naturally look around.

The motivation for this study stems from the extensive use of CAVE systems in various fields, e.g. in the industry. We argue that HMD systems offer many unique and new application areas requiring immersion, and our results seem to support this argument. The benefits of HMDs come from their portability, as they are often small and mobile, and scalability, as they are less dependent on specific equipment or physical setup. Omnidirectional content could prove useful for example in situations where several people manipulate large objects (such as skylifts) at the same time, as they can show relevant information in multiple directions. CAVE systems also have their uses, for example in situations where the information needs to be presented to multiple persons at the same time.

In the following, we first analyze and summarize the related work in this field of research, which is then followed by a comprehensive description of both applications and their differences. Next, we introduce the methodology used in this study and then report the results of the evaluation along with the discussion on the main findings. We conclude the paper by discussing how these results could be used in designing more immersive interactive ODV applications that offer a better user experience.

2 Related Work

2.1 Immersion in Virtual Environments

The term immersion has many definitions in the scientific community, and there is clearly some discrepancy on what the term actually means. There are no prior evaluations on immersion in interactive ODV applications, and therefore the related work presented below is based on studies on immersion in VR applications. Immersion is an important aspect of virtual reality applications, as it is believed to affect user's behavior with and in these applications [31]. Based on Slater [26], the level of immersion is dependent only on the system's rendering software and display technology. By this definition, immersion is objective and measurable. What some researchers refer to as immersion, Slater defines as presence. According to them, presence is "an individual and context-dependent user response" [26], as in the experience of 'being there'. In short, *immersion* is defined as objective level of sensory fidelity the system provides, whereas *presence* refers to the user's subjective experience and response to the system. Using Slater's definitions, the level of immersion is easier to measure, but restricts the evaluation so that it can be made only on the technological level. This includes only the technical aspects such as field of view (FOV), field of regard (FOR), display size and resolution and the use of stereoscopy. There are several evaluation methods for measuring immersion/presence (based on the definition used), for example the ones by Witmer & Singer [32], Schubert, Friedmann & Regenbrecht [25] and Usoh et al. [30].

Immersion has also been studied extensively in the context of video games, and Brown and Cairns [5] attempted to resolve the disparity with the term. They conducted a qualitative study amongst gamers and talked to them about their experience on playing video games. The study resulted in a grounded theory where immersion was used to describe a person's "degree of involvement with the game". This finding supported the idea that immersion is a cognitive phenomenon. The theory also identified restrictions that could limit the degree of user's involvement, including engagement, engrossment and total immersion.

As the related work shows, immersion can be defined in several ways, depending on many factors such as the emphasis on technology, the research domain and the method of evaluation. With VR related studies, Slater's [26] division of immersion and presence is more prevalent, whereas in video game related studies the term immersion is used more often. In this paper, immersion is referred to as a perceptual phenomenon that is dependent on the individual and the context.

2.2 Omnidirectional Videos

Lot of scientific research has been done to enable the use of omnidirectional video. There exists a large variety of algorithms and devices to capture, construct, project, compress, display and automatically analyze omnidirectional video content.

Application domains, where omnidirectional video has received wider interest include remote operation and telepresence applications [4][20][8], some of which include

automatic situation tracking based on the omnidirectional imagery and directional audio. Another application field identifiable in literature is remote operation of unmanned machines and vehicles, for example drones by using omnidirectional video. Applications where omnidirectional video is used by consumers [17][13][3] provide immersive experiences to cultural contents, e.g., in museums [15][14][19] and theatre [9]. Other application domains include education, e.g., teaching sign language [12], and health care, e.g., relieving stress during medical care [10], and therapy [23]. There has been little research on using ODV in industrial use, for example in demonstrating or training purposes.

From the human-computer interaction perspective, augmenting omnidirectional video with interactive content [2] and UI elements [22] are crucial features in many applications. Another field is multisensory augmentations of video content, e.g., simulated wind [22], to further immerse the viewer and improve sense of presence. Interaction studies have also looked at gesture-based interaction [34][24] and second screen interfaces [33] to interact with omnidirectional video content. For example, Benko and Wilson [1] present the Pinch-the-Sky Dome, which projects a full 360 view of omnidirectional data onto an inner side of a dome-shaped structure. The view is controlled using mid-air gestures from anywhere inside the space, and it supports several simultaneous users. They found that mid-air gestures could enhance immersion in an omnidirectional context.

3 iODV Applications

In this section, we introduce the iODV applications that were built for this study. Both applications used the same ODV content with length of 60 seconds. When the content is finished, it starts again from the beginning. Both applications have two types of user interface elements: exits and hotspots. When activated, an exit takes the user to another video that is linked to that particular exit element, and hotspots provide contextual information about the environment. First, we introduce the video production procedure used for content creation, and then explain the basic features and interaction techniques for both applications. Finally, we compare the main differences between these two applications.

3.1 Video Production

The videos used in this study were recorded with six GoPro 4 cameras attached to a Freedom360 mount on top of a tripod. The resulting six videos from each shot were converted into 4k omnidirectional videos by using AutoPano Video Pro 2 and AutoPano Giga 4 software. Panoramic images and videos are usually divided into either cylindrical (limited vertical field of view – VFOV) or spherical ($360^{\circ} \times 180^{\circ}$) views.

For this study, we produced a total of six videos, three of which were shot indoors, and three in an outdoor environment. Each video was roughly one minute long. Indoor videos were recorded in an industrial hall used for repairing and maintenance of sky-lifts. Each video contains some movement, such as people walking around and working,

and a forklift riding around the hall. Two of the indoor videos were recorded from a top of a ladder to offer a better view of the surroundings. The outdoor videos were recorded in downtown Tampere, Finland. These videos were recorded during quiet hours, but nonetheless contained a relatively large amount of movement, i.e., people walking on the streets.

3.2 cCAVE

For our first experiment, we implemented a multimodal CAVE application, circular CAVE (cCAVE), where the user can explore omnidirectional videos via eight displays set in the form of an octagon. A cylindrical view where the horizontal FOV is 360 degrees and vertical 150 degrees was used in the application. In this system, the user is located at the center of the octagon, sitting on a rotating chair (see Figure 1). The chair has a rotating sensor that sends the rotational axis to the computer. This sensor data is used to update user interface elements on one of the displays, e.g. when the chair is pointing at specific coordinates. The application was developed with Vizard virtual reality software. The omnidirectional video content is then displayed on a 3-dimensional cylinder that is divided between the displays so that each monitor covers 45 degrees of the content.

Each interface element (exits and hotspots mentioned earlier) has a coordinate range (i.e. when the rotating chair is pointed at this range) in which they are shown on the screen. The interface elements are triggered by dwelling, i.e. by focusing an element in the center of the view (by turning the chair towards it) and waiting for five seconds. Dwelling is a relatively common technique for selecting targets with e.g. gaze and mid-air gestures, which is utilized by a number of applications (e.g. [16]). Before the hotspots are activated, they are presented on the screen as blue circles with an exclamation mark inside. Exits are presented as green arrows. During the activation period, the element is scaled up in order to visualize that it is being selected. Users can cancel the activation process by turning away from the element. Similarly, a hotspot dialog is closed by turning away from it.

We used a set of eight Eyevis Eye-LCD 4000 M/W monitors. Each monitor has a screen diagonal of 40 inches with full HD resolution and they were raised 77 cm from the ground. They were 91 cm high, 53 cm wide and 13 cm thick. The bezel between two monitors was 28 mm (14mm in one monitor). These monitors were set up so that they covered an area of 360 degrees around the user. The rotating chair's seat height was adjusted to 50 cm and the distance from the user's head to the monitors was approximately 60 cm. The outer walls of the cCAVE installation were 175 cm wide and 192 cm high. The total resolution for the application was 4320 x 3840 pixels. The monitors were connected to AMD HD 7870 display adapter with 1 GHz processor and 2 gigabytes of GDDR5 memory.



Fig. 1. The cCAVE system. The rotating chair used for interaction is at the center of the system. Eight monitors (only 6 shown in the image) each show 45 degrees of the omnidirectional video content. The two monitors in front are attached to the doors that are opened for entrance and closed during use.

3.3 Amaze360

Amaze360 is an iODV application for HMDs that allows the user to freely observe omnidirectional videos by simply turning one's head in the desired direction. The screen is divided into two separate viewports in order to create a stereoscopic effect, thus creating a sense of depth. This effect is done with the spherical presentation of the video content, as the video content itself is not stereoscopic. The video content used by the application has 360-degree horizontal and 180-degree vertical field of view and the video is projected on a virtual sphere. The viewport's field of view is 60 degrees.

Interface elements (exits and hotspots) in Amaze360 are also triggered by dwelling, but with slight differences. These elements are activated by focusing on an element in the center of the view (by turning the head towards it) and waiting for two seconds. The hotspot and exit icons in Amaze360 are similar to the ones used in cCAVE (blue circle with an exclamation mark inside for the hotspots, and green arrows for the exits). The entire set up and a screenshot of the Amaze360 application with hotspot activation can be seen in Figure 2.

Amaze360 is C# application built on the Unity platform, and it utilizes the Oculus Mobile SDK 1.0.0.0 for iODV features. The application also uses the Easy Movie Texture plugin to enable smooth video playback on mobile devices. It is run on Samsung Note 4 and utilizes the Samsung GEAR headset.



Fig. 2. Top: Amaze360 physical setup. **Bottom:** Amaze360 application view. The video is shown as a stereoscopic presentation. Activated hotspot is shown at the center of screen.

3.4 Differences between the applications

Even though the two applications are intended for the same purpose, there are obvious differences ranging from physical setup and display devices to interaction mechanics. These differences further affected some design choices for both applications. A general overview of the features and differences can be seen in Table 1.

	CAVE	Amaze360
Application Field of View (Horizontal*Vertical, in degrees)	180*150	60*60
Interaction Method	Rotational chair (sensor)	Head/device orientation based activation
Contextual information activation range	X-axis	X- and Y-axis
Contextual information location on the screen	Bottom center	Center
UI Element Activation Time (seconds)	5	2

Table 1. Differences between the two applications

The primary difference between the two applications is in how content is presented – cCAVE shows the ODV in multiple monitors whereas the Amaze360 uses a stereoscopic presentation on a mobile device. In other words, cCAVE always physically displays the full 360-degree view of the content. Therefore, the user sees the content with the full field of view of the human eye. Amaze360, on the other hand, is limited to a 60-degree sector of the content at any given time.

Another major difference is in how the applications are interacted with, i.e. how hotspots and exits are activated. The cCAVE system utilizes the rotation of the chair, and therefore only uses the X axis (chair's rotation relative to the screens) for activating UI elements. Amaze360 relies on head orientation, and hence uses both X and Y axes. For illustration on these differences, see Figure 3.



Fig. 3. The hotspot activation sectors illustrated in both applications. The gray coordinate area represents the coordinate range of hotspots in cCAVE, and the circular area represents the X- and Y- coordinate range used in Amaze360.

Due to the difference in how UI elements are activated, both applications vary in how contextual information is presented. In cCAVE, textual content is shown (when a hotspot is activated) at the bottom of the screen. This design choice was made so that the textual content would not obscure the object it is referring to. In Amaze360, textual information was presented on top of the corresponding hotspot (see Figure 4). This was due to the interaction method: as the user activates hotspots by turning their head towards them, it makes sense that the displayed information is displayed in the same position so that the user does not need to adjust the head once more. Furthermore, this allows closing activated hotspots by turning the head away from them, similar to closing hotspots in cCAVE by rotating the chair to another position.

