memoAR

An Augmented Reality Application on Urban Story Telling

A Medialogy Master's Thesis by

Martin Havsager Nielsen

Aalborg University, Spring 2013

AALBORG UNIVERSITET

Title:
memoAR – An Augmented Reality Application on Urban Story Telling

Project Period:
Med10

Supervisors:
Ann Morrison
Mark Billinghurst

Project group:
mta131034

Members:
Martin Havsager Nielsen

Copies: 4
Pages: 102 (+ appendices)
Appendices: 9
Attachments: DVD
Finished: 23/05/13

Abstract:

This thesis presents a location-aware augmented reality (AR) application, which involves audio-based stories and interviews geo-tagged to locations in a cityscape. The project is partly a collaboration with New Zealand Historic Places Trust in Christchurch, NZ, where earthquakes have left the city in ruins, and created an interesting potential for urban stories.
An Android application is developed, which uses AR and spatial audio techniques to let users explore city stories and related multimedia directly from the urban environment.
Two experiments are conducted: a comparative usability test of AR interaction methods, as well as an empirical field observation of the final application in use.
Results show that users are positive towards the usability of the application, overall and at a task-level.
Further tests are planned in order to address and evaluate the user experience of the spatial audio sources.

Copyright@2013. This report and/or app ended material may not be partly or completely published or copied without prior written approval from the authors. Neither may the contents be used for commercial purposes without this written approval.

Preface

This thesis is submitted in partial fulfilment of the requirements for the Master of Science degree in Medialogy at Aalborg University. The 40 ECTS thesis is written in the period 4 February - 23 May 2013.

The report describes the development of *memoAR* - an audio-driven Android application on urban story telling. The work is part of a long-term project, which seeks to investigate the combined potential of social networking and mobile augmented reality. While the main focus of this thesis is on the development of an audio-based tour application, aspects from the long-term goal of the project are included in this report as well. Hence the report should be read with the long-sighted goal of integrating social networking and mobile augmented reality in mind.

References are cited as (Author Date) e.g. (Azuma 1997), and a full reference list can be found near the end of the report.

Enclosed in the report is a DVD containing the following:

- Project report in PDF version.
- An audio-visual production documenting the project.
- The Android source code of the developed application.

The author wishes to thank the following people for their contributions to the thesis:

- Main supervisor Ann Morrison for input and feedback throughout the project period.
- Gun A. Lee for providing technical support on the Android coding.
- Staff and students at the HITLab New Zealand for a wonderful stay and heaps of inspiration.
- Lucy Holland for a nice collaboration on major design decisions.
- Zoe Roland from New Zealand Historic Places Trust for providing me the opportunity to participate in the High Street Stories project, which is the foundation for this thesis.
- Test participants for helping with the evaluation of the project.

Glossary/Abbreviations

API / Application Programming Interface

An interface for accessing services and applications of a particular program.

AR / Augmented Reality

A novel technology, where digital content is superimposed over the user's view of the real world, to give the impression that the real and virtual world coexist. Definition from Azuma (1997).

GPS / Global Positioning System

A technology, which uses signals from satellites orbiting the earth to provide navigation information and estimate the geographical position of a device capable of receiving GPS signals.

HITLabNZ / Human Interface Technology Laboratory New Zealand

A research institute at the University of Canterbury, New Zealand, who is involved in the High Street Stories project, which this thesis is a part of. Further, they have developed the OutdoorAR framework, which is the technical foundation for the implementation of this thesis' product.

HMD / Head-mounted display

Computer display, which presents digital content to the user. Worn on the head, so the user maintains a view of the surrounding environment. HMD's can generally be split in optical and video see-through displays, dependent on how the real and virtual world are composited.

MARS / Mobile Augmented Reality System

Term covering wearable augmented reality systems from the Columbia University, who pioneered the area of mobile/wearable computing and augmented reality in the mid 1990's.

NZHPT / New Zealand Historic Places Trust

A national heritage agency, seeking to ensure the survival and appreciation of New Zealand's cultural and architectural history. The primary stakeholder of the High Street Stories project and a collaborative partner in the design and development of the thesis' product.

PDA / Personal Digital Assistant

A mobile device with a number of functionalities for assisting the user. Typically contains a touchscreen for user input, memory card for storage of data and software such as calendars, to-do lists and calculators.

SUS / System Usability Scale

A questionnaire for measuring the usability of a system. Contains 10 different statements, to which participants rate their level of agreement. The scale aims at measuring the overall usability of a system by looking at the combined rating of respondents' answers. Invented by John Brooke in 1996.

UI / User interface

The space between the user and a computer system, in which the interaction occurs. Definition from http://www.techterms.com/definition/user_interface

UGC / User-generated content

Various media content generated and shared by actual users of a system. Popular in modern communication technologies, for instance in social media and networking services.

List of Figures

Figure 1 Location-based services cover a wide range of application areas (Nil 2013). 5

Figure 2 The context of urban story telling is the starting point for investigating the combined potential of mobile computing, AR and social media. 6

Figure 3 Dedicated geo-social networking systems. **A**: Foursquare presents interesting locations around the user based on other users' recommendations. **B**: Microsoft Research's Geo-Life recommends popular locations and travel sequences based on GPS trajectories (Microsoft 2013). 9

Figure 4 Mobile AR systems. **A**: Layar overlays virtual content on print media (Layar 2013). **B**: AR channel created in junaio (Junaio 2013). **C**: Wikitude AR scene showing interesting locations of a city (Wikitude 2013). 10

Figure 5 Combining AR and social media. **A**: Twitter 360 - users geotag their tweets and locate Twitter friends in the nearby environment (Twitter 2013). **B**: SocializAR - AR for conference and business networking (Twnkls 2013). 11

Figure 6 Envisioning a future of social networking and mobile AR, where a quick scan of a person reveals insight into this person's life (Buckland 2010). 12

Figure 7 Roadside Stories tour. **A**: The user can download individual tours related to New Zealand's major travel routes. **B**: A map view shows the locations of the tour stops. **C**: A stop is associated with an audio file telling historic facts about the location, and an image gallery. 14

Figure 8 Apps for city stories and museum tour guides. **A**: Soho Stories provides immersive stories tagged to the historic Soho district in London. **B**: "The O" provides an interface for exploring art galleries at Museum of New and Modern Art in Hobart, Tasmania (Mona 2013). 15

Figure 9 Earthquakes have left Christchurch in ruins. **A**: The WestPac Trust Building in the High Street area is facing demolition. **B**: The cathedral, the city's trademark building, lost its spire following the February 2011 earthquake (PlaybackTheatre 2011). 19

Figure 10 Rebuilding Christchurch. **A**: The Re:START container mall in Cashel Street (Inhabitat 2011). **B**: GapFiller minigolf event appears at various vacant sites of the city. **C**: Community chess at vacant site in Colombo Street. (GapFiller 2013). 20

Figure 11 Using media technology to explore the city's past and future. **A**: CityViewAR shows 3D models of the demolished buildings in Christchurch (FutureGov 2012). **B**: CCDU 3D presents the future development plan for Christchurch's central business district (3News

2013). 21

Figure 12 Major changes in the cityscape. **A**: Aalborg's harbourfront is undergoing dramatic changes these years (AalborgKommune 2010). **B**: The old power station buildings at Nordkraft are now centers for concerts and venues. 22

Figure 13 Pervasive games from the iPerg project. **A**: In "Hitcher", users control a hitchhiker and interact with other hitchhikers through a location-aware mobile device (MutualFriend 2008). **B**: In the game "Interference", AR technology is used to create an immersive urban adventure game (Iperg 2008). 24

Figure 14 Mobile augmented reality systems. **A**: User performs a task on the first Touring Machine (Feiner et al. 1997). **B**: Virtual points of interest are embedded in the environment. Interacting with the digital flags presents narrated multimedia presentations at a university campus (Höllerer et al. 1999).. 25

Figure 15 Mobile AR games. **A**: The award-winning Mozzies game running on the Siemens SX1 phone (Eden 2010). **B**: GeoBoids integrates a location-aware gameplay and spatial audio techniques (Lindeman & Lee 2012). 25

Figure 16 Audio in the field. **A**: User explores the narrative of the Oakland Cemetery mixed reality experience (Dow et al. 2005). **B**: Equipment of Sotto Voce, which emphasises social interaction at museums (Aoki et al. 2002). 26

Figure 17 In their work on sound gardens, Vazquez-Alvarez and his colleagues use proximity and activation zones to present audio content to the user. The use of spatial sound techniques helps managing overlapping sound sources in the audio environment. (Vazquez-Alvarez et al. 2012). 28

Figure 18 Classification of social media and networking sites based on social presence, media richness theory and self-presentation (Kaplan & Haenlein 2010). 30

Figure 19 Facebook for iOS 5.0. **A**: The News Feed shows the most recent updates from friends. **B**: The "Nearby" feature depicts interesting locations and friends' recommendations near the user. 31

Figure 20 Flickr for iOS **A**: The user profile shows the user's photostream and lets him/her browse photos from groups and contacts. **B**: Browsing images generated and tagged to locations near the user's location. 32

Figure 21 Nokia City Lens uses the orientation of the mobile device to decide which screens to present to the user. **A**: When device is in landscape mode and tilted frontwards, an AR view is displayed. **B**: Tilting the phone towards the ground shows a map view. **C**: Rotating the device to portrait mode depicts a list view. (Nokia 2012). 34

Figure 22 Conquering mobile interaction design challenges. **A**: LucidTouch provides an additional input surface behind the mobile display to conquer occlusion issues (Wigdor et al. 2007). **B**: The SHARK word-gesture keyboard relies on drawing marks and gestures rather than typing text (Zhai & Kristensson 2003). 35

Figure 23 Overview of the thesis scope and future work.. 38

Figure 24 Conceptual overview of the tour application. Through a mobile device, the user can interact with stories tagged to buildings and regions in a city. 41

Figure 25 Mobile phone as pointing device. **A**: A red icon indicates that a building nearby has an audio source attached to it. **B**: Pointing the device directly at the icon plays the audio source.. 45

Figure 26 Occlusion issues. **A**: Blue dots are audio sources that are playable from the user's position. Red dots indicate audio files that cannot currently be activated. **B**: Scene from the camera view.. 46

Figure 27 3D sound and no visual interface. If the user enters a sound's "activation area", the sound is played back. Sounds fade in and out as the user navigates. If multiple audio sources are near each other, the user will have to walk closer to one of the sources to hear the audio, as illustrated by the four sources clustered towards the right in the figure. 47

Figure 28 Visualising the geo-tagged city stories. **A**: AR view with coloured icons superimposed on camera view of a smartphone over locations in the physical environment. **B**: Ortographic map view providing exocentric overview of the scene. **C**: List view with distance and direction indicators. 49

Figure 29 When a story is selected, additional information about a story is shown, while its story is played back. A media player interface lets users control the audio. 50

Figure 30 When a scene is selected, a new layout appears on the screen, while a spatial audio snippet is played back. From here, the user decides whether he/she wants to hear the full interview and explore additional content related to the story.. 51

Figure 31 Views switch when device is rotated. If the device is in portrait mode, a List view of the scenes is shown. If the device is in landscape mode, either the AR or the Map view is shown.. 52

Figure 32 UI widgets for filtering scenes in AR and Map view, and accessing the "Info" screen and the application's web site. A reticle in the center of the screen helps users aim at the scenes. 53

Figure 33 The foundation of the system is a native Android application, which links to the OutdoorAR framework provided by HITLabNZ. This framework integrates important librar-

ies and API's such as Google Maps, OpenGL and FMOD. 56

Figure 34 Activity diagram showing the procedure of deciding which scenes to display in the AR, map and list views. 57

Figure 35 In the AR view, scenes are superimposed over the camera view of the phone. **A**: A reticle helps users aim at the scenes. **B**: When a scene is in focus, a seperate layout appears while a spatialised snippet of the story is played back. Tapping the Play button on the layout takes the user to the SelectedActivity. 58

Figure 36 The map view shows all stories' icons superimposed on top of a map provided by the Google Maps API.. 59

Figure 37 The two list views. **A**: The regular list view displays stories from a particular category or, as in this example, all stories sorted by their distance to the user. **B**: Tapping the "Themes" button takes the user to another list view, displaying a list of the themes. Tapping a theme or the "Nearby" button takes the user back to the regular list view, now displaying the stories based on the user's selection. 60

Figure 38 UML activity diagram depicting the orientation switch effect, here using the list view for illustrative purposes.. 61

Figure 39 The screen appearing when a story is selected for full playback contains a large image of the building, a short textual description and a basic media player for audio control. **A**: Landscape mode. **B**: Portrait mode. 62

Figure 40 Demographic information about participants of the usability experiment.. 68

Figure 41 Results from the usability test. **A**: Task completion times. **B**: Error rates. 69

Figure 42 Increasing the area of the screen, which the scene should intersect to be in focus. **A**: Originally, a scene needed to intersect the exact centerpoint of the screen. **B**: In the improved version, a scene should intersect a virtual square, which occupies an area of 20x20 pixels around the center of the screen. 72

Figure 43 Location of scenes in the two settings for the experiment. **A**: Sindal setting around the city's main street. **B**: Aalborg setting at the central harbourfront. 75

Figure 44 Breakdown of test participants. 76

Figure 45 Participants conducting the experiment in the two settings. **A**: Participant in Aalborg with busy traffic (and a weekend market) in the background. **B**: Participant in the more quiet setting in front of the church in Sindal. 77

Figure 46 Subjective measurements from the field trial. Participants rate their level of agreement to six different statements related to errors, time-on-task and ease. Responses are converted

to a numeric scale from 1-5 with 1 being the negative extrema, and 5 being the positive extrema. 79

Figure 47 Mean score for the SUS questionnaire. 80

Figure 48 Expanding the audio tour to integrate user-generated stories. **A**: When aiming at a building, icons representing different media de-cluster from the building's standard icon (a blue or green dot). Notice the "Author" button in the UI hud, which lets users create their own city stories. **B**: Touching one of the media icons plays back the latest city story of this medium and lets the user socially engage by responding to the story through audio, video or photo captures. 88

Figure 49 An alternative approach to engage social interaction. **A**: Pointing at a building depicts the latest media updates from friends, tagged to this location. **B**: Tapping the "Author" button on the UI allows a user to create his/her own city story. The location can be exactly pinpointed, and photo, audio or video captures can be attached to the story. 89

Figure 50 Combining pre- and user-generated stories. **A**: Aiming at a blue icon tagged to a building plays back its "official" story - a story generated by the city's council. A green icon represents a source that is currently playing. **B**: When a scene is in focus, media icons representing user-generated stories de-cluster from the green icon. **C**: Tapping the Photo icon depicts a list of user-generated photos left at this location. The user can capture a photo and generate his/her own city story. 90

Figure 51 Milgram's reality-virtuality continuum. AR can be placed near the centre of the figure, since it enhances the user's perception of the real world instead of fully replacing the real world (Milgram & Kishino 1994). 105

Figure 52 Creating the impression that the real and virtual world coexist is the fundamental challenge of augmented reality. In this example, the table is real and the lamp and chairs are virtual. (Azuma 1997) 105

Figure 53 AR applications serving three different purposes: **A**: AR chemistry for educative purposes (Paradox 2011). **B**: Heinz Tomato Ketchup recipe book appearing when scanning a bottle of ketchup with a smartphone (Blippar 2011). **C**: Qualcomm demonstration of AR for a magazine advertisement for footwear (Qualcomm 2012). 106

Figure 54 AR interaction methods. **A**: Mobile information browser in Wikitude (RVNews 2012). **B**: 3D interface using interpretation of a user's hand gestures (Piumsomboon et al. 2011). **C**: Tangible AR using a paddle-shaped object to manipulate digital content (Kato et al. 2000). 110

Table Of Contents

Preface v
Glossary/Abbreviations vii
List of Figures ix
Table of Contents 1
1. Introduction 5

PART 1: ANALYSIS 7

2. Related work 9
2.1 Mobile computing and social media 9
2.2 Mobile computing and augmented reality 10
2.3 Augmented reality and social media 11
2.4 Limitations of existing solutions 12
2.4.1 Invasion of privacy 12
2.4.2 Narrow social scope 13
2.5 Mobile tours, mixed reality and audio experiences 13
2.5.1 Discover Christchurch's Inner City Heritage 14
2.5.2 Roadside Stories 14
2.5.3 Soho Stories 15
2.5.4 "The O" in MONA 15
2.5.5 FreshAir 16
2.5.6 Mixed reality experiences and situated documentaries 16
2.5.7 Audio stickies: audio-based AR 16
2.6 Thesis contributions 17
3. City stories 19
3.1 2010-2012 Christchurch earthquakes 19
3.1.1 Stress, trauma and rebuild delays 19
3.1.2 Rebuild and recover: sparkling creativity 20
3.1.3 Media technology in the recovery process 20
3.2 New harbourfront in Aalborg 21
3.2.1 From industry to knowledge and experience 21
3.2.2 Remembering the old industry 22
3.3 Chapter summary 22
4. Audio, mobility and augmented reality 23
4.1 Ubiquitous computing and mobile devices 23
4.2 Mobile and wearable AR 23
4.3 (Spatial) Audio in the field 25
4.4 Challenges of audio interfaces 27
4.5 Chapter summary 28
5. Social media and social networking 29

5.1 Definitions 29
5.2 Mobile social media 30
5.2.1 Facebook for iOS and Android 30
5.2.2 Flickr for iOS and Android 31
5.2.3 Geo-social networking 31
5.3 Chapter summary 32
6. Mobile interaction design 33
6.1 Mobile contexts in urban environments 33
6.2 Physical restrictions, challenges and solutions 33
6.2.1 Limited display size 33
6.2.2 Occlusion and the "fat-finger" problem 34
6.2.3 Multi-modality and limited I/O technologies 34
6.3 Chapter summary 35
7. Problem formulation 37
7.1 Research goals 37
7.2 Limitations 38
7.3 Methodology 38
PART 2: PRODUCT DEVELOPMENT 39
8. Conceptual development 41
8.1 Overview of the problem space 41
8.2 Customer requirements 42
8.3 Target group 42
8.4 Conceptual model 43
8.5 Personas and use case scenarios 43
8.5.1 Persona 1: Amy, 45, Christchurch, NZ 43
8.5.2 Persona 2: Christian, 29, Aalborg, DK 44
8.6 Chapter summary 44
9. Design 45
9.1 Audio interface considerations 45
9.1.1 Mobile phone as pointing device 45
9.1.2 3D audio interface 46
9.1.3 Multimodal interface 47
9.1.4 Nearest audio file plays 47
9.1.5 Single-touch interaction 48
9.2 Designing the audio tour 48
9.2.1 Audio content: Interviews tagged to historic buildings 48
9.2.2 Browsing the stories 49
9.2.3 Selecting and playing back stories 50
9.2.4 Navigating between ego- and exocentric perspectives 51
9.2.5 Other screens 52
9.2.6 UI elements and user feedback 52

9.2.7 Icons and buttons 53
9.3 Chapter summary 53
10. Android implementation 55
10.1 System components 55
10.2 Android framework 55
10.3 Data structure & system flow 56
10.4 Views 57
10.4.1 ARViewActivity 57
10.4.2 MapViewActivity 58
10.4.3 ListViewActivity and CategoryListViewActivity 59
10.4.4 Navigating between AR, map and list views 60
10.4.5 SelectedActivity 61
10.5 3D audio using FMOD 62
10.6 Chapter summary 63
PART 3: EVALUATION 65
11. Comparative usability test of interaction methods 67
11.1 Purpose 67
11.2 Setup 67
11.3 Observations 69
11.3.1 Task completion times and errors 69
11.3.2 Impact of icon size 70
11.3.3 Personal preference 70
11.4 Results 70
11.5 Suggestions for improvements 71
11.6 Chapter summary 71
12. Empirical evaluation 73
12.1 Purpose 73
12.2 Setup 73
12.2.1 The tasks 75
12.2.2 Data collection 75
12.2.3 Participants 76
12.3 Observations 76
12.3.1 General observations and interview comments 76
12.3.2 Comparing Aalborg and Sindal contexts 78
12.3.3 Post-task ratings 78
12.3.4 System usability 79
12.3.5 Technical issues 79
12.4 Chapter summary 80
13. Summary of findings 81
14. Discussion 83

14.1 Social desirability bias 83
14.2 Low fidelity stories 83
15. Conclusion **85**
16. Future work - integrating social networking **87**
16.1 The problems of privacy: the interface is the issue - not the content 87
16.2 Expanding urban stories - sociAR 87
16.3 Engaging social interaction 88
16.4 Combining pre- and user-generated stories 89
16.5 AR 2.0: In-situ authoring 90
16.6 Stand-alone system or Facebook app? 91
16.7 The near future part 1: Improving and evaluating the audio experience 91
16.8 The near future part 2: Binding the stories together 92
16.9 Chapter summary 92
References **93**

PART 4: APPENDICES **103**

A. Augmented Reality 105
B. Storyboard: Amy, Christchurch 113
C. Storyboard: Christian, Aalborg 115
D. System navigation 117
E. Android activity state paths 119
F. UML class diagram of ARViewActivity 121
G. Stories for empirical field trial 123
H. Questionnaire : Post-task ratings 125
I. System Usability Scale (SUS) 127

1. Introduction

Every city has a story to tell. Walk through a city, and you can feel the hand of history as you pass by its historic sites. Major cities like Berlin, New York or Paris are associated with major historical events of global importance, and while other cities might not possess the same importance from a world-wide perspective, their history can still be of great value to their local inhabitants.

Recent advances in location-aware technology have made *geo-social networking*, where social media content is associated with geographic information, popular. On social networking and media services like Facebook[1] and Flickr[2], status updates and user-generated content are geo-tagged. Users check in with their current location and share location-based information with their friends and the public. Modern smartphones with GPS technology and various sensors turn mobile devices into particularly powerful tools for location-based services, which cover a wide range of application areas (Figure 1).

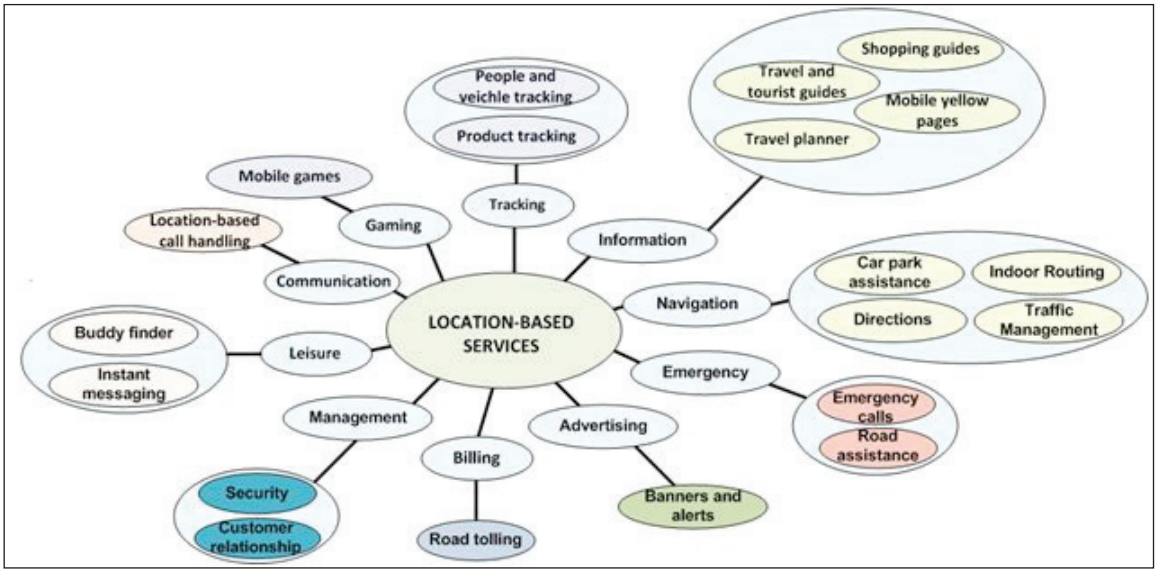


Figure 1: Location-based services cover a wide range of application areas (Nil 2013).

Today, the user of a smartphone has access to a lot of functionalities, and is able to solve many of his/her daily tasks directly from the phone (Iwata et al. 2010). Smartphones act as minicomputers and in 2011, the number of smartphones sold superseded the number of personal computers sold (Mashable 2012a). Many of the tasks people perform while "mobile" are related to location-based information, such as navigating, searching and locating (Reichenbacher 2003), and improvements in location-aware technology have made mobile devices highly capable of solving these tasks.

The improving technology has also opened for mainstream adoption of mobile augmented reality (AR) applications, where a user's perception of the real world is enhanced by superimposed digital content. On the mobile platform, AR can provide a powerful interface in location-aware systems (Höllerer et al. 1999), and be a valuable aid for the mobile user in many different scenarios (Höllerer & Feiner 2004).

1 http://www.facebook.com

2 http://www.flickr.com

An area, which AR has yet to embrace is social media, where users share their own digital content and articulate personal relationships (Gervautz & Schmalstieg 2012). With the current focus on location-based services in social networking, and if the following assumptions hold, this comes as a surprising fact:

- Mobile users enjoy geo-tagging content and share location-based information with their social networks.
- Location-aware technology is rapidly improving on mobile platforms.
- Augmented reality provides a powerful interface for location-based services.

The long-term goal of this work is to investigate the combined potential of AR and social media. A system, which takes advantage of AR as a location-based service, while maintaining the richness of desktop-based social networking and media services.

The scope of social media is diverse, so to delimit the project *urban story telling* might provide a context that makes it natural to integrate mobile AR and social media into one, coherent system. Grounded in stories about interesting locations and events of a city, users can share media content tagged to the physical environment, and AR technology can be used to register the user-generated content to the buildings.

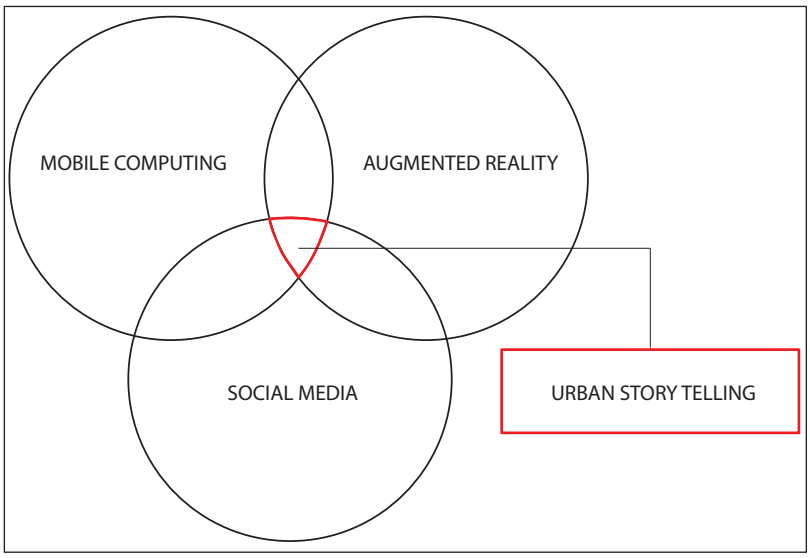


Figure 2: The context of urban story telling is the starting point for investigating the combined potential of mobile computing, AR and social media.

Urban story telling in mobile contexts also makes it appealing to look at the potential of audio as a communication medium. In both AR and social media, the role of audio is relatively under investigated when compared to the visual domain, but can help overcome some limitations of mobile devices. Further, audio can be used as a medium to present immersive stories about a city's interesting locations, which could be a foundation for user-generated content, in the shape of audio comments, images and video.

Therefore, this thesis will investigate the potential of a system, which delivers an audio-based mobile experience, and uses an AR interface to present digital content to the user. The long-term goal is to expand this experience to integrate user-generated content and aspects of social networking.

PART 1: ANALYSIS

In Part 1 the theory and problem domain of the thesis is analysed. First, chapter 2 presents related work related to the short- and long-sighted goals of the thesis. Chapter 3 presents two interesting contexts for urban story telling, while chapters 4, 5 and 6 deal with audio, augmented reality and social media, in particular their affordances in mobile computing.

2. Related work

This chapter presents existing work within the scope of the thesis. First, existing systems combining mobile computing, augmented reality and social media are described to provide an overview of the long-term challenges and opportunities for the thesis. Then, examples of mobile tour applications and audio-driven experiences are presented, which are relevant for the short-term goal: urban story telling in mobile settings. Finally, the limitations of existing solutions and the thesis contributions are presented.

2.1 Mobile computing and social media

Systems combining mobile computing and social media often take advantage of location-aware technology to add geospatial information to traditional social media services. This phenomena is called geo-social networking, a popular trend in recent days, where a range of location-aware social networking applications have been released (Cnet 2010). Dedicated geo-social networking applications include Loopt[3] and Foursquare[4] (Figure 3A), who let mobile users find friends nearby and identify interesting locations around them. At the same time, established social networking applications like Facebook and Flickr for iOS and Android allow users to interact and share media based on their physical location. Users associate content with a geographic position and "check in" with their own location. Foursquare even lets users earn badges when they check in at different locations, turning the location-aware feature into an entertaining game or treasure hunt experience.

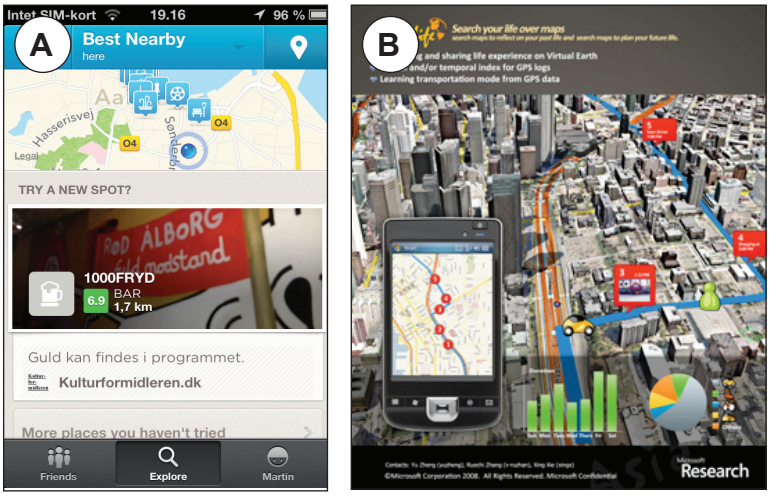


Figure 3: Dedicated geo-social networking systems. **A**: Foursquare presents interesting locations around the user based on other users' recommendations. **B**: Microsoft Research's Geo-Life recommends popular locations and travel sequences based on GPS trajectories (Microsoft 2013).

According to Chow et al. (2010), geo-social networking services often serve specific services, such as searching for nearby friends, sharing geo-tagged messages or using trajectory data to provide recommendations for multiple users. Microsoft Research's Geo-Life (Figure 3B) is a recommendation system,

3 http://www.loopt.com
4 http://www.foursquare.com

which collects GPS trajectories (time-stamped locations) from mobile users. These trajectory data are mined to identify popular locations and travel sequences within a region. For instance, when a tourist visits a city, its popular locations and travel sequences are presented as recommendations on the mobile display, which helps visitors navigate in the city and find the locations they are assumed to be interested in. In more general terms, Geo-Life lets mobile users share travel experiences. (Zheng et al. 2010)

An example of a system where users share geo-tagged messages is Google Buzz[5]. Originally a desktop-based service, the mobile version lets users share messages, images, videos and links and associate the content it with the user's location. The Buzz project was heavily criticized over privacy issues, and abandoned in 2011 following many lawsuits (WebProNews 2013).