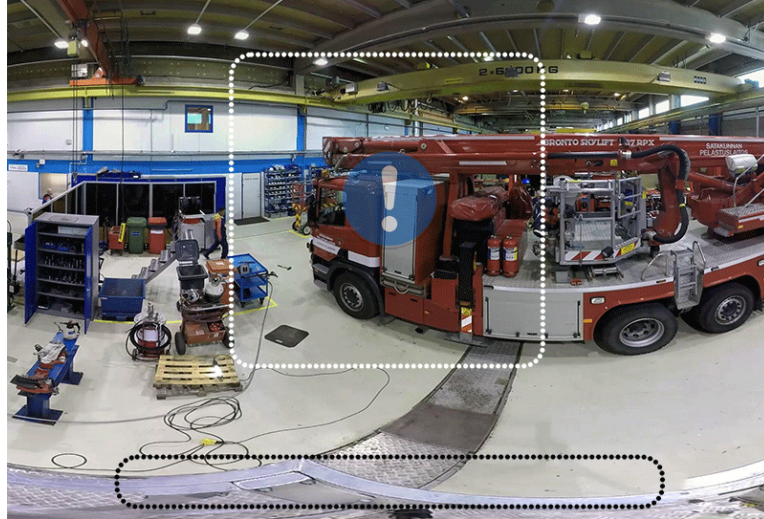


Fig. 4. Hotspot locations in the two applications. HMD hotspot location is presented in white dotted line and CAVE system hotspot location in black dotted line.

Finally, the activation time for UI elements was also different between the applications because of the conclusions made during pilot testing: a short activation time sometimes caused accidental activations in the CAVE system, whereas with Amaze360 these were not as prevalent. This was caused by the slower interaction with the chair – turning one’s head is much faster and more precise than turning on a chair. The pilot tests verified that the Amaze360 application could have a significantly shorter activation time (2 seconds) for the UI elements than the CAVE system (5 seconds).

4 Experiment of CAVE and HMD

For this study, we conducted two separate experiments which evaluated the user experience, level of immersion and spatial abilities in immersive virtual environments that utilize omnidirectional videos. Experiment 1 was conducted with the CAVE system and Experiment 2 was conducted with a HMD and the Amaze360 application.

4.1 Participants

A total of 34 participants took part in the study, both experiments having 17 participants. The cCAVE was evaluated by 8 females and 9 males aged 30.9 on average (SD = 5.46) and the Amaze360 system also by 8 females and 9 males with an average age of 30.7 (SD = 5.43). They were recruited from around a university campus and were compensated with a movie ticket for their participation. All participants were naïve with respect to interacting with omnidirectional videos, as in they had not use CAVE, HMD or other type of applications that utilize these type of videos.

4.2 Procedure

In the evaluation scenario the participants were asked to explore the virtual environments that consist of omnidirectional videos. Both indoor and outdoor environments were presented to the user as separate scenarios (one could not move from inside locations to the outside locations, and vice versa). They could move from one location to another after they had spent thirty seconds in one location. The time limitation was set in order to encourage exploration and looking around the scenery instead of just moving quickly from one scenery to another. Each location also contained two hotspots which, when activated, offered contextual information about the object they were referring to. Both indoor and outdoor video content consisted of three different locations and the last location led the user back to the first one, which made it possible for the participant to explore the locations indefinitely.

No specific tasks were given to the participants because we wanted to emphasize the explorative nature of the experiment. This way the participants could concentrate solely on experiencing the virtual environment. The users could use the system under evaluation as long as they wanted to. They informed the researcher when they were finished with each scenario (indoor and outdoor). Participants used each system (both indoor and outdoor scenarios combined) for approximately 10 minutes on average.

In Experiment 1 the participants used the cCAVE system in a laboratory setting while sitting on the rotating chair. In Experiment 2 they used the Amaze360 application also in a laboratory setting while standing and wearing the HMD. Both locations were approximately the same size. For both experiments, conditions were balanced so that half of the users started using the system in indoor locations and the remaining half in outdoor locations. A researcher was present during the procedure for support in case of a technical fault or other disturbance, but did not otherwise intervene with the evaluation.

4.3 Data collection and analysis

We gathered general information from all participants, including age, gender, and experience level with the iODV applications. For the user experience evaluation, we used the SUXES [29] method. It is an evaluation method for collecting subjective user feedback of multimodal systems. In this method, the participants fill out a subjective feedback form about their expectations and experiences on using the system. The form consisted of 9 user experience related claims to which the participants responded on a 7-point Likert scale, where 1 = "Totally disagree", 4 = "Neither agree nor disagree" and 7 = "Totally agree".

Participants filled the expectations form after the user had been informed of the procedure and had been shown to the basics of the system, but before the user personally experienced the system. Then, after they had used the application, users filled out their experiences on a similar form. In addition, after the experiment, participants answered to question regarding their level of immersion during the use of the system ("While using the system, I felt like I was actually standing on the streets/industrial hall"). The same 7-point Likert scale was used for the questions regarding immersion. We decided

to disregard the existing evaluation methods for measuring immersion for practical reasons – our custom-made questionnaire allows us to compare the results with the UX results for different modalities using the SUXES method [29]. Finally, we logged basic interactions with timestamps in both systems, such as start and end times of the application, activations of hotspots, and movements from one video to another. We also considered adding the Santa Barbara Sense-of-Direction questionnaire [11] to the evaluation, but decided against it as the evaluation itself was not about measuring spatial ability.

5 Results

The main research interests in this study were the feeling of immersion and the user experience with the two applications. In addition, we report the results from logged interaction data. For all results, a Bonferroni-corrected independent t-test was conducted to compare the results between the two systems. Here, we treat the disagree-agree-like scale to be equidistant, which is why the t-test for analyzing the results was used. For the statistical analysis, an average UX score of both indoor and outdoor video content was used.

5.1 Expectations versus Experiences

When comparing the UX results of the two experiments, statistically significant differences were discovered between the expectations and the actual user experience on both applications, especially with the HMD. For average UX ratings on all statements in both systems, see Figure 5.

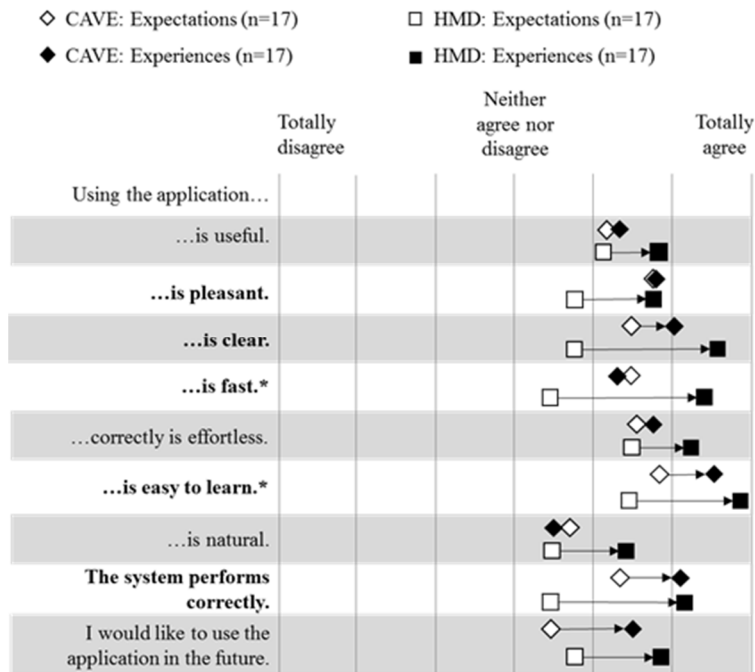


Fig. 5. Average UX ratings for expectations and experiences on both systems. Arrows indicate the direction of the change between expectations and experiences. The statements in bold had statistically significant differences between the applications regarding expectations, and those marked with asterisk in experience.

In almost all metrics the actual use experience exceeded the expectations, especially so with the HMD. Using both systems were considered to be very easy to learn by the participants. All participants except for two in the first experiment and one in the second one agreed (scored either 5, 6 or 7 on the Likert scale) with the statement that the system is useful (Experiment 1, $M = 5.29$, $SD = 1.047$ and Experiment 2, $M = 5.82$, $SD = .883$).

Participants had higher expectations on the cCAVE system used in the first experiment. Statistically significant effects were detected in expectations on pleasantness (Experiment 1, $M = 5.71$, $SD = .920$ and Experiment 2, $M = 4.76$, $SD = 1.251$); $t(32) = 2.499$, $p < 0.05$, and clarity (Experiment 1, $M = 5.53$, $SD = 1.125$ and Experiment 2, $M = 4.71$, $SD = .849$); $t(32) = 2.410$, $p < 0.05$, where the users anticipated more from the CAVE system. cCAVE users were also more optimistic on how fast the system is (Experiment 1, $M = 5.47$, $SD = 1.179$ and Experiment 2, $M = 4.41$, $SD = .939$); $t(32) = 2.896$, $p < 0.05$, and if it performs correctly (Experiment 1, $M = 5.29$, $SD = 1.263$ and Experiment 2, $M = 4.47$, $SD = .874$) was found; $t(32) = 2.210$, $p < 0.05$.

Regarding the user experience, the questionnaire results on both applications were generally positive. With cCAVE, 88 % of the users gave positive feedback (scored either 5, 6 or 7 on the Likert scale) on the system's usefulness. 82 % of the users thought that the system was pleasant to use, and 100 % of the users felt that the use of the system

is easy to learn (where 2.9 % ranked it at 5, 29.5 % ranked it 6 and 67.6 % ranked it at 7 on the Likert scale). The HMD application received even more positive results, where 94 % of the users thought that the system is useful and pleasant to use. Like with cCAVE, all of the HMD users felt that the system is easy to learn.

Comparing the results from the two experiments, some statistically significant findings were discovered. The HMD application ($M = 6.35$, $SD = .862$) was considered to be faster than the cCAVE system ($M = 5.29$, $SD = 1.532$); $t(32) = -2.484$, $p < 0.05$. Participants also felt that the HMD application ($M = 6.88$, $SD = .332$) is easier to learn than cCAVE ($M = 6.53$, $SD = .514$); $t(32) = -2.376$, $p < 0.05$.

5.2 Immersion and System Interaction

The main interest in addition to the user experience was the feeling of immersion experienced during the use. Between the two applications, statistically significant differences were observed with both indoor video content (Experiment 1, $M = 5.18$, $SD = 1.629$ and Experiment 2, $M = 6.18$, $SD = .883$); $t(32) = -2.225$, $p < 0.05$, and outdoor video content (Experiment 1, $M = 5.18$, $SD = 1.510$ and Experiment 2, $M = 6.29$, $SD = .686$); $t(32) = -2.779$, $p < 0.05$. The immersion level of participants for both applications with indoor and outdoor videos can be seen in Figure 6.