2.2 Mobile computing and augmented reality

Rapid improvements in mobile technology have made AR applications popular in recent years, and these are now moving from research laboratories into larger consumer markets (Van Krevelen & Poulman 2010). The goal of mobile and wearable AR applications is usually to enhance the mobile user's perception of the real world by overlaying visual, auditory or haptic material (Feiner et al. 1997). In outdoor environments, location-aware technology (often GPS) and inertial sensors are used to track the user's viewpoint and register virtual content to a geographic location. The virtual content is superimposed on the camera view of a phone to present information seemingly tagged to the physical surroundings. One recent example is Nokia's City Lens[6], which presents information about places in the user's surrounding environment in categories such as Food, Shopping and Fun. The user can see the distance to the location, share information with his/her friends and check out user-generated reviews.

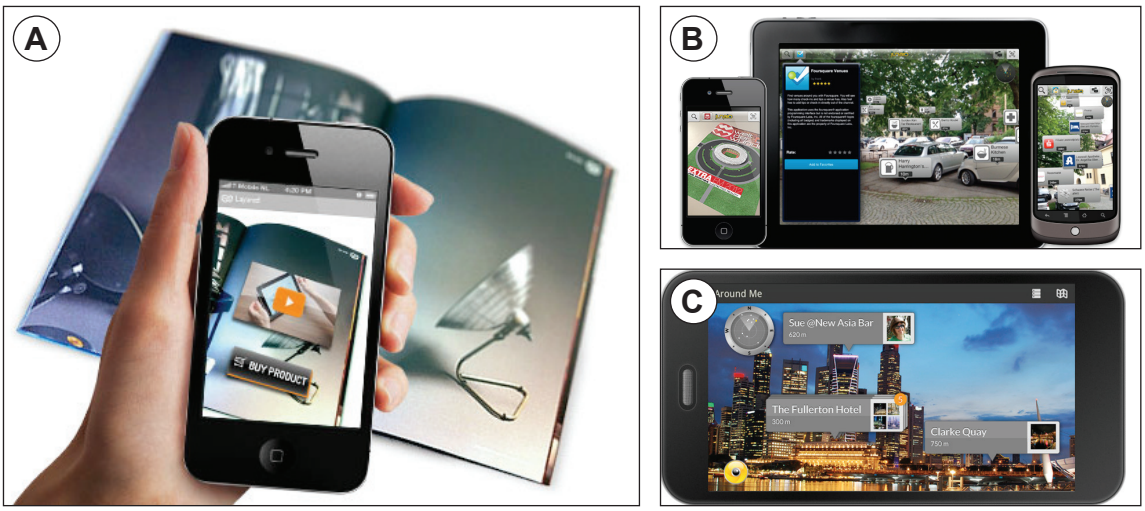


Figure 4: Mobile AR systems. **A**: Layar overlays virtual content on print media (Layar 2013). **B**: AR channel created in junaio (Junaio 2013). **C**: Wikitude AR scene showing interesting locations of a city (Wikitude 2013).

5 http://www.google.com/buzz

6 http://betalabs.nokia.com/trials/nokia-city-lens-for-windows-phone

Layar[7] specialise in vision-based AR by superimposing digital content on print media, such as magazines or postcards. Through a desktop interface (the Layar Creator) the user can associate a print page with digital content, such as images, videos or links to web pages. The digital content can be displayed and interacted with through the mobile Layar application, by scanning the print page through the camera view of the phone (Figure 4A).

junaio[8] (Figure 4B) is an AR browser, which lets users generate their own channels of information for others to download and experience. Examples of existing junaio channels include "London Bus Live", which provides information about bus stops around the user, and channels for identifying local shopping deals in a specific region. junaio provides the opportunity for creating both vision-based and geo-location based AR applications.

Finally, Wikitude[9] (Figure 4C) is an AR browser, where users choose from a range of categories, such as restaurants, accommodation or sights, to browse superimposed information about interesting locations around them. AR scenes can be browsed through either an AR view or a map or list view, which provide exocentric perspectives of the scenes.

2.3 Augmented reality and social media

Existing systems combining AR and social media are still rare, but the existing systems are mobile as well. Twitter 360[10] is a supplement to the social networking service Twitter, allowing users to geo-tag their tweets and identify Twitter friends in the nearby environment. When a user navigates in the environment, the distance to his/her friends (who have activated the geolocation option in the settings) can be seen on the display, which allows users to use the feature as a friend-finder tool (Figure 5A).

Another interesting concept is SocializAR[11], which is aimed at connecting people at conference and business events. A conference visitor wears an AR badge, which can be scanned by other visitors. When a visitor's badge is scanned, his/her LinkedIn profile appears on top of the camera view of the "scanner's" smartphone, along with the latest social networking activities surrounding the visitor. Further, the scanner can invite the other visitor to connect on LinkedIn (Figure 5B).

Figure 5: Combining AR and social media. **A**: Twitter 360 - users geotag their tweets and locate Twitter friends in the nearby environment (Twitter 2013). **B**: SocializAR - AR for conference and business networking (Twnkls 2013).

7 http://www.layar.com
8 http://www.junaio.com
9 http://www.wikitude.com
10 http://www.twitter-360.com
11 http://twnkls.com/socializar/

Other examples of combinations of mobile AR and social media/networking are limited to conceptual ideas, demonstrating important issues that must still be overcome before AR/social media applications will turn mainstream. In 2010, South African media entrepreneur Matthew Buckland and designer Phillip Langley envisioned a future of social networking as integrated with AR (Figure 6). Relying on accurate GPS, facial and object recognition and reasonable resolution, they imagine mobile users pointing their phone towards another person to instantly retrieve information about that person, including their friends and social networking updates. The obvious inherent privacy issues are discussed, but Buckland believes that it is comforting that users could simply refuse to provide access to social networks, and that "everyone will mostly be in the same boat" (Buckland 2010).

Figure 6: Envisioning a future of social networking and mobile AR, where a quick scan of a person reveals insight into this person's life (Buckland 2010).

2.4 Limitations of existing solutions

A few obvious challenges are inherent in existing systems combining mobile AR and social networking, which must be conquered before these ideas turn mainstream. Below, the issues are described in more detail, including possible solutions to help overcome them.

2.4.1 Invasion of privacy

As the systems presented above suggest, privacy issues in geosocial networking applications are usually addressed by letting the user explicitly switch location-aware features on or off or asking the user for explicit permission to use his/her personal location information everytime it is needed. This can make users suspicious, and cause them to simply avoid using the location-aware feature at all. This naturally limits the number of users and thereby the overall feasibility of the service. For instance, an application like Twitter 360 is likely more interesting when a large number of users actually activate the location settings. Another privacy issue, inherent in SocializAR and Matthew Buckland's vision, is obtrusion from strangers. A concept like SocializAR is feasible in a conference setting, where visitors agree on the

premises. It would, however, be inappropriate to scan random people on the street by holding up a phone in front of them, as this behaviour would be considered socially unacceptable and highly obtrusive.

The context of urban story telling can help overcome some of these obstacles in existing systems. If the user scans a location instead of a person, and by doing so gets access to information about other users' content associated with this location, users do not need to fear direct intrusion from strangers or reveal their current location. This could work under the assumption that users to an increasing extent geo-tag their content to locations, the "key" to solving privacy-aware and rich geosocial networking being the location itself, and not the user who associates content with it. A user would still have to reveal their location when geo-tagging content, for instance when taking a photograph of the Eiffel Tower in Paris to share it with his/her friends. However, they can still be in charge of what the location is used for, and does not need to reveal his/her current position.

2.4.2 Narrow social scope

A social networking service like Facebook is rich in its scope, allowing users to share media content as well as respond to friends' content through comments, replies and the "like" button. Further, a user can create special interest groups with a selection of friends or join groups beyond his/her own social network, and use the instant messaging tool for live chatting. Although mobile social networking applications have become richer in recent years, the systems integrating AR are still limited in the social features they offer. SocializAR allows conference visitors to get a quick preview of other visitors' LinkedIn profiles and suggest to connect, which is appropriate for a conference setting but not usable or appropriate on the street. Twitter 360 allows users to get information about the location of friends' tweets and eventually use the tool as a buddy-search tool, but not much more. When compared to their web-based counterparts, mobile AR systems integrating social media are still narrow in their social scope. In the ideal world, and in the context of urban story telling, mobile users would hold up their phone over locations in the cityscape, see content generated by friends, and do the same actions as they would on standard social networking services. That is, discussing, replying, commenting and "liking" friends' content.

2.5 Mobile tours, mixed reality and audio experiences

Mobile tour applications might provide the starting point for a system, which combines mobile AR, social media and urban story telling. Tour apps often make use of location-aware technology to help visitors of a region identify and get information about interesting locations around them. They are often targeted at tourism, museum and heritage contexts, and present information about locations as the user navigates in the environment. Tour applications usually contain a number of scenes with content tagged to geographical locations, browsed through some kind of geospatial visualisation, for instance a digital map. Companies like 7scenes[12] and MyTours[13] help non-programmers develop their own tour applications through web interfaces, and from their respective apps, a user has access to their tour databases. Below, a few of these tours are presented, as well as stand-alone systems, where the audio experience itself is the main feature.

12 http://www.7scenes.com

13 http://www.mytoursapp.com

2.5.1 Discover Christchurch's Inner City Heritage

This tour is developed in 7scenes, and can be accessed through the 7scenes application from the App Store or Android Market. The tour presents information about Christchurch's inner city heritage. Information about a site can be browsed through a map, which plots out the different locations of the tour as blue "place media tabs". Tapping on a scene displays the media content associated with the scene, and also notifies the user if he/she needs to be located at the spot to access the information. The issue of how to help the user navigate and synchronise his/her location with the story of a tour is an important challenge in mobile tour applications. In this application, the "story" can be joined from anywhere and browsed in any order, however, the associated content is not unlocked unless the user is within 25 meters of a site. In this case, the media content related to the site will pop up, and the user is asked to perform a task related to the site. This makes it necessary for the user to be in the area to fully explore the tour. The application also has a "Nearby" feature which lists the tour sites nearby, including their distance to the user.

2.5.2 Roadside Stories

Roadside Stories is an application created in MyTours. It provides an audio tour experience of New Zealand's major travel routes, and contains 13 different tours with a total of 140 stops. Each stop is associated with an audio file and images related to the location. Further, each tour includes a map with the locations of its stops. The tours can be downloaded individually before going on the road. When on the road, the user can choose whether to play the whole tour, or select individual stops to play back at any time. The user can browse the scenes through a map or a list containing the stops within the tour. Each stop is associated with an audio file, which provides interesting facts about the location. The user can enjoy an image gallery and read a short description of the location. Unlike the 7scenes tour described above, users can enjoy the scenes from anywhere, and do not need to be near the locations to hear the stories. Figure 7 depicts screenshots from the application.

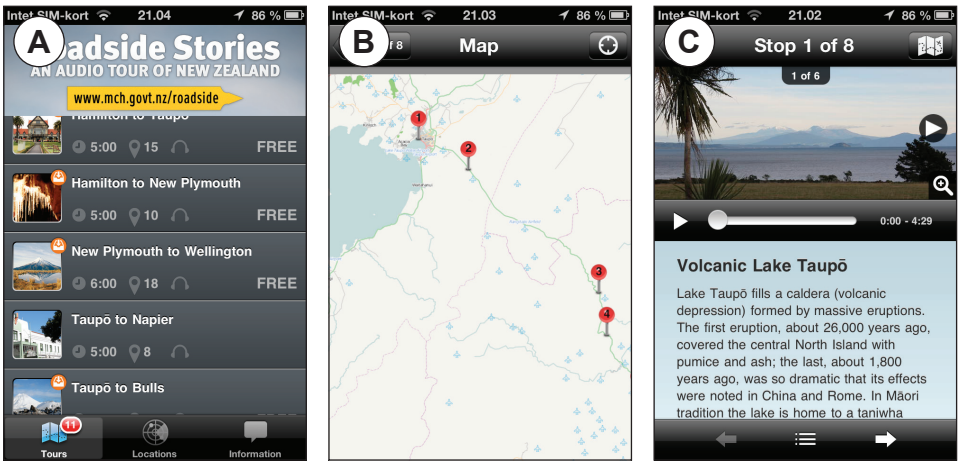


Figure 7: Roadside Stories tour. **A**: The user can download individual tours related to New Zealand's major travel routes. **B**: A map view shows the locations of the tour stops. **C**: A stop is associated with an audio file telling historic facts about the location, and an image gallery.

2.5.3 Soho Stories

This application provides an immersive audio tour of London's historic Soho area. As users walk through the area, short tales linked to the immediate surroundings play in the user's earphones. The stories fade in and out as the user walks towards or away from the sound landmarks. The user does not need to be in Soho to explore the stories. By entering the "Virtual" mode and touching a spot on a map view of the Soho area, a user far away can hear the corresponding story as well. This application is particularly interesting because of the immersive "field recordings", complete with background noise from the streets and the audio fading in and out as the user navigates in the area. Figure 8A depicts a screenshot of the application's interface.

2.5.4 "The O" in MONA

Instead of using labels on the wall to present their exhibition pieces, the Museum of Old and New Art (MONA) in Hobart, Tasmania has developed a mobile tour guide for their visitors. When they arrive, visitors are handed an iPod Touch on which an application runs, called "The O" (Figure 8B). This application provides an interface for accessing information and engaging multimedia related to the gallery. A list view shows the art objects near the user, which the user can select to get more information, as well as browse images and read responses and stories about the art pieces. The application also provides raw, low fidelity audio interviews with the artists (FreshAndNew 2011). A visitor can further "love" or "hate" the individual art pieces, share the artwork with friends, and save his/her tour to access information about the visit later on.

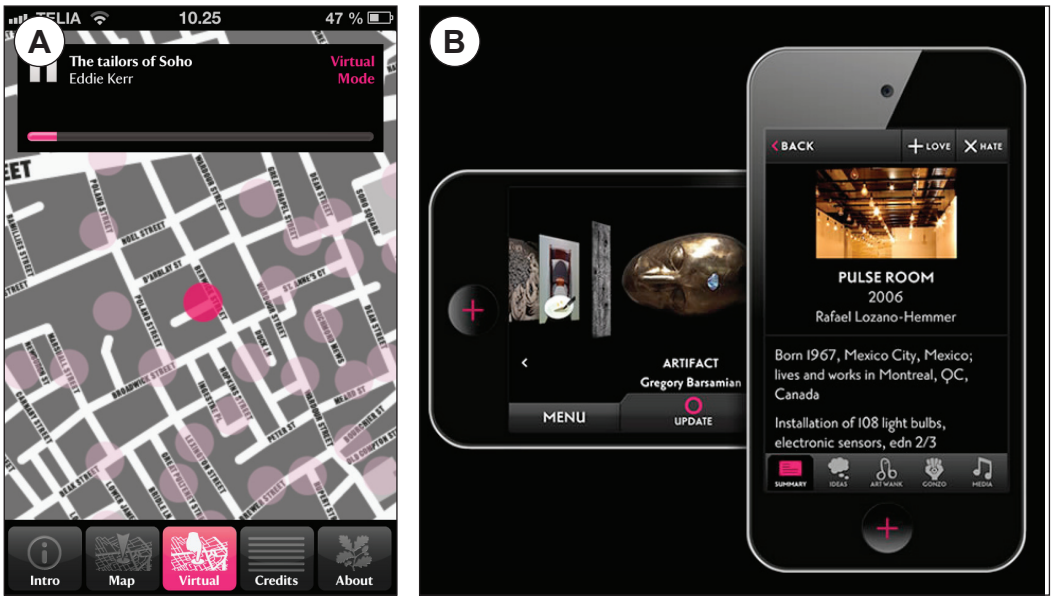


Figure 8: Apps for city stories and museum tour guides. **A**: Soho Stories provides immersive stories tagged to the historic Soho district in London. **B**: "The O" provides an interface for exploring art galleries at Museum of New and Modern Art in Hobart, Tasmania (Mona 2013).

2.5.5 FreshAir

FreshAir[14] is a tour application, which integrates AR and lets users author personal city stories and present different "realities" tagged to the physical environment for visitors of a region. Through a web interface, users can create and edit tours with their own, personal stories, and tag them to physical locations, for instance their home university. These can then be browsed by visitors to the area through an AR interface, by pointing towards the locations in the region through the camera view of an iPhone or Android device.

2.5.6 Mixed reality experiences and situated documentaries

Stephen Dow and his colleagues present an audio-based mixed reality experience for visitors at the historic Oakland Cemetery in Atlanta, Georgia, USA. Visitors at the cemetery can explore a dramatic narrative, which informs about the deceased and the periods in which they lived. A handheld interface lets users explore additional content related to the gravesites. (Dow et al. 2005) This work builds upon early work in wearable computing and AR, which was pioneered by Feiner et al. (1997), and expanded into research on "situated documentaries", where users explore a documentary of the history of the Columbia University campus in situ through a wearable setup (Höllerer et al. 1999). The user wears a head-mounted display, which presents points of interest around the user. A handheld device allows the user to browse the material embedded in the environment, for instance video clips, images and other multimedia content. Chapter 4 and Appendix A describe wearable and mobile AR in more depth.

2.5.7 Audio stickies: audio-based AR

An interesting extension to the Oakland Cemetery experience is the use of spatial audio techniques to geo-tag the individual stops on the tour. Langlotz et al. (2012b) introduce Audio Stickies - spatially aligned and user-generated audio annotations. The authors argue that existing AR solutions either use audio or visual media, but do not consider their combined potential. The user-generated audio stickies are precisely linked to the environment using panorama-based tracking (Wagner et al. 2010), and enhanced by visual cues on the mobile display. An audio sticky can be played by pointing towards it using the camera view of a smartphone. Tapping the screen on a designated position lets the user create a new audio sticky. In the system, a simulation of 3D sound is implemented by panning the sound from left to right in the stereo channels and adjusting the volume.

The Audio Stickies project builds upon earlier experiments on augmented audio. The earliest example is Benjamin Bederson's system, which sought to enhance the social experience of a museum experience by playing an audio description of a piece as a visitor, or a group of visitors, walks up to the piece (Bederson 1995). This opens for a more social and user-controlled experience, which is different from the traditional approach in museums. Usually visitors rent audio tapes telling about the artworks, which is more limited in the social scope and provides a linear experience with less user freedom. Haller et al. (2002) combine sound and graphics in an AR application, which lets users generate and manipulate 3D audio sources using a pen as input device and observing the audio sources through a head-mounted display. The audio sources can be moved freely and users can ob-

14 http://www.playfreshair.com

serve the results immediately. Rozier (2000) presents linked audio sources tied to a location in space. When a user is in the vicinity of the location of an audio source, the sound is played back to the user.

2.6 Thesis contributions

The limitations of existing systems combining mobile AR and social media have already been described and discussed. This thesis will seek to present a foundation for enhancing the scope and user experience of these systems, through an audio-based experience in urban settings, which provides the following:

1. A mobile audio tour, which tells stories geo-tagged to urban locations.
2. The possibility to naturally expand the audio experience through the integration of social media and networking aspects.

The work presented in the remaining chapters will briefly touch the areas of mobile social media and networking, which is part of the long-term goal of the thesis. However, the main focus will be to investigate the role of audio and AR in mobile computing in the process towards an experience on urban story telling. This experience will serve as the foundation for an eventual integration of social networking and media in the future.

The next chapter presents two interesting urban contexts, which can provide the stories and stand as immersive settings for a mobile audio experience on urban story telling.

3. City stories

This chapter presents two cities, where recent events have provided interesting contexts for social story telling. The first stop is Christchurch, New Zealand's third largest city, where major earthquakes have left the city's buildings in ruins and its inhabitants struggling to recover from the traumatic events. The second city is Aalborg, Denmark, whose harbourfront is currently undergoing dramatic changes due to the transition of the city from an industrial to a research oriented city.

3.1 2010-2012 Christchurch earthquakes

In September 2010 the Canterbury region of New Zealand was hit by a magnitude 7.1 earthquake, causing widespread damage to Christchurch, New Zealand's third largest city. The earthquake was followed by a 6.3 magnitude aftershock in February 2011, which had its epicenter much closer to the city, leaving the city's central business district in ruins, and causing severe damage to the city's suburbs. This earthquake caused 185 fatalities. Since those two large earthquakes, the city has experienced ongoing aftershocks continuing to remind the inhabitants of the tragic events, which occurred in their city.

Figure 9: Earthquakes have left Christchurch in ruins. **A**: The WestPac Trust Building in the High Street area is facing demolition. **B**: The cathedral, the city's trademark building, lost its spire following the February 2011 earthquake (PlaybackTheatre 2011).

3.1.1 Stress, trauma and rebuild delays

The ongoing earthquakes have had severe effects on Christchurch's inhabitants. Many people living in affected areas feel stressed, anxious or depressed following the events (Dorahy & Kannis-Dymand 2013). The city has lost many of its trademark buildings, including the famous cathedral (Figure 9B), and is struggling to recover its identity. Further, the reccuring aftershocks constantly remind Canterbury inhabitants of the shaky ground beneath them. The aftershocks and decisions on land also cause significant delays in the rebuild process (NZHerald 2012). Christchurch has many historic heritage buildings, and conflicting interests make the rebuild process long and tiresome.

3.1.2 Rebuild and recover: sparkling creativity

The events in Christchurch have also seen the community rise in a joint effort to reshape the city's identity. Following the earthquakes, many creative efforts and projects have been put forward. Weekly "Gap Filler" events activate the vacant sites in Christchurch with concerts, dance events and football, providing great opportunities for people in Christchurch to have fun and experience normalcy in their difficult everyday lives.

The Re:START mall on Cashel Street is another attempt to get people back to the city centre. This is a pedestrian shopping mall, built in eight weeks using shipping containers, providing a delightful heartbeat to the city and the first steps in a long rebuild process (Inhabitat 2011). The exciting process of rebuilding Christchurch has even caused Lonely Planet to include the city in their Best in Travel 2013 guide, emphasising Christchurch's "unique opportunity to rethink urban form" (LonelyPlanet 2012).

Figure 10: Rebuilding Christchurch. **A**: The Re:START container mall in Cashel Street (Inhabitat 2011). **B**: GapFiller minigolf event appears at various vacant sites of the city. **C**: Community chess at vacant site in Colombo Street. (GapFiller 2013).

3.1.3 Media technology in the recovery process

At the University of Canterbury, which is based in Christchurch, researchers have explored unique ways to use technology in Christchurch's recovery process. This has provided some interesting and creative application, which help Christchurch inhabitants explore their city's past and future. CityViewAR[15] (Figure 11A) is a mobile AR application, which displays 3D models of the destroyed or demolished buildings superimposed over the camera view of a smartphone at the empty sites. This application provides a way

15 http://www.hitlabnz.org/index.php/products/cityviewar

for the city's inhabitants to remember what the city used to look like, and has become quite popular.

Another product from the Canterbury University is the CCDU 3D application[16] (Figure 11B). This application is targeted at tablets, and visualises the Christchurch Central Recovery Plan, which outlines the future development of Christchurch (CCDU 2012). Those two applications offer unique opportunities for Christchurch inhabitants to explore their city's past and future while navigating its streets.

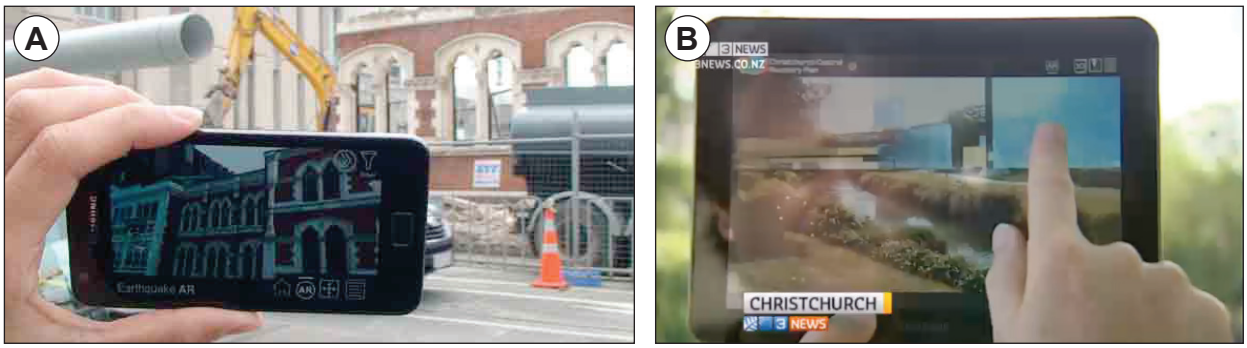


Figure 11: Using media technology to explore the city's past and future. **A**: CityViewAR shows 3D models of the demolished buildings in Christchurch (FutureGov 2012). **B**: CCDU 3D presents the future development plan for Christchurch's central business district (3News 2013).

3.2 New harbourfront in Aalborg

The harbourfront in Aalborg, Denmark's fourth largest city, is undergoing major changes in these years, as part of the transition of the city from an industrial into a knowledge-oriented city. Due to its central location in a city, which is an important transport junction, the shape of the harbourfront has been affected by many political strategies and discourses throughout the years (Galland 2011). Traditionally, the city has been dominated by industry, such as C.W. Obels tobacco fabric, Dansk Andels Cementfabrik (DAC, cement factory) and the power station buildings at Nordkraft. However, soon the industry will be largely replaced by knowledge-centered urban spaces, but will still be remembered as a vital part of the city's history.

3.2.1 From industry to knowledge and experience

The harbourfront in Aalborg is seen as a key component in transitioning the city into a knowledge oriented city. Combined with the postmodern wave of the late 20th century, the closure of several of Aalborg's historic industry along the harbourfront has caused the city to rethink its urban regeneration, shifting emphasis towards urban living conditions and recreation (Galland 2011). While previously focusing on expanding the important industry, the city's focus is now on creating attractive urban spaces, which provide opportunities for creativity and experiences (Aalborg-Kommune 2010). An example of this is the Utzon Center at the central harbourfront, which is now a dynamic center for exhibition and science. Also, the old power station buildings at "Nordkraft" are now turned into a center for culture and leisure, hosting a lot of concerts and venues (Figure 12B).

16 https://play.google.com/store/apps/details?id=com.hitlabnz.c3d

3.2.2 Remembering the old industry

In spite of the exiting transition of the city, Aalborg has had a long history of industry and manufacturers, and the old buildings are surrounded by nostalgia and stories, which are an important part of the city's identity. For many inhabitants in Aalborg, the transition of the city is a rapid and dramatic change, and it can be difficult to remember what the city used to look like. To this extent, stories about the past, present and the future of the cityscape will be be a valuable contribution to the city's identity, and a natural object for discussing and sharing urban stories.

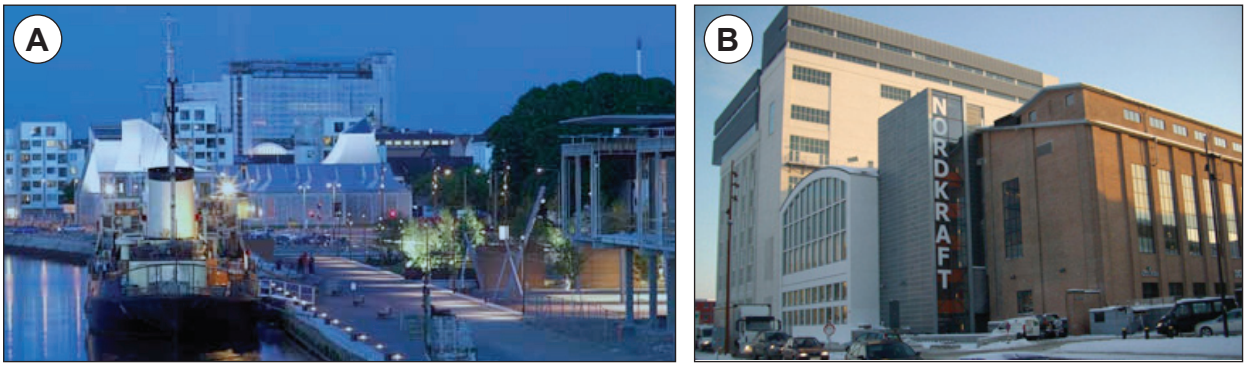


Figure 12: Major changes in the cityscape. **A**: Aalborg's harbourfront is undergoing dramatic changes these years (AalborgKommune 2010). **B**: The old power station buildings at Nordkraft are now centers for concerts and venues.

3.3 Chapter summary

This chapter has presented two cities where recent events have created a potential for social, urban story telling. Events like those presented in this chapter may provide a foundation for mobile experiences on city stories, and it is easy to imagine many other cities, where events and occasions make urban story telling interesting. The two contexts presented here are exciting because of the nostalgia surrounding historic buildings that are no longer present in the cityscape, and the curiosity of what might be there in the future. The next chapter digs deeper into the potential of mobile computing in urban environments.

4. Audio, mobility and augmented reality

With rapid improvements in mobile technologies, systems that increase users' awareness of and interaction with their surrounding environment, have seen release in recent years. Mobile devices, in particular smartphones and tablets, possess sensors, which to various extents are capable of measuring the user's context of usage, and present information to the user based on this context. Location-awareness is the most widespread example of integrating contextual factors in mobile systems. In recent years, AR has become a popular interface to location-aware systems, in particular in the visual domain. Audio-driven AR experiences have been invented as well, often in the shape of mixed reality narratives or location-based storytelling. This chapter presents concepts and applications, which provide mobile, context-aware experiences for mobile users, with a special focus on audio and AR.

4.1 Ubiquitous computing and mobile devices

In 1991, Mark Weiser envisioned a future, where computers would disappear and become seamlessly integrated into people's everyday life, with no one noticing their precense. The idea of allowing computers and devices to fade into the background, interconnected in ubiquitous networks, led to the famous term *ubiquitous computing* (Weiser 1991). Weiser envisioned tabs, pads and boards as ubiquitous input devices, although the true power of ubiquitous computing would come from the interplay of all these devices.

Smartphones and tablets of today can be considered the first truly ubiquitous computers, helping users to manage everyday tasks, keep in touch with others, and seeing, hearing and sensing the environment (Ballagas et al. 2006). Camera-enabled and location-aware smartphones provide opportunities to integrate the urban environment into mobile experiences. One example is *pervasive games*, where regular games are expanded spatially, temporally or socially (Montola 2005). *iPerg* was a EU funded project on pervasive games running from 2004 to 2008. The objective was to create new game experiences, "tightly interwoven with our everyday lives through the objects, devices and people that surround us and the places we inhabit."[17] The games of the iPerg project cover treasure hunt games, which use cellular positioning on mobile phones in location-based gaming situations, alternate reality games, where everyday life is woven in narratives that layer additional interaction upon the real world, and urban adventure games where AR technology is used to create gaming experiences immersed in the real world. Figure 13 depicts two examples of pervasive games from the iPerg project.

4.2 Mobile and wearable AR

Augmented reality (AR) is a particularly interesting technology for mobile environments, since it can provide a powerful interface for location-aware mobile computing (Höllerer et al. 1999). In AR, virtual objects are superimposed over a user's view of the real world, to give the impression that the real and digital world coexist (Azuma 1997). For mobile contexts, the challenge of AR is to provide a service, which integrates virtual information into a person's environment without constraining the individual's location to a specially

17 http://iperg.sics.se/index.php

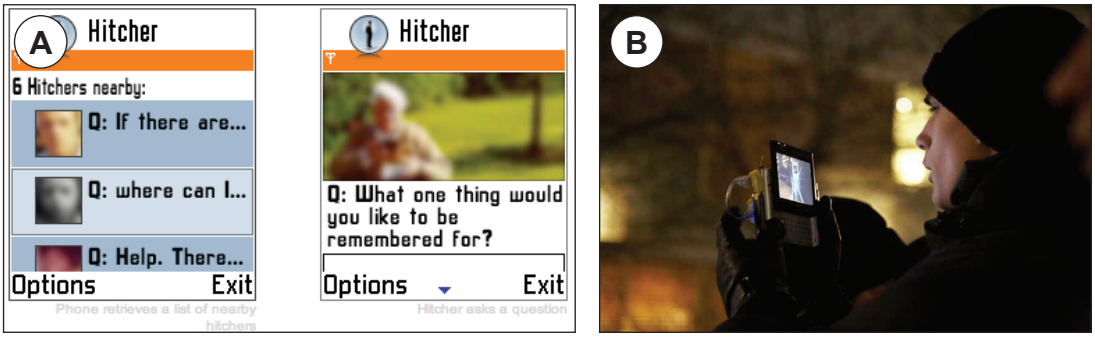


Figure 13: Pervasive games from the iPerg project. **A**: In "Hitcher", users control a hitchhiker and interact with other hitchhikers through a location-aware mobile device (MutualFriend 2008). **B**: In the game "Interference", AR technology is used to create an immersive urban adventure game (Iperg 2008).

equipped area (Höllerer & Feiner 2004). Some of the important challenges in AR include input, display and tracking technologies. Those challenges are discussed in more depth in Appendix A. Usually, AR on mobile platforms relies on GPS technology and inertial sensors, such as compass, accelerometers and gyroscopes to track the user's position and orientation, and register virtual content to locations in the real world.

Mobile AR was pioneered by researchers in wearable computing systems in the 1990's. The *Touring Machine* (Feiner et al. 1997) is the first mobile AR system (MARS), where the user wears a backpack containing a computer setup, and interacts with the system using a stylus on a handheld display. Further the user wears a see-through head-mounted display (HMD), where digital labels depict names of real buildings in the user's surrounding environment, and appear to float over the buildings. Through the handheld display, or using a trackpad to cycle through elements appearing on the HMD, the user can access basic information about the buildings in the environment. The Touring Machine was improved in a more complex system on *situated documentaries* - narrated multimedia presentations combining audio, still images, video sequences and graphics (Höllerer et al. 1999). The context was a university campus tour, telling stories about events that took place at the Columbia University campus, with virtual flags denoting interactive points of interests in the user's surroundings.

Since the mid 1990's AR researchers have also explored the use of personal digital assistants (PDA's) and mobile phones as less bulky alternatives to MARS. The NaviCam system is a well known example of displaying virtual annotations embedded in the environment through a PDA (Rekimoto & Nagao 1995). In 2003, the first untethered mobile AR system to overcome the bulky hardware and socially unacceptable nature of MARS was presented (Wagner & Schmalstieg 2003).

With location-aware and camera-enabled smartphones of today, mobile AR is now reaching larger industries and consumer markets (Van Krevelen & Poulman 2010). AR browsers, such as junaio, Wikitude and Layar provide the opportunity to create outdoor AR experiences, for instance to present information about locations in the surrounding urban environment. Also, dedicated AR games where users play relative to their physical location have been developed. In Mozzies, virtual mosquitoes are superimposed on the camera view of a Siemens SX1 phone, the task being to kill the mosquitoes by aiming and shooting at them[18] (Figure 15A). In 2012 another game, GeoBoids, was presented, where

18 http://shkspr.mobi/blog/2010/11/augmented-reality-games-how-far-have-we-come-in-7-years/

users have to travel to outdoor locations to find and collect flocks of GeoBoids, special virtual creatures visible and hearable through the mobile device (Lindeman & Lee 2012) (Figure 15B). This project is of particular interest, since it uses spatial audio techniques to help the user orient and locate the flocks.

Also in 2012 Google released Ingress[19], which is a massively multiplayer online video game incorporating AR. Users join one of two forces (Resistance or Enlightened) and battle to hack portals centred around public arts and parks in large cities. Ingress relies on users actually moving in the real world, which is necessary to claim portals for their team and protect them from their enemy. Users browse the portals on a map and through simple AR overlays over the real world (Geekosystem 2013).

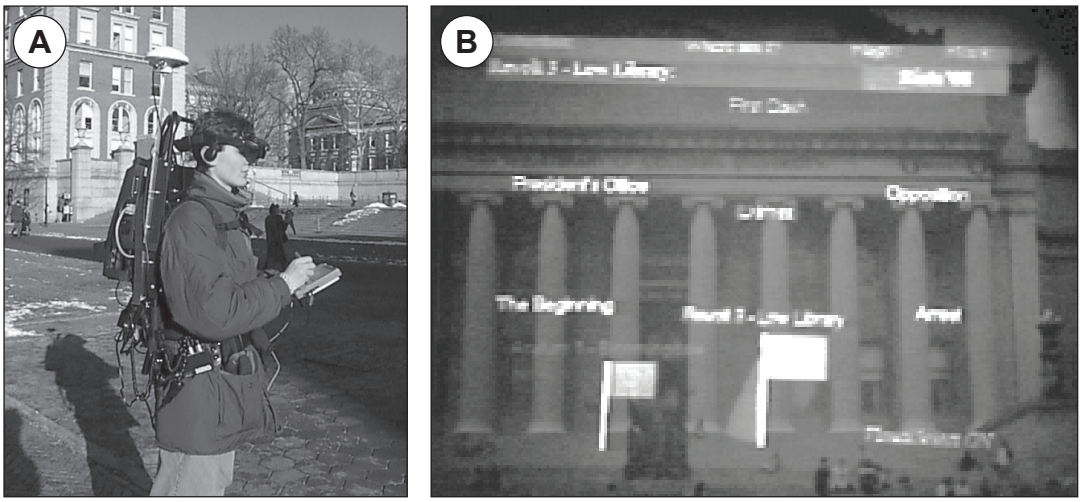


Figure 14: Mobile augmented reality systems. **A**: User performs a task on the first Touring Machine (Feiner et al. 1997). **B**: Virtual points of interest are embedded in the environment. Interacting with the digital flags presents narrated multimedia presentations at a university campus (Höllerer et al. 1999).

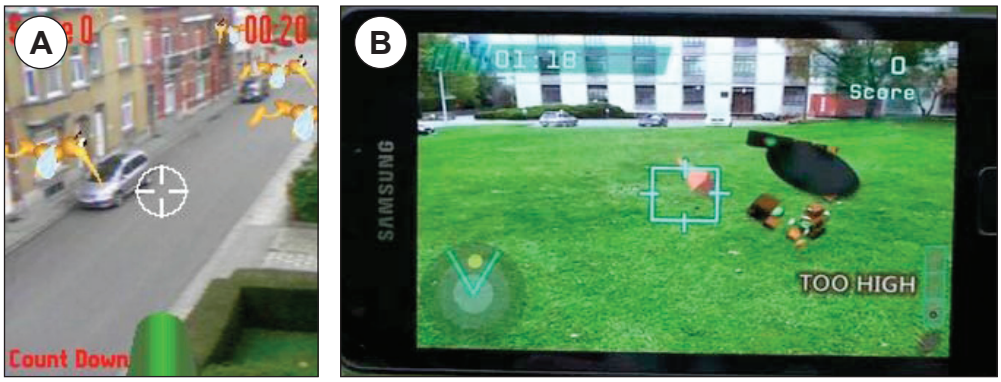


Figure 15: Mobile AR games. **A**: The award-winning Mozzies game running on the Siemens SX1 phone (Eden 2010). **B**: GeoBoids integrates a location-aware gameplay and spatial audio techniques (Lindeman & Lee 2012).

4.3 (Spatial) Audio in the field

The role of audio in mobile environments has previously been limited, partly due to poor audio output options (Dunlop & Brewster 2002). However, some researchers have developed immersive audio ex-

19 http://www.ingress.com/

periences in mobile and outdoor environments, where users wear head- or earphones and explore audio content in outdoor settings. Often these systems use handheld, visual interfaces to support the audio experience. One example is a mixed reality experience, which uses audio as its primary medium for visitors at the Oakland Cemetery in Atlanta, Georgia, USA (Dow et al. 2005) (Figure 16A). While navigating at the cemetery, users listen to a dramatic narrative telling stories about the deceased and the period in which they lived. The authors use a spatial narrative supported by a linear story, which provides an overall structure for the user, and ties the content together. At each site (a grave) the visitor can explore additional content related to the site. In this way, the authors both provide system control through the narrative, and user control by letting the user take over the system when he/she wants to explore specific content at a site. In the spatial narrative, a narrator guides the visitor from grave to grave. For the additional content related to each stop, the user can interact through a handheld controller to explore different topics or receive more content related to the site. Further, the controller has buttons for repeating and pausing the audio, and adjusting the volume. The issue of agency in location-aware applications is thoroughly discussed in this project. Should the user's location control the narrative, or should the story guide the user around to the different locations? In this example, if the user leaves the narrative, he/she is guided by statements such as "Turn around" to help the user return to the path of the narrative.

Other mobile audio experiences, similar in nature to the Oakland Cemetery experience, include an audio tour of the Alcatraz Prison in San Francisco[20] and various museum tour guides, such as Bederson's early example of audio-driven AR (Bederson 1995). Another example is the Sotto Voce project (Figure 16B), which emphasises social interaction in museums. Two museum visitors can share audio content, that is descriptions about objects in the environment, through single-ear headsets and using an abstract concept for audio sharing, which the authors refer to as "eavesdropping" (Aoki et al. 2002).

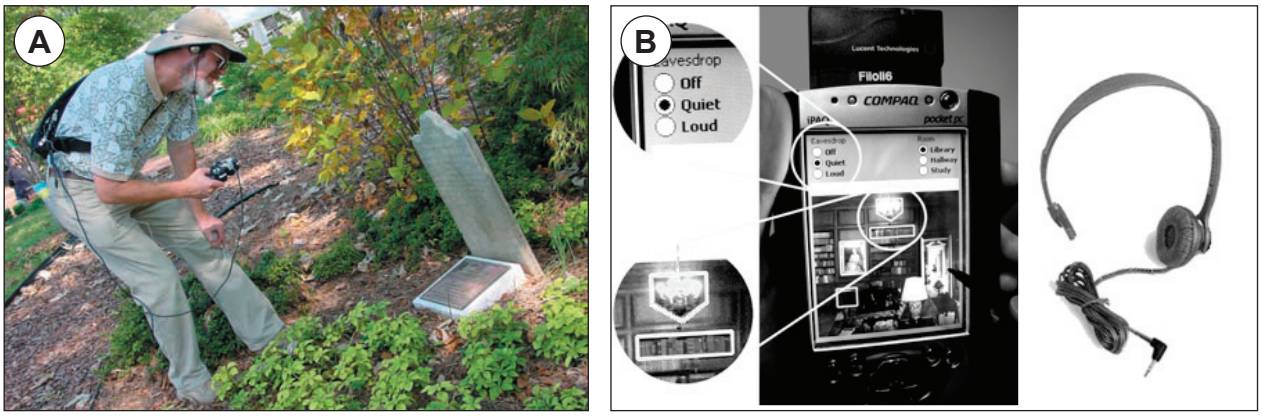


Figure 16: Audio in the field. **A**: User explores the narrative of the Oakland Cemetery mixed reality experience (Dow et al. 2005). **B**: Equipment of Sotto Voce, which emphasises social interaction at museums (Aoki et al. 2002).

While many mobile audio experiences rely on the use of standard earphones and playback similar to early walkman experiences, more intelligent and interactive audio techniques provide opportunities for spatialised sound events, which can be hard to perceive in standard headphones (because of front-back confusion and lateralisation - the perception of a virtual sound as existing inside the head). The

20 http://www.nps.gov/alca/index.htm

ability to localise a virtual sound source is very important in *augmented reality audio*, the audio equivalent to visual AR, where virtual sounds are combined with a representation of the real acoustic environment, "so that virtual sounds are perceived as an extension to the natural ones" (Härmä et al. 2004). One technique to achieve sophisticated audio AR is so-called "Microphone-Hear-Through AR", where microphones mounted on the user's earphones capture sounds from the real world and mix the sounds with digitally created audio objects (Lindeman et al. 2007). Other techniques to overcome some of the issues of audio perception in traditional headphones and enhance users' perception of virtual sounds include binaural recordings and bone conduction techniques. In bone conduction, sound signals are delivered through the bony structure of the skull, which has the advantage that the virtual sounds do not block the signals from sounds from the real world (Lindeman et al. 2007).

Some researchers have investigated the combined potential of visual and audio AR. Haller and his colleagues demonstrate an interface, where tangible AR techniques is used to place and manipulate virtual 3D sound sources in physical space, using a pen as input device (Haller et al. 2002). On mobile systems, the Audio Stickies project allows users to leave precisely positioned audio annotations in the physical space. The audio comments are combined with visual cues on the mobile display, and an audio sticky can be activated by pointing towards it (Langlotz et al. 2012b). The project relies on a simulation of 3D sound, where audio sources closer to the user's physical location play louder, and the sounds are panned in the stereo channels of the headphones, corresponding to their position on the mobile screen.

4.4 Challenges of audio interfaces

While limited input and output options have previously limited the role of audio in mobile contexts, another important challenge is to support multiple sound streams at once without overloading the user (Vazquez-Alvarez 2010). While visual displays require the designer to consider how to avoid information clutter, this is even more important for an auditory interface. Researchers have found that the use of spatialised sounds in earphones can improve users' ability to perceive multiple sound streams simultaneously (Vazquez-Alvarez & Brewster 2010).

Another design issue regards when a sound should be activated and played back for the user. Some researchers have used proximity zones, where a sound is activated once a user is closer to the source than a threshold value. In their work on sound gardens, Vazquez-Alvarez and his colleagues use two levels of audio feedback as the user navigates in an audio environment: audio cues in an audio landmark's proximity zone, which guide users towards its activation zone - a narrower zone, which activates and presents additional content for the user. The sounds are spatialised in order to manage overlapping audio landmarks (Vazquez-Alvarez et al. 2012).

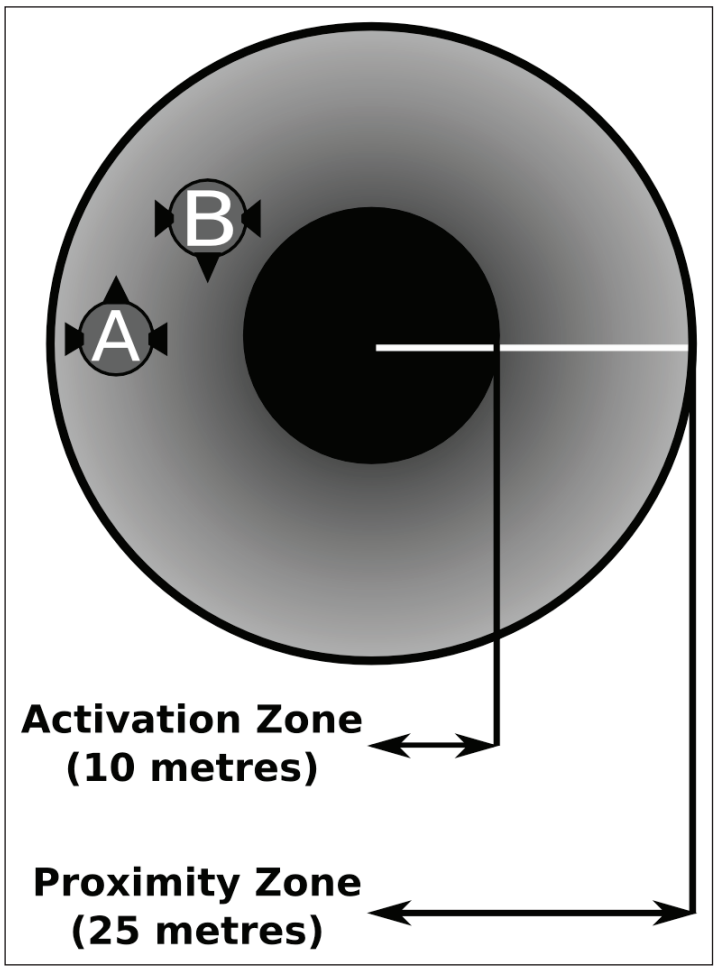


Figure 17: In their work on sound gardens, Vazquez-Alvarez and his colleagues use proximity and activation zones to present audio content to the user. The use of spatial sound techniques helps managing overlapping sound sources in the audio environment. (Vazquez-Alvarez et al. 2012).

4.5 Chapter summary

This chapter has described recent developments in mobile computing, which is a fast growing area due to rapid technology improvements. Context-awareness is an important feature of modern mobile devices, and location-aware technology such as GPS and inertial sensors have made pervasive games and location-based services popular. AR can provide a powerful interface to location-aware systems, but the use of audio has been relatively limited in mobile systems due to limited input and output options. Existing systems have mostly been in the shape of audio tours, where users explore audio landmarks in a physical environment, usually wearing earphones. Some researchers have explored the use of spatial sound techniques to improve the audio experience in mobile systems, and approach augmented experiences that resemble visual AR. The next chapter presents social media and networking, which are the focal points for the long-sighted goal of this thesis.

5. Social media and social networking

This chapter presents social media and social networking, since an integration of AR and social networking is the long-sighted goal of this thesis. Definitions of the two concepts are provided, including their differences and similarities. Examples of existing systems are provided, after which the focus is changed to mobile contexts, specifically the opportunities and challenges involved in social networking on mobile platforms.

5.1 Definitions

According to Lerman (2007), social media and social networking are two related, but distinguishable concepts. People use pure social networking sites like Facebook and LinkedIn to articulate personal and professional relationships, and to keep in touch with friends. Users of social media sites create personal networks to express their personal interests, and share those with people they do not neccesarily know offline. As for social networking sites, Lampe et al. (2006) find that users of Facebook mostly use the service to learn more about a person, with whom they share on offline connection, as opposed to finding persons online with whom they would like to connect offline.

An example of a social media service is Flickr, a social photo-sharing site, where users share their own images and participate in special interest groups and discussions. Other social media sites include SlideShare[21] (presentations and slides) and Digg[22] (news content). Independent of which content their users share, social media sites share the following characteristics (Lerman 2007):

- Users create or contribute content.
- Users annotate content with descriptive keywords, or tags.
- Users evaluate content, either passively or actively.
- Users create social networks by designating other users with similar interests as friends or contacts.

Kaplan & Haenlein (2010) classifies social media according to two well established theories, social presence and media richness. Social presence theory says that media differ in the degree of contact (acoustical, visual and physical) they allow to emerge between two communication partners (Short et al. 1976). Media richness theory is grounded in the idea that media differ in the level of information they allow to be transmitted in a given time interval, and that therefore, different media are not equally effective in resolving ambiguity and uncertainty, which is the assumed goal of all communication (Daft & Lengel 1986). Combined with Goffman's (2002) theory of self-presentation, which says that people wish to control the impressions they give to others in social interactions, social media can be classified according to Figure 18.

Of highest relevance to this thesis are content communities and social networking sites, since these two groups are expected to possess the largest potential in a context of urban story telling: browsing, discussing and sharing media content tagged to geographical locations in a city-

21 http://www.slideshare.net

22 http://www.digg.com

scape. The focal point is on mobile social media, where users use the services while "on the go".

		Social presence/ Media richness		
		Low	Medium	High
Self-presentation/ Self-disclosure	High	Blogs	Social networking sites (e.g., Facebook)	Virtual social worlds (e.g., Second Life)
	Low	Collaborative projects (e.g., Wikipedia)	Content communities (e.g., YouTube)	Virtual game worlds (e.g., World of Warcraft)

Figure 18: Classification of social media and networking sites based on social presence, media richness theory and self-presentation (Kaplan & Haenlein 2010).

5.2 Mobile social media

Today, many of the established social media and networking services are accessible as mobile applications as well. Many challenges exist when moving a web-based service to mobile platforms, and in social media contexts, these challenges are often related to usability and performance (Mashable 2012b), while maintaining the scope and richness of the original web-based service. However, as mobile technologies improve and mobile interaction design becomes more sophisticated, recent social media applications offer nearly all the features that users expect from their experience on the web, as well as additional features particularly interesting in mobile contexts. Below, analyses of two established social media and networking applications are provided, focusing on the limitations and opportunities offered by mobile platforms as compared to the web experience.

5.2.1 Facebook for iOS and Android

The social networking site Facebook has been offered on the iOS and Android platforms since 2009. When Facebook for iOS 5.0 was released in 2012, the Facebook app was for the first time built natively in iOS, while previously written in HTML5. This change caused a significant increase in the app's performance, which has traditionally been one of the pain points of the mobile version of Facebook (Verge 2012). Also, the user now has access to most of the features they know from the desktop version of Facebook. In the top of the screen, the user sees friend requests, messages and notifications via three buttons, which most Facebook users are familiar with from the desktop version. The app opens directly into the News Feed which continuously refreshes to show the most recent updates (Figure 19A). A left swipe takes the user to the sidebar, which lets the users choose what to view (user profile, messages, news feed etc.), while a right swipe reveals their contact list.

An important extension to the web-based version is the "Check-in" feature, a location-based feature, which allows users to inform about their current location in a status update. Further, a "Nearby" function can be accessed from the sidebar, which shows geo-tagged local landmarks in the vicinity of the user (Figure 19B). The user can see who of his friends like or have visited a certain location, as well as the distance to the location.

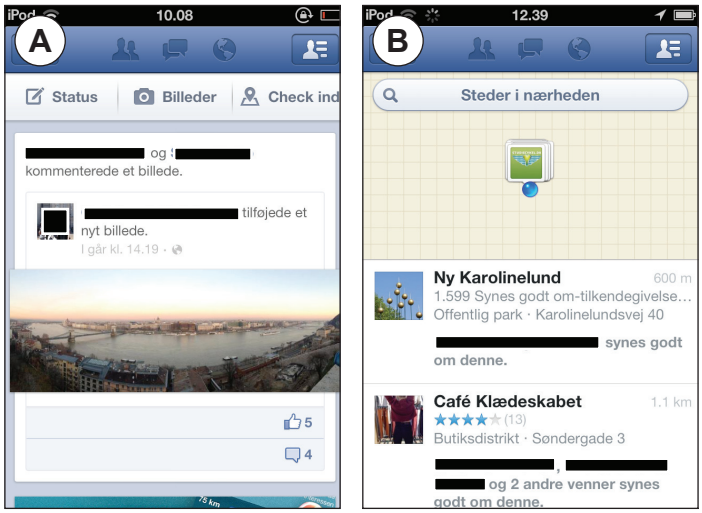


Figure 19: Facebook for iOS 5.0. **A**: The News Feed shows the most recent updates from friends. **B**: The "Nearby" feature depicts interesting locations and friends' recommendations near the user.

5.2.2 Flickr for iOS and Android

The image-hosting and social photo-sharing site Flickr is also available on mobile platforms (Figure 20). Like Facebook, Flickr for iOS and Android offers basically the same functionality as the web version, but with additional location-aware features. From a tab bar at the bottom of the screen, users can capture and edit their own pictures taken with the phone's camera, upload photo streams, browse photos from contacts or groups, comment on those, as well as edit photos captured through the phone's camera. A "Nearby" feature allows users to see photos uploaded close to his/her own location, providing a location-aware feature working similar to Facebook's Nearby feature. A News Feed stream similar to that of Facebook is integrated as well, refreshing with recent updates on a user's swipe command.

5.2.3 Geo-social networking

While Facebook, Flickr and other established social media and networking services expand their existing services with location-aware features on mobile platforms, dedicated geo-social networking applications focus primarily on the location element itself. Foursquare, who was already mentioned in chapter 2, is an important player in geosocial networking, allowing users to interact socially based on their location. A user can explore interesting places nearby, and see tips and recommendations from friends. Further, it is possible to earn a number of badges by being active.

There are many privacy concerns associated with geo-social networking, in particular when users need to reveal their own location. This information can be sensitive, and in serious cases open for threats such as stalking or burglary. Traditionally, privacy issues in location-aware services have been addressed under the assumption that the user is anonymous, but this can not be assumed in social networking contexts (Freni et al. 2010). The solution to privacy issues has often been to let the user provide explicit permission for the service to use his/her location information. However it is still possible, for instance on Facebook,

for other users to tag their location in a status update, and associate their friends with this location, without asking for permission. Privacy issues continue to be a major obstacle in geo-social networking applications.

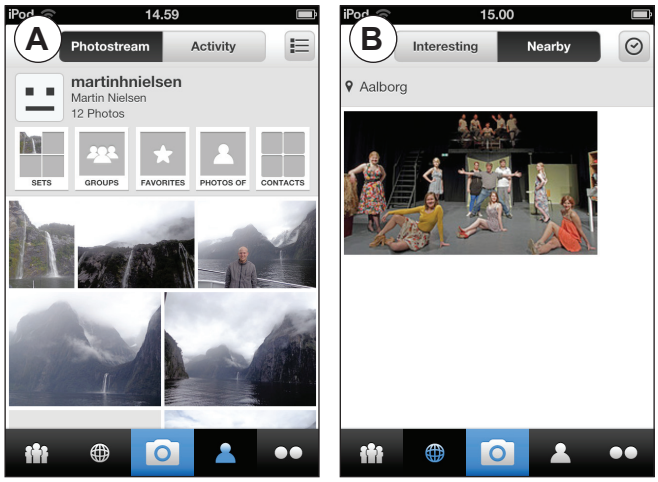


Figure 20: Flickr for iOS **A**: The user profile shows the user's photostream and lets him/her browse photos from groups and contacts. **B**: Browsing images generated and tagged to locations near the user's location.

5.3 Chapter summary

This chapter has briefly described social media and social networking as two related, but distinguishable concepts. The challenges of moving a web-based social media service onto mobile platforms have been described, but also the quality of recent social media apps, which are starting to integrate most of the features from the web-based versions. Improvements in performance and interaction design have made social media and networking apps more rich in content and social scope. Most mobile services include a number of location-aware features, which is an advantage of mobile social media over their web-based counterpart, but contain inherent challenges, primarily related to privacy. The next chapter highlights some of the important challenges in mobile interaction design, which are relevant for systems integrating social networking aspects and for a system relying on AR as the primary interface.

6. Mobile interaction design

While rapid technological improvements have made mobile devices very powerful, the contexts in which people use mobile devices are still different from those of traditional desktop interaction. Further, mobile devices are limited in input and output options, battery consumption etc., making the task of designing mobile interactions difficult. This chapter introduces relevant interaction design challenges, which are particularly interesting for AR systems.

6.1 Mobile contexts in urban environments

Users of mobile devices are often involved in other tasks than the mobile activity itself, and in contexts that are more complicated than desktop environments. To be successful, mobile applications should be aware of the contextual factors involved in mobile interactions. Recent smartphones with location-aware technology, such as GPS and compass, have made location-based services very popular. This is partly due to the fact that many of the tasks mobile users carry out on their devices are related to location-based information (Reichenbacher 2003). However, context in mobile interaction design involves more than information about the user's location. Social, psychological and physical factors are important as well. Tamminen et al. (2004) investigate these contextual factors, and find that mobile users in urban environments often have a mental plan on how to navigate from A to B, which gets interrupted by unplanned situational acts caused by context changes. Also, navigating through an urban environment causes users to pay attention to the immediate environment, requiring attentional resources from other tasks (such as a phone call or text message). These factors make it necessary to design mobile experiences, which demand minimal attention, and put minimal cognitive and physical load on the user. Further, in order to avoid embarrassment and disruption of the immediate environment, some researchers argue that there is a need for more subtle, discreet and "intimate" interfaces (Costanza et al. 2005). The social acceptability factor continues to be an important challenge for AR applications to conquer before turning mainstream.

6.2 Physical restrictions, challenges and solutions

Apart from contextual factors challenging mobile interaction designers, mobile devices are also restricted in their physical properties, which need to be taken into account when designing a mobile experience. First and foremost, the display size of mobile devices is restricted, because a mobile device in its nature should always be portable.

6.2.1 Limited display size

Even though touch-enabled smartphones maximise the visual display, the device itself still needs to be portable. This makes it important for the interface designer to be economic with the amount of visual information to present, in order to avoid overloading the user. Stephen Brewster has investigated the use of non-speech sound as a means to increase the usability of mobile button interaction and to reduce the size of buttons and other widgets on a mobile display (Brewster 2002). Mobile phones with inertial

sensors such as accelerometers and gyroscopes allow users to rotate or tilt the device to perform tasks, which would otherwise be accomplished through button interaction. This is seen in Nokia's City Lens application, where users navigate between different views by tilting and rotating the device (Figure 21).

Figure 21: Nokia City Lens uses the orientation of the mobile device to decide which screens to present to the user. **A**: When device is in landscape mode and tilted frontwards, an AR view is displayed. **B**: Tilting the phone towards the ground shows a map view. **C**: Rotating the device to portrait mode depicts a list view. (Nokia 2012).

6.2.2 Occlusion and the "fat-finger" problem

Even though touch screens help overcome many of the traditional issues with mobile interfaces, such as poor performance on pointing tasks (Wigdor et al. 2007), they possess challenges in their own nature. With direct touch, fingers can occlude the screen, potentially leading to misinterpretations and errors (Potter et al. 1988). The LucidTouch project addresses occlusion issues by providing an additional multi-touch input surface behind the mobile display, visualised through a pseudo-transparent screen (Wigdor et al. 2007) (Figure 22A).

Derived also from the limited display size, the "fat finger" problem is a well known issue with touch screens, in particular when using a touch-based keyboard with small buttons. Solutions to conquer this problem has included the use of tactile feedback (Brewster et al. 2007) and special word-gesture keyboards, which rely on drawing rather than typing (Zhai & Kristensson 2003) (Figure 22B).

In AR settings, researchers have started to investigate gesture-based interaction using depth-sensing cameras, such as Microsoft's Kinect sensor (Clark & Piumsomboon 2011), where users do not need to touch the screen to interact with the AR content, but interacts through hand or finger gestures. Although still futuristic in mobile contexts, depth-sensing cameras on mobile phones are realistic in the future, and could become an interesting tool to overcome occlusion issues in mobile AR interfaces.

6.2.3 Multi-modality and limited I/O technologies

While touch-enabled smartphones seem to have conquered some of the traditional challenges in mobile interaction design, some challenges still remain. As mentioned above, text-based input remains a big challenge for researchers, with tactile feedback and word-gesture keyboards being prominent attempts to overcome these issues. While tactile feedback is already widely used, word-gesture keyboards have yet to reach mainstream adoption. Also the most common input techniques do not always complement the diverse context of mobile interaction. Some researchers in mobile interaction design have investigated multimodal interfaces, where the input and output modalities change depending on different contexts of usage, e.g. walking or driving

(Lemmelä et al. 2008). Audio feedback is especially useful in contexts where users have their visual attention elsewhere, for instance when driving and using a GPS. Also, audio output can be useful to overcome limited screen space, for instance when using speech as a replacement for large amounts of text-based information.