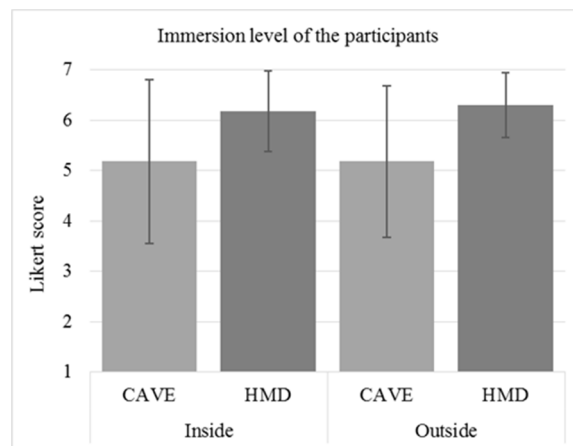


Fig. 6. Average immersion level of the participants in both applications with indoor and outdoor video content

Based on interaction log data, some statistically significant differences in the application use times were observed. cCAVE was used for longer periods of time (in seconds) in total (both outdoor and indoor scenarios combined) than the HMD application (Experiment 1, $M = 884.47$, $SD = 357.91$ and Experiment 2, $M = 561.41$, $SD = 214.52$); $t(32) = 3.193$, $p < 0.05$. Participants also used the CAVE system for longer periods with the indoor video content (Experiment 1, $M = 502.82$, $SD = 303.77$ and Experiment 2, $M = 260.76$, $SD = 92.96$); $t(32) = 3.142$, $p < 0.05$. There was no observed effect with

outdoor video content. The total times spent with both indoor and outdoor video content can be seen in Figure 7.

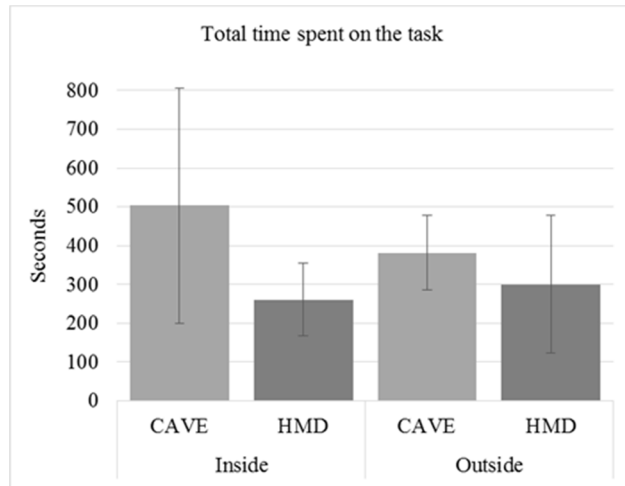


Fig. 7. Total mean time spent on task with indoor and outdoor video content

6 Discussion

6.1 Expectations Versus Experiences

The most interesting finding regarding the user experience was that in almost all metrics the actual use experience exceeded the expectations, especially so in the second experiment with the HMD. The interaction method in the cCAVE experiment can be the reason for the difference in expectations on pleasantness – sitting and interacting with a chair can be expected to be more comfortable for users than standing up while wearing the HMD. In addition, it might be difficult to make any estimates on the pleasantness and clarity on the sort of a “black box” HMD device, which offers no cues on the method of interaction to the user. The cCAVE set up might be more impressive and futuristic looking than HMD devices in general. Another factor to consider is the physical set up of the two applications: cCAVE is a large installation built on a metallic rig with eight monitors, whereas the headset is using the smaller Samsung Gear headset and basic Samsung Note 4 mobile device. The system size difference itself might indicate that the cCAVE is more powerful than the compact HMD device. In addition, a desktop computer can be presumed to have better performance than a smaller mobile device, which might implicate to some participants that the system itself is also better graphics and performance-wise.

The HMD application was considered to be faster than cCAVE, which can be at least partly explained by the interaction method: head turning used with the HMD is much faster to perform than rotating the on cCAVE. As mentioned earlier, both systems were

regarded as easy to learn, but for the HMD this metric was significantly higher. The intuitive method where the viewport is rotated based on the user's head orientation offers the user an efficient way to start interacting with the virtual world immediately after they wear the device. The UI elements draw the user's attention and when they concentrate on these elements, they are activated and animated which again hints the user that something is happening.

The implications of these results are that both CAVE systems and HMD applications utilizing ODVs are regarded as both useful and easy to learn. Both of the applications had very simple interaction methods which were based on dwell-time. This seems to be a meaningful way of interacting with these types of systems, especially when the interaction is kept simple. Nevertheless, more research is required in order to understand the relationship between complex UI elements and different interaction methods.

Overall, the positive feedback on both applications validates their use on this study. The applications were also very robust and had no technical faults during evaluations, which might have also affected the participant's feedback on the user experience. The actual user experience was much more positive than the user's expectations with both systems, but especially so with the HMD application.

6.2 Immersion and User Interaction

Our evaluation suggests that the Amaze360 application is more immersive than the cCAVE with both video content types. There are many possible explanations for this result. First, the headset obscures any other visual stimuli from the view, only showing the contents of the application to the user, whereas in the cCAVE the user can still observe objects outside the screens, including the bezels of the monitors. Second, the HMD provides a stereoscopic effect (coming from the spherical projection, not the video itself) which creates an illusion of depth. This is not provided in cCAVE. Third, since interaction with the Amaze360 is based on head orientation, it does not require any external devices which might enhance the feeling of immersion even further. In the first experiment the aim was to make the interaction as simple as possible with the use of the rotational chair, but it is still not as natural as interaction with the headset. In future implementations a combination of body tracking and gaze tracking could be combined to produce a similar interaction solution as in the HMD application.

Despite the unique advantages of the HMD application, the positive feedback for this application is interesting when considering the current limitations of the technology. For instance, Amaze360 offers a relatively limited 60-degree field of view, which is much smaller than that of the human eye, whereas cCAVE had no such physical limitations. However, none of the users reported this as an issue.

Some cCAVE users had trouble finding the textual content from the bottom of screen even when they were informed about the location beforehand, during the introduction. Participants had no trouble finding or activating the hotspots with the Amaze360 application, but three participants noted that the hotspot text box obscures the visibility of the actual object behind it. One solution for this could be an opaque text box that does not hide the content. These findings indicate that the optimal location for the contextual information is somewhere around the center of the screen where the user is most often

looking at, but that it also should not dominate the viewport. It should also be located close to the actual UI element activating it, so that the user quickly finds it.

One participant using cCAVE remarked that the reflection on the monitors broke the feeling of immersion, as the participant could see the monitors behind him reflected in the monitors in front of him. Four Amaze360 users reported that they felt dizzy during the indoor scenes, which were filmed on a ladder. This interesting finding and its connection to acrophobia could be an interesting topic of research, and has also been looked into by Coelho et. al [6]. None of the participants did not report any motion sickness effects in either applications. Three participants using cCAVE and one user using the Amaze360 stated that the resetting of the omnidirectional video content back to the beginning (due to looping videos) broke the immersion somewhat. In addition, some video production errors that caused distortions were breaking the feeling of immersion for one cCAVE user. These distortions can be eliminated with careful planning of the recording and editing phase of the ODV content. The biggest hurdles in the post-production phase are the color level differences between the cameras, stitching errors where the content between the cameras are not overlapping properly or displaying of the camera equipment in the recording. Also, if the content needs to loop, some attention should be paid to how smoothly the end of the content loops back to the beginning. These problems will likely dissipate once the ODV recording and production technologies advance. When comparing the results between outdoor and indoor video content, there was no significant difference in the feeling of immersion.

The difference in use times with the indoor video content is also an interesting finding. As this same effect was not observed in outdoor environments, one explanation for the difference between the systems could be in the claustrophobic nature of the indoor environment and the limited field of view used in the HMD application. Another explanation could be the filming location of the indoor video content. Two out of three of these videos were recorded from a higher ground, i.e., from a ladder. Four HMD users said that they felt dizzy during these scenes, which might affect the total time used with the indoor videos.

We also note that CAVE systems are diverse and may significantly vary between setups. The cCAVE system was unique but also relatively limited in regards to the rotating chair. It would be interesting to research immersion further with CAVE systems, in particular with larger installations inside which users could walk freely. Also, there are factors that should be taken into account in the future evaluations. For example, evaluating the participant's spatial abilities with Santa Barbara sense-of-direction scale [11] before they use the application.

6.3 Implications for iODV Applications

In the past, CAVE systems have been used extensively in many areas such as the industry [28][7]. However, we argue that HMD systems offer many unique, new application areas because of two reasons. First, due to their small size and easy physical setups, HMDs are easily portable. Second, they are more scalable and adjustable, i.e. less dependent on specific equipment and a specific physical setup. These features could make HMDs a valuable asset in many situations. For instance, we recorded the

indoor video content used in our experiments in a skylift maintenance hall. However, maintenance on skylifts is often conducted in the field. Field technicians could carry HMDs with them and access informative content on-the-spot, in case they needed additional guidance on e.g. how to conduct some specific maintenance procedure on a skylift model unfamiliar to them. We believe omnidirectional video content could prove useful in such situations, as a potentially complicated procedure may be difficult to fully document (and view) on a regular camera, especially if the procedure involves large objects.

7 Conclusion

In this paper, we investigated the user experience and level of immersion in iODV applications that utilize omnidirectional videos. We conducted a comparative study between two applications: a CAVE system, cCAVE, and a head-mounted display application, Amaze360. We collected and analyzed interaction logs and questionnaire data to gain insight on similarities and differences between these two systems and on the feeling of immersion and user experience in iODV applications in general.

Our main findings suggest that in regards to user experience in interactive ODV applications, the experiences exceed the user's expectations. These differences were especially evident with the HMD system, as the users' expectations were exceeded in many aspects such as pleasantness, clarity and performance of the system. Both the CAVE and the HMD applications were considered very easy to learn. Some of the differences in user experience between these two iODV applications can be explained by the different user interaction methods. Head orientation-based interaction used with the HMD is much faster to use than the rotating chair of the CAVE system.

Another interesting take away from our study is that ODV is a very immersive medium. Overall, the HMD application was considered to be more immersive than the CAVE system. This effect was observed with both indoor and outdoor video content. We primarily attribute the immersiveness of the HMD application to a) the head-mount that effectively blocks outside visual stimuli and allows concentration on the content, b) the stereoscopic view creating a sense of depth and c) the viewport on the display is based on head orientation, allowing the user to naturally look around.

As interactive ODV applications are becoming more available in the consumer market, further research on the possibilities of this medium is necessary. For future work, it would be meaningful to study the feeling of immersion on a video content with different heights (skyscraper versus a cave) and different types of background movement (crowded street versus peaceful forest), as these properties were not within the scope of this study. Also, the effect of a moving camera (e.g. a roller coaster or a racing car) and its effects on immersion should be evaluated. This could provide more insight on what kind of ODV content offers the most immersive experience to the user.

8 REFERENCES

1. Benko, H. and Wilson, A. D. (2010). Multi-point interactions with immersive omnidirectional visualizations in a dome. In ACM International Conference on Interactive Tabletops and Surfaces (ITS '10). ACM, New York, NY, USA, pp.19-28.
2. Berning, M., Yonezawa, T., Riedel, T., Nakazawa, J., Beigl, M., and Tokuda, H. (2013). 360 degree interactive video for augmented reality prototyping. In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, UbiComp '13, ACM (New York, NY, USA, 2013), pp. 1471-1474.
3. Bleumers, L., Van den Broeck, W., Lievens, B., and Pierson, J. (2012). Seeing the bigger picture: a user perspective on 360° TV. In Proceedings of the 10th European conference on Interactive TV and video (EuroITV '12). ACM, New York, NY, USA, pp. 115-124.
4. Boult, T. E. (1998). Remote Reality via omnidirectional imaging. In ACM SIGGRAPH 98 Conference abstracts and applications (SIGGRAPH '98). ACM, New York, NY, USA.
5. Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. CHI 2004, ACM Press, pp.1279-1300.
6. Coelho, C.M., Waters, A.M., Hine, T.J. & Wallis, G. (2009). The use of virtual reality in acrophobia research and treatment. *Journal of Anxiety Disorders*, Volume 23, Issue 5, pp. 563-574.
7. Cruz-Neira, C. (1995). Virtual reality based on multiple projection screens: the cave and its applications to computational science and engineering. Doctoral Dissertation, University of Illinois, Chicago, United States.
8. De la Torre, F., Vallespi, C., Rybski, P. E., Veloso, M., & Kanade, T. (2005). Omnidirectional video capturing, multiple people tracking and identification for meeting monitoring. Technical report. <http://repository.cmu.edu/robotics/128/>
9. Decock, J., Van Looy, J., Bleumers, L. and Bekaert, P. (2011). The Pleasure of Being (There?). An Explorative Study into the Effects of Presence and Identification on the Enjoyment of an Interactive Theatrical Performance using Omnidirectional Video. In Proceedings of the International Society for Presence Research Annual Conference. ISPR 2011. Edinburgh Napier University, Edinburgh, Scotland, UK.

10. Fassbender, E. & Heiden, W. (2014). Atmosphaeres - 360° video environments for stress and pain management. In: Ma, M., Oliveira, M. & Hauge, J. (Eds.) "Serious Games Development and Applications", Vol. 8778 of Lecture Notes in Computer Science, Springer International Publishing, pp. 48-58.
11. Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30, pp. 425-448.
12. Järvinen, A. & Ekola, L. (2014). Turning point – A practical assessment of using 360 video in sign language interpreting studies. Diaconia University of Applied Sciences. Degree Programme in Sign Language Interpretation. Bachelor Thesis.
13. Kasahara, S., Nagai, S., Rekimoto, J. (2015). First Person Omnidirectional Video: System Design and Implications for Immersive Experience. In Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, June 03-05, 2015, Brussels, Belgium.
14. Kwiatek, K. (2012). How to preserve inspirational environments that once surrounded a poet? Immersive 360° video and the cultural memory of Charles Causley's poetry. In *Virtual Systems and Multimedia (VSMM)*, 2012 18th International Conference, pp. 243-250.
15. Kwiatek, K., Woolner, M. (2010). Transporting the Viewer into a 360° heritage story: Panoramic interactive narrative presented on a wrap-around screen. In *Virtual Systems and Multimedia (VSMM)*, 2010 16th International Conference, pp. 234-241.
16. Mäkelä, V., Heimonen, T., Luhtala, M., & Turunen, M. (2014). Information wall: evaluation of a gesture-controlled public display. In Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14). ACM, New York, NY, USA, pp. 228-231.
17. Neng, L. A. R. and Chambel, T. (2010). Get around 360° hypervideo. In Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '10). ACM, New York, NY, USA, pp. 119-122.
18. Neumann, U., Pintaric, T., and Rizzo, A. (2000) Immersive Panoramic Video. Proceedings of the Eighth ACM, pp. 493-494.

19. Okura, F., Kanbara, M., Yokoya, N. (2011). Fly-through Heijo palace site: augmented telepresence using aerial omnidirectional videos. In ACM SIGGRAPH 2011 Posters (SIGGRAPH '11). ACM, New York, NY, USA, Article 78.
20. Onoe, Y., Yamazawa, K., Takemura, H. Yokoya, N. (1998) Telepresence by Real-Time View-Dependent Image Generation from Omnidirectional Video Streams. *Computer Vision and Image Understanding*, Volume 71, Issue 2, August 1998, pp. 154-165.
21. Peñate W., Pitti C.T., Bethencourt J.M., de la Fuente J., Gracia R. (2007). The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology*, Volume 8, Issue 1, pp. 5-22.
22. Ramalho, J., and Chambel, T. (2013). Windy sight surfers: sensing and awareness of 360° immersive videos on the move. In *Proceedings of the 11th european conference on Interactive TV and video, EuroITV '13*, ACM (New York, NY, USA, 2013), pp. 107-116.
23. Rizzo, A. A., Ghahremani, K., Pryor, L. & Gardner, S. (2003). Immersive 360-degree panoramic video environments: research on creating useful and usable applications. In *Human-computer interaction: theory and practice*, vol 1, J. Jacko and C. Stephanidis, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp. 1233-1237.
24. Rovelo Ruiz, G. A., Vanacken, D., Luyten, K., Abad, F., and Camahort, E. (2014). Multi-viewer gesture-based interaction for omnidirectional video. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, CHI '14*, ACM (New York, NY, USA, 2014), pp. 4077-4086.
25. Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10, pp. 266-281.
26. Slater M. (2003). A note on presence terminology. *Presence Connect* 3: 3.
27. Slater, M., Alberto, C., Usoh, M. (1995). In the Building or Through the Window? An Experimental Comparison of Immersive and Non-Immersive Walkthroughs, *Virtual Reality Environments in Architecture*, Leeds 2-3rd November 1994, Computer Graphics Society.

28. Talaba, D., Amditis, A. (2008). Product Engineering: Tools and Methods based on Virtual Reality. Springer Science & Business Media. pp. 75-77.
29. Turunen, M., Hakulinen, J., Melto, A., Heimonen, T., Laivo, T., and Hella, J. (2009). SUXES – User Experience Evaluation Method for Spoken and Multimodal Interaction. In Proceedings of Interspeech 2009. ACM, New York, NY, USA, pp. 109-112.
30. Usoh, M., Catena, E., Arman, S. & Slater, M. (2000). Using Presence Questionnaires in Reality. Presence: Teleoper. Virtual Environ. 9, 5 (October 2000), pp. 497-503.
31. Welch, R.B. (1999). How Can We Determine if the Sense of Presence Affects Task Performance? Presence: Teleoperators and Virtual Environments. 8(5): pp. 574-577.
32. Witmer, B. G., & Singer, M. J. (1994). Measuring immersion in virtual environments. (ARI Technical Report 1014). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.
33. Zoric, G., Barkhuus, L., Engström, A., and Önnvall, E. (2013). Panoramic Video: Design Challenges and Implications for Content Interaction. Proceedings of the 11th European Conference on Interactive TV and Video, ACM, pp. 153–162.
34. Zoric, G., Engström, A., Barkhuus, L., Ruiz-Hidalgo, J., and Kochale, A. (2013). Gesture interaction with rich TV content in the social setting. In Exploring and Enhancing the User Experience for Television, Workshop of ACM SIGCHI Conference on Human Factors in Computing Systems, CHI'13 (Paris, France, 04/2013 2013).



Publication VII

Kallioniemi, P., Keskinen, T., Hakulinen, J., Turunen, M., Karhu, J., and Ronkainen, K. (2017). Effect of gender on immersion in collaborative iODV applications. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA, pp. 199-207. doi:10.1145/3152832.3152869

© ACM, 2017. Reprinted with permission.

Effect of Gender on Immersion in Collaborative iODV Applications

Pekka Kallioniemi, Tuuli Keskinen, Jaakko Hakulinen, Markku Turunen, Jussi Karhu, Kimmo Ronkainen

University of Tampere, Tampere, Finland
{firstname.lastname@uta.fi}



Figure 1. A snapshot from an omnidirectional video recorded in downtown Tampere, Finland. The image has been slightly cropped from top and bottom to better fit the media.

ABSTRACT

Interactive omnidirectional video (iODV) is a media format that allows the user to explore and interact with a 360-degree view of the recorded scenery. Recently, novel collaborative applications for presenting iODV content have emerged. Often their goal is to offer as immersive as possible experience to the users. Previous studies suggest that gender affects the feeling of immersion in virtual environments and other media, but there has been only little research on immersion in iODVs, and nothing in the context of collaborative interaction. In this research, we studied gender effect with participants (N=30, 15 pairs) performing a collaborative wayfinding task. Subjective data gathered with a customized immersion questionnaire showed statistically significant differences between male and female participants in *Spatial Immersion* and *Involvement* subscales. There were no statistical differences in *Interaction*, *Realness*, *Physical* and *Auditory* subscales. Several possibly affecting factors were observed during the task completion. Our results also indicate that performing interactive, collaborative tasks in iODV applications helps building a shared understanding between the users

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

MUM 2017, November 26–29, 2017, Stuttgart, Germany
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5378-6/17/11...\$15.00
<https://doi.org/10.1145/3152832.3152869>

Author Keywords

Immersion; gender differences; interactive omnidirectional videos; collaborative virtual environments.

ACM Classification Keywords

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities.

INTRODUCTION

Omnidirectional videos (ODVs) are becoming more prevalent in interactive media due to enhancements in technologies of producing and recording such content. Big companies such as YouTube, Vimeo and Facebook are offering their own platforms and distribution channels for ODV content. ODVs are typically recorded with a camera (or set of cameras) that cover 360 degrees of the recorded scene and they can be viewed via computers, mobile devices or head-mounted displays (HMDs). In the most basic form, the only interactions are to pan, tilt or roll the video. Our study focuses on collaborative iODVs, which are ODV applications with more interactive elements than just looking around the scene [19]. Some examples of such interactions include activation of UI elements or transitioning from one scene to another. iODV applications can be used on different platforms, but our study concentrates on content displayed on HMDs. *Immersion* (sometimes referred as *presence*), the feeling of ‘being there’ [31], is one of the most important features in Virtual Reality applications utilizing iODVs [19]. Slater, Alberto and Usoh [32] concluded that individuals with a higher sense of immersion performed better overall in a given task. Witmer [40] suggested that in the case where a sufficient level of realism is provided, performance in virtual environments is comparable to actual environments. When the users are provided with sufficient levels of realism and

immersion, in addition to meaningful content, virtual environments can be useful tools in many contexts. One example of these contexts is virtual reality exposure therapy.

One characteristic has been repeatedly suggested to be responsible for differences between individuals in the feeling of immersion: gender [15]. Differences between genders in immersion have been reported in activities such as watching television [8], playing video games [22] and interacting with virtual environments [33]. Gender differences have not yet been studied in the context of collaborative iODVs which is the main reason for conducting this study.

In addition to the lack of studies on the effects of gender, the motivation for this research stems from two issues: first, collaborative iODV applications are a new concept and therefore has been studied relatively little, therefore more research is needed to find out what elements are crucial for creating a meaningful experience with them. Second, immersion and user experience are crucial elements in meaningful VR applications, and we know little about them in interactive and collaborative settings.

In this study we examined the differences between genders in the feeling of immersion in a collaborative iODV application. The evaluation was done with application called CityCompass VR [18], in which two users were collaborating in physically remote locations, navigating through interactive ODV city landscapes in order to find a common goal. One of the users has to guide the other through these scenes until they reach their destination. The users communicated via audio connection. This application was originally created for collaborative language learning.

Our main research questions for study were as follows:

- Are there differences in immersion between the genders while performing collaborative tasks in iODV applications?
- Are there any gender differences in the task performance (task completion time, navigational mistakes)?