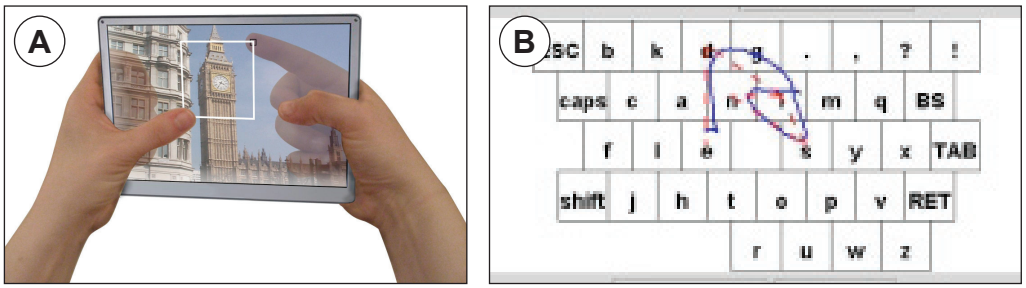


Figure 22: Conquering mobile interaction design challenges. **A**: LucidTouch provides an additional input surface behind the mobile display to conquer occlusion issues (Wigdor et al. 2007). **B**: The SHARK word-gesture keyboard relies on drawing marks and gestures rather than typing text (Zhai & Kristensson 2003).

6.3 Chapter summary

In this chapter, some of the major issues of mobile interaction design have been presented. While the invention of modern smartphones has helped conquer many of these challenges, some still remain. Because mobile devices need to be portable, screen size continues to be limited, and occlusion and fat-finger problems remain. Some of the clever approaches to help overcome these issues can also be relevant in AR settings, for instance by using depth-sensing cameras to interpret user gestures or use the internal sensors of a mobile device to navigate between different visualisations of geo-spatial content.This chapter ends Part 1 of the thesis. In the next chapter, the problem formulation of the thesis is put forward.

7. Problem formulation

So far, this thesis has presented two different cities, where events have created a potential for exciting urban stories: the New Zealand city of Christchurch, where earthquakes have left many historic buildings demolished, and the new harbourfront in Aalborg, Denmark, which is undergoing dramatic changes in the transition of the city from an industry to a knowledge-oriented city. Aspects of AR and spatialised audio have been described, in particular how they link to mobile computing in urban environments, a well as their potential in location-aware systems. Social media and networking services have been described as well, since these are the focal points of the long-sighted goal of the thesis, and important to keep in mind for this report. Finally, challenges in mobile interaction design have been described, including the use of audio and multimodal interaction to overcome some of the limitations of mobile interaction. In Part 2 of the thesis, a mobile audio experience on urban story telling will be developed, which integrates AR and which eventually can be used for integrating social networking concepts. This leads to the final problem formulation.

Is it possible to develop an audio-driven AR system for presenting urban stories, which could eventually work as a foundation for integrating mobile social networking services?

7.1 Research goals

To evaluate the impact of the developed product, the following questions are of main interest and will be treated throughout the rest of this report:

- Which interaction method is suitable for a mobile, audio-based AR experience?
- How can a good audio experience be ensured in a busy, urban environment?
- What impact do contextual factors of the urban setting have on the usability of the application?
- What technical issues do end users face, when they use the application in a real urban setting?
- How is the overall usability of the developed product rated by users?
- How can social media and networking expand the user experience of the developed product?

Those questions are relatively explorative in nature, and to evaluate them, a number of experiments will be carried out towards the end of the report. Of particular interest is an empirical field trial of the final application in use, since involvement from real users has become increasingly important for AR interfaces, in order to bring the technology out of the research labs and into people's everyday life (Dünser & Billinghurst 2011). Further, comparative usability evaluations of some of the basic functionalities, such as interaction with the digital audio content and quick-and-dirty design tests, will also be conducted to ensure a broad evaluation of the application.

7.2 Limitations

Since this thesis is part of a long-sighted goal of integrating mobile AR and social media, it is necessary to delimit the project to a more specific context. Part 2 of this thesis treats the development of a mobile audio experience on urban story telling, called *memoAR*. This work includes conceptual ideas as well as the design, implementation and evaluation of a usable product prototype. Pointing towards future work, Chapter 16 presents considerations as well as low fidelity conceptual sketches of an eventual integration of social networking - which build upon the already developed mobile audio experience. Figure 23 depicts a diagram of this thesis' scope in the context of the long-sighted goal.

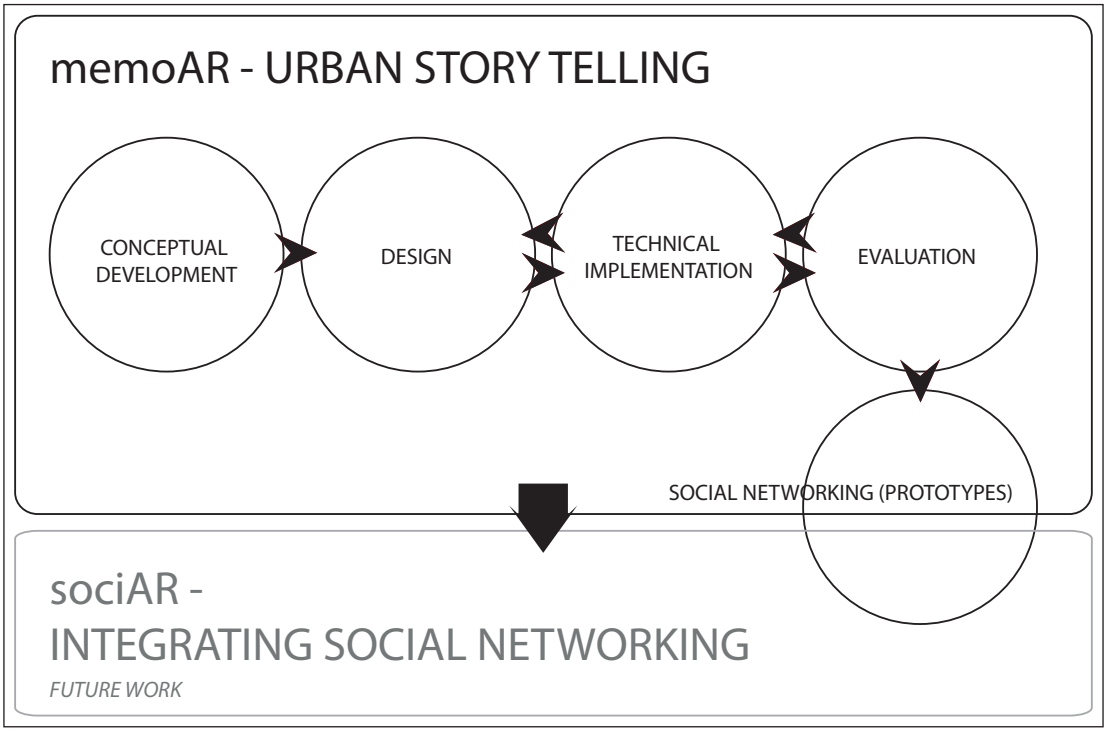


Figure 23: Overview of the thesis scope and future work.

7.3 Methodology

The mobile audio experience on urban story telling will be a collaboration between the author, the Human Interface Technology Laboratory (HITLabNZ) at the University of Canterbury and New Zealand Historic Places Trust (NZHPT). Their wish is to develop a full-working audio tour application with interviews and stories for inhabitants to enjoy while navigating in the city of Christchurch, New Zealand. This application, called "High Street Stories", will serve as the foundation for the memoAR project, and will be iteratively designed and developed in close collaboration with the stakeholders (NZPHT) to keep their wishes and requirements in focus throughout the project period. The High Street Stories project can be seen as a specific implementation of the memoAR application, which should be adjustable to the contexts of other cities.

PART 2: PRODUCT DEVELOPMENT

In Part 2, the product of the thesis is developed. Chapter 8 presents conceptual thoughts about the product to be developed, including a target group analysis and the development of a conceptual model. Chapters 9 and 10 describe the design and implementation of the product.

8. Conceptual development

This chapter presents conceptual thoughts for the memoAR audio application and an eventual integration of social networking ideas. First, the overall problem space is defined after which target group considerations, user requirements and a conceptual model are presented. Finally, two personas are developed, demonstrating two realistic use case scenarios for the application.

8.1 Overview of the problem space

The task at hand is to develop a mobile application for urban story telling, which lets users listen to interviews, geo-tagged to buildings in an urban setting, and eventually author and share their own content grounded in these audio files. The foundation of the system is an audio-based tour application developed in collaboration with the organisation New Zealand Historic Places Trust (NZHPT), whose wish is to present geo-tagged interviews with people telling about the historic buildings in the High Street precinct in Christchurch, New Zealand. Users will browse and explore the stories from their mobile phone while navigating in the environment, through a visual interface, which helps the users orient and navigate. Additional information about a story related to a building or an area will be accessible from the mobile interface by providing options to select it for further exploration. The urban setting of Christchurch is the focus for this thesis, but the application should be easy to adapt to other cities' contexts, where events make urban stories interesting. Figure 24 depicts a conceptual overview of the system to be developed.

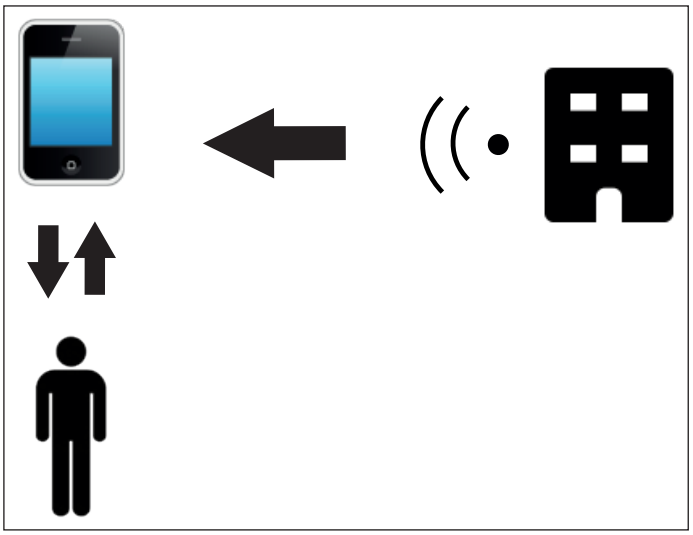


Figure 24: Conceptual overview of the tour application. Through a mobile device, the user can interact with stories tagged to buildings and regions in a city.

An obvious assumption is that users will expect a simple and easy way to interact with the audio content, and a minimum amount of effort should be required to learn the system. Further, since the system revolves around interacting with objects tagged to the physical environment, it should be easy for users

to navigate and orient themselves in their surroundings. To this extent, it will be necessary to provide different visualisations of the stories, to allow users to immerse in specific scenes and get an overview of the larger geographical space in which the stories exist. Finally, issues such as user vs. system control, interacting with the audio files, audio clutter, visual cues and information filtering are important issues to address, and mobile interaction design principles should be obeyed. By providing interesting stories tagged to locations in a city, it is suggested that these stories will provide interesting contexts for mobile users to create and share urban stories themselves. The following claims are made for the project:

- Augmented reality is a natural interface candidate to experience geo-tagged city stories on mobile devices. The camera view of a smartphone provides a "window to the world" and an egocentric perspective, from which geo-tagged stories can easily be browsed.
- Spatial audio can provide an additional immersive element to the audio tour and thus improve the user experience.
- The combination of egocentric and exocentric visualisations will provide the best user experience, because a user will both need access to detailed information about a scene and an overview of the environment, in which the stories exist.

8.2 Customer requirements

NZHPT are preserving New Zealand's cultural and architectural heritage. Their wish is a mobile tour application, which lets inhabitants of Christchurch remember and explore their city's cultural, architectural and social history after many buildings have been or are facing demolition following the earthquakes in 2010 and 2011. The tour application is intended as a supplement to the High Street Precinct website[23], which contains images of the historic buildings and a sound map related to the area. This project is funded by the Vodafone Foundation through the Christchurch Earthquake Appeal Fund and will serve as the foundation for the memoAR mobile application. The application will be downloadable from this website and contain interviews, which are edited by NZHPT and organised in different themes. In addition to the sound of the interview, each story will be associated with an image of the building and text describing the contents of the interview.

8.3 Target group

NZHPT sees the audio tour in Christchurch as an experience to be enjoyed by everybody interested in the city's history. They expect families and elderly as much part of the target group as young, tech-savvy mobile users. This makes it important to provide an experience, which is immediately enjoyable and easy to learn for all users. Although approximately 50% of smartphone users are between 25 and 34 years old (AnsonAlex 2012), elderly people are rapidly adopting smartphones as well. Applications targeting elderly are often in the areas of telecommunication and health (Socyberty 2011), but a system offering information on a city's cultural heritage and social history is likely to attract elder users as well. Therefore, the interface should aim at being easy to use for an elder audience. Based on studies on universal interaction design, some of the issues

23 http://www.highstreetstories.co.nz

elderly mobile users face are related to font size, small key/button space and the fragility of mobile devices (Plos & Buisine 2006). This implies that text-based information should be limited, and interactive UI widgets should be of a reasonable size. Further, the system should not require much effort to learn and master.

8.4 Conceptual model

Since the main challenge of the system will be to provide a good audio experience, the user's interaction with the audio content will be the most important feature. The system will primarily consist of information geotagged to physical locations, and to this extent the metaphor of pointing towards an object using AR can help the user orient, and reveal whether information is attached to a specific location. As an interaction method for basic navigation in the environment, this is expected to be relatively easy to learn because of the camera analogy, but could be more challenging for more complex tasks, such as selecting or manipulating a digital story. Since the user is expected to spend some time navigating and orienting to locate stories, there should be a tight interplay between the navigational and the explorational tasks.

In order to "access" the audio experience, several options exist. In theory the user should not need to look at the display or pay attention to his/her mobile device to explore audio content. Sounds can be spatially registered to the geographical locations of buildings, and played back in 3D in headphones. Each sound from a story could play at all time, but with the volume and stereo channel positions dependent on the user's location and distance relative to the story. This way, a user can be guided towards a building of interest and enjoy the audio experience related to the building by following the spatial location of the sound, without looking at the mobile display. However, since each interview will be associated with visual content, such as text descriptions and images, a visual interface needs to be implemented as well. Regardless of the use of sound techniques, the system should carefully integrate the audio and visual interface and be aware of what information to present at any given time.

8.5 Personas and use case scenarios

Below, two fictive users and use case scenarios for the audio tour and a possible integration of social media respectively are described. Storyboards for the scenarios can be found in Appendix B and C. Personas can be useful for maintaining a focus on the end users when developing a product (Nielsen 2013), which is expected to be a good point of departure for the actual design decisions.

8.5.1 Persona 1: Amy, 45, Christchurch, NZ

Amy is a 45-year-old woman, who lives and works in the healthcare in central Christchurch. Her world was ruined when the earthquakes in 2010 and 2011 destroyed her beloved city, in which she has lived most of her life. After a period of personal turbulence, struggling to recover from her psychological scars, she is now looking towards the future. She finds hope in the community gathered around reshaping the city's identity, and is excited about what the city will look like in the future. At the same time, she is afraid to forget the memories of what her beautiful city used to be like. So many memories, so many stories to share, making her emotional and sad when walking up and down the city streets.

On most Sundays, Amy goes to the High Street precinct in central Christchurch to take a walk in the ruined central business district. She usually went here often to enjoy this lively part of town, but now all shops and cafés are closed, and most buildings are facing demolition. Because of her interest in this area, she often goes to the High Street Precinct Facebook page[24], which contains stories and photos from the city. Further, the Facebook page introduced her to a website, from which the High Street Stories application could be downloaded. This application, she is told, contains interviews with persons telling about the historic buildings in the High Street area. The idea of using her telephone to listen to stories and help her remembering is appealing to her, so she downloads the app, but is unsure about what to expect from it. At the same time, she is unsure about how to use the application because of her lack of knowledge about technology. However, having recently purchased an Android smartphone, she has become more familiar to interacting with mobile applications, and hopes she can benefit from this knowledge when using the application. *A storyboard of a specific usage scenario for Amy can be found in Appendix B.*

8.5.2 Persona 2: Christian, 29, Aalborg, DK

Christian lives in Aalborg, Denmark, where he works as a lawyer. His work occupies most of his time, and he usually spends most waking hours at the office. Social media is important to him, and he likes to articulate his personal relationships and share his interests on social networks like Facebook and Flickr. Often he will take photographs of interesting things he sees on his way home from work, and upload these photos to his Facebook and Flickr accounts to share with his online friends. This way of providing small anecdotes of his observations is a pleasant break from his busy schedule. He enjoys following the updates of his friends as well, which is a nice way to keep track of what is going on in their lives.

On his way to work, Christian passes by the harbour front, which is currently undergoing dramatic changes because of Aalborg's transition from an industrial to a research-oriented city. Recently Christian heard of the Aalborg Harbour Front application, which lets people hear stories about the buildings of the historic harbour front, as well as comment and discuss the on-going process of restructuring the cityscape. Due to his natural interest in new technologies and social media he downloads the application. As he is critical towards the limitations of mobile systems and often experiences frustrations in existing social media apps, his main concern is that the application will not provide him the social experience that he knows from Facebook and Flickr. *A storyboard of a usage scenario for Christian is found in Appendix C.*

8.6 Chapter summary

This chapter has presented conceptual thoughts regarding the development of the mobile audio experience, which is based on customer requirements from the NZHPT for the High Street Stories audio experience. Target group considerations have been discussed, primarily in relation to ensuring a good user experience for elderly and non-tech savvy users. Two personas have been developed, which will help ensuring a user-centered focus for future design decisions, which are presented in the next chapter.

24 https://www.facebook.com/Highstreetprecinctproject?fref=ts

9. Design

This chapter describes the design process of the application, focusing on the decisions for the High Street Stories implementation in Christchurch, but also for the memoAR application in general. The application will contain audio sources geo-tagged to historic buildings in the area, explored through a mobile interface. Considerations regarding the audio playback and interaction for mobile tours in general are firstly discussed, as well as issues of agency, information filtering and visual guidance in a system, which uses audio as its primary medium. Then, the actual design decisions of the application are presented, iteratively developed in collaboration with the customer (NZHPT) and the HITLabNZ.

9.1 Audio interface considerations

Under the assumption that an AR view is a good interface to provide an egocentric perspective of the stories in the user's surrounding environment, choosing the right interaction technique for accessing the audio content is vital for a good user experience, partly since elderly users are part of the target group. As stated in chapter 4, audio interfaces can be challenging, in particular when it is necessary to support multiple, simultaneous sound streams without overloading the user. In the following, different approaches to interaction methods and presentation of the audio content are discussed, focusing on how to combine the audio with a visual AR interface. This will help laying out the different approaches to the interaction method for consideration before the final design decisions.

9.1.1 Mobile phone as pointing device

This idea is based on the metaphor of the smartphone as a pointing device. Through the camera view of a smartphone, the user sees the real world with digital icons representing the audio sources superimposed over the physical locations. The audio sources are activated by pointing the camera view towards a location associated with an icon. The audio stops when the user no longer aims at the building. Visual cues in the shape of colour codes help the user notice which audio source is currently in focus and played in his/her earphones (Figure 25).

Figure 25: Mobile phone as pointing device. **A**: A red icon indicates that a building nearby has an audio source attached to it. **B**: Pointing the device directly at the icon plays the audio source.

A problem with this approach is the ergonomic challenge of having to stretch out the arm for a longer period of time to listen to a interview. One aid to overcome this issue is to implement a seperate command for stopping the audio file again, allowing users to continue listening to an interview even after pointing away from the building. Another possibility is to have short audio snippets playing when the mobile device is pointed towards a building, suggesting users to perform an explicit command to listen to the full interview, such as touching a button for activation. The user can then sit down and enjoy the audio file without worrying about the pointing aspect.

Another issue is the possibility of activating an audio file by mistake. If audio files occlude each other on the screen, how will the user be able to select the audio file of interest and not the wrong one? One possible approach is to only let the user activate the nearest scene on a straight path away from him/her. For instance, if the left side of a road contains five buildings with playable audio files, these five icons are potentially all near the center of the display when the device is pointed in the direction of the buildings. However, only the nearest of those audio files would be playable from the users location, allowing the user to immerse only in the very near environment. If the user needs access to occluded stories farther away, those would need to be accessed from an exocentric view, such as a map view of the environment.

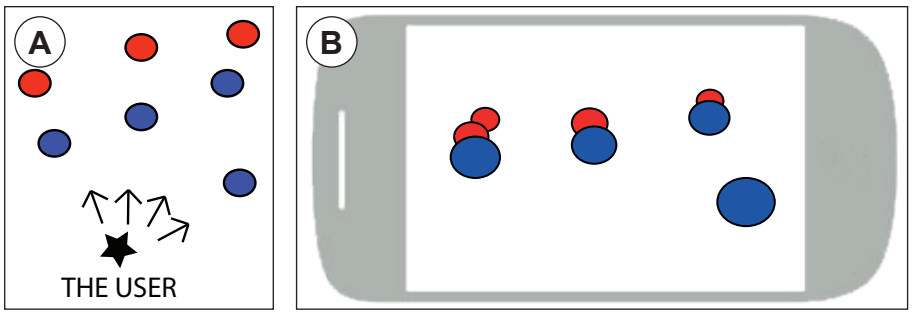


Figure 26: Occlusion issues. **A**: Blue dots are audio sources that are playable from the user's position. Red dots indicate audio files that cannot currently be activated. **B**: Scene from the camera view.

9.1.2 3D audio interface

Another approach to the interaction method will require no explicit action from the user to play back the audio sources. Instead the user's location is used to adjust the playback of the audio sources surrounding the user. Sounds are spatially registered to physical locations, and a sound's position in the stereo channel of a user's earphones is based on the mobile device's position and orientation relative to the audio sources. The volume of the audio sources fades up and down as the user walks towards or away from a scene. This provides an illusion of 3D sound, similar to the approach described by Pulkki & Karjalainen (2001), and as seen partly in the Soho Stories application (see Chapter 2). To be convincing, this approach could track the user's head orientation for instance through an HMD setup to register the sounds' positions based on the user's head orientation instead of the orientation of the phone.

A challenge of this approach is to avoid audio clutter. In audio environments, it is important to manage multiple audio streams without overloading the user (Vazquez-Alvarez 2010).

One option is to let the radius of virtual, circular activation zones around the audio source decrease in densely populated areas to ensure that a maximum of two sound files are played at full volume at the same time. An advantage is that the user can carry the device in the pocket and still get an audio experience, maybe even more immersive than standard playback in earphones.

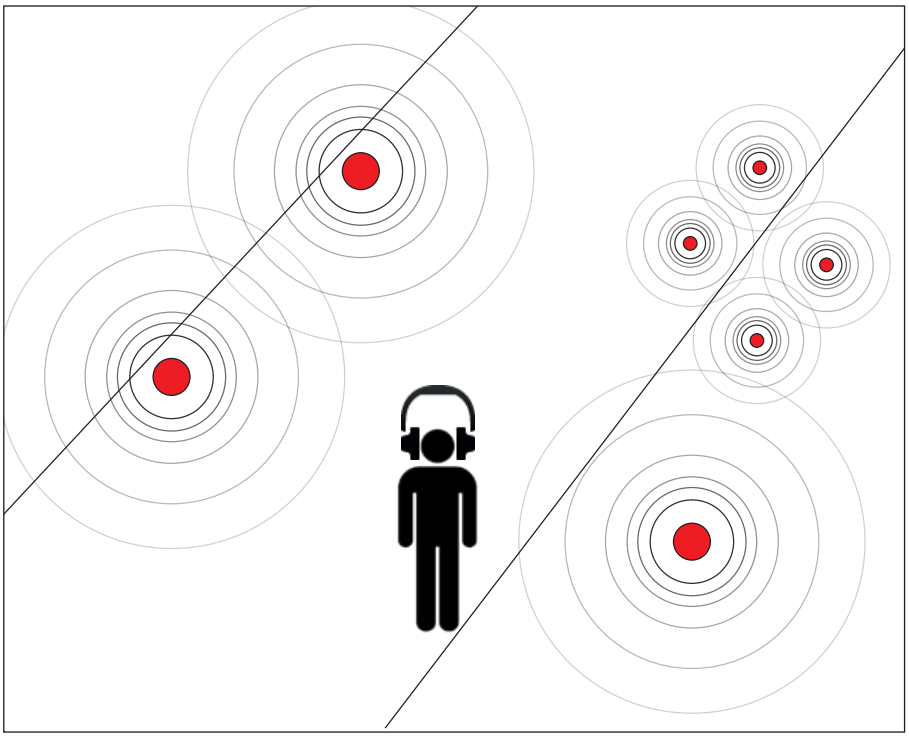


Figure 27: 3D sound and no visual interface. If the user enters a sound's "activation area", the sound is played back. Sounds fade in and out as the user navigates. If multiple audio sources are near each other, the user will have to walk closer to one of the sources to hear the audio, as illustrated by the four sources clustered towards the right in the figure.

9.1.3 Multimodal interface

Combining input modalities potentially opens up for richer user experiences in AR systems, for instance through speech in combination with tangible or gestural commands (Irawati et al. 2006). Multimodal interaction could be integrated in the memoAR application, for instance by using touch commands for selecting an audio file, and speech commands for controlling the audio, such as "Play", "Stop", "Pause" and "Volume up". Further, it could be interesting to investigate speech to filter the stories on the mobile display and playing back audio sources that are not currently visible in the camera's field-of-view.

9.1.4 Nearest audio file plays

Another idea is to always play the audio source closest to the user. For all sources, the distance to the mobile device is calculated, and the one closest to the user is the one played. An audio source fades in and out as the user walks towards or away from it. This approach takes away all control from the user, who needs to stop at a location to keep an audio file in focus. This also means that the user does not need to bother about the interaction, but is provided a fixed audio experience.

9.1.5 Single-touch interaction

This approach is based on standard touch interaction, where audio sources are visualised as touchable icons on the screen, and the user taps the icon to play the file. When tapping, the user activates the audio and receives detailed information about the scene and the content related to it. The advantage of this approach is the use of a well established and for most users familiar interaction style, which does not require much learning. It might however not be as immersive an audio experience as the proximity-based interaction methods, where users interact with the audio sources by navigating in the environment, and where the incitement for using spatial audio techniques is greater.

9.2 Designing the audio tour

Having presented different approaches to interacting with the audio sources, this section walks through the actual design decisions in close collaboration with the customers (NZHPT). First, the content of the audio sources is presented, after which the decisions regarding interface design, layout, screens and navigation of the audio tour are discussed.

9.2.1 Audio content: Interviews tagged to historic buildings

The backbone of the audio tour are approximately 90 stories with 40 subjects, who have a historic connection to the buildings in Christchurch's High Street precinct, which are now demolished or facing demolition. The stories are divided into eight different themes, covering both interviews and music sessions: *High Street Regeneration, Red Light District, Architectural Heritage, Music, Art Scene, Manufacturing, High Street Yarns, Ngai Tahu History*. The stories will be browsed by inhabitants and tourists through smartphones with standard earphones, since these are most convenient for wearable audio applications (Härmä et al. 2004). The style of the interviews are raw recordings with ambient background noise, not unlike the style in Soho Stories. Some important usability challenges are immediately identified:

- With 90 stories in 8 different categories, the interface should allow the user to easily find particular stories he/she wants to hear.
- A user interested in stories from a particular theme should be able to filter the information shown on the screen, and browse content from this theme only.
- If a user listens to a story of particular interest, he/she should be guided to stories, which might be of similar interest.

The most important concern is to avoid overloading the user with visual and audio information, which has an impact on the design of the visual cues, and the way the audio is played back. In line with traditional mobile interaction design guidelines, users should receive proper feedback on their actions, and always know what is going on during their experience with the tour.

9.2.2 Browsing the stories

The essential part of the application is the user's interface to hearing and exploring the stories tagged to the environment. An AR view can visualise the stories through superimposed icons on the camera view. This way, the user gets an egocentric view of the stories within the tour, with icons of varying size, indicating their distance from the user (Figure 28A). Adjusting the size of the icons to indicate that some stories are farther away from the user is a way to minimise information clutter on the mobile display, and provide a visual cue to the exact location of the stories.

With 90 stories covering a relatively large area, the user should also have access to an overview of the whole area the tour covers, and to be able to browse stories, which are not necessarily within the camera's field of view. Existing mobile AR systems, such as Wikitude, Layar and Nokia City Lens, use map and/or list views to provide overviews of the scenes. These are useful, because most users are familiar with those visualisations. Since this system is to be enjoyed by non-tech-savvy users, the exocentric views of the scenes will be visualised through a map view and a list view, similar to some of the already popular AR applications. In the AR view, the icons representing the stories will scale in size to indicate the varying distances to the user, and provide visual cues to the category of a story. The list view will provide the user additional navigation aids through arrows pointing in the direction of the scenes, as well as text-based distance indicators (Figure 28C). The map view shows a top-down 2D map with icon overlays on the locations of the stories as well as the user's current position (Figure 28B).

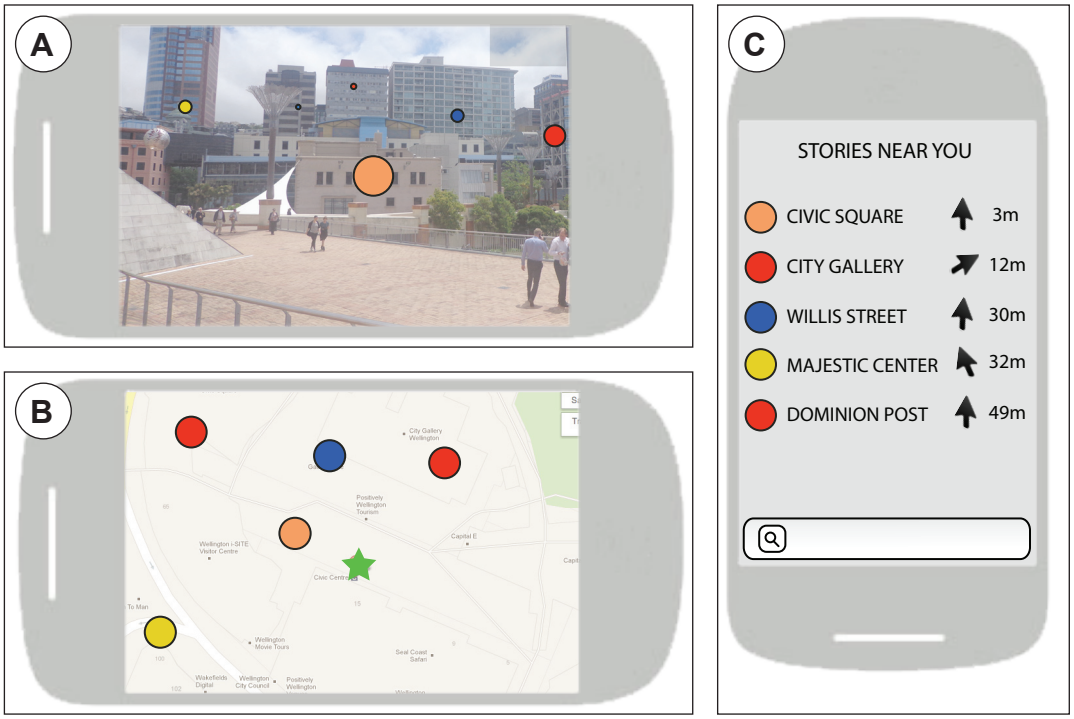


Figure 28: Visualising the geo-tagged city stories. **A**: AR view with coloured icons superimposed on camera view of a smartphone over locations in the physical environment. **B**: Ortographic map view providing exocentric overview of the scene. **C**: List view with distance and direction indicators.

Since the stories will have associated additional content, a new screen should appear when a story is selected and played back. This screen will show additional information about a scene, which consists of:

- Thumbnail image
- Title of story and name of interview subject
- Duration of interview
- Short textual description of interview content

The user will also be able to control the audio through a media player interface for stopping, pausing and rewinding the audio clip. Figure 29 depicts a mockup of this screen.

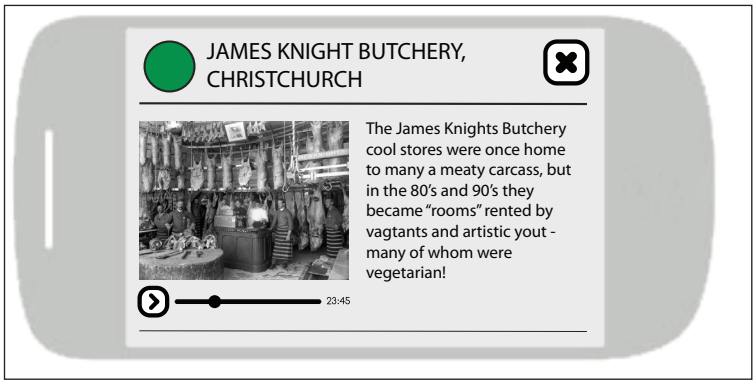


Figure 29: When a story is selected, additional information about a story is shown, while its story is played back. A media player interface lets users control the audio.

9.2.3 Selecting and playing back stories

When a user is in the AR, map or list view, he/she has access to the stories of the application. As described in section 9.1, several options exist in regards to interacting with the audio content. Assuming a touch-enabled device, tapping an icon on the screen would be natural for activating the audio. However, this approach can be troublesome in an AR context, since a handheld mobile device combines input and display technology (Bai et al. 2012). Touching a floating AR icon on the screen can cause camera motion, which affects the user experience negatively. An alternative approach is to use a point-to-select analogy, since users are already pointing towards the buildings when they orient themselves in the AR view. When a scene is in the center of the mobile display, this automatically plays back a localised snippet of the story, providing a hint as to what to expect from the full story. While localising the sound does provide a hint to the distance to the sound sources, pointing at an icon to play it back is not the most immersive way to use spatial sound techniques, since the up-down and left-right position will always be in the center of the mobile screen when the sound is activated. A more obvious scenario for using spatial sound would be when having multiple sounds played back simultaneously from directions all around the user, with the audio position providing hints to the location of their sources in the environment. This would however make the audio experience more complicated than is intended for the High Street Stories

project. The effect of spatialising the sounds in the AR view is a feature, which should be tested further on.

A user might not be interested in listening to the full interview, so when a scene is pointed at and its snippet is played back, a seperate layout appears on the screen, on which the user can tap a "Play" button to hear the full interview and explore the story's additional content. By using this layout, the issues of touching the floating AR icons on the screen are reduced. In the map and list views, standard touch input is more beneficial, since those views do not contain a camera view, and all icons are fixed on the screen. The snippets will however also be used in the map view, to let the user hear a short audio file before deciding to listen to the full story. Figure 30 depicts mockups of the AR and map views with seperate layouts appearing when a scene is selected and its audio snippet is played back.

Figure 30: When a scene is selected, a new layout appears on the screen, while a spatial audio snippet is played back. From here, the user decides whether he/she wants to hear the full interview and explore additional content related to the story.

9.2.4 Navigating between ego- and exocentric perspectives

Since the audio tour is intended to be experienced in situ, it should be fast and easy to navigate between the ego- and exocentric views. Further, if a user listens to a story of a particularly interesting topic, he/she should immediately be able to access stories of the same topic. Because of the limited screen size of mobile devices, it is important to avoid overwhelming the user with visual information on the screen. One opportunity to overcome the lack of screen space is to imitate the function of Nokia City Lens, where the orientation of the device is used to switch between the AR, map and list views. Nokia uses analogies from the real world in regards to how people normally interact with maps and lists. This is expected to be useful in urban environments, where users pay limited attention to the mobile interaction. At the same time, this approach leaves more screen space for important UI elements, such as buttons that link to important screens. This method to navigate between the visualisations bears some resemblance to the work by Mulloni et al. (2010), who use the tilt and rotation of the device to present sophisticated "zooming interfaces" to smoothly navigate from the egocentric AR view to exocentric panoramic and top-down exocentric views.

Another important design decision regards the list view. When a user chooses "Start Audio Tour" from the main screen, he/she will start the tour in a list view, which shows all scenes nearby. From here the user can choose to browse the stories around him/her, or tap a "Themes" button, which displays another list view containing the different themes. Tapping one of the themes shows a list of stories of this particular theme. The user can flip the device to go to the AR or Map view (Fig-

ure 31). From each of the views, an individual scene can be selected to explore the content of the story in details. Appendix D shows a diagram of the navigation between the application's main screens.

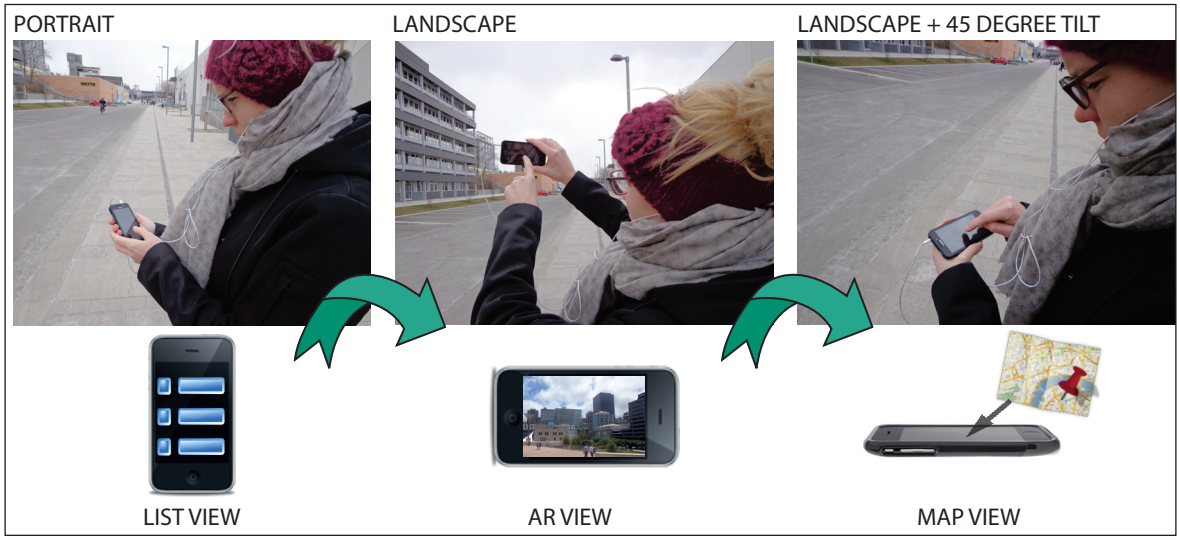


Figure 31: Views switch when device is rotated. If the device is in portrait mode, a List view of the scenes is shown. If the device is in landscape mode, either the AR or the Map view is shown.

9.2.5 Other screens

Apart from the vital screens for visualising and browsing the stories, the application will further contain a startup screen, where the user can choose to start the audio tour, get information about the application or a link to a website (in this case the High Street Stories project web site). The information screen informs the user about how to use the app, a signature explanation, credits, project partners etc. For the High Street Stories application, this screen shows all sponsors of the project, the HITLabNZ development team etc., but the content will differ for other application contexts.

9.2.6 UI elements and user feedback

Since no buttons are needed for navigating between the AR, map and list views, this leaves some space for few, important user interface elements to occupy screen space. The information screen is the user's "quick guide" for valuable information, and should be accessible from everywhere in the app. This makes it reasonable to include a button for accessing this screen in the AR, map and list views. Further, the app should be integrated with the High Street Stories web site, making it important to provide the user easy access to the web site from the app. Therefore, a hud with two buttons for these features is included in each of the three main views (Figure 32).

In the AR and map views, a "Filter" button in the top of the screen provides another mechanism for choosing to display only a subset of all stories. The button takes the user to a list of the themes, which can be checked on and off, to only see stories of particular categories. Further, the AR view will contain a radar to help the user orient and see surrounding scenes. The UI buttons occupy 80x48 pixels on the screen, and are thus within Google's standard recommendations for touchable UI components (Google 2013b).

Another imporant interface element is a reticle in the center of the screen, which helps the user aim towards the scenes in the environment. The reticle is made

in two colours: black when no scene is in focus, and green when a scene is in focus.

Figure 32: UI widgets for filtering scenes in AR and Map view, and accessing the "Info" screen and the application's web site. A reticle in the center of the screen helps users aim at the scenes.

9.2.7 Icons and buttons

The graphics for the High Street Stories project are designed in collaboration with New Zealand design company NV Interactive. They are responsible for implementing the project's website, which contains a soundmap of the High Street precinct as well as a link to download the High Street Stories application. The application should have a design coherent with this website, which is still a work in progress. So far the design decisions have implied that the icons representing the different themes will be distinguishable by their colours, and stylistically kept fairly simple. The same goes for the UI buttons, which should be simple and easy to interpret. The buttons will be designed to have pressed and an unpressed states to provide proper visual feedback for users, and will follow Google's recommendations for the size of touchable UI objects. In the following chapter, which describes the implementation of the application, stylistically simple buttons and widgets are implemented as placeholder icons to demonstrate the design ideas, but those will eventually be replaced by graphics resembling the High Street Stories website.

9.3 Chapter summary

This chapter has presented a walkthrough of the major design decisions for the mobile audio experience for the memoAR audio experience, in the particular context of the High Street Stories project. First, important considerations regarding the audio interface has been described, which are the most important feature of the application. Then, a walkthrough of the interface design, layouts and interaction method has been presented. The design is still work in progess, due to a collaboration with the New Zealand company NV Interactive, who are currently working on the High Street Stories website, which the application's design will rely on. The next chapter walks through the final implementation of the mobile application, which is developed on the Android platform.

10. Android implementation

Having presented and discussed the design decisions of the memoAR application, this chapter presents the implementation of the audio tour, which will be carried out on the Android platform.

10.1 System components

The application is developed on the Android platform and will be built for smartphone devices. Tablets would be an option for the audio tour as well, but the interaction methods and usage scenarios make smartphones more handy. In the future, it will be natural to expand the application to support iPhones as well, but the AR framework used for the application (described shortly) does not currently support the iOS platform.

The application will be developed using an OutdoorAR framework[25] provided by the HITLabNZ. This framework is targeted at mobile AR applications, and provides the fundamental tools to implement AR experiences on the Android platform. The framework allows AR developers to focus on the system interface and user experience, which is appealing for the focus of this thesis. For the audio, the FMOD Ex audio engine[26] provides opportunities for generating spatialised sound sources, and is already integrated in the outdoor AR framework. For the map view, the Google Maps API is used, since this is easy to implement on the Android platform and is integrated in the OutdoorAR framework as well. For creating the visual graphics, the open source API OpenGL[27] is also integrated in the AR framework. Figure 33 depicts a diagram of the system components and their interplay. In short the application is built as a native Android application, which integrates additional libraries for AR, map, graphics and audio.

10.2 Android framework

The Android framework is an open source platform, which makes it possible for developers to implement mobile applications on the Android platform. Developers can access device hardware and the same framework API's as used by core Android applications (Google 2013a). The framework consists of a set of services and systems, which includes the following:

- Template *layouts* for ordering UI elements on the screen.
- A *Resource Manager*, which provides access to resources such as images, strings and layout files.
- An *Activity Manager*, which handles the lifecycle of applications and activities.

The lifecycle of Android activities can be viewed as a stack, where the running activity calls a new activity, is suppressed and lets the new activity run on top. When the activity on top is destroyed, the activity below is resumed (Google 2013c). A full overview of the Android activity states can be found in Appendix E.

25 http://hitlabnz.org/index.php/products/mobile-ar-framework
26 http://www.fmod.org/
27 http://www.opengl.org/

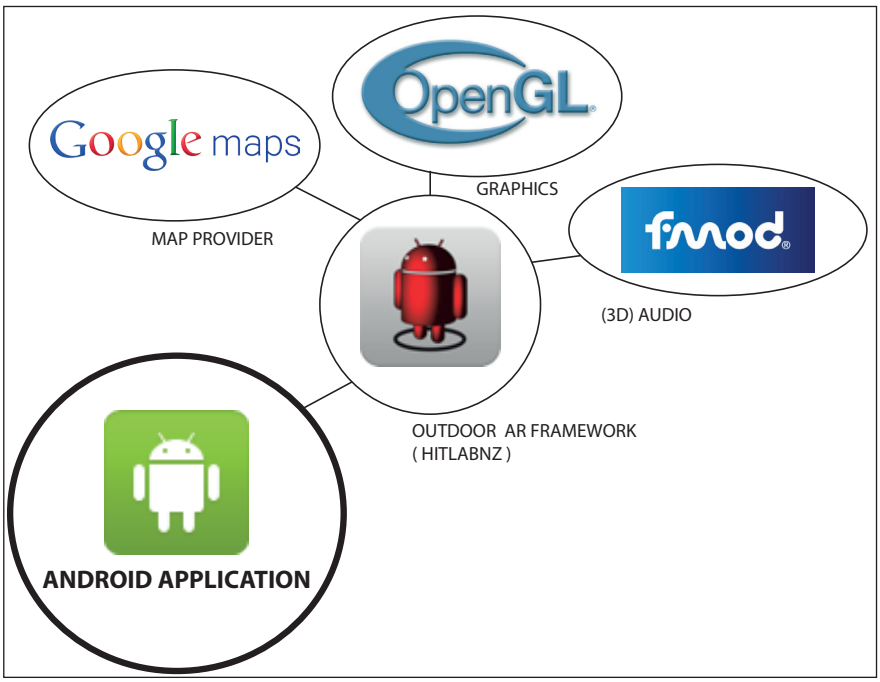


Figure 33: The foundation of the system is a native Android application, which links to the OutdoorAR framework provided by HITLabNZ. This framework integrates important libraries and API's such as Google Maps, OpenGL and FMOD.

10.3 Data structure & system flow

The individual stories are set up as instances of the `OAScene` class from the OutdoorAR framework. This class stores information about a scene, such as its name, location and category. It is also possible to add additional metadata specific for this particular application, such as the name of the interview subject, the audio files (the full file and the short snippet as .wav-files) and thumbnail images to visualise the buildings and areas related to the stories. The scenes are added to a temporary list in a custom data manager. This data manager extends the `OADataManager` class, which is useful for storing and managing the data of each scene. This class stores the scenes and allows them to be accessed from the AR, map and list views for visualisation. Further it links to the Assets folder of the application, which stores icon sources and audio files related to each scene. In our custom data manager, we further add mechanisms for information filtering by storing a temporary array of all the scenes, which are not released for visualisation before explicit permission is provided through the user's commands on the UI.

When a user decides to filter the scenes through the "Filter" button in the AR or map view, this is done in the `FilteringActivity`. Here the user checks or unchecks a checkbox for each category, dependent on whether he/she wants stories from this category visualised or not. The filtered categories are saved as a string to the `AppPreferences` class (e.g. "Historic, Food, Business"), which stores universal data that should be accessible from anywhere in the app. Whenever the AR, map or list view is started, this string is checked to ensure that only stories from the categories the user has selected for browsing, are displayed. Another way of filtering the stories displayed on the screen is achieved by selecting scenes of a particular category or scenes nearby in the `CategoryListActivi-`

ty. Figure 34 illustrates the procedure of deciding which scenes to display for the user in the AR, map and list view. If the list view is entered, first it is checked whether the user entered the activity via the "Nearby" button in the CategoryListActivity. If this is the case, scenes should be sorted by their distance to the user and only the X nearest stories should be displayed. Otherwise, the AppPreferences class is checked to see, which categories have been checked by the user. In the AR and map views, there is no "Nearby" function, so here we only need to check AppPreferences everytime the view is entered.

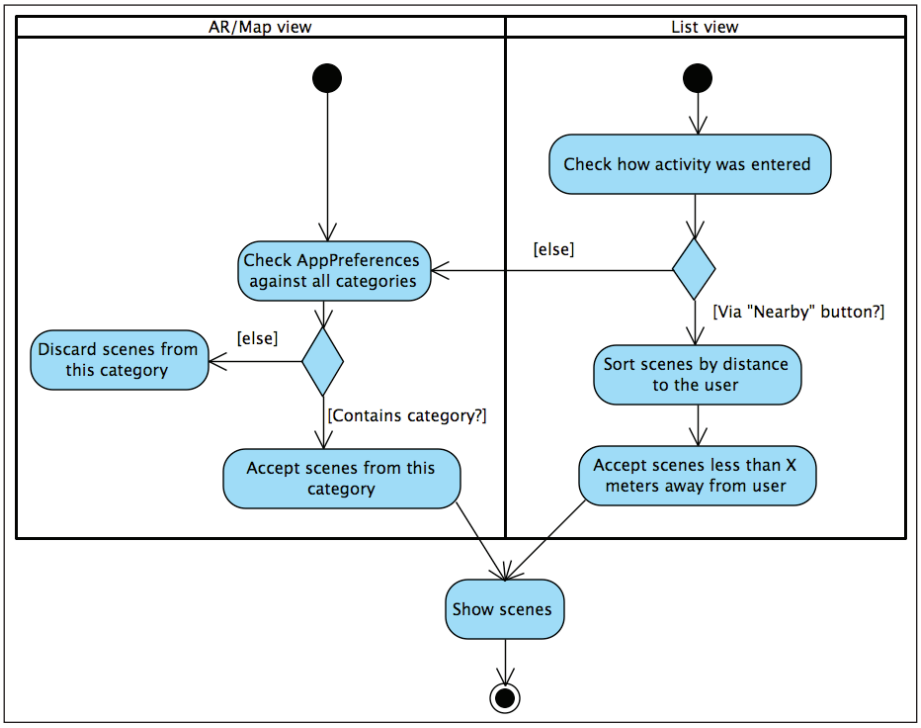


Figure 34: Activity diagram showing the procedure of deciding which scenes to display in the AR, map and list views.

10.4 Views

10.4.1 ARViewActivity

The `ARViewActivity` extends the OutdoorAR framework's `OAARComponentBase` class, which contains the important funtionalities for AR related activities. First of all, it sets up an AR view with a camera background and a scene view augmented on top. The scene view extends Android's `GLSurfaceView` class, which provides access to displaying OpenGL rendering. The scene view uses a scene graph structure to store the visual representation of the AR scenes, which are represented graphically and rendered through OpenGL commands on the scene graph. The graphics are of the `BillboardModelData` class from the Outdoor AR framework. This class provides a 2D graphic, which takes an Android drawable resource and its dimensions (width and height) as input parameters. Graphics from this class have the special feature of always pointing towards the camera, which is useful for this application. Also the icons' sizes depend on their distance to the user, providing a visual cue to how far away the scenes are the AR view. The placeholder icons are developed in

Adobe Photoshop, and refer to the themes of the stories (see section 9.2).

Another feature of the `OAARComponentBase` is a compass, which provides an additional overview of scenes in the surrounding environment. This will help the user orient without having to navigate to the map or list views. Further, the class contains accessor methods to a sensor manager, which manages information about location and orientation, and a data manager, which is customised to handle the filtering mechanisms and decide which information to display on the screen.

In the activity itself, a custom UI layout with three buttons (Filter, Info and WWW) and a reticle image in the center of the screen to help the user aim at the scenes are laid out over the AR view (Figure 35A). When a scene intersects the center of the screen, the story's suggestive audio snippet is played back, using the FMOD Ex audio library to simulate 3D sound and give the impression that the sound is played from the scene's location. Further, an event is triggered, which displays a new layout attached to the icon in focus. This layout is an instance of the `NotificationBubble` class, depicting the name and a short content description of the story in focus, while the snippet is played back. It further shows valuable information about the scene such as its duration and subject, as well as an interactive "Play" button (Figure 35B). This button takes the user to the `SelectedActivity` class, where the user can explore the story in detail, and listen to the full story. A UML class diagram of the ARViewActivity and its interplay with the OutdoorAR framework can be found in Appendix F.

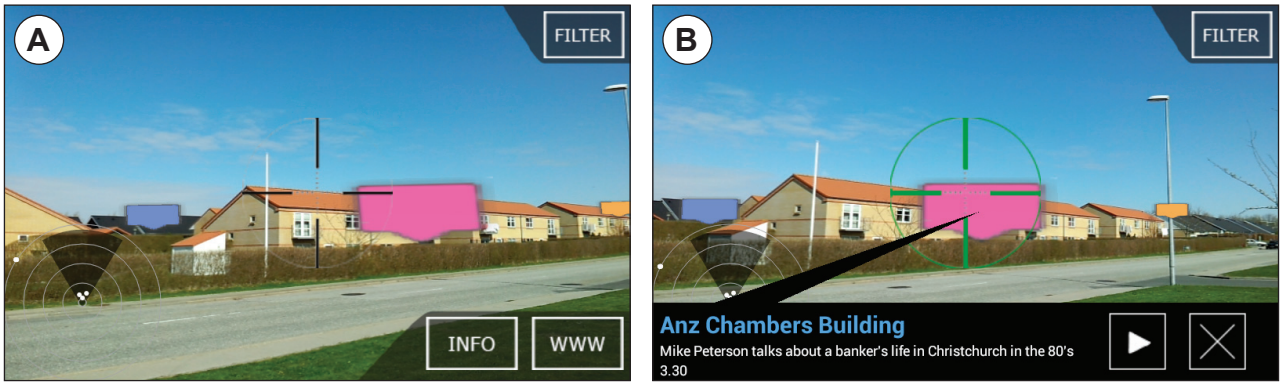


Figure 35: In the AR view, scenes are superimposed over the camera view of the phone. **A**: A reticle helps users aim at the scenes. **B**: When a scene is in focus, a seperate layout appears while a spatialised snippet of the story is played back. Tapping the Play button on the layout takes the user to the SelectedActivity.

10.4.2 MapViewActivity

The `MapViewActivity` extends the `OAMapComponentBase` class of the OutdoorAR framework, which provides all necessary methods to set up a map view. This includes integration with the Google Maps API, the possibility to overlay additional layers on top of the map, create points of interest (scenes) and events for selecting the scenes. It sets up a map view component with the scenes' icon sources by linking to the `OAIconSource` class. This class looks in the data manager to locate the icon source that should visualise the scenes from different categories. These icons are overlaid on the scenes' geographical locations on the map. Similar to the AR view, when a user selects a scene by touching it on the map, the audio snippet of the scene is played, and a seperate layout appears with a "Play" button for entering the `SelectedActivity`.

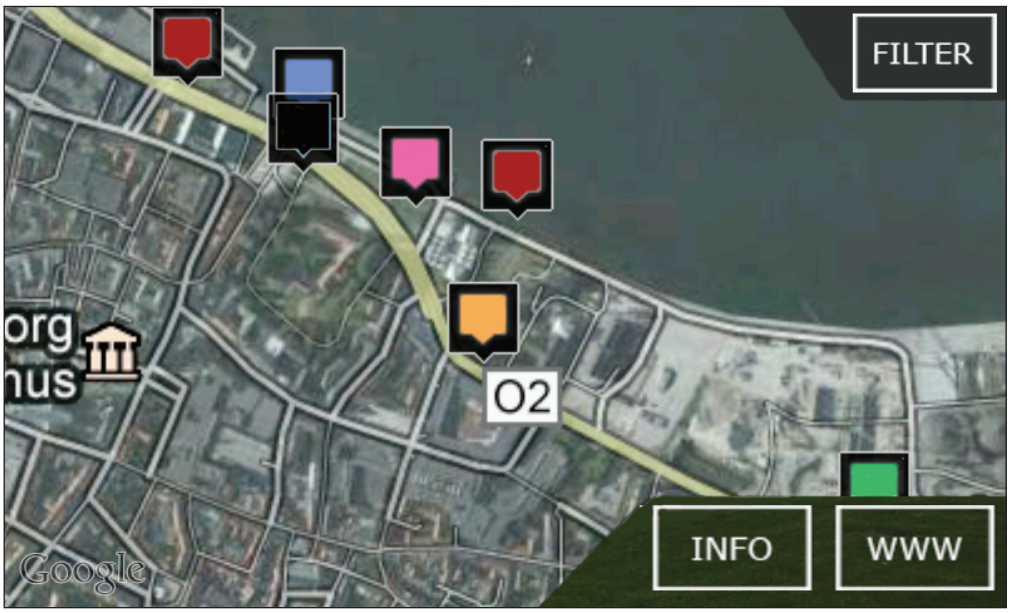


Figure 36: The map view shows all stories' icons superimposed on top of a map provided by the Google Maps API.

10.4.3 ListViewActivity and CategoryListViewActivity

The `ListViewActivity` extends the `OAListComponentBase` class, which provides a foundation for creating list views linked to the data manager. The list view will implement some sorting options for letting scenes swap position in the list, for instance when the view is accessed through the "Nearby" button, at which point the scenes should be sorted based on their distance to the user. This is implemented by comparing the GPS positions of all scenes against the user's position using the Euclidean distance formula as expressed in the following pseudocode:

```
distanceX = scene.getLatitude - currentPosition.getLatitude;
distanceY = scene.getLongitude - currentPosition.getLatitude;
distEuclid = Squareroot((distanceX*distanceX)+(distanceY*distanceY));
```

As the user navigates in the environment, the scenes swap position in the list view. Further the user can see the distance to each scene in kilometers or meters, dependent on how close the user is to the scene. Finally a graphic of an arrow informs about the direction to each scene. This arrow rotates as the user navigates in the environment to always point in the direction of each scene. The rotation is calculated using the GPS coordinates of each scene, which are compared to the location of the mobile device. Combined with the data from the compass in the mobile device (which provides the heading, or yaw, value), this makes it possible to make the arrows point in the exact direction of the scenes. Figure 37A depicts the regular list view.

The `CategoryListViewActivity` is a simplified version of the regular list view, not showing distance indicators or arrows. This only displays the different themes of the application, their icons and the number of stories of this particular scene (Figure 37B). Tapping a theme takes the user to the regular list view, now displaying only scenes from this particular category. In the `CategoryListViewActivity`, the user can also

select the "Nearby" button, which visualises stories relatively close to the user - regardless of their theme

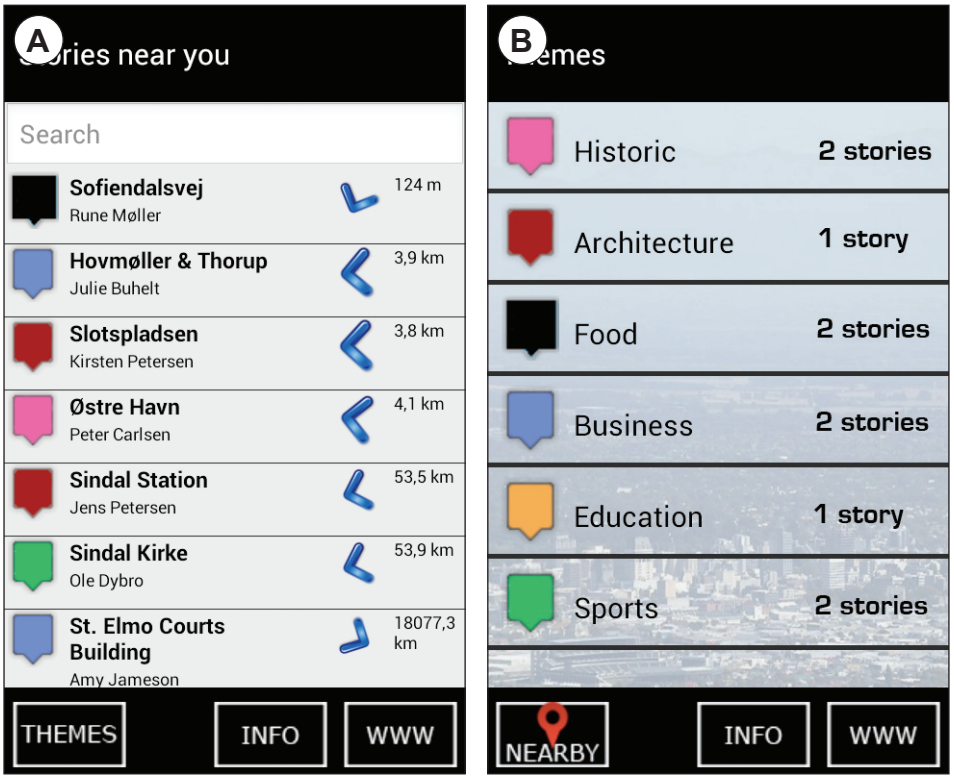


Figure 37: The two list views. **A**: The regular list view displays stories from a particular category or, as in this example, all stories sorted by their distance to the user. **B**: Tapping the "Themes" button takes the user to another list view, displaying a list of the themes. Tapping a theme or the "Nearby" button takes the user back to the regular list view, now displaying the stories based on the user's selection.

10.4.4 Navigating between AR, map and list views

For implementing the navigation between the ego- and exocentric perspectives, the `OASensorManager` class is useful. Through this class, we can access information about the internal sensors of the mobile device, such as its GPS status as well as accelerometer and gyroscope values. We use the gyroscope and accelerometer data to get the roll, pitch and yaw values of the device in Euler angles. These values are derived from the three axes of an airplane, and describe the rotation of a rigid body in 3D space. Roll usually refers to rotation around the forward axis, which is equivalent to a transition of the mobile device from portrait to landscape mode. Pitch refers to rotation around the side axis, and yaw refers to the up axis. (Van Verth & Bishop 2008) For navigating between the AR, map and list views, we only need info about the roll and pitch values of the device. The values are measured in degrees, and we need to define a proper threshold for the orientation switch, when a change in roll or pitch is detected. 45 degrees is convenient for the threshold, as this value is the halfway point between portrait and landscape mode. Figure 38 depicts an activity diagram of the orientation checks in the list view. If changes in the roll values of the device are detected (the device is rotated towards landscape mode), either the AR or map view will start, which is decided by checking the pitch values, which equals the tilt of the device. This procedure is similar in the AR and Map views.

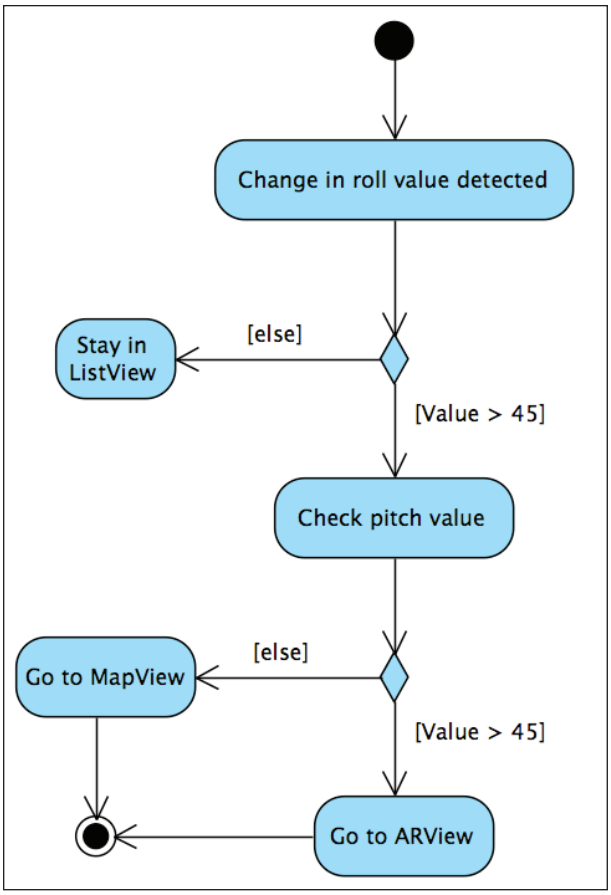


Figure 38: UML activity diagram depicting the orientation switch effect, here using the list view for illustrative purposes.