The participants filled in a questionnaire regarding the feeling of immersion after performing the collaborative task with the CityCompass VR application. This questionnaire consisted of six different subscales: *Spatial immersion* (the sense of being physically present in a virtual environment), *Interaction* (interacting with the virtual environment), *Involvement* (measurement of awareness devoted to the virtual environment), *Realness* (depicting the realism experienced in the virtual environment), *Physical* (physical effects of the virtual environment, such as nausea and dizziness) and *Auditory* (auditory aspects of the application, e.g., ambient soundscapes or audio communication aspects of the virtual environment). Three subscales (*Spatial immersion*, *Involvement* and *Realness*) were extracted from Takatalo, Nyman & Laaksonen [35] and *Interaction*, *Auditory* and *Physical* subscales were added based on our previous experiences and studies with collaborative virtual environments [e.g., 18]. These customized subscales and questions provide a simple questionnaire that is suitable for

many age groups, including elementary pupils. The goal was to avoid questionnaires with too many and/or too complex questions/statements. Males reported significantly higher scores in *Spatial immersion* and *Involvement* subscales. No statistically significant effects were observed with the other subscales. No gender differences were detected in any of the wayfinding metrics (task completion time, activation of dead ends), either. In order to gain more insights on the reason behind these detected gender differences, we asked the participants which elements increased and/or decreased their feeling of immersion in the iODV applications. Most common phenomena decreasing the feeling were errors in video looping sequences (situations where the video is reset and objects suddenly disappear and/or appear), blurring of the lenses (caused by sweating) of the HMD device and stitching errors in the ODV video.

Regarding collaboration, we found out that that performing interactive, collaborative tasks in iODV applications helps building a shared understanding between the users. This same phenomena was not evident for example in a study by Tang and Fakourfar [34], perhaps because their experiment was more explorative and a common goal for the users was missing.

Next, we present related work on the topics of this study, which is then followed by a comprehensive description of the CityCompass VR application that was used in our evaluation. Then, we introduce the methodology and report the results. We conclude this paper by discussing the results and their implications and talk briefly about the future work on this topic.

RELATED WORK

Immersion in Virtual Environments

When using or playing video games within virtual environments (VEs), it is quite common that certain sense of “being there” is developed. This phenomenon is quite rare in the traditional media (excluding cinema, where it is known as the diegetic effect [10]), but is very commonly observed in interactive media that presents 3-dimensional space for the user. Some of these include Virtual Reality (VR) applications and 3-dimensional games. For this reason, VR applications are offering a new domain for entertainment and even treating different kind of phobias [11, 25]. This sense of “being there”, depending on the context and the field of research, can be defined with two terms: *immersion* and *presence*.

Immersion has been defined in many ways in the scientific community, and there is still some discrepancy on what this term means. It is a very important aspect of virtual reality applications, and it has also been widely studied in gaming. It is believed to affect the user’s behavior with and in VR applications [32, 38] and in video games [9]. Slater [31] stated that the level of immersion is solely dependent on the system’s rendering software and display technology, making immersion an objective and measurable variable, whereas “an individual and context-dependent user response” [31] he

used the word presence. By Slater's definitions, immersion can be easier to measure, but also restricts the evaluation to the technological details, including field of view (FOV), field of regard (FOR), display size and resolution and stereoscopy, if it is used.

There are several methods for evaluating this subjective experience of immersion, ranging from short questionnaires with 2 to 8 items regarding the user's feeling of immersion [3, 19, 20] to longer ones with more than 30 items [4, 29, 39]. Factor analysis in immersion have revealed a structure of three factors, or *The Big Three of Presence* [35]: (1) Spatial presence (or immersion), (2) Involvement and (3) Realness.

In the context of video games, immersion has been studied extensively. Brown and Cairns [9] tried to resolve the disparity with the term by conducting a qualitative study amongst gamers and asked them about their experiences with video games. They created a grounded theory based on a qualitative study amongst gamers. In this theory, immersion was described as person's "degree of involvement within the game", supporting the idea that immersion is a cognitive phenomenon. They also defined restrictions that could limit the user's involvement, including engagement, engrossment and total immersion.

To summarize the previous work done in these fields, immersion/presence can be defined in several ways. In Virtual Reality applications, the division between immersion and presence by Slater [31] is more prevalent and commonly used, whereas in the field of video game research, the term immersion is used more frequently. For the scope of this study, the term immersion will be used and it is defined as a "perceptual phenomenon that is dependent on the individual and the context" [19].

Omnidirectional Videos

There is a great deal of previous work on omnidirectional imaging and videos in different domains. For example, in industrial use they have been studied in the context of remote operations [28] and telepresence applications [13, 24]. For consumer markets, they have been used and studied in the context of museums [28] and theatre [12]. Other uses include education [17], health care and therapy [27].

Adding interactive content [7] and UI elements [26] are important aspects when designing applications that utilize ODVs. Kallioniemi et al. [19] grounded the term iODV, interactive omnidirectional videos, for this type of content. There are already some guidelines for producing and recording content for interactive purposes, e.g., in Saarinen et al. [28] and Argyriou et al. [2]. In order to enhance the user experience, multisensory augmentations such as simulated wind [26] have been added to support the ODV content.

Interaction with omnidirectional content has been studied extensively, and for example gesture-based interaction [41] and second screen interfaces [42] have been evaluated. Benko and Wilson [6] presented a system that supports mid-air gestures and several simultaneous users. For their

evaluation with iODVs, Kallioniemi et al. [19] used device position-based dwell-timer for HMD and a rotating chair with a rotating sensor for CAVE.

Collaborative applications utilizing ODV have been developed in the recent years. Singhal and Neustaedter [30] implemented BeWithMe, a collaborative telepresence system designed for long-distance couples. The application was created for sharing their daily activities and experiences in the form of omnidirectional video feed. Tang and Fakourfar [34] studied watching a "guided tour" ODV content together and pointed out that it is challenging to build "a shared understanding of what was being looked at and discussed".

Gender Differences in the Feeling of Immersion

Gender differences in immersion have been studied previously in several contexts. For example, males and females report different levels of immersion when watching television [8], when interacting with virtual environments [33] and when playing video games [22]. With television, females tend to have stronger emotional reactions and therefore higher levels of immersion. The results in a study by Slater et al. [33] suggest that males reported higher levels of immersion than females in a task which required them to remember and count virtual cues. Lachlan and Krmar [22] stated that males felt higher levels of immersion in video games regardless of their previous video game experience. In this study, the male participants expressed more sensory presence and control over their environment than their female counterparts.

The results from a study by Kallioniemi et al. [18] suggested that males are more immersed while performing collaborative tasks in virtual environments. Felnhofner et al. [15] studied emotional traits and gender differences in virtual reality applications. They found statistically significant differences between male and female participants in all of their subscales (Spatial presence, Realness, Sense of Being There) except for Involvement.

CITYCOMPASS VR APPLICATION

For this evaluation, we utilized a collaborative iODV application called CityCompass VR [18]. In this application, two remotely located users are communicating via built-in audio connection and collaborating in order to navigate to a shared goal through scenes, i.e., sequential real-life locations presented in omnidirectional videos. During this task, one of the users acts as a *tourist* who has arrived to a new location and tries to find their way to a tourist attraction. The second user acts as a *guide*, helping the tourist on their way (the correct route can be seen on the guide's view). Each user has their own view of the application and they can freely explore their surroundings, i.e., they can be looking at different parts of the video than the other user.

Interaction Design

CityCompass VR is used with a HMD device that can be seen in Figure 2 and it uses a head-position based interaction where the viewport is updated based on the position of the

HMD device, i.e., the head. The application screen is divided into two separate viewports in order to create a stereoscopic effect, thus creating a sense of depth (Figure 2). The video content itself is not stereoscopic, but the effect is done with the spherical presentation in the application. The field of view in the video content used in the application is 360 degrees on the horizontal plane and 180 degrees on the vertical plane. The video is projected on a virtual sphere. The viewport's field of view is 60 degrees. Users are also wearing an audio headset for spoken communication purposes.

The application has two types of UI elements: *exits* and *hotspots*. Once activated, an exit takes the user to another video that is linked to that particular exit element, and hotspots offer contextual information about the environment it is connected to (see Figure 2). These elements in the application are activated with dwelling, i.e., by focusing an element on the center of the screen and waiting for two seconds. Dwelling is a relatively common technique for selecting targets with e.g., gaze and mid-air gestures, which is utilized by a number of applications (e.g., [23]).

CityCompass VR is implemented with C# and it is built on the Unity platform using version 5.6.1f1. Playback of the ODV content is done with Unity's native video player. Application is run on Samsung Galaxy S7 and utilizes the Samsung GEAR headset. The application uses a client-server

model. Client-side has a separate logger application implemented for Windows. This application tracks the user's actions and also plays back the audio from microphones from both clients. Audio and other messages between the clients are handled with Photon Unity Network and Photon Voice plugins. Communication between clients is mostly about synchronizing user's locations within route. Photon Voice is used to VoIP communication between the users. See Figure 3 for the CityCompass VR architecture diagram. ODV content was filmed with 6 GoPro Hero 4 camera's that were mounted in Freedom 360 rig.

Wayfinding Scenario

For the evaluation, we created a wayfinding scenario that the participants need to complete. In this scenario, the users need to find their way from a railway station to a nearby business district. In real life, the distance is about 1 kilometer. The route consists of eight or nine (depending on the route selected) different scenes and during the task two different routes can be taken (the route can be seen on the map in Figure 4). The users can communicate by speaking freely (in this study, in English) with each other via headsets. Only the user who acts as the tourist can move along the scenes and once they activate an exit, both the tourist and the guide are taken to the next scene. Each scene has three to four exits and



Figure 3. Top: CityCompass VR physical setup. Bottom: CityCompass VR application view. iODV content is shown as a stereoscopic presentation. Exit icon can be seen on the left, hotspot icon is shown on the right.

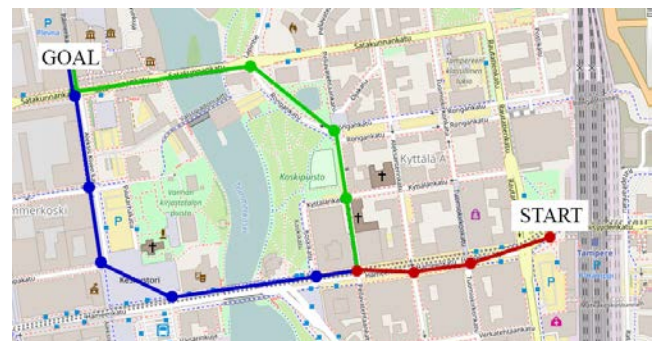


Figure 4. The route for the collaborative task. After the first four intersections the tourist can choose from two different routes (the one leading west or the one leading north). Image was exported from OpenStreetMap.com.

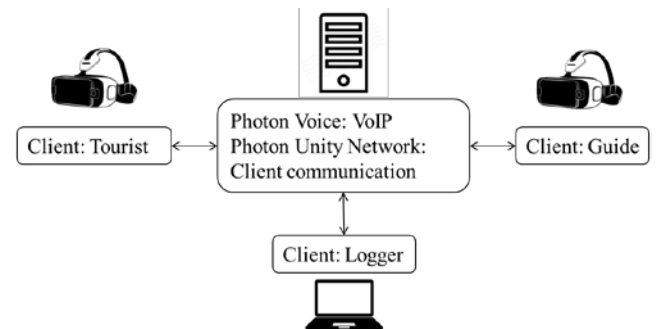


Figure 2. CityCompass VR architecture.