10.4.5 SelectedActivity

When a scene is selected, a new activity is started, which lets the user explore the story in depth. This is the `SelectedActivity`, which grabs the selected scene's information and presents this information to the user while playing back the full story. First of all, a `MediaPlayer` object is set up for controlling the interview file related to the story. This object provides access to basic audio controls, such as pausing, stopping and adjusting the audio file to a specific playback position. Using an Android SeekBar UI widget and a single button which stops or starts the audio, this provides handy controls for the user to enjoy the story in full. When this activity is in focus, an image of the building or area related to the story is displayed as well. Also, a short text description is implemented, providing information about the content of the story. Since the activity can be accessed from either the AR, map or list view, two seperate layouts are implemented for optimal display in both landscape and portrait mode. Figure 39 depicts the SelectedActivity in both landscape and portrait mode.

Figure 39: The screen appearing when a story is selected for full playback contains a large image of the building, a short textual description and a basic media player for audio control. **A**: Landscape mode. **B**: Portrait mode.

10.5 3D audio using FMOD

While the `SelectedActivity` simply uses an instance of Android's `MediaPlayer` class for audio playback, a more complex audio playback is implemented in the AR view. When a user explores the stories in the surrounding environment from the AR view before deciding to select a particular story, the story snippets are localised using the FMOD audio library. The location of a sound is based on a three-dimensional vector, (X,Y,Z) which stores the scene's position on the mobile screen. X corresponds to the left-right axis, Y to the up-down axis and Z represents depth, equivalent to the scene's distance to the user. When using screen coordinates combined with a point-based interaction method, localising the sound does not have an effect in the X and Y coordinates, since a sound source will always be in center of the screen, when it is played back. If the sounds were not localised based on the device's orientation, but instead used the user's head orientation and position in relation to the scenes, the 3D sound effect could be more convincing. However, the depth coordinates make the volume increase and decrease as the user walks towards or away from a scene, which is still assumed to provide a distance cue and a powerful user experience when interacting with the audio snippets. For future work - perhaps when investigating alternative interfaces - the 3D sound effect could be explored in greater depth.

The FMOD library is linked to the OutdoorAR framework using the `OASoundManagerFMOD` class, which accesses the sounds attached to each scene. The sounds are stored as .wav files in the Assets folder of the project, which also stores the icons of the scenes. When a scene is pointed at in the AR view, the screen coordinates are calculated to set the volume and position of the sound in the stereo channels of the user's earphones. This implies that a sound snippet source can be very low in volume, but when the story is actually selected, the full audio file plays back normally in the `SelectedActivity`. The temporary audio clips used before the final stories from New Zealand arrive are stereo, 16-bit and sampled at 44.1 kHz.

10.6 Chapter summary

This chapter has described an implementation of the memoAR application on the Android platform, in the specific context of the High Street Stories project. The application uses an OutdoorAR framework provided by the HITLabNZ, which integrates functionalities for AR experiences, including a map view (Google Maps API) and 3D audio playback (FMOD). The different screens through which the users interact with the application have been presented, as well as diagrams depicting important logic and navigational features. In Part 3, the application is evaluated through two seperate experiments. Further, a section on future work discusses the opportunities for integrating social networking, which is the long-sighted goal of the thesis.

PART 3: EVALUATION

This part of the report evaluates the product of the thesis. First, a usability experiment is conducted, comparing different interaction methods for the AR view. Then an empirical field trial is carried out in a real world setting, before the thesis is concluded and discussed. A final section on Future Work discusses an eventual integration of social media and networking.

11. Comparative usability test of interaction methods

This section treats the conduction of a usability test, which compares three different interaction methods for selecting scenes in the AR view. Since an initial design decision has been to focus on aim-based interaction methods to present different levels of information to the user, it makes good sense to evaluate those interaction methods by comparing them against more established AR interaction methods.

11.1 Purpose

Recall section 7.1, an important user experience factor regards the user's interaction with the audio content from the egocentric AR view. The purpose of this experiment is to compare different ways of interacting with AR scenes, more specifically the basic task of selecting a scene. Unlike the list and map views, in the AR view scenes are superimposed on top of a camera view, which possesses some usability challenges such as camera motion when touching the screen and jumping icons following unstable tracking. The AR view provides an egocentric view of the scenes, where users need to point their mobile device directly towards the scenes of interest, making it a different challenge than when using the map and list views, where scenes appear fixed on the screen, and can be touched directly. This experiment compares three approaches to selecting an AR scene:

- a "single touch" method, where the user selects the scene by directly touching its icon on the screen.
- an "aim-and-touch" method, where the user aims at a scene and taps a fixed play button, which appears when the scene is overlapping the center of the screen.
- an "aim-and-wait" method where the user aims at a scene and keeps the scene in the center of the screen for two seconds to select it.

Other researchers have compared alternative ways of interacting with augmented content, for instance through a Freeze-Set-Go technique, which allows users to freeze the AR view and manipulate the scene on the frozen image (Lee et al. 2009), and a Snap-To-Feature technique, where the interaction point from a mouse or cursor is snapped to image features when the point get close enough (Lee & Billinghurst 2012). These alternatives rely on computer vision-based tracking methods, which are not suitable for open, outdoor environments. Ballagas et al. (2005) present the idea of using the camera phone as a pointing device to interact with large, public displays, but do not discuss the challenges of outdoor AR settings.

Also under observation in the experiment is the impact of the icon's size on the task performance, which is expected to be an important indication of the system's usability. Having the icons vary in size on the display is an aid to overcome information clutter, and can provide hints to the distance to the scenes.

11.2 Setup

Seven participants participate in the study, which is a 3x3 within-subjects, factorial design. The participants come with different backgrounds, and novices as well as experts in AR

are recruited. Figure 40 depicts selected demographic information about the participants.

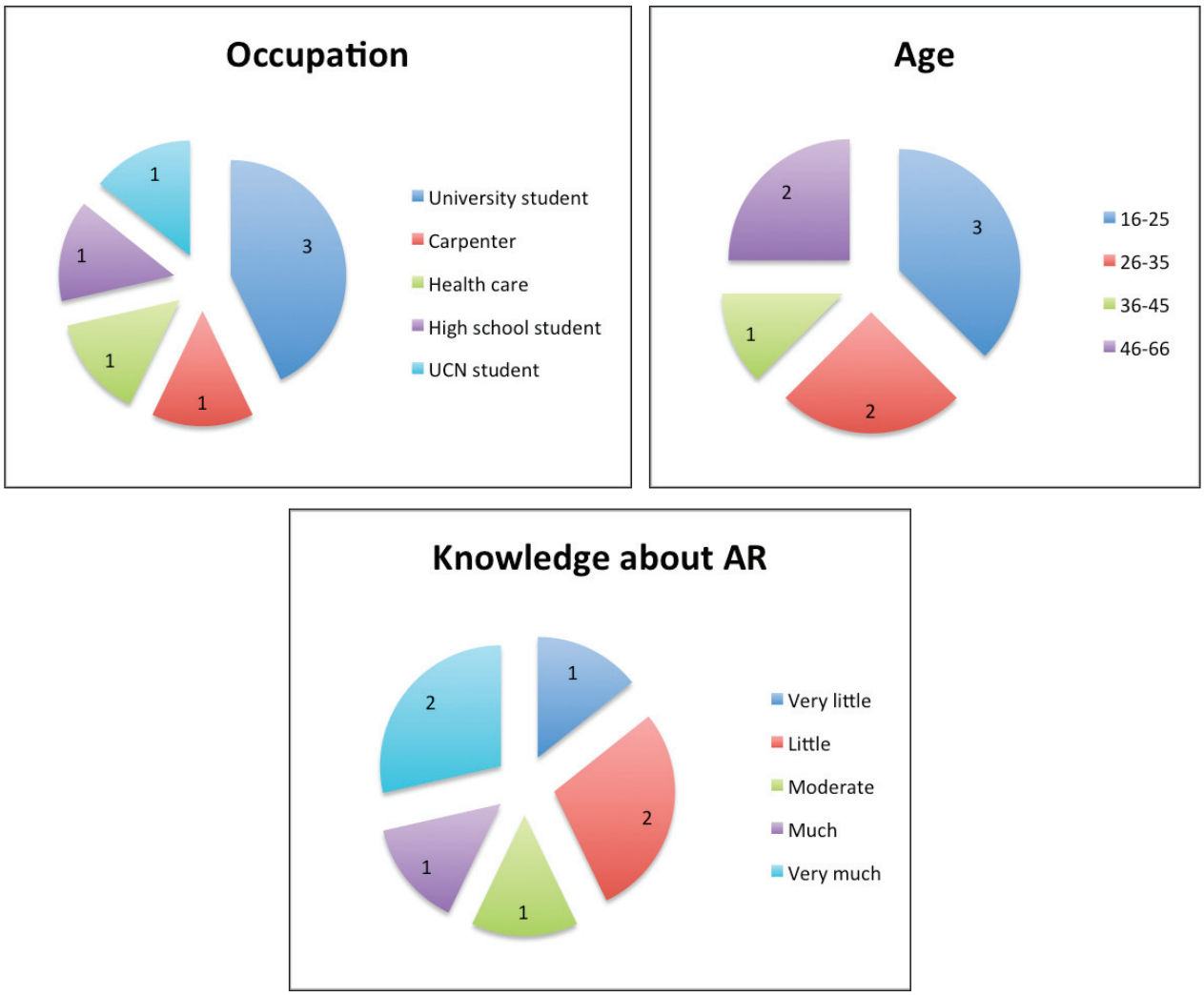


Figure 40: Demographic information about participants of the usability experiment.

The independent variables are the interaction method (single-touch, aim-and-click and aim-and-wait) and the size of the icons on the screen (90, 60 and 30 density-independent pixels, dp). These sizes are based on Google's recommendations, which say that touchable UI components should be laid out around 48 dp units, corresponding to a target size of approximately 9 mm. This size is within the recommended target sizes of touchable UI components (7-10 mm). (Google 2013b) We choose icon sizes above and below this recommendation, to test the impact in the AR view, where icons are smaller the farther away a scene is from the user. The three icons are all visible in the phone camera's field of view when the test is started and further distinguishable by their color (blue, yellow and pink) to aid the participants during the experiment.

The dependent variables of the experiment are task completion times and errors, which will be logged automatically through a small Android program written specifically for the test. The participants perform a simple task, where they select each of the three scenes two times for each of the three interaction methods. This gives 7 x 3 x 3 x 2 = 126 individual trials, where, for each trial, errors and task completion times are logged. An error is defined as a failed attempt to select a scene. For the aim-and-touch and aim-and-wait

interfaces, this happens when a participant has the right scene in focus, but focus is lost before selection is completed. For the single touch method, an error occurs when a scene is not correctly selected when the participant touches the screen. Before each trial the user taps a "Start" button on the screen, after which the test begins and the logging of task completion times and errors starts. The data is printed as a text log for further analysis. The order in which the participants perform the tasks is counterbalanced to compensate for eventual learning effects. Further, three participants perform the task for icons in order pink-yellow-blue, three in order blue-yellow-pink, and three in order yellow-blue-pink. Before starting the trial for each interaction method, the participant spends three minutes getting familiar with the method. During the test, a facilitator overlooks the participant performing the tasks, and writes down eventual issues that should be taken into consideration during further analysis. After the test, the participants are asked about their general performance, and asked to rank the interaction methods after personal preference and ease of use.

The experiment is conducted on a Huawei Honor U8860 smartphone, and uses a mock GPS location to position the mobile device correctly in relation to the icons on the screen. This ensures that the icons are always in the right distance to the participant.

11.3 Observations

11.3.1 Task completion times and errors

The experiment reveals that for both task completion times and errors, the single touch method outperforms the two aim-based methods. The average task completion time for the single touch method, regardless of icon size, is 2.2 seconds, while it is 5.8 and 6.52 for aim-and-touch and aim-and-wait respectively. The average error rate for the single touch method is 0.4, while it is 0.98 and 0.74 for aim-and-touch and aim-and-wait. Figure 41 depicts diagrams of the experiment results.

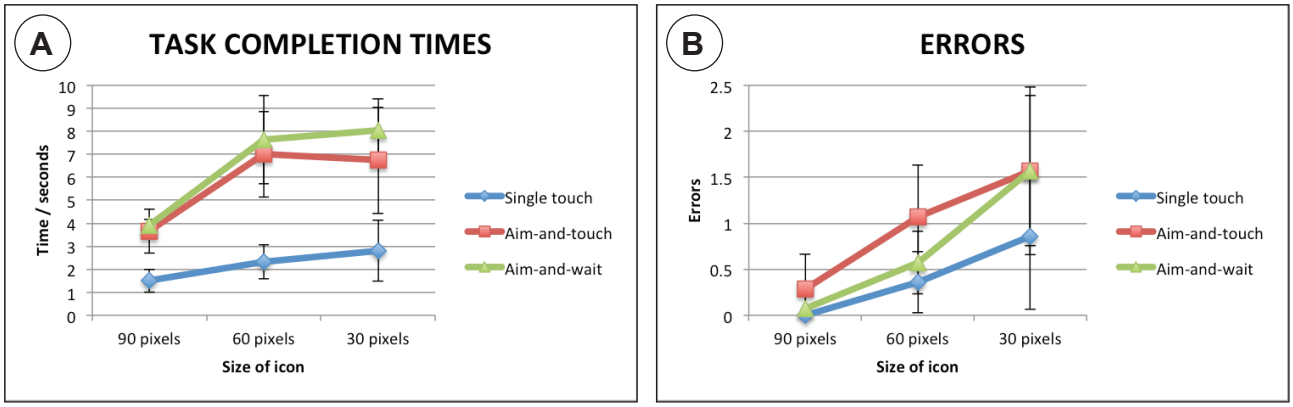


Figure 41: Results from the usability test. **A**: Task completion times. **B**: Error rates.

Comparing the two aim-based methods, the error rates for the 90dp and 60dp icons are higher with the aim-and-touch method, but almost identical for the 30dp icon. Based on the participants' comments to the task completion, the multi-tasking approach of the aim-and-touch method made it difficult to keep the icons in focus. Some participants remarks that they had to take their eyes off the crosshair to focus on the play button, making it easy to lose the icon's focus, which results in an error. Keeping the icon in focus

is generally easier with the aim-and-wait method, because the user can use both hands for aiming and do not have to take their eyes away from the icon. Because of the camera tracking, however, when the icon is very small, it gets difficult to keep the small icons in focus for two consecutive seconds. The timer aspect of the aim-and-wait method also causes the task completion time to be higher than the aim-and-touch method, where the participant could touch to select immediately after getting an icon in focus. The problems with the multi-tasking approach of aim-and-touch are more outspoken among novice participants without knowledge about AR, while some experienced users find aim-and-touch more natural and intuitive, and appreciate that they do not have to wait for a system-defined amount of time to select an icon.

11.3.2 Impact of icon size

Comparing the aim-based methods, it appears that between 90 and 60 pixels, the task completion times and error rates increase rapidly. An error rate of more than 1.0 per trial can probably be considered high in a controlled lab experiment like this. For task completion time, the gap between 60 and 30 pixels does not seem to make a huge difference, but the aim-and-wait method seems much more difficult for an icon of size 30 pixels than for an icon of size 60 pixels. This again suggests that a timer-based interaction method on small icons can be troublesome in AR settings because of the camera motion and unstable viewpoint.

11.3.3 Personal preference

After the experiment, the participants are asked for their personal preference of interaction method, particularly putting weight on ease of use. Participants rate the methods from 1 to 3, where 1 is the most preferred method, and 3 is the least preferred method. Table 1 depicts the average rankings for the three methods.

Single touch	Aim-and-click	Aim-and-wait
1.29	2.57	2.14

Table 1: Average personal preference rate for each of the three interaction methods. Participants rate the interaction methods from 1-3, where 1 is the most preferred, and 3 the least preferred method.

The single touch method is overall the most preferred method. Most participants remark that this method is the most easy in this context, where the unstable viewpoint causes the objects to jump around on the screen. Keeping the icon in focus for 2 seconds (aim-and-wait) and having to perform multiple actions (aim-and-touch) to select the icons are the most outspoken usability issues of the aim-based methods. One participant notes the usefulness of the aim-and-wait method in outdoor settings, for instance when wearing gloves in winter times, where touching the screen is not possible.

11.4 Results

This experiment has provided insight into challenges and opportunities of aim-based interaction methods for the AR view. The following points sum up the major observations:

- Aim-based methods are slower than single touch and more prone to errors
- A “multi-tasking” approach, with both aiming and touching is difficult for novices, but appealing to experts
- Aim-and-wait is only feasible for icons larger than approximately 50 pixels
- Novices emphasise safety and simplicity - experts emphasise efficiency

While the experiment speaks in favour of the single touch method, the context of usage is necessary to think into the results. The aim-and-touch approach is the least favoured method for this experiment, which only tested the simple task of selecting a scene. The results could be different when testing more complicated scenarios, for instance when having different levels of content to display for the user. For the memoAR application, two levels of content are provided: suggestive content, which provides a quick overview of what to expect of a story, and in-depth content for playing back the full story and present additional media. Single touch might have some limitations in these more complicated contexts, where aim-based methods can be more “direct”, since the user is already pointing towards the environment in the AR view. The aim-based methods do however need improvements to be useful.

11.5 Suggestions for improvements

The experiment did reveal that improvements are necessary to make the aim-based methods feasible. First it is necessary to ask: which error rates and task completion times can be considered a success for aim-based interaction methods? Since this experiment has been conducted in an indoor lab setting, the task completion times and errors will most certainly increase in outdoor, real world contexts. Sunlight, the walking user and the external environment will all contribute to reduce users’ performance. From this perspective, an error rate of more than 1.5 (aim-and-touch + aim-and-wait for an icon of 30 pixels) is high in a lab setting. One solution to improve the aim-based methods is to scale the icons (make them larger), which would reduce the error rates and make it easier to select a scene far away. On the other hand, this approach would also increase the information clutter on the mobile display. Another option is to increase the area of the screen, which a scene should intersect to be selected. For the experiment, the icons had to intersect the exact centerpoint of the screen, but the area could be increased to, say, 20x20 pixels around the center of the screen. Finally, inspired by the freeze-frame approaches described earlier, once a scene is in focus, the user could perform an explicit action (touching a de-freeze button, shaking the device etc.) to de-select it. This approach could however imply a lot of frustrating false positives.

11.6 Chapter summary

This chapter has presented a usability test with the purpose of comparing and evaluating alternative approaches to AR interaction against each other and against the more established single touch method. Results has shown that at least for basic tasks, aim-based methods are outperformed by single touch. Novices find it difficult to multitask to perform a selection, and the unstable viewpoint makes it difficult to integrate the aim-and-wait method. The two aim-based methods are however expected to possess advan-

tages when it comes to more complicated tasks. In the next chapter, a field trial is carried out, seeking to investigate the application in real use. This will further highlight some of the opportunities and challenges of aim-based methods in AR visualisations, and help evaluating the overall usability of the application.

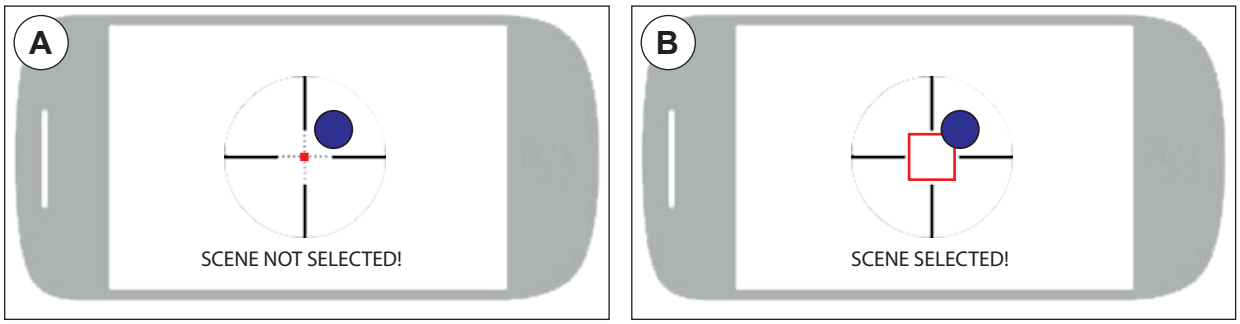


Figure 42: Increasing the area of the screen, which the scene should intersect to be in focus. **A**: Originally, a scene needed to intersect the exact centerpoint of the screen. **B**: In the improved version, a scene should intersect a virtual square, which occupies an area of 20x20 pixels around the center of the screen.

12. Empirical evaluation

This chapter treats the conduction of an explorative field trial of the memoAR application in use. In the previous chapter, a usability experiment on different interaction methods was conducted, which highlighted some interesting challenges of AR interaction. What is even more interesting for a mobile audio experience like memoAR, is the users' actual experience of the application in the field, not constrained to a single task or interaction method in a lab setting. This chapter broadens the evaluation by setting up an empirical field trial in two urban settings: a relatively noisy and busy environment on the harbourfront of Aalborg, DK, and a more quiet setting in the smaller village of Sindal, DK.

12.1 Purpose

To this day, actual user evaluations of AR systems outside lab settings have been quite rare (Dünser & Billinghurst 2011). Since memoAR contains rather complicated usage scenarios and different ways of visualising the scenes, the goal of this experiment is to take a rather explorative approach to evaluate the application in a realistic usage scenatio. Also, a field trial is expected to help shed light on contextual factors affecting the interaction, such as weather, traffic, other people etc. The following questions are of particular interest for the experiment:

- What issues, technical and contextual, arise when the application is used in realistic usage scenarios?
- Do contextual factors such as weather conditions and traffic affect users' performance and audio perception during the completion of tasks using the application?
- How do users perform tasks related to basic usage of the AR, map and list views?
- How do users orient themselves, when they need to identify an unknown story?
- How do users rate the overall usability of the application?

The most important aspect to clarify for this experiment is the overall feasibility of the application's features, primarily whether users intuitively understand how to navigate between the visualisations, know how to orient themselves using the app, and are capable of browsing and selecting scenes in each of the visualisations. This experiment will help evaluate the application at its current state and identify areas for optimisation and improvements before future experiments, which will focus more intensely on the audio experience.

12.2 Setup

Before the experiment, some changes in the application are implemented based on the usability experiment in the previous chapter. The area, which a scene should intersect to be in focus, is increased to a 20x20 pixel region around the center of the mobile scree to compensate for the unstable tracking. Further, the "Play" button in the AR and map view is increased in size, and the feedback of UI widgets is improved by implementing simple visual cues to indicate when a button is pressed.

A Danish version of the memoAR application is implemented, called "Nordjyske Historier" ("Stories from Northern Jutland"), containing stories tagged to buildings and areas in Aalborg and Sindal. The categories are *Historic, Architecture, Food, Business, Education* and *Food*, and each setting has one story in each category. The stories from both settings are included in the same app, which means that Aalborg participants can browse the Sindal stories and vice versa. The stories are tagged to the following locations:

Aalborg:

- Aalborghus Castle (Historic)
- Utzon Center (Architecture)
- Nordkraft (Sports)
- Toldbod Square (Business)
- John Bull Pub (Food)
- Damsø og Damsø Language School (Education)

Sindal:

- Sindal Stadium (Sports)
- Sindal School (Education)
- Sindal Church (Architecture)
- Sindal Station (Historic)
- Mammas Pizza (Food)
- Jyske Bank (Business)

Further, each setting has a "Story X", which is included to test participants' ability to identify an unknown building, which has a story attached. The environments are chosen because they provide two different contexts, which could affect the observations, in relation to the way participants use the application. While the Aalborg stories are set in a relatively noisy and crowded setting, the Sindal context is more quiet and deserted.

The stories are written by the author in a descriptive, advertising style. All stories are in Danish and recorded by a male or female speaker, who resemble fictive placeholder subjects used for the stories. This is of course not as realistic and believable as in the High Street Stories context, but a necessary tradeoff because of time constraints. The speaks are inspired by descriptions of the sites from web pages and can be found in Appendix G (translated to English). All buttons and text in the application are translated to Danish as well.

Before the experiment, participants spend 10 minutes exploring the app and getting familiar with all its functionalities. They are encouraged to ask questions, should those arise, and the facilitator makes sure that all functionalities of the application have been presented for the participants. The participants are taken to a starting point, which is within a distance of 50 meters for most of the stories in each setting. The starting point is chosen to ensure that user's will need to walk a short distance (<200 m) to complete some of the tasks, while other tasks do not require walking. Figure 43 depicts a geographical overview of the location of the stories as well as the participants' starting point in each setting. At the starting point, the user is handed the first of seven tasks.

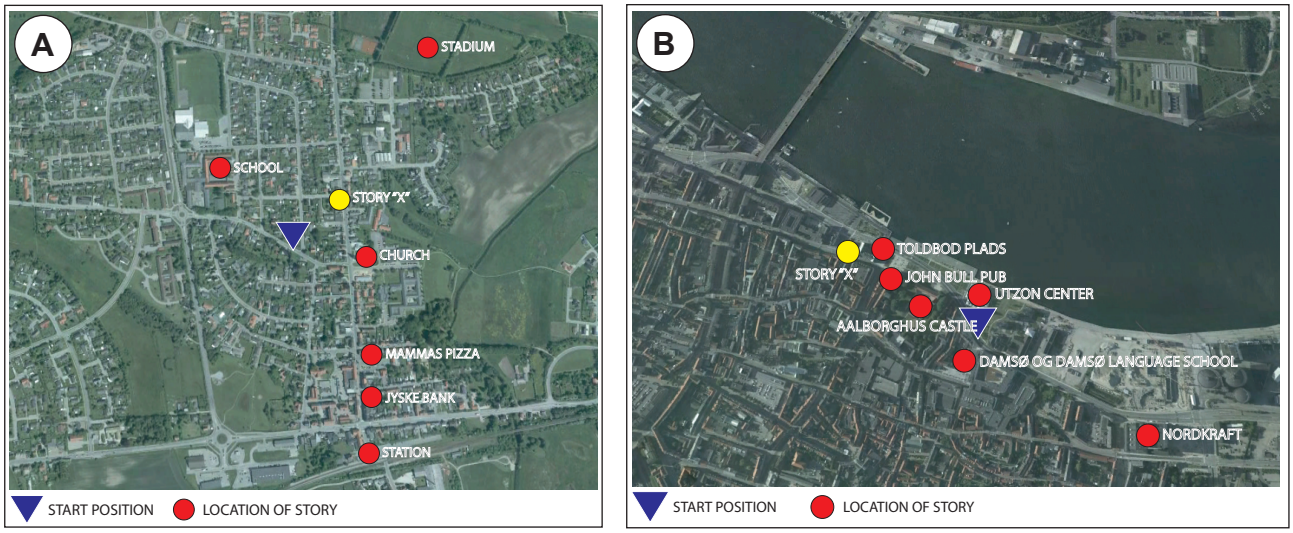


Figure 43: Location of scenes in the two settings for the experiment. **A**: Sindal setting around the city's main street. **B**: Aalborg setting at the central harbourfront.

12.2.1 The tasks

When the experiment is started, the participants are asked to complete the following tasks, which aim at testing the most important features of the application. Those include the AR, map and list views, as well as the filtering and visualisation switch mechanisms:

- Task 1: Start the tour and find out which story is closest to you. How far away is this story?
- Task 2: Sindal Church / Utzon Center is associated with a story. Use the AR view to hear the extract of this story. Recite the speech to the facilitator.
- Task 3: Find out how far away the nearest story of the category "Business" is.
- Task 4: Find the story related to Sindal School / Nordkraft and use the AR view to access and play back the full story.
- Task 5: Use the map view to identify the nearest story of the category "Food".
- Task 6: Use the list view to find the story named "Story X". Navigate to this story's location and identify its building.
- Task 7: Use the "Filter" mechanism to visualise stories from the categories "Sports", "Education" and "Business" only.

Those tasks ensure that the participants are forced to explore all functionalities of the application during their task completion.

12.2.2 Data collection

During the experiment, an observer takes notes of all relevant observations regarding the participants' choices and actions. After each task the participant is asked to complete a short questionnaire regarding their task performance. The questions ask for the participant's subjective judgement of the application's usability at a

task level, more specifically the time-on-task, errors and ease of the task. This questionnaire ensures feedback on both effectiveness and efficiency. The questionnaire template can be found in Appendix H. After the experiment the participants are asked to complete another questionnaire, which treats the overall usability of the application. This questionnaire is based on John Brooke's System Usability Scale (SUS) (Brooke 1996), which is known to provide reliable and valid measures of a system's usability, even at small sample sizes (Sauro 2011). The SUS questionnaire template can be found in Appendix I. Finally, participants are briefly interviewed about some of the observations by the facilitator to provide a better insight into the participants' choices and actions during the experiment. Overall, the approach ensures both quantifiable, subjective measurements, qualitative data and a measurement of the overall usability of the application.

12.2.3 Participants

10 participants are recruited for the experiment, all second-degree contacts to the author, ensuring that they would not have a personal tie to the author. The exception is one participant, the author's cousin (Partic. # 7), who is chosen because he has a technical background and is expected to provide valuable additional information regarding the user interface and technical issues. When asked to participate, participants have no knowledge about the purpose of the application, and it is deliberately chosen to recruit participants with little or no expertise in AR. Figure 44 depicts a breakdown of the participants.

Partic. #	Gender	Age	Owns smartphone?	AR knowledge	Occupation
1 - SIN	F	18	Yes	No	High school st.
2 - SIN	F	17	Yes	No	High school st.
3 - SIN	F	17	Yes	No	High school st.
4 - SIN	F	17	Yes	Little	High school st.
5 - SIN	F	17	Yes	No	High school st.
6 - SIN	F	44	Yes	Little	Social/health care worker
7 - AAL	M	22	Yes	Little	University st. (Electrical eng.)
8 - AAL	F	20	Yes	No	University st. (Social worker)
9 - AAL	F	20	Yes	No	High school st.
10 - AAL	F	22	Yes	Little	Pedagogue

Figure 44: Breakdown of test participants.

12.3 Observations

12.3.1 General observations and interview comments

Overall subjects appear to have a good understanding of the features of the application. Most tasks are completed relatively easily and with little hesitation from the participants. Even though the participants have received a thorough introduction to all features of the application and the experiment requires participants to use most of the features, participants generally seem to re-

member all the functionalities, and are not overwhelmed with the complexity of the tasks.

Not surprisingly, most problems occur in the AR view (Task 2 and 4), primarily caused by the GPS tracking making icons appear a bit off the buildings. Two participants hold the phone still over the buildings, waiting for the icon to "fly in" over the build-

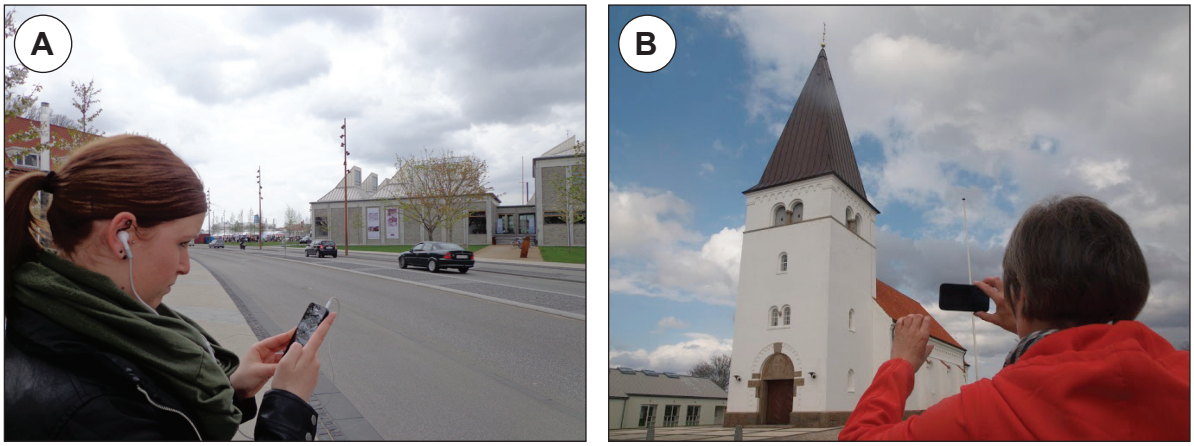


Figure 45: Participants conducting the experiment in the two settings. **A**: Participant in Aalborg with busy traffic (and a weekend market) in the background. **B**: Participant in the more quiet setting in front of the church in Sindal.

ing. The remaining participants understand that they sometimes need to scan the near environment surrounding the building, and they identify the right icon after a maximum of 5 seconds.

For Task 4, where participants must activate a story which is located 150-200 meters away, participants most often enter the AR view already from the beginning of the task, as opposed to using the list or map to identify the location of the story, and then walk a certain distance to approach the setting of the story. This causes the story's icon to appear small on the screen, making it difficult to keep the icon in focus. As compared to the usability experiment, these issues are however not as severe as expected, and only one participant spends more than 20 seconds solving the task.

An interesting observation is that users do not seem to take advantage of the benefits of each visualisation to locate an unknown story. For instance, most subjects in the Sindal experiment are unaware of the location of Sindal School. Instead of identifying the story's location on the map or list before starting the AR view (Task 4), they seem to solve the task by trial-and-error, by immediately starting the camera view and scanning the environment, checking the name on the layout appearing when a random icon is in focus, before eventually finding the right icon. A few participants even use the filter mechanism to filter out all other categories than Education (which they *guess* is the category of the Sindal School story). These observations suggest that there could be a need for a tighter interplay between the visualisations, and that users need more encouragement to use the visualisations as supplements to each other.

One participant struggle to remember the interaction method of the AR view, and attempts to touch the icons of the stories instead of aiming at them. In the successive interview, the participant remarks that this was because "it seemed obvious". The visual cue of the reticle did not aid the participant in this case, with her remarking that she did not actually take notice of the reticle.

The list and map views seem to work really good for their basic functionalities, although partici-

pants do not show a deeper intuitive feel for these exocentric perspectives, for instance when locating the position of a story at an unknown location (Task 4 - see above). Participants are able to orient and identify the building of the unknown "Story X" using the arrows in the list view with relative ease (Task 6). Due to a weak (and occasionally missing) GPS signal, sometimes arrows do not update fast, or simply point in the wrong direction. This causes participants to stop for a few seconds, wait for the arrows to update and then walk on. This is mostly an issue in the Aalborg setting, probably caused by larger buildings blocking the GPS signal and slightly more overcast weather conditions. Participants do not seem to use the distance indicator to achieve a mental idea of where the story is located. This implies that even though the subjects are familiar of the area, they occassionally take a slight detour through obvious environmental obstacles, while seeming to focus blindly on the arrows and not using their common sense to take the distance to the story into account.

The Filter mechanism seems to work good, and participants understand intuitively how to solve the filtering task (Task 7). A few participants struggle to remember that they need to be in the map or AR view to activate the filter mechanism. The one tech-savvy participant notes the inconsistency in calling the mechanism "Filter", since it is somewhat similar to the "Themes" mechanism in the List view, which also lets users decide what themes to display. This is an important observation, and for follow-up design work, the consistency between the "Filter" and the "Themes" funcionalities needs improvements.

12.3.2 Comparing Aalborg and Sindal contexts

In spite of contextual differences, especially related to the level of traffic and ambient noise, no major differences are observed between the Sindal and Aalborg settings. As an example, in Task 2 participants have to recite the spoken message played when pointing towards a scene in the AR view. This task do not cause any problems for the participants in either setting. Further, no differences in regards to the way participants completed the different tasks are observed.

12.3.3 Post-task ratings

For the post-task questionnaire, participants are asked to rate their level of agreement with six statements (Appendix H). Questions A and D asks for their perception of whether they made mistakes during the task, questions B and E regard the time on task, and C and F regard whether participants perceive the task as easy. The responses are converted to numbers from 1-5, with 1 being the negative extrema and 5 the positive extrema. The three pairs of questions are averaged with the results across all participants depicted in Figure 46.

For all parameters, the score is well above medium for most tasks, and no significant differences are found between the parameters. This suggests that participants generally feel they solved the tasks without major complications and within a reasonable amount of time, which goes well with the observations. The smallest ratings are provided for the tasks related to the AR view (Task 2 and 4) as well as Task 6, where participants used the list view to identify the location of a story. As the observations revealed, the AR view did indeed cause the most trouble to the participants, but these were not as severe as expected after the usability evaluation in Chapter 11. The "free" tasks where participants could solve the tasks as they preferred (Task 1, 3 and 7), all score high in the subjective data. Observations supported that once participants understood the task, they seemed very deliberate and from the beginning knew how they would solve the task.

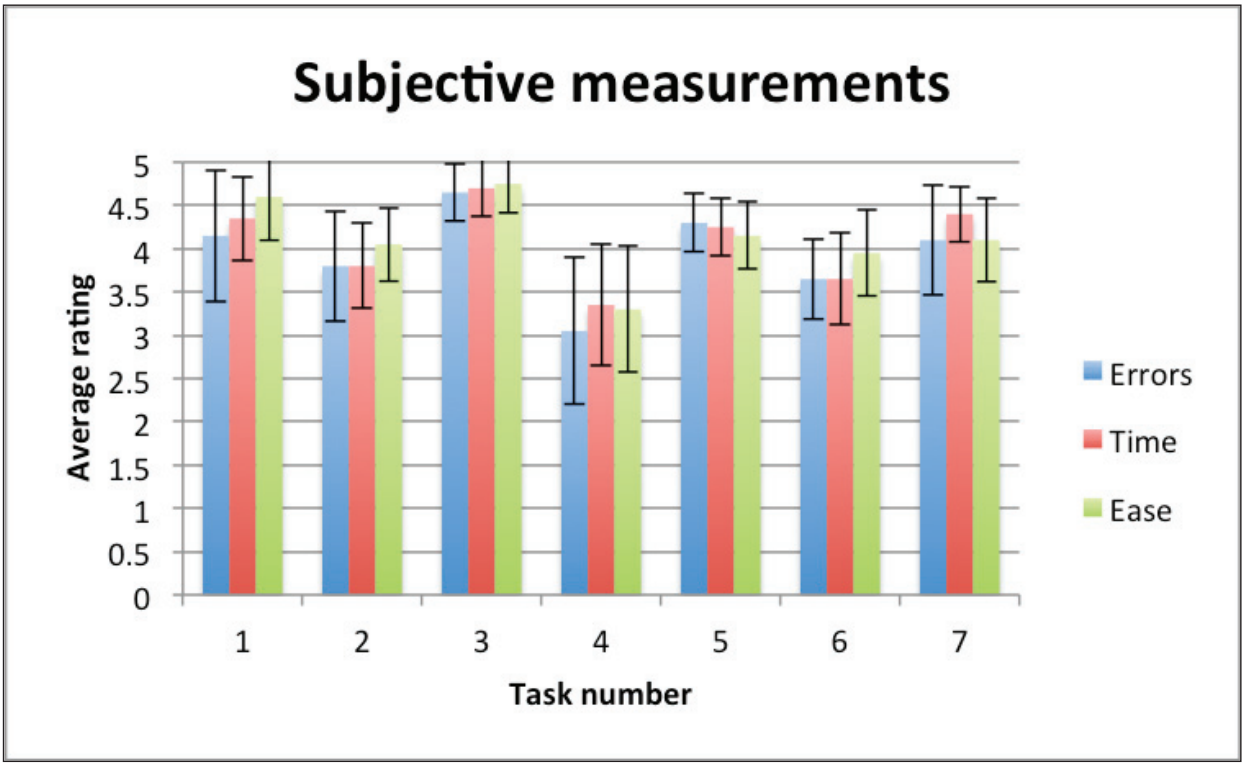


Figure 46: Subjective measurements from the field trial. Participants rate their level of agreement to six different statements related to errors, time-on-task and ease. Responses are converted to a numeric scale from 1-5 with 1 being the negative extrema, and 5 being the positive extrema.

12.3.4 System usability

The results of the SUS questionnaire reveal that participants are overall positive towards the usability of the application. Averaging the SUS responses, memoAR ends at a score of 79.75 (Figure 47). Based on the recommendations by Tullis & Albert (2010), who reviewed 50 usability studies using SUS, this can be considered pretty good. Bangor et al. (2008) argue that better products will receive a SUS score from the high 70s to upper 80s. The average score goes well with the observations, which suggested that participants generally solved most tasks easily and did not struggle to remember or use the application's functionalities.

12.3.5 Technical issues

The experiment further revealed some technical issues, which should be addressed in further work on the application. It is not surprising that the GPS tracking was a bit unprecise, which implied that in the AR view participants often needed to scan away from a building to find its icon, and further occasionally lost tracking of an icon. This also caused occasional confusion in the list view, when participants used the arrows to guide them in the right direction of a story.

The application occassionally crashed during the experiments, which seems related to memory issues and bugs in the implementation of the orientation switch mechanism. These bugs are important to fix for future work before releasing the application for people in Christchurch.

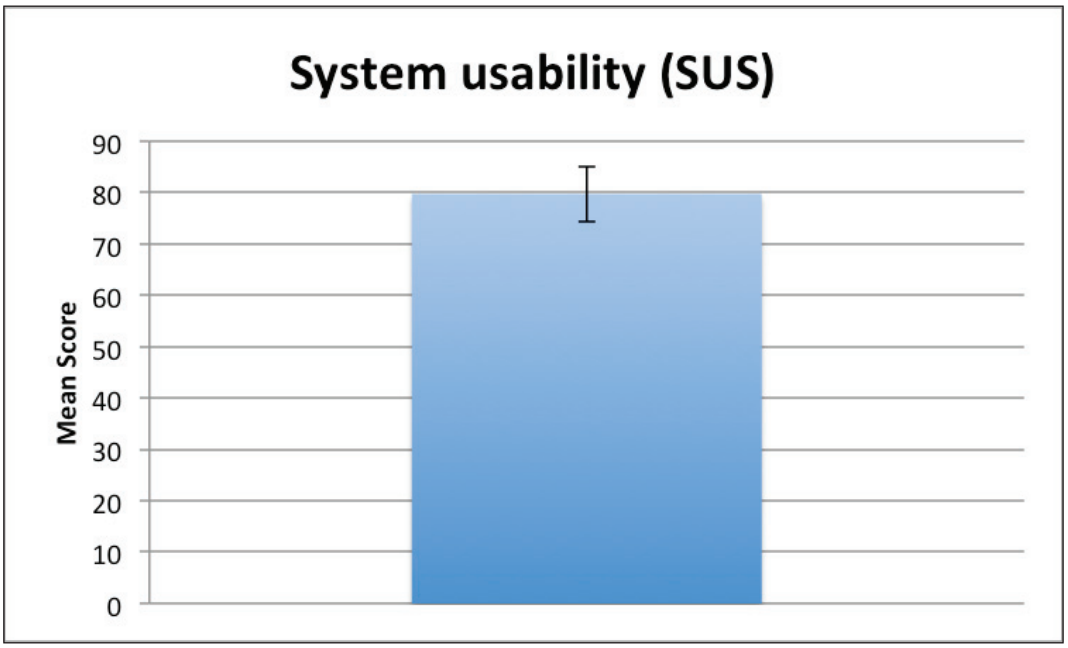


Figure 47: Mean score for the SUS questionnaire.

12.4 Chapter summary

This chapter has described the conduction of an empirical field trial aiming at testing the functionalities of the memoAR application. A Danish version of the app has been implemented, containing stories geo-tagged to buildings in Aalborg and Sindal. 10 participants has taken part in the experiment, in which they completed seven tasks using the application. An observer took notes throughout the experiment, and participants completed post-task and post-session questionnaires on the usability of the application at task-level and for the overall experience of the application. Results has shown that the application possesses a good overall usability, and its functionalities are easy to learn and understand. The AR view causes the biggest troubles for the user because of the unstable tracking and point-based interaction method.

13. Summary of findings

The experiment conducted in the previous chapter has provided positive insight into the usability of the application, as well as areas that need improvements in future work. Summing up the results from the empirical field trial, the following points remain as the main findings:

- The functionalities of the application are easy to understand. After a quick introduction to the application's main features, users have few problems remembering how to use them, including the use of the device's orientation to navigate between visualisations.
- The amount of information displayed on the individual screens is appropriate. Participants generally seemed to know intuitively how to solve the tasks, and were not cognitively overloaded when using the application.
- More emphasis should be put on the use of map and list views to support the AR view - and vice versa. Participants did not seem to take full advantage of each visualisation to solve more complex tasks, such as locating the building of a story.
- The AR view needs improvements to be highly useful. While most participants solved the tasks of the AR view with relative ease, interacting with scenes far away remains as a problem, which needs more thorough addression.
- There is a need for more consistency between the "Filter" and "Themes" buttons. While both the AR, map and list views provide similar mechanisms for filtering the scenes visualised on the display, the function has a different name in the list view (Themes) than in the AR and map views (Filter). Currently, this is because users do not have the opportunity to select multiple categories from the category list view, making the mechanism slightly different in the list view.
- Contextual factors, such as traffic and other pedestrians, do not seem to affect users' interaction with the application.

14. Discussion

This chapter discusses the validity of the test results of the previous chapter. Further, it briefly discusses ideas for further testing, which have been left out of this thesis' experiments, but should be conducted to ensure an even broader evaluation of the memoAR application.

14.1 Social desirability bias

While the empirical field trial did provide positive feedback to the application, some potential biases need to be considered. The following points speak for a tendency towards the so-called social desirability bias (Tullis & Albert 2010):

- Most participants were a second degree contact to the author. While the author did not know the participants in person, the fact that the author was, say, "my friend's brother" could have had an impact on their responses.
- The author conducted the experiment without assistance. This meant that the author took the role as the facilitator, technical support, observer and, not the least, the person responsible for handing out and gathering questionnaires. While the author ensured full anonymity, encouraged participants to be as honest as possible and took great care to look away when the questionnaires were completed, the experiment setup could still have caused participants to provide more positive responses than they would in a less obtrusive setup.
- The author did not do anything to hide that the application was his own work. While participants were not directly informed that the application was developed by the author, this was most likely the impression participants had during the trials.
- The application occassionally crashed during the trials, which could indeed be a source of frustration for the users. The experiment participants might well have ignored those errors when they filled in the questionnaires.

With these important points mentioned, the results should certainly not be interpreted as invalid. The results overall go well with the observations and participants' comments off the record. There is a noticeable difference between the participant's subjective judgements of the usability for the AR view and the list/map views, which is in correlation with the observations. Further, some participants provided particularly constructive feedback, for instance when suggesting improvements of the filter mechanism, pointing out usability issues of the AR view as well as the usage patterns of the list view, where some participants blindly followed the arrows and ignored the distance indicators.

14.2 Low fidelity stories

It is easy to perceive the stories used for the experiment as artificial. Stories were recorded in a low fidelity format, and it was obvious that the stories were not "real" interviews. Once the ap-

plication is completed with all stories from Christchurch, an experiment here could reveal different results. Generally, though, participants did seem to understand that the application was a prototype, and seemed to have a good idea of the application's potential in a more polished version.

Once the application is updated with improvements from the experiments, another experiment should be conducted in Christchurch, letting users interact more freely with the application to gain an in-depth understanding of the user experience of the application. To this extent, a very important aspect will be the audio experience, which has not been tested in the field trials of this thesis. The only aspect on the audio experience revealed in the experiment, was that participants could easily recite the speak occurring when pointing at an icon, even though the environment was noisy. As previously mentioned, the use of spatial audio is not so relevant in the current implementation, as the user cannot perceive left-right and up-down differences because of the point-based interaction method. Chapter 16 briefly discusses ideas for improving and evaluating the audio experience.

While in chapter 8 it was put forward that the application should target elder users, whether the application has succeeded in this sub-goal remains to be tested. While it has been ensured that touchable UI components are within recommended sizes and some of the interactive features are expected to appeal an elder audience, evaluating the application on elder users is vital before releasing the application.

15. Conclusion

This thesis has described the development of a mobile audio application on urban story telling, called memoAR. This application is intended for use in Christchurch, New Zealand, as part of the High Street Stories project, which is run by New Zealand Historic Places Trust in collaboration with HITLabNZ.

An Android application has been developed, which integrates augmented reality and spatial audio techniques to present stories tagged to buildings and areas in the physical environment. While the city of Christchurch is the primary target for the application, memoAR is adaptable to other cities' contexts, where events have created a potential for urban story telling.

Two experiments have been conducted. A usability test has investigated the potential of point-based interaction methods and compared those against a single touch approach. The point-based methods are outperformed by the single touch method in terms of task completion time and errors, but are still assumed to possess advantages for more complicated tasks in an audio interaction scenario. Suggestions for improving the point-based interaction methods have been registered in the usability test as well, and have since been implemented in the application.

An empirical field trial of the application in use has also been conducted in two urban scenarios: A busy and noisy setting in Aalborg as well as a more rural and quiet setting in the village of Sindal. Through observations combined with post-task and post-session questionnaires, the experiment has shown that the application is highly usable, both at a task-level and in users' general attitude towards the application. memoAR has received a score of 79.75 on John Brooke's System Usability Scale, which is an indication of a good and usable product. Further, the post-task questionnaires have revealed that participants felt they had solved most tasks in the application with few errors and within a reasonable amount of time. This is in correlation with the observations from the experiment. Finally the field trial has revealed technical bugs and interface issues, which will be implemented before the application sees its final release in Christchurch, New Zealand.

16. Future work - integrating social networking

As stated several times throughout this thesis, memoAR is part of a long-term goal of integrating mobile AR and social networking and media. This chapter seeks to elaborate on this goal and present ideas that expand the audio experience to let users generate their own content - city stories - and share these stories with each other directly within an urban environment.

16.1 The problems of privacy: the interface is the issue - not the content

In chapter 2, Matthew Buckland's vision of an integration of AR and social media was presented. His vision relies on computer vision techniques (for instance facial recognition), great camera resolution and accurate GPS, to scan friends and strangers on the street using the camera of a smartphone, immediately revealing personal details and social networking updates. This vision has triggered a lot of debate, with debatters calling the vision thrilling and exciting, but also scary and terrifying because of the obvious privacy issues. Who would want a stranger to "scan" them on the street with a smartphone to gain access to their personal details? What is interesting is that AR, in this context, is only an interface to reveal information that is already easy to find otherwise. Everyone can find personal details about a stranger through a Google search, which would immediately reveal details from the person's Facebook and Twitter accounts, as well as job information, family ties etc.

On the social networks, users can decide for themselves which content they want to share with friends and which content they want to share with the public, and an AR interface would not change this. So when people talk about Buckland's vision as scary and terrifying, they are not referring to the revealing of personal content, but the thought of being scanned by a stranger with a smartphone. This behaviour is unacceptable from a social perspective, and it is unlikely that the public attitude towards this obtrusion of privacy will change in the near future. Should AR and social networking merge together naturally and reach a broad consumer market, the idea of using AR to scan a person is not feasible. To this extent, the context of urban stories possesses interesting ideas for overcoming the privacy issues of social media and AR.

16.2 Expanding urban stories - sociAR

Urban stories provide an interesting context for an eventual integration of social networking, media and AR. The main feature of memoAR is that the content is tagged to locations and buildings, and not to the persons who create it. Therefore, the first step towards an integration of social networking could be the implementation of an in-situ authoring tool to let people generate their own urban stories and tag them to locations in the cityscape. These stories would contain photos, videos and, not least, audio recordings generated by a user "on the go". The user decides whether to share a city story with his/her friends or the public, similar to the options available on Facebook. When a user of sociAR aims at a building from an AR view, he/she can browse the stories generated by his/her friends tagged to the building.

Information visualisation is an important aspect of such a system, especially when a building has many stories attached to it (imagine the Eiffel Tower). How does the user decide what content to browse,

and from which medium? How does the user filter friends' content from content generated by the public? A popular interface feature of touch-based systems is the "de-cluster" effect, where overlapping content is pulled away from each other, when they are selected. A building with user-generated content associated could be represented by a blue icon on the screen. When the icon is pointed at, new icons representing the various types of media de-cluster from the building's icon (Figure 48A). Tapping the audio icon takes the user to hear the latest audio file left at this position. The user can engage socially in the story by providing replies in the shape of audio, video or photo captures, or simply leave a text message to the story (Figure 48B).

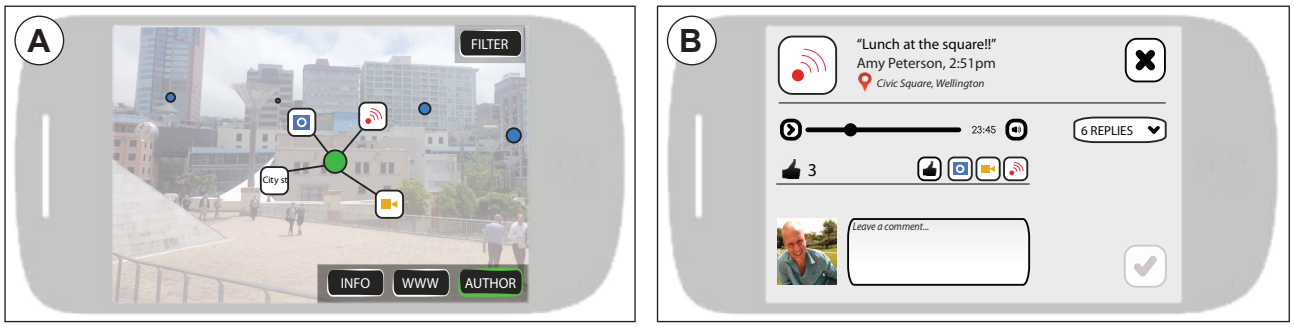


Figure 48: Expanding the audio tour to integrate user-generated stories. **A**: When aiming at a building, icons representing different media de-cluster from the building's standard icon (a blue or green dot). Notice the "Author" button in the UI hud, which lets users create their own city stories. **B**: Touching one of the media icons plays back the latest city story of this medium and lets the user socially engage by responding to the story through audio, video or photo captures.

The advantage of an AR interface over a traditional interface is that the user can socialise and enjoy the stories of friends in situ. This provides an opportunity for users to become more aware of their immediate environment, and opens for direct, social engagement in the city.

16.3 Engaging social interaction

On social networking and media services users either share content such as images, news and slides, or articulate personal relationships. Many social networking services let users socialise by commenting and replying on the user-generated content and create special interest groups with users with similar interests. In the context of urban stories, it makes sense to provide ways for users to interact through comments and replies to content from friends. These replies could be in the shape of audio, video or photo captures, similar to YouTube's video response system, but here associated with a physical location in the cityscape.

Instead of de-clustering the icons representing the different media, another approach to engage social interaction is to display a notification list, when a building is scanned. This list shows the most recent user-generated content tagged to this location and continuously updates itself to show the most recent social media updates from this location (Figure 49A). Tapping a story plays back its content and provides the opportunity to comment and reply on the story. The "Author" screen provides a crisp interface, focusing on capturing media and the ability to pinpoint the exact location of the story (Figure 49B).

Figure 49: An alternative approach to engage social interaction. **A**: Pointing at a building depicts the latest media updates from friends, tagged to this location. **B**: Tapping the "Author" button on the UI allows a user to create his/her own city story. The location can be exactly pinpointed, and photo, audio or video captures can be attached to the story.

16.4 Combining pre- and user-generated stories

Is is possible to imagine a system where pre- and user-generated stories merge into one, coherent user experience, which delivers information about the interesting locations of a city? A social tourist guide, where official stories from the city's council inform about the must-see destinations, and the user can explore the destinations in depth through user-generated stories, reviews and recommendations. Earlier in this report, the aim-and-touch based interaction method was described, and it was emphasised that a good incitement to take this multitasking approach was necessary. Combining pre- and user-generated stories as an art tourist guide might be a context, where this interaction method could be particularly useful. In this idea, a visitor to a city scans the cityscape in the AR view. Blue icons represent the most known buildings or areas in the city, where the city's council has left official stories. Aiming at one of these buildings or areas plays back the

official story (Figure 50A). While listening to the story, the user can touch the icon (which has now turned green, indicating that this building's story is now in focus). This makes media icons representing different types of user-generated content decluster from the green icon (Figure 50B). Tapping one of these icons depicts the latest UGC of this media type in a list format (Figure 50C). This screen also provides a short-cut for the user to generate his/her own content of this particular medium, for instance a camera capture.

Figure 50: Combining pre- and user-generated stories. **A**: Aiming at a blue icon tagged to a building plays back its "official" story - a story generated by the city's council. A green icon represents a source that is currently playing. **B**: When a scene is in focus, media icons representing user-generated stories de-cluster from the green icon. **C**: Tapping the Photo icon depicts a list of user-generated photos left at this location. The user can capture a photo and generate his/her own city story.

16.5 AR 2.0: In-situ authoring

The authoring tool will be a very important aspect of a future integration of mobile AR and social networking where millions of users not only view, but also create and share augmented content. First of all, it would be appealing to provide an easy way to generate stories on the spot from the mobile UI. Many authoring tools for AR content exist, usually as web services or desktop-based creator interfaces. While these systems allow users to create AR content without great programming efforts, they do not target smartphone users, which is expected to be the main interface to AR content in the future, where "every consumer of AR is potentially also an author" (Langlotz et al. 2012a). As an example, BirdsView[28] of-

28 http://www.birdsview.de

fers a desktop creator to author additional information to real-world locations, and publish the content to AR browsers, such as junaio or Layar, but the user cannot create content from the mobile device.

Some researchers have investigated in-situ authoring tools directly from mobile devices. Guven et al. (2006) present a system where users freeze the frame from a smartphone's camera view and create textual annotations to overlay objects in the frame. The annotations are then interpolated to their locations in physical space when the frame is defrozen, making the annotations appear fixed in space. For precise depth positioning, authors can use an additional rotating input device, to reel an object along a ray on a virtual Z axis. Langlotz et al. (2012a) also present a system for in situ authoring, modifying and manipulation of augmented 2D and 3D models directly from a mobile device. They further present the idea of Augmented Reality 2.0, where millions of users will share a massive amount of augmented content. In this context, the authors present thoughts on how to share the content with other users. In AR 2.0 the smartphone is an important device, because of its low cost and AR capabilities,

The task of creating a useful authoring tool for sociAR might not be as difficult as the systems mentioned above. The goal is not to provide tools for authoring complex 3D models or objects, but to capture photos, video or audio and link these captures to a physical location. The content will be represented by simple, pre-generated icons, which the user does not need to worry about. For sociAR, it might be enough to provide a decent interface for capturing the media, defining which persons to share the content with, and pinpoint an exact location, which could be achieved through a combination of an AR view for approximate location, and a map view for pinpointing the exact location.

16.6 Stand-alone system or Facebook app?

The implementation of sociAR first and foremost relies on a well developed plan for publishing the content to all target users and devices as well as accessing the users' social networks. One option is to implement the application as a Facebook plugin, which would provide access to people's social networks and potentially mobilise a large number of users. This would (if successful on Facebook's AppCenter) make the experience more interesting, because more users would make the social aspect more immersive and engaging. The downside of this approach is that the application would need to be implemented as a web application in languages such as PHP, Python or C#, which would require a total reimplementation of the app. The advantage of keeping the application as a standalone system is more freedom in relation to design, layout and the overall user experience. It would however require more thoughts regarding the handling of user profiles, opportunities for adjusting personal settings and filtering the content to target personal friends or the public.

16.7 The near future part 1: Improving and evaluating the audio experience

The big question that remains after the experiments is whether the interaction method justifies the use of spatial sound techniques. Having to point at icons to activate their sound sources clearly reduces the effect of the spatialised audio. Dependent on the number of sound sources in the environment,

using proximity-based interaction and tracking both the user's location and orientation to locate the sound sources more precisely in the user's headphones, might provide a more convincing audio experience. Like in the Soho Stories application, a sound source would fade in and out as the user enters or leaves the sound's proximity zone. Sounds are positioned in the ear channels as in Vazquez-Alvarez and Brewster's sound gardens. As a supplement to the audio experience, one could imagine the use of visual cues in the AR view to indicate a sound that is currently playing. The user does not need to point at the sound sources to hear their audio, but can use the AR view to scan the environment and get a visual indication of where the sound sources come from. In this way, the visual interface might provide a better supplement to the audio experience than is the case in the current version of the application.

16.8 The near future part 2: Binding the stories together

In its current state, the memoAR application (at least in the High Street Stories implementation) does not provide any narrative to guide the user to other stories in the application. While it is appealing to let the user move around freely and explore the content on his/her own tempo, a linear narrative, which binds the stories together, has some advantages as well. The narrative would fully depend on the overall goal of the story context. Since the application is intended to be meaningful for many different urban contexts, some cities would benefit from a linear story to guide the user in the environment. One way to make this alternative interesting is to make the application resemble a treasure hunt, with an overall story unveiling as the user progresses. When a user has reached the first point of the story (its activation zone), he/she gets access to the content of the story, and a speak guides the user towards the next point. By checking the user's location, audio cues would suggest the user to "turn around", "go back 10 meters" or "turn right after 10 meters".

16.9 Chapter summary

This final chapter of the thesis has sought to put the High Street Stories/memoAR application into perspective by presenting plans for integrating social networking. In the context of city stories, it is only natural to think of ways to socially engage users by providing opportunities for them to generate their own content - actively - instead of simply enjoying pre-generated content - passively. The overall goal of future mobile experiences is likely to increase users' awareness of their surrounding environment and let them interact with friends as well as strangers in relation to locations and events of their city. Having developed a static mobile tour experience on city stories, this next step is obvious to widen the social scope of an audio tour in particular, as well as mobile experiences in general.

References

AalborgKommune (2010): New Harbourfronts - Aalborg and Nørresundby. Accessed March 13, 2013. http://www.aalborgkommune.dk/Om_kommunen/Byplanlaegning/Havnefront/Documents/folder_engelsk_tryk.pdf

AnsonAlex (2012): Smartphone Usage Statistics 2012. Accessed March 15, 2013. http://ansonalex.com/infographics/smartphone-usage-statistics-2012-infographic/

Aoki, P. M., Grinter, R. E., Hurst, A., Szymanski, M. H., Thornton, J. D., & Woodruff, A. (2002). Sotto voce: exploring the interplay of conversation and mobile audio spaces. In Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves (pp. 431-438). ACM.