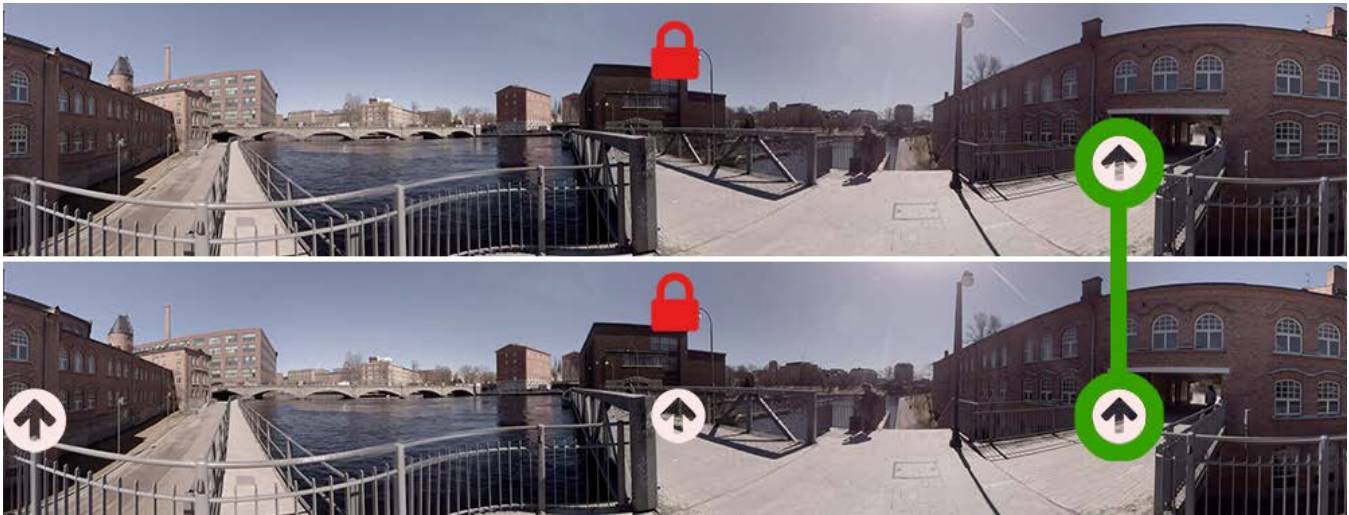


Figure 5. Dead end scene with tourist's view on top and guide's below. Lock icon indicates that the scene is a dead end. The correct route can be seen on both views on the right side of the screen.

only one of them is the correct one (takes the users closer to the goal).

If the tourist chooses an incorrect exit, both users are moved either to the previous scene or to a dead end scene. In dead end scenes the tourist needs to describe their surroundings and the guide needs to find a correct location out of four different scenes (See Figure 5). After finding the correct location, the guide can activate it after which both users are taken back to the previous scene.

METHOD

Participants

16 males and 14 females participated (15 pairs; age 21–37, $M=27.5$, $SD=5.8$) in the evaluation after providing informed consent. The dyads consisted of five male-male pairs, five female-female pairs and five mixed pairs. None of the participants had used applications that utilize interactive omnidirectional videos and only seven participants had heard of them. 14 out of 30 participants had used VR applications a few (1–3) times, 7 had heard of them and only one user had used them several times. 15 out of 30 participants felt very relaxed before the evaluation, 10 felt moderately relaxed, 4 did not feel relaxed nor tense, and one felt moderately tense. All participants were speaking Spanish as the first language, but the task itself was conducted in English. English was chosen because another key purpose of the application is second-language learning, and data towards this aspect were collected at the same time. The participants received no compensation for their participation in the study.

Procedure

In our evaluation scenario the participants worked in pairs where one of the users took the role of a tourist and the other one acted as a guide. It was up to the participants to decide which role they took during the evaluation. The users were communicating with each other via audio connection integrated into the CityCompass VR application. Each user was completing the task in their own room with a researcher

working as an assistant, giving instructions and also doing observations during the task completion. The guide was giving instructions for the tourist and the goal of the task was to find a tourist attraction.

The route was located in Tampere, Finland, and it was unfamiliar to the participants. Participants started their task from the railway station and the goal for this task was to find Finlayson industrial area. There was no time limit for the completion of the task. Both of the users were given instructions on how to use the application (looking around, activating the UI elements and resolving dead ends) and how to communicate with their partner.

Data Collection

We gathered background information from all participants, including basic information (age, gender and first language) and experience level with iODV and VR applications. After completing the task, participants filled out a questionnaire regarding the immersion in the system. This questionnaire was comprised of ten items from Barfield, Baird & Bjorneseth [5], Baños et al. [4] and Jennett et al. [16]. These were responded on a 7-point Likert-like scale. Eight of the questions were immersion-related, one was about the audio quality and one was about the feeling of nausea during the task. As the quality of the audio connection is crucial for the task completion, it was added to the questionnaire. Also, a question about the feeling of nausea was added, as the symptoms of VR sickness (headaches, stomach awareness, sweating, vomiting, etc. [21]) will probably affect the user's sense of immersion. We further categorized the questionnaire items into six subscales: *Spatial immersion*, *Involvement*, *Realness*, *Interaction*, *Auditory* and *Physical*. *Spatial immersion* describes the sense of physically immersed in a virtual space, *Involvement* measures the awareness devoted to a virtual environment, *Realism* depicts the realism attributed to and experienced within the virtual environment, *Interaction* depicts the interaction within the environment, *Auditory* measures the audio aspects of the

application and *Physical* measures the user's physical well-being while using the application. The questionnaire items, corresponding subscales and answering scale extremes can be seen in Table 1.

Sub-scale	Item	Answering scale	
		1	7
Spatial immersion	How strong was your sense of immersion (i.e., feeling like you were there) in the virtual environment?	Very weak	Very strong
	Somehow I felt that the virtual world surrounded me.	Not at all	Very much so
Interaction	To what extent did your movements in the virtual world seem natural to you?	Not at all	Very much so
	With what degree of ease were you able to navigate within the virtual environment?	With great difficulty	With great ease
Involvement	What was your overall enjoyment in navigating throughout this environment?	I did not enjoy at all	I enjoyed very much
	To what extent did you feel as though you were separated from your real-world environment?	Not at all	Very much so
Realness	In your opinion, how was the quality of the videos in the virtual environment?	Very low quality	Very high quality
	How real did the virtual world seem to you?	Not real at all	Very real
Auditory	How was the quality of the audio connection?	Very bad quality	Very good quality
Physical	Did you feel nauseous while using the system?	Not nauseous at all	Very nauseous

Table 1. Questions, their subscale categories and answering scales from our immersion questionnaire.

Participants were also observed while using the system. These observations concentrated on the participants' non-task related comments (e.g., comments about the application and immersion) during the task completion and their actions and body movements during the evaluation. We also logged the users' interactions during the task, including transitions from one location to another, activation of dead ends and activations of hotspots. Finally, we performed an interview where we asked the participants a following interview question: *What aspects of the application decreased your sense of immersion (i.e., the feeling of 'being there')?* This

question was discussed verbally after which the researcher wrote the answers down for each participant.

RESULTS

Immersion

For analyzing the immersion related data, we conducted a Mann-Whitney U test in order to detect gender differences in the subscales of immersion. Our tests indicated that the *Spatial immersion* was statistically significantly greater for males ($M=5.97$) than for females ($M=5.14$), $U=38$, $p=.001$ (Figure 6).

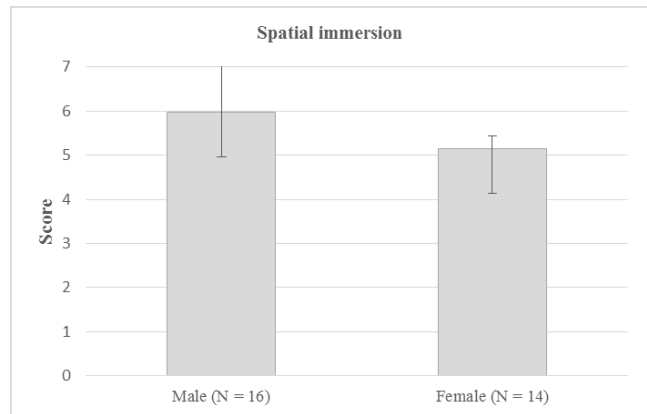


Figure 6. Means in the *Spatial immersion* subscale for males and females.

A Mann-Whitney U test was also conducted to determine whether there was a difference in the feeling of *Involvement* between males and females. Results of the analysis indicate that males ($M=5.81$) scored significantly higher in *Involvement* than females ($M=5$), $U=57.5$, $p=.022$ (Figure 7).

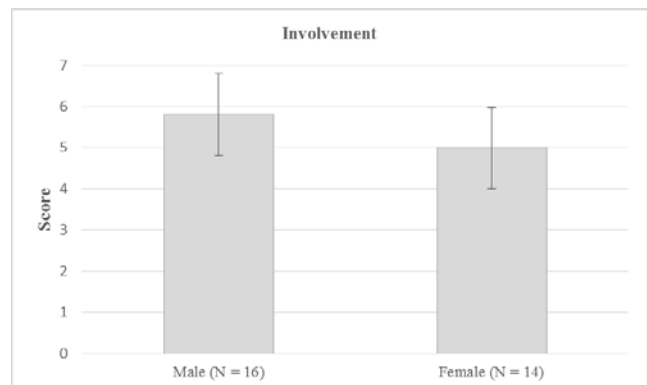


Figure 7. Means in the *Involvement* subscale for males and females.

For *Interaction*, the analysis indicated no difference between the genders (males: $M=5.63$, $SD=.992$, females: $M=5.11$, $SD=1.003$), $U=79$, $p=.179$ (Figure 8). No statistically significant effect between males ($M=5.94$, $SD=.911$) and females ($M=5.75$, $SD=.981$) was detected for *Realness*, either: $U=104.5$, $p=.759$ (Figure 9).

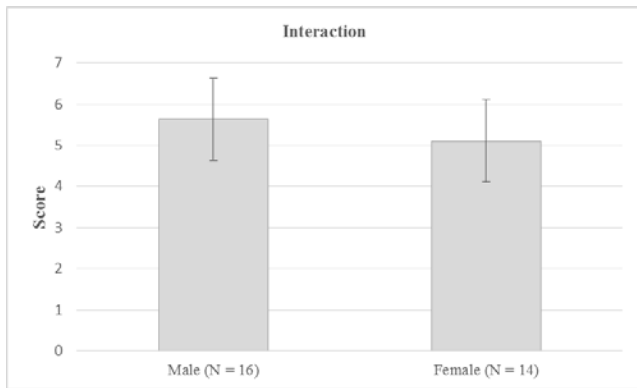


Figure 8. Means in the *Interaction* subscale for males and females.

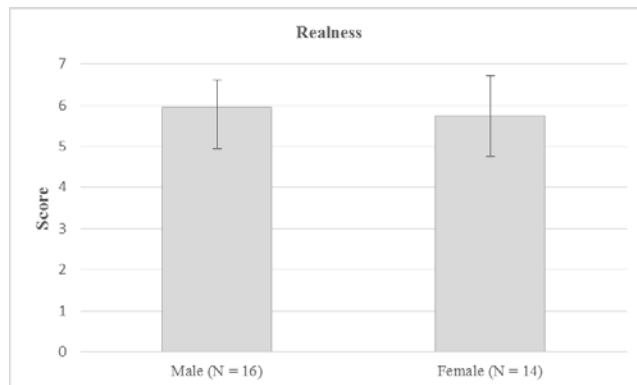


Figure 9. Means in the *Realness* subscale for males and females.

No statistically significant effect was detected between genders in the *Physical* subscale (males: $M=1.88$, $SD=1.025$, females: $M=1.64$, $SD=.842$), $U=99.5$, $p=.608$, nor in the *Auditory* subscale (males: $M=6.19$, $SD=.981$, females: $M=6.50$, $SD=.760$), $U=93$, $p=.448$.

For the task completion time (seconds) and number of dead ends, we conducted an independent t-test in order to see if there are any differences between the genders. There was no statistically significant difference between males ($M=678$, $SD=203.07$) and females ($M=746.21$, $SD=240.88$) in the task completion time; $t(28)= -.832$, $p=.413$. There was also no statistically significant effect in the number of dead ends between males ($M=1.56$, $SD=1.315$) and females ($M=2.93$, $SD=2.369$); $t(28)= -1.987$, $p=.057$.

Observations and Interview Questions

Users often resorted to their first language when they did not understand the current state of the task (usually in a dead end situation where the roles were switched). 12 out of 30 participants (5 males, 7 females) spoke in their first language in order to resolve these situations. Regarding the interview question, 9 participants (3 males, 6 females) complained that the blurring of the lenses in the Gear VR device decreased their feeling of immersion during the use. 4 participants (all males) stated that the video looping sequence caused a breaking of immersion as some of the dynamic objects in the scene disappeared and others appeared. 2 participants (both

males) claimed that the stitching errors (artefacts in the areas where the cameras intertwined) decreased their sense of immersion.

DISCUSSION

Feeling of Immersion and Task Completion with Collaborative iODVs

As gender has been repeatedly suggested to be responsible for individual differences in the feeling of immersion, the current study explored the differences between males and females in immersion while using a collaborative iODV application. Our study supports previous findings in different use contexts [8, 15, 18, 22, 33] where gender was suggested to be a possible explaining factor in the formation of immersion. We studied this phenomena with six different subscales: *Spatial immersion*, *Interaction*, *Involvement* and *Realness*. **Males reported significantly higher scores in *Spatial immersion* and *Involvement* subscales.** Several explaining factors for these findings have been suggested before. Waller, Hunt and Knapp [37] stated that this difference could be due to computer experience, but for example Felnhofer et al. [15] detected differences in immersion between genders even when there was no statistically significant difference in computer experience between males and females that partook in their study. Kallioniemi et al. [18] and Lachlan & Krcmar [22] stated that video game experience may have positive influence on immersion and user experience in general. Former study examined this effect in collaborative virtual environments, which is also relevant in the context of this study. Felnhofer et al. [15] suggested the possibility of self-efficacy in handling computer hardware and software as an explaining factor for the higher feeling of immersion among men. This relation between self-efficacy and immersion should be studied further for more concrete evidence.

Tang and Fakourfar [34] claimed that it is difficult to build a shared understanding between the users while they are observing ODV content collaboratively. This was not evident in our study, and one of the explaining factor could be in the collaborative task itself – the users share a common task and goal, rather than just viewing the content passively. This active collaboration keeps the participants more focused on the task at hand and also have to consider the other user’s viewpoint and situation. To summarize, our results suggest that **performing interactive, collaborative tasks in iODV applications help building a shared understanding between the users.**

The interview question regarding immersion offered some possible explanations and insights on the subject. The blurring of the lenses because of sweating was a common phenomenon which might also partly explain the difference in the feeling of immersion between males and females – 6 females experienced blurring (versus only 3 males). Errors in the looping sequence (objects suddenly disappearing and/or appearing) was reported to be immersion-breaking event by 4 male participants. This effect can be difficult to prevent when the iODV content is recorded in a busy

environment, e.g., in a city. One solution for this could be adding fade in – fade out effects to the beginning and end of the video, as this would prevent the disappearing dynamic objects (e.g., pedestrians, cars, cyclists) in the videos. Although, these added effects might break the illusion of continuity in the iODV content. Stitching errors at the camera lens intersections was also reported as immersion-breaking event by two male participants. It is worth noting that only the first out of these phenomena can be explaining the differences in the feeling of immersion between genders (the last two were only detected in males). **Stitching errors and errors in the looping sequence can be fixed with better production design, i.e., by following iODV content guidelines** (e.g., [2, 28]). One possible solution for blurring of the lenses could be anti-fog sprays that are often used by divers.

No statistically significant differences were found between genders in the task completion metrics. Participants were not encouraged to complete the task as quickly as possible, and therefore the task completion time is quite irrelevant in any comparison between males and females. Some participants were also more prone to explore the iODV content in the scenes rather than hurrying to find the correct route. These explorations may also explain some of the dead end activations during the task. In this regard, it is meaningless to compare gender differences in immersion and its effects on task completion in this study. Similar findings regarding collaborative wayfinding tasks in virtual environments were previously reported by Kallioniemi et al. [18], where there were no statistically significant differences in wayfinding metrics between males and females. Even though spatial cognition and abilities are out of the scope of this study, it is meaningful to point out that collaborative wayfinding tasks seem to somewhat negate the differences in spatial abilities between genders, which are sometimes referred as being significant, see e.g., Voyer et al. [36].

As the results showed no difference between males and females in *Interaction* subscale, creating any gender-specific guidelines for interaction in iODV applications is not necessary. Regarding *Spatial immersion* and *Involvement*, further research is required in order to find out which elements improve or reduce these feelings. Interestingly, females feel more immersed while watching television [8], whereas males tend to be more immersed while playing video games [22] or while interacting with virtual environments [15, 18, 33]. Bracken [8] stated that “women evaluate some types of nonfictional television content as more real than men”. They suggest that this may also apply to fictional programming. Further research between the main differences of these mediums is required, but we suggest that by **designing virtual environments with television-type content could be one solution to narrow down the immersion gap between the genders.** One example of using these dramatized scenarios is the *Bollywood Method* [1]. This method has been mainly used for feedback purposes in the context of Indian culture. Bollywood creates movies that

usually have a larger-than-life fantasy, which are met with great excitement and engagement, especially within the local Indian populations. In this method, users are presented with a dramatized scenario that requires them to take on the role of a character with a specific goal. This method was tested with an airlines ticket booking application, where the users were asked to imagine that their niece is unknowingly getting married to an underground gangster, who is actually already married. Then the users must book flight tickets to Bangalore with the incriminating evidence in their sole possession to stop the wedding [1]. This kind of dramatized scenarios could be used in order to immerse the users more with the given task.

Measuring the Feeling of Immersion with Customized Immersion Subscales

The reason for using a custom questionnaire and subscales for the sense of immersion was the general complexity of the validated methods. Our initial pilot tests indicated that many of the questions used in these methods are too complex (as both in English and as translated versions) and also that many of them consisted of too many items. As we are planning to conduct further studies with different age groups, starting with elementary level pupils, these complex questionnaires could not be used. Therefore, we constructed our own questionnaire with the usual subscales and also added one that can be considered crucial for the sense of immersion in iODV applications – interaction. Interaction is considered to be one of the prime causes of immersion in virtual environments (e.g., [14]) and therefore should be considered as one of the subscales when evaluating it.

The *Auditory* and *Physical* subscales consisted of only one item each, which might not be sufficient amount for any meaningful analysis on these subjects. In this study, they were added because the effect of nausea and low quality audio connection could affect the overall use experience and also the feeling of immersion. The low score for nausea ($M=1.77$, $SD=.919$) and high score for audio quality ($M=6.33$, $SD=.884$) suggest that these factors were not negatively affecting the feeling of immersion of the participants in this particular study. For future studies, the relaxation item from the basic information could be added to the *Physical* subscale and an item about the (not yet implemented) ambient soundscape of the application could be part of the *Auditory* subscale. This way each subscale would consist of 2 items.

CONCLUSION AND FUTURE WORK

In this research, we studied the gender differences in collaborative iODV applications. For this, a collaborative virtual environment called CityCompass VR was implemented. Participants cooperated in order to complete a wayfinding task in the application. After the task, the users filled out a questionnaire consisting of questions in six immersion-related subscales.

We detected significant differences in two of the subscales: *Spatial immersion* (the sense of being physically present in a virtual environment) and *Involvement* (measurement of

awareness devoted to the virtual environment). Further interviews revealed three main phenomena decreasing the feeling of immersion: 1) blurring of the lenses of the HMD device while use, 2) errors in the looping sequence (objects suddenly disappearing and/or appearing) and 3) stitching errors (e.g., overlapping video streams) in the videos. The first effect was mostly reported by females whereas the last two were reported by males.

Another contribution of this study is the customized questionnaire that expands on the previous work in the field. In addition to the more traditionally used subscales (*Spatial immersion*, *Realness* and *Involvement*) we added three more: *Interaction*, *Physical* and *Auditory*. In addition, we simplified the previous questionnaires in order for them to be more suitable for different age groups, including elementary pupils.

Regarding collaboration, our results suggest that when provided a common task, rather than just viewing the ODV content passively, users are more focused on the given task. This finding indicates that these interactive, collaborative tasks help building a shared understanding between the users.

No clear indicators as to why males reported higher feeling of *Spatial immersion* and *Involvement* than females was detected. We suggest adding dramatized, story-like elements in to iODV applications in order to immerse the users more to the given task. In addition to questionnaires, gaming experience, self-efficacy and technology acceptance should be considered when evaluating immersion. Also, more objective measures such as heart rate or even EEG could be used to further determine the origins of these differences.

REFERENCES

1. Apala L. Chavan. 2005. Another culture, another method. In *Proceedings of the 11th International Conference on Human-Computer Interaction*. 21(2).
2. Lemonia Argyriou, Daphne Economou, Vassiliki Bouki & Ioannis Doumanis. 2016. Engaging immersive video consumers: Challenges regarding 360-degree gamified video applications. In *IEEE International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security, IUCC-CSS '16*. IEEE 2016, pp. 145–152. <http://dx.doi.org/10.1109/IUCC-CSS.2016.028>
3. Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 10, 6 (December 2001), 583-598. <http://dx.doi.org/10.1162/105474601753272844>
4. Rosa M. Baños, Cristina Botella, Azucena Garcia-Palacios, H Villa, Concepción Perpiñá, Mariano Alcaniz 2000. Presence and reality judgment in virtual environments: A unitary construct? *CyberPsychology and Behaviour*, 3, 327–335. <https://doi.org/10.1089/10949310050078760>
5. Woodrow Barfield, Kevin M. Baird, Ove J. Bjorneseth. 1998. *Presence in virtual environments as a function of type of input device and display update rate*. *Displays*, 19, pp. 91–98. [https://doi.org/10.1016/S0141-9382\(98\)00041-9](https://doi.org/10.1016/S0141-9382(98)00041-9)
6. Hrvoje Benko and Andrew D. Wilson. 2010. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. ACM, New York, NY, USA, 19-28. <https://doi.org/10.1145/1936652.1936657>
7. Matthias Berning, Takuro Yonezawa, Till Riedel, Jin Nakazawa, Michael Beigl, and Hide Tokuda. 2013. pARNorama: 360 degree interactive video for augmented reality prototyping. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (UbiComp '13 Adjunct)*. ACM, New York, NY, USA, 1471-1474. <http://dx.doi.org/10.1145/2494091.2499570>
8. Cheryl C. Bracken. 2005. Presence and Image Quality: The Case of High-Definition Television. *Media Psychology*, 7, pp. 191–205. http://dx.doi.org/10.1207/S1532785XMEP0702_4
9. Emily Brown and Paul Cairns. 2004. A grounded investigation of game immersion. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, New York, NY, USA, 1297-1300. <http://dx.doi.org/10.1145/985921.986048>
10. Noël Burch. 1979. *To the distant observer. Form and meaning in the Japanese cinema*. London: Scholar Press. <http://hdl.handle.net/2027/spo.aq5060.0001.001>
11. Carlos M. Coelho, Allison M. Waters, Trevor J. Hine, Guy Wallis. 2009. The use of virtual reality in acrophobia research and treatment, In *Journal of Anxiety Disorders*, Volume 23, Issue 5, Pages 563-574, ISSN 0887-6185. <https://doi.org/10.1016/j.janxdis.2009.01.014>
12. Jan Decock, Jan Van Looy, Lizzy Bleumers and Philippe Bekaert. 2011. The Pleasure of Being (There?). An Explorative Study into the Effects of Presence and Identification on the Enjoyment of an Interactive Theatrical Performance using Omnidirectional Video. In *Proceedings of the International Society for Presence Research Annual Conference. ISPR 2011*. Edinburgh Napier University, Edinburgh, Scotland, UK. <https://doi.org/10.1007/s00146-013-0487-6>
13. Fernando De la Torre, Carlos Vallespi-Gonzales, Paul Rybski, Manuela Veloso, and Takeo Kanade. 2005. Omnidirectional video capturing, multiple people tracking and identification for meeting monitoring. Technical report. <http://repository.cmu.edu/robotics/128/>

14. John V. Draper, David B. Kaber, and John M. Usher. 1998. Telepresence. *Human Factors*, 40, 354–375. <https://doi.org/10.1518/001872098779591386>
15. Anna Felnhofner, Oswald D. Kothgassner, Leon Beutl, Helmut Hlavacs, Ilse Kryspin-Exner. 2012. Is Virtual Reality made for Men only? Exploring Gender Differences in the Sense of Presence. *Proceedings of the International Society on presence research*, pp. 103–112.
16. Charlene Jennett, Anna L. Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *Int. J. Hum.-Comput. Stud.* 66, 9 (September 2008), 641–661. <http://dx.doi.org/10.1016/j.ijhcs.2008.04.004>
17. Anu-Maria Järvinen, and Leena Ekola. 2014. Turning point – A practical assessment of using 360 video in sign language interpreting studies. Diaconia University of Applied Sciences. Degree Programme in Sign Language Interpretation. Bachelor Thesis. <http://urn.fi/URN:NBN:fi:amk-201404295266>
18. Pekka Kallioniemi, Tomi Heimonen, Markku Turunen, Jaakko Hakulinen, Tuuli Keskinen, Laura Pihkala-Posti, Jussi Okkonen, and Roope Raisamo. 2015. Collaborative navigation in virtual worlds: how gender and game experience influence user behavior. In *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology (VRST '15)*, Stephen N. Spencer (Ed.). ACM, New York, NY, USA, 173–182. <https://doi.org/10.1145/2821592.2821610>
19. Pekka Kallioniemi, Ville Mäkelä, Santeri Saarinen, Markku Turunen, York Winter, and Andrei Istudor. 2017. User Experience and Immersion of Interactive Omnidirectional Videos in CAVE Systems and Head-Mounted Displays. In: Bernhaupt R., Dalvi G., Joshi A., K. Balkrishan D., O'Neill J., Winckler M. (eds) *Human-Computer Interaction – INTERACT 2017*. INTERACT 2017. Lecture Notes in Computer Science, vol 10516. Springer, Cham, 299–318. https://doi.org/10.1007/978-3-319-68059-0_20
20. Taevong Kim, and Frank Biocca. 1997. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, 3 (2). <https://doi.org/10.1111/j.1083-6101.1997.tb00073.x>
21. Eugenia M. Kolasinski. 1995. Simulator Sickness in Virtual Environments. 68. U.S. Army Research Institute for the Behavioral and Social Sciences, Vol 4, Issue 1027.
22. Kenneth A. Lachlan, and Marina Krmar. 2011. Experiencing Presence in Video Games: The Role of Presence Tendencies, Game Experience, Gender, and Time Spent in Play. *Communication Research Reports*, 28 (1), 27–31. <https://doi.org/10.1080/08824096.2010.518924>
23. Ville Mäkelä, Tomi Heimonen, Matti Luhtala, and Markku Turunen. 2014. Information wall: evaluation of a gesture-controlled public display. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia (MUM '14)*. ACM, New York, NY, USA, 228–231. <http://dx.doi.org/10.1145/2677972.2677998>
24. Yoshio Onoe, Kazumasa Yamazawa, Haruo Takemura, and Naokazu Yokoya. 1998. Telepresence by Real-Time View-Dependent Image Generation from Omnidirectional Video Streams. *Comput. Vis. Image Underst.* 71, 2 (August 1998), 154–165. <http://dx.doi.org/10.1006/cviu.1998.0705>
25. Wenceslao Peñate, Carmen T. Pitti, Juan M. Bethencourt, Juan de la Fuente, Ramon Gracia. 2007. The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology*, Volume 8, Issue 1, pp. 5–22.
26. João Ramalho and Teresa Chambel. 2013. Windy sight surfers: sensing and awareness of 360° immersive videos on the move. In *Proceedings of the 11th European Conference on Interactive TV and Video (EuroITV '13)*. ACM, New York, NY, USA, 107–116. <https://doi.org/10.1145/2465958.2465969>
27. Albert A. Rizzo, Kambiz Ghahremani, Larry Pryor, and Susanna Gardner. 2003. Immersive 360-degree panoramic video environments: research on creating useful and usable applications. In *Human-computer interaction: theory and practice*, Vol 1, J. Jacko and C. Stephanidis, Eds. Lawrence Erlbaum Associates, Mahwah, NJ, USA, pp. 1233–1237.
28. Santeri Saarinen, Ville Mäkelä, Pekka Kallioniemi, Jaakko Hakulinen, and Markku Turunen. 2017. Guidelines for Designing Interactive Omnidirectional Video Applications. In: Bernhaupt R., Dalvi G., Joshi A., K. Balkrishan D., O'Neill J., Winckler M. (eds) *Human-Computer Interaction – INTERACT 2017*. INTERACT 2017. Lecture Notes in Computer Science, vol 10516. Springer, Cham, 263–272. https://doi.org/10.1007/978-3-319-68059-0_17
29. Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10 (3), 266–281. <https://doi.org/10.1162/105474601300343603>
30. Samarth Singhal, and Carman Neustaedter. 2017. BeWithMe: An Immersive Telepresence System for Distance Separated Couples. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17 Companion)*. ACM, New York, NY, USA, pp. 307–310. <https://doi.org/10.1145/3022198.3026310>

31. Mel Slater. 2003. A note on presence terminology. *Presence Connect* 3: 3.
32. Mel Slater, Christina Alberto, Martin Usoh. 1995. In the Building or Through the Window? An Experimental Comparison of Immersive and Non-Immersive Walkthroughs, *Virtual Reality Environments in Architecture*, Leeds 2-3rd November 1994, Computer Graphics Society.
33. Mel Slater, Anthony Steed, John McCarthy, and Francesco Maringelli. 1998. The Influence of Body Movement on Subjective Presence in Virtual Environments. *Human Factors*, 40, 469–477. <https://doi.org/10.1518/001872098779591368>
34. Anthony Tang and Omid Fakourfar. 2017. Watching 360° Videos Together. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4501-4506. <https://doi.org/10.1145/3025453.3025519>
35. Jari Takatalo, Göte Nyman, Leif Laaksonen. 2008. Components of human experience in virtual environments. *Computers in Human Behavior*, 24 (1), pp. 1–15. <https://doi.org/10.1016/j.chb.2006.11.003>
36. Daniel Voyer, Susan Voyer, and Philip Bryden. 1995. Magnitude of sex differences in spatial abilities: A meta-analysis consideration of critical variables. *Psychological Bulletin*, 117, pp. 250–270. <https://doi.org/10.1037/0033-2909.117.2.250>
37. David Waller, Earl Hunt, and David Knapp. 1998. The Transfer of Spatial Knowledge in Virtual Environment Training. *Presence: Teleoper. Virtual Environ.* 7, 2 (April 1998), 129-143. <http://dx.doi.org/10.1162/105474698565631>
38. R.B. Welch. 1999. How Can We Determine if the Sense of Presence Affects Task Performance? *Presence: Teleoperators and Virtual Environments*. 8(5): pp. 574–577.
39. Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The factor structure of the presence questionnaire. *Presence: Teleoper. Virtual Environ.* 14, 3 (June 2005), 298-312. <http://dx.doi.org/10.1162/105474605323384654>
40. Bob G. Witmer, John H. Bailey, Bruce W. Knerr, and Kimberley C. Parsons. 1996. Virtual spaces and real world places: transfer of route knowledge. *Int. J. Hum.-Comput. Stud.* 45, 4 (October 1996), 413-428. <http://dx.doi.org/10.1006/ijhc.1996.0060>
41. Goranka Zoric, Louise Barkhuus, Arvid Engström, and Elin Önnvall. 2013. Panoramic video: design challenges and implications for content interaction. In *Proceedings of the 11th European Conference on Interactive TV and Video (EuroITV '13)*. ACM, New York, NY, USA, 153-162. <https://doi.org/10.1145/2465958.2465959>
42. Goranka Zoric, Arvid Engström, Louise Barkhuus, Javier Ruiz-Hidalgo, and Axel Kochale. 2013b. Gesture interaction with rich TV content in the social setting. In *Exploring and Enhancing the User Experience for Television, Workshop of ACM SIGCHI Conference on Human Factors in Computing Systems, CHI'13 (Paris, France, 04/2013 2013)*.

Details of the dissertations are available at
<http://www.uta.fi/sis/tauchi/dissertations.html>.

1. **Timo Partala:** Affective Information in Human-Computer Interaction
2. **Mika Käki:** Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula:** Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari:** Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi:** Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen:** Software Tutoring in Speech User Interfaces
7. **Harri Siirtola:** Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen:** Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov:** iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova:** Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta:** Text Entry by Eye Gaze
12. **Ying Liu:** Chinese Text Entry with Mobile Phones
13. **Toni Vanhala:** Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen:** Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves:** Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku:** Face Interface
17. **Juha Leino:** User Factors in Recommender Systems: Case Studies in e-Commerce, News Recommending, and e-Learning
18. **Joel S. Mtebe:** Acceptance and Use of eLearning Solutions in Higher Education in East Africa
19. **Jussi Rantala:** Spatial Touch in Presenting Information with Mobile Devices
20. **Katri Salminen:** Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli
21. **Selina Sharmin:** Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts
22. **Tuuli Keskinen:** Evaluating the User Experience of Interactive Systems in Challenging Circumstances
23. **Ogunbase Adewunmi Obafemi:** Pedagogical Design and Pedagogical Usability of Web-Based Learning Environments: Comparative Cultural Implications from Africa and Europe
24. **Hannu Korhonen:** Evaluating Playability of Mobile Games with the Expert Review Method
25. **Jani Lylykangas:** Regulating Human Behavior with Vibrotactile Stimulation
26. **Ahmed Farooq:** Developing technologies to provide haptic feedback for surface based interaction in mobile devices
27. **Ville Mäkelä:** Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments
28. **Pekka Kallioniemi:** Collaborative Wayfinding in Virtual Environments