Azuma, R. T. (1997). A survey of augmented reality. Presence-Teleoperators and Virtual Environments, 6(4), 355–385.

Azuma, R., Weon Lee, J., Jiang, B., Park, J., You, S., & Neumann, U. (1999). Tracking in unprepared environments for augmented reality systems. Computers & Graphics, 23(6), 787–793.

Bai, H., Lee, G. A., & Billinghurst, M. (2012). Freeze view touch and finger gesture based interaction methods for handheld augmented reality interfaces. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand (pp. 126-131). ACM.

Ballagas, R., Rohs, M., & Sheridan, J. G. (2005). Mobile phones as pointing devices. In Pervasive 2005 Workshop on Pervasive Mobile Interaction Devices (PERMID 2005).

Ballagas, R., Borchers, J., Rohs, M., & Sheridan, J. G. (2006). The smart phone: a ubiquitous input device. Pervasive Computing, IEEE, 5(1), 70-77.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction, 24(6), 574-594.

Bederson, B. B. (1995). Audio augmented reality: a prototype automated tour guide. Conference companion on Human factors in computing systems. ACM, 1995, 210-211.

Billinghurst, M., Poupyrev, I., Kato, H., & May, R. (2000). Mixing realities in shared space: An augmented reality interface for collaborative computing. In Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on (Vol. 3, pp. 1641-1644). IEEE.

Billinghurst, M., Kato, H., & Myojin, S. (2009). Advanced interaction techniques for augmented reality applications. Virtual and Mixed Reality, 13–22.

Billinghurst, M. (2012). AR Interaction [PowerPoint slides]. Retrieved from http://www.slideshare.net/ marknb00/426-lecture6b-ar-interaction

Blippar (2011). Heinz tomato ketchup recipe book [Video screenshot]. Accessed May 1, 2013. http://www.youtube.com/watch?v=GbplSdh0lGU

Bolt, R. A. (1980). "Put-that-there": Voice and gesture at the graphics interface (Vol. 14, No. 3, pp. 262-270). ACM.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. Personal and Ubiquitous Computing, 6(3), 188-205.

Brewster, S., Chohan, F., & Brown, L. (2007). Tactile feedback for mobile interactions. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 159-162). ACM.

Brooke, J. (1996). SUS-A quick and dirty usability scale. Usability evaluation in industry, 189, 194.

Buckland, M. (2010). The future of social networking - a concept investigation with augmented reality. Accessed March 7, 2013. http://matthewbuckland.com/?p=1041.

CCDU (2012): The Recovery Plan. Accessed April 5, 2013. http://ccdu.govt.nz/the-plan

Chow, C. Y., Bao, J., & Mokbel, M. F. (2010). Towards location-based social networking services. Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, 31-38.

Clark, A., & Piumsomboon, T. (2011). A realistic augmented reality racing game using a depth-sensing camera. In Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry (pp. 499-502). ACM.

Cnet (2010). Reporters' Roundtable: Checking in with Facebook and Foursquare. Accessed March 7, 2013. http://www.cnet.com/8301-30976_1-20015615-10348864.html

Cohen, P. R. (1992). The role of natural language in a multimodal interface. In Proceedings of the 5th annual ACM symposium on User interface software and technology (pp. 143-149). ACM.

Costanza, E., Inverso, S. A., & Allen, R. (2005). Toward subtle intimate interfaces for mobile devices using an EMG controller. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 481-489). ACM.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. Management Science, 32(5), 554–571.

Dorahy, M. J., & Kannis-Dymand, L. (2013). Psychological Distress Following the 2010 Christchurch Earthquake: A Community Assessment of Two Differentially Affected Suburbs. Journal of Loss and Trauma, 17(3), 203–217.

Dow, S., Lee, J., Oezbek, C., MacIntyre, B., Bolter, J. D., & Gandy, M. (2005). Exploring spatial narratives and mixed reality experiences in Oakland Cemetery. Proceedings of the 2005 ACM SIGCHI International Conference, 51-60.

Dunlop, M., & Brewster, S. (2002). The challenge of mobile devices for human computer interaction. Personal and ubiquitous computing, 6(4), 235-236.

Dünser, A., & Billinghurst, M. (2011). Evaluating Augmented Reality Systems. Handbook of Augmented Reality, 289-307.

Eden, T. (2010). Augmented Reality Games. How Far Have We Come in 7 Years? Accessed May 1, 2013. http://shkspr.mobi/blog/2010/11/augmented-reality-games-how-far-have-we-come-in-7-years/

Feiner, S., MacIntyre, B., Hollerer, T., & Webster, A. (1997). A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. Personal and Ubiquitous Computing, 1(4), 208–217.

Fitzmaurice, G. W. (1993). Situated information spaces and spatially aware palmtop computers. Communications of the ACM, 36(7), 38-49.

Freni, D., Ruiz Vicente, C., Mascetti, S., Bettini, C., & Jensen, C. S. (2010). Preserving location and absence privacy in geo-social networks. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 309-318). ACM.

FreshAndNew (2011). Experiencing The O at MONA. Accessed April 29, 2013. http://www.freshandnew.org/2011/10/experiencing-the-o-at-mona-a-review/

FutureGov (2012). CityViewAR [Digital photograph]. Accessed May 1, 2013. http://www.futuregov.asia/media/photologue/photos/cache/cityviewar_gallery_display.jpg

Galland, D. (2011). Waterfront redevelopment processes in Aalborg, Denmark: Making sense of different planning logics. In AAG Annual Meeting.

GapFiller (2013). Images from GapFiller events [Digital photographs]. Accessed May 1, 2013. http://www.gapfiller.org.nz/

Geekosystem (2013). Inside Ingress, Google's Augmented Reality Android Game. Accessed April 26, 2013. http://www.geekosystem.com/google-ingress-review/

Gervautz, M., & Schmalstieg, D. (2012). Anywhere Interfaces Using Handheld Augmented Reality. Computer, 45(7), 26-31.

Goffman, E. (2002). The presentation of self in everyday life. 1959. Garden City, NY.

Google (2013a). Android, the world's most popular mobile platform. Accessed April 3, 2013. http://developer.android.com/about/index.html

Google (2013b). Metrics and Grids. Accessed April 10, 2013. http://developer.android.com/design/style/metrics-grids.html

Google (2013c). Activity class documentation. Accessed May 17, 2013. http://developer.android.com/reference/android/app/Activity.html

Guven, S., Feiner, S., & Oda, O. (2006). Mobile augmented reality interaction techniques for authoring situated media on-site. In Mixed and Augmented Reality, 2006. ISMAR 2006. IEEE/ACM International Symposium on (pp. 235-236). IEEE.

Haller, M., Dobler, D., & Stampfl, P. (2002). Augmenting the reality with 3D sound sources. In ACM SIGGRAPH 2002 conference abstracts and applications (pp. 65-65). ACM.

Härmä, A., Jakka, J., Tikander, M., Karjalainen, M., Lokki, T., Hiipakka, J., & Lorho, G. (2004). Augmented reality audio for mobile and wearable appliances. Journal of the Audio Engineering Society, 52(6), 618-639.

Henrysson, A., Billinghurst, M., & Ollila, M. (2005). Face to face collaborative AR on mobile phones. In Mixed and Augmented Reality, 2005. Proceedings. Fourth IEEE and ACM International Symposium on (pp. 80-89). IEEE.

Höllerer, T., Feiner, S., & Pavlik, J. (1999). Situated documentaries: Embedding multimedia presentations in the real world. In Wearable Computers, 1999. Digest of Papers. The Third International Symposium on (pp. 79-86). IEEE.

Höllerer, T., & Feiner, S. (2004). Mobile augmented reality. Telegeoinformatics: Location-Based Computing and Services. Taylor and Francis Books Ltd., London, UK.

Inhabitat (2011): Re:START Shipping Container Mall Opens In Christchurch But Faces Lawsuit. Accessed March 7, 2013. http://inhabitat.com/restart-shipping-container-mall-opens-in-christchurch-but-faces-lawsuit/cashel-mall-4/

Iperg (2008): Interference game [Digital photograph]. Accessed May 1, 2013. http://iperg.sics.se/images/gallery_new/int6.jpg

Irawati, S., Green, S., Billinghurst, M., Duenser, A., & Ko, H. (2006). An evaluation of an augmented reality multimodal interface using speech and paddle gestures. Advances in Artificial Reality and Tele-Existence, 272-283.

Iwata, T., Yamabe, T., & Nakajima, T. (2010). Towards a Mobility Enhanced User Interface Design for Multi-task Environments: An Experimental Study on Cognitive Workload Measurement. In Proceedings of the 2010 Sixth International Conference on Intelligent Environments (IE '10). IEEE Computer Society, Washington, DC, USA, 106-111.

Junaio (2013). Junaio platform [Digital photograph]. Accessed May 1, 2013. *h*ttp://www.junaio.com/fileadmin/upload/images/IMG-header/HEAD-download_junaio.png

Kato, H., & Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on (pp. 85-94). IEEE.

Kato, H., Billinghurst, M., Poupyrev, I., Imamoto, K., & Tachibana, K. (2000). Virtual object manipulation on a table-top AR environment. In Augmented Reality, 2000.(ISAR 2000). Proceedings. IEEE and ACM International Symposium on (pp. 111-119). IEEE.

Lampe, C., Ellison, N., & Steinfield, C. (2006). A Face (book) in the crowd: Social searching vs. social browsing. In Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work (pp. 167-170). ACM.

Langlotz, T., Mooslechner, S., Zollmann, S., Degendorfer, C., Reitmayr, G., & Schmalstieg, D. (2012a). Sketching up the world: in situ authoring for mobile augmented reality. Personal and ubiquitous computing, 16(6), 623-630.

Langlotz, T., Regenbrecht, H., Zollmann, S., & Schmalstieg, D. (2012b): Audio Stickies: Visually-guided Spatial Audio Annotations on a Mobile Augmented Reality Platform. Unpublished.

Layar (2013). Layar platform [Digital photograph]. Accessed May 1 2013. http://www.layar.com

Lee, G. A., Yang, U., Kim, Y., Jo, D., Kim, K. H., Kim, J. H., & Choi, J. S. (2009). Freeze-Set-Go interaction method for handheld mobile augmented reality environments. In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology (pp. 143-146). ACM.

Lee, G. A., & Billinghurst, M. (2012). Assistive techniques for precise touch interaction in handheld augmented reality environments. In Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry (pp. 279-285). ACM.

Lemmelä, S., Vetek, A., Mäkelä, K., & Trendafilov, D. (2008). Designing and evaluating multimodal interaction for mobile contexts. In Proceedings of the 10th international conference on Multimodal interfaces (pp. 265-272). ACM.

Lerman, K. (2007). Social browsing & information filtering in social media. arXiv preprint arXiv:0710.5697.

Lindeman, R. W., Noma, H., & de Barros, P. G. (2007). Hear-through and mic-through augmented reality: Using bone conduction to display spatialized audio. In Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (pp. 1-4). IEEE Computer Society.

Lindeman, R., & Lee, G. (2012). GeoBoids: mobile AR for exergaming. In Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction (pp. 100-100). ACM.

LonelyPlanet (2012). Best in Travel 2013 - Top 10 cities. Accessed March 13, 2013. http://www.lonelyplanet.com/themes/best-in-travel-2013/top-10-cities/

Mashable (2012a). Smartphone Sales Overtake PCs for the First Time. Accessed March 7, 2013. http://mashable.com/2012/02/03/smartphone-sales-overtake-pcs/

Mashable (2012b). New Facebook for iOS: Same Coat of Paint, Better Engine. Accessed March 7, 2013. http://mashable.com/2012/08/23/facebook-ios-5-hands-on/

Microsoft (2013). Geo-Life [Digital photograph]. Accessed May 1, 2013. http://research.microsoft.com/en-us/people/yuzheng/projects.aspx

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems, 77(12), 1321-1329.

Mellor, J. P. (1995). Enhanced Reality Visualization in a Surgical Environment (No. MIT-AITR-1544). MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB.

Mona (2013). The O [Digital photograph]. Accessed May 1 2013. http://www.mona.net.au/theo/

Montola, M. (2005). Exploring the edge of the magic circle: Defining pervasive games. In Proceedings of DAC (p. 103).

Mulloni, A., Dünser, A., & Schmalstieg, D. (2010). Zooming interfaces for augmented reality browsers. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 161-170). ACM.

MutualFriend (2008). Hitchers screenshot [Digital photograph]. Accessed May 1, 2013. http://www.amutualfriend.co.uk/html/hitchers_about.html

Nielsen, L (2013). Personas. In: Soegaard, Mads and Dam, Rikke Friis (eds.). "The Encyclopedia of Human-Computer Interaction, 2nd Ed.". Aarhus, Denmark: The Interaction Design Foundation. Available online at http://www.interaction-design.org/encyclopedia/personas.html

Nil (2013). Applications of Location-based Services. Accessed May 7, 2013. http://stack.nil.si/ip-corner/LocationBasedServices/#chapter2

Nokia (2012). Nokia City Lens screenshots [Digital photographs]. Accessed May 1, 2013. http://betalabs.nokia.com/trials/nokia-city-lens-for-windows-phone

NZHerald (2012). How (and where) will Christchurch be rebuilt? CERA boss answers readers' questions. Accessed March 7, 2013. http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10786562

Paradox (2011). AR Chemistry [Video screenshot]. Accessed May 1, 2013. http://www.youtube.com/watch?v=IpNrWKQFq6Q

Piekarski, W., & Thomas, B. (2002). ARQuake: the outdoor augmented reality gaming system. Communications of the ACM, 45(1), 36-38.

Piumsomboon, T., Clark, A., & Billinghurst, M. (2011). Physically-based interaction for tabletop augmented reality using a depth-sensing camera for environment mapping. In Proceedings of the 26th International Conference on Image and Vision Computing New Zealand.

PlaybackTheatre (2011). ChristChurch Cathedral [Digital photograph]. Accessed May 1, 2013. http://www.playbacktheatre.org/wp-content/uploads/2011/04/Christchurch-cathedral-after-the-earthquake.jpg

Plos, O., & Buisine, S. (2006). Universal design for mobile phones: a case study. In CHI'06 extended abstracts on Human factors in computing systems (pp. 1229-1234). ACM.

Potter, R. L., Weldon, L. J., & Shneiderman, B. (1988). Improving the accuracy of touch screens: an experimental evaluation of three strategies. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 27-32). ACM.

Pulkki, V., & Karjalainen, M. (2001). Localization of amplitude-panned virtual sources I: stereophonic panning. Journal of the Audio Engineering Society, 49(9), 739-752.

Qualcomm (2012). Augmented Reality - Marketing Footwear [Video screenshot]. Accessed May 1, 2013. http://www.youtube.com/watch?v=kLIX5rP_6n0

Reichenbacher, T. (2003). Adaptive methods for mobile cartography. In Proceedings of the 21st International Cartographic Conference ICC: Cartographic Renaissance (pp. 1311-1321).

Rekimoto, J., & Nagao, K. (1995). The world through the computer: Computer augmented interaction with real world environments. In Proceedings of the 8th annual ACM symposium on User interface and software technology (pp. 29-36). ACM.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1994). A comparison of optical and video see-through head-mounted displays. In SPIE Telemanipulator and Telepresence Technologies (Vol. 2351, pp. 293-307).

Rozier, J. M. (2000). Hear&There: An augmented reality system of linked audio (Doctoral dissertation, Massachusetts Institute of Technology).

RVNews (2012). Wikitude AR browser [Digital photograph]. Accessed May 1, 2013. http://rvnews-net.com/2012/04/10/app-the-reality-of-wikitude/

Sauro, J. (2011). Measuring Usability With The System Usability Scale (SUS). Accessed May 12, 2013. http://www.measuringusability.com/sus.php

Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications.

Socyberty (2011). The Elderly Smartphone Users. Accessed March 15, 2013. http://socyberty.com/people/the-elderly-smartphone-users/

Sutherland, I. E. (1968). A head-mounted three dimensional display. In Proceedings of the December 9-11, 1968, fall joint computer conference, part I (pp. 757-764). ACM.

Tamminen, S., Oulasvirta, A., Toiskallio, K., & Kankainen, A. (2004). Understanding mobile contexts. Personal and Ubiquitous Computing, 8(2), 135-143.

Thomas, B., Demczuk, V., Piekarski, W., Hepworth, D., & Gunther, B. (1998). A wearable computer system with augmented reality to support terrestrial navigation. In Wearable Computers, 1998. Digest of Papers. Second International Symposium on (pp. 168-171). IEEE.

Tullis, T., & Albert, W. (2010). Measuring the user experience: collecting, analyzing, and presenting usability metrics. Morgan Kaufmann.

Twitter (2013). Twitter interface [Digital photograph]. Accessed May 1 2013. http://cdn.iphone-news.org/wp-content/uploads/2009/12/twitter-360-iphone-augmented-reality.jpg

Twnkls (2013). SocializAR [Digital photograph]. Accessed May 1 2013. http://twnkls.com/social-izar/

Van Verth, J. M., & Bishop, L. M. (2008). Essential Mathematics for Games & Interactive Applications: A Programmer's Guide. Morgan Kaufmann Publishers.

Vazquez-Alvarez, Y. (2010). Designing spatial audio interfaces for mobile devices: supporting multitasking and context information. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 481-482). ACM.

Vazquez Alvarez, Y., & Brewster, S. A. (2010). Designing spatial audio interfaces to support multiple audio streams. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (pp. 253-256). ACM.

Vazquez-Alvarez, Y., Oakley, I., & Brewster, S. A. (2012). Auditory display design for exploration in mobile audio-augmented reality. Personal and Ubiquitous Computing, 16(8), 987-999.

Verge (2012): Facebook for iOS goes native, waves goodbye to HTML 5. Accessed March 7, 2013. http://www.theverge.com/2012/8/23/3262782/facebook-for-ios-native-app

Wagner, D., Mulloni, A., Langlotz, T., & Schmalstieg, D. (2010). Real-time panoramic mapping and tracking on mobile phones. In Virtual Reality Conference (VR), 2010 IEEE (pp. 211-218). IEEE.

Wagner, D., & Schmalstieg, D. (2003). First steps towards handheld augmented reality. In Proceedings of the 7th International Symposium on Wearable Computers (ISWC'2003) (pp. 127-137).

Weiser, M. (1991). The computer for the 21st century. Scientific American, 265(3), 94-104.

WebProNews (2013). Report: Google Sued Again Over Google Buzz Privacy. Accessed March 7, 2013. http://www.webpronews.com/report-google-sued-again-over-google-buzz-privacy-2013-01

Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., & Shen, C. (2007). Lucid touch: a see-through mobile device. In Proceedings of the 20th annual ACM symposium on User interface software and technology (pp. 269-278). ACM.

Wikitude (2013). Wikitude platform [Digital photograph]. Accessed May 1 2013. http://www.wikitude.com

You, S., Neumann, U., & Azuma, R. (1999). Hybrid inertial and vision tracking for augmented reality registration. In Virtual Reality, 1999. Proceedings., IEEE (pp. 260-267). IEEE.

Zhai, S., & Kristensson, P. O. (2003). Shorthand writing on stylus keyboard. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 97-104). ACM.

Zheng, Y., Xie, X., & Ma, W. Y. (2010). GeoLife: A collaborative social networking service among user, location and trajectory. IEEE Data Engineering Bulletin, 33(2), 32-40.

3News (2013). App shows Chch pre- and post Quake [Video screenshot]. Accessed May 1, 2013. http://www.3news.co.nz/App-shows-Chch-pre--and-post-quake/tabid/309/articleID/289706/Default.aspx

PART 4: APPENDICES

A. Augmented Reality

This appendix walks through theory behind augmented reality technology, including the fundamental challenges involved in creating impressive AR systems, different application areas, as well as a section on mobile and wearable AR.

A1. Background

In its essence, augmented reality is the superimposition of computer-generated contene onto the real world, to create the impression that the real and virtual world coexist (Azuma 1997) (Figure 51). It differs from the concept of virtual reality in that it seeks to enhance the real world instead of fully replacing it. AR can be placed near the middle of Milgram and Kishino's real-virtual continuum (Milgram & Kishino 1994) (Figure 52).

Figure 51: Creating the impression that the real and virtual world coexist is the fundamental challenge of augmented reality. In this example, the table is real and the lamp and chairs are virtual. (Azuma 1997)

Figure 52: Milgram's reality-virtuality continuum. AR can be placed near the centre of the figure, since it enhances the user's perception of the real world instead of fully replacing the real world (Milgram & Kishino 1994).

Since early pioneering work by Ivan Sutherland and his students (Sutherland 1968) using a head-mounted display and a mechanical tracker mounted to the roof, researchers have been attracted by the idea of using computer graphics to enhance people's perception of the real world. Early application areas include visualisating hidden brain structures in surgery operations (Mellor 1995), annotations of objects in the real world (Rekimoto & Nagao 1995), and later on, entertainment and gaming contexts as well, for instance ARQuake (Piekarski & Thomas 2002) and ARTennis (Henrysson et al. 2005). Today, with smartphones and AR browsers, AR has started to gain wider popularity and is moving into larger consumer markets (Van Krevelen & Poulman 2010). Figure 53 depicts four recent, AR systems for educative and marketing purposes.

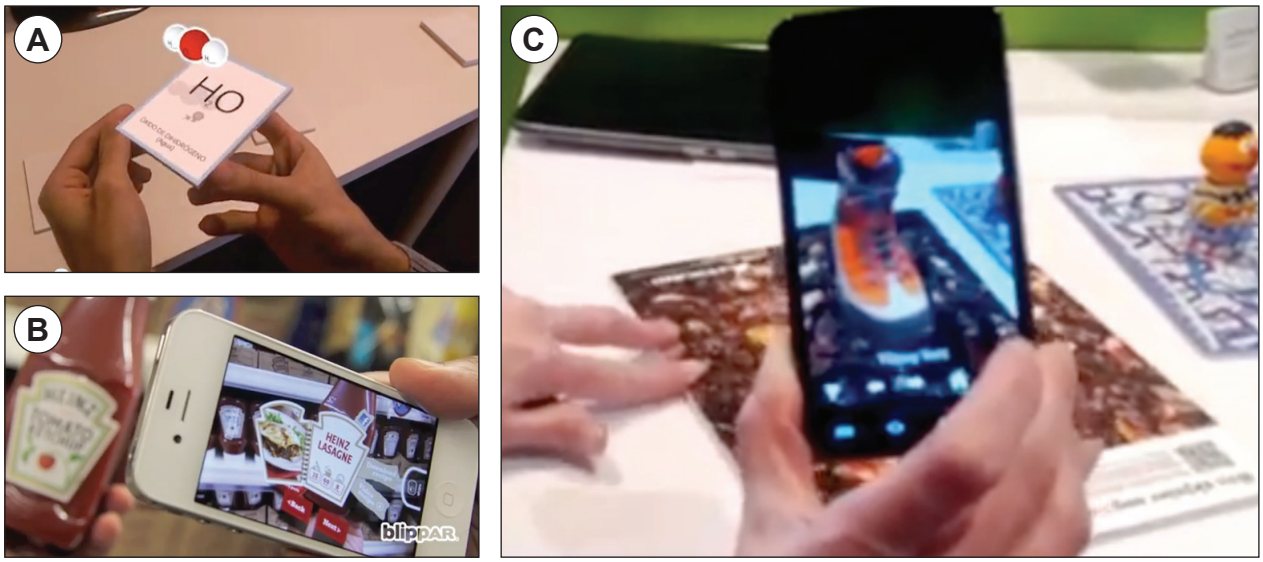


Figure 53: AR applications serving three different purposes: **A**: AR chemistry for educative purposes (Paradox 2011). **B**: Heinz Tomato Ketchup recipe book appearing when scanning a bottle of ketchup with a smartphone (Blippar 2011). **C**: Qualcomm demonstration of AR for a magazine advertisement for footwear (Qualcomm 2012).

Three fundamental challenges exist in AR systems: Tracking, display and input technology. In the following sections, these challenges are described in depth.
.

A2. Tracking and registration

One of the key challenges in AR is how to continuously track the user's viewpoint and register the virtual objects correctly to the physical environment. Several approaches to tracking exist, including:

- Magnetic tracking
- Mechanical tracking
- Inertial tracking
- Optical tracking
- Location-based tracking (GPS)

A detailed walkthrough of all of these systems is beyond the scope of this thesis. Below, optical and

location-based tracking are described, since these are by far the most used tracking methods in AR:

A2.1 Optical tracking

Optical tracking uses vision-based techniques to track a user's viewpoint and achieve correct registration. Traditionally this was done by placing fiducials in the physical environment, and using image processing techniques to detect their location. An early example of this approach can be seen in Mellor (1995). In 1999, ARToolKit was invented (Kato & Billinghurst 1999). This is a software library for building rapid AR prototypes, which has since become widely popular in AR communities. ARToolKit uses special black and white markers with patterns easily detectable by a computer. The system determines the camera pose in relation to the markers and superimposes 3D virtual models over the markers. Advantages of marker-based tracking is its precision, possibly down to pixel-level, and that the markers can be used for creative purposes. Some of the challenges include occlusion, motion blur and change in lighting conditions. An alternative is to use marker-less tracking, where natural features in an image, such as edges and corners, are tracked and compared against a database of keypoints. This approach is slower and more difficult than marker-based tracking, but removes the need to develop specialised (and artificial) patterns, turning any 2D texture into a potential marker.

A2.2 Location-based tracking

Aligning the real and the virtual world is the main goal of AR systems. Marker-based AR systems usually achieve pixel-accurate registration, but can be considered as "prepared" environments, which is useful in indoor, controlled settings (Azuma et al. 1999). In larger, outdoor environments, vision-based approaches are often not feasible, since it is much more difficult to prepare such an environment in advance. The most common approach in outdoor AR is to use GPS technology to track a user's location in the environment, and combine this with inertial sensors, for instance compasses and gyroscopes, to achieve orientation data. This provides a tracking system with 6 degrees of freedom: position in X-Y-Z and orientation in pitch, roll and yaw angles (Höllerer & Feiner 2004). GPS is, however, only accurate to 10-30 meters and dependent on signals from satellites. This challenge can be addressed through assisted or differential GPS. Assisted GPS uses data and signals from external cell towers to get the precise satellite position. This approach is useful in cases where it is difficult to get a satellite fix, for instance in a city or in poor weather conditions. Differential GPS uses a network of fixed reference stations on the ground, which can improve GPS accuracy to 1-3 meters (Höllerer & Feiner 2004).

A2.3 Hybrid approaches

Some AR systems try to combine different tracking technologies to compensate for each others weaknesses and benefit from each other's strengths. Azuma et al. (1999) adds video processing methods to a system, which relies on inertial tracking (compass and gyroscope). Compared to their base system, the hybrid approach greatly reduces registration errors. You et al. (1999) present a hybrid approach, where vision-based tracking is used for correction of drift and lack of accuracy of the inertial sensors, which on the other hand compensate for the computationally expensive vision-based tracking.

A3. Display

Traditionally, most systems have applied visual displays for presenting AR experiences. The most used approaches for visual displays are optical and video see-through head-mounted displays (HMD's). These two methods are distinguishable by the way the user senses the real world, and possess their own advantages and disadvantages, described briefly below.

A3.1 Video and optical see-through head-mounted displays

In video see-through displays, the user's view of the real world is provided through cameras mounted on the head. The graphical content is composited with the input from the cameras, and sent to monitors in front of the user's eyes (Azuma 1997). In an optical see-through display, the user has a direct view of the real world, and virtual images are shown through transmissive, optical combiners bounced off from head-mounted monitors (Azuma 1997). Video see-through makes it easier to manipulate objects, since the real world is digitised, and brightness and contrast can be matched with the real world. Optical see-through displays are safer than video see-through displays, since the user can still see the real world, even when the power fails (Van Krevelen & Poulman 2010). The approach is however not so suitable for occlusion of real or virtual objects difficult, and it is not as easy to mediate objects as when using video see-through. For a detailed walkthrough and comparison of see-through head-mounted displays, see Rolland et al. (1994).

A3.2 Handheld displays

Some disadvantages of head-mounted displays exist, primarily because of limitations of head-worn technology (Van Krevelen & Poulman 2010). Further they are cumbersome and clumsy to wear, which prevents mainstream adoption. A less bulky, and more socially acceptable display alternative, is the use of handheld displays, where video see-through techniques are applied, and a small, ergonomic device works as a "magnifying glass" on the real world. This idea was first implemented on PDA´s in the mid 90's, for instance in the NaviCam project (Rekimoto & Nagao 1995). Due to improvements in mobile technology, handheld AR has now found a broad usage platform on camera-enabled phones, a development foreseen by Wagner & Schmalstieg (2003).

A3.3 Other display techniques

Bimber & Raskar (2006) present alternative, spatial approaches to display techniques, seperating the user from the display: Spatial see-through displays, where images are aligned with the physical environment, through the use of optical components such as beam splitters, transparent screens or holograms, and projective displays, where images are front-projected directly on physical objects, instead of displayed on an image plane within the user's field of view. Unlike head-mounted displays, spatial displays do not require special eye-wear and can cover large surfaces (Van Krevelen & Poulman, 2010).

A4. Interaction

Unless other sources are referenced, his section is largely based on a lecture on AR interaction by professor Mark Billinghurst (Billinghurst 2012).

Traditionally, the tracking and display technologies have gained the most attention in AR research, and not the user's interaction with the AR content (Billinghurst et al. 2009). In this section, some of the established approaches to AR interaction are presented.

A4.1 Information browsers

Many of the popular mobile applications today are information browsers, where the digital content is manipulated to enter a special information mode. This method has often been the preferred interaction style in handheld and mobile applications, with obvious examples being recent AR browsers, such as junaio, Wikitude and Layar. Earlier PDA-based systems have used input devices such as styluses to select the digital objects (Wagner & Schmalstieg 2003), while touch-based interaction is mostly used on modern mobile systems.

A4.3 3D interfaces

3D interfaces in AR contexts let users select and manipulate 3D objects in physical space through traditional 3D user interface interaction. Usually, the user performs explicit gestures, which are interpreted by the system to perform selection and manipulation of the virtual objects. Recently researchers have investigated the use of depth-sensing cameras, such as Microsoft Kinect, to create AR experiences with environment awareness and physical interaction, and which afford natural interaction with the virtual content (Piumsomboon et al. 2011). A major disadvantage of 3D interfaces is that they usually do not provide tactile feedback for the user. They can however be particularly interesting in expressive and creative contexts.

A4.2 Tangible AR

A promising direction for AR interaction is so-called Tangible AR, where AR display technology is combined with input methods from tangible user interfaces, and real objects are used to interact with virtual content. The benefits of tangible input are particularly powerful in AR contexts, where there is a close connection between the real world and the digital content (Billinghurst et al. 2009). Real objects possess physical properties and constraints, which make them easy and natural to use. Further, AR display technology can help overcome some of the limitations of tangible user interfaces, for instance the seperation between task and display space. The natural potential of Tangible AR has been explored in collaborative table-top settings, for instance the Shared Space (Billinghurst et al. 2000) and in improved systems with more complex interactions, such as the MagicCup interface (Billinghurst et al. 2009) or the VOMAR project, which interprets gesture input from a paddle-shaped object, with which the user can manipulate digital objects (Kato et al. 2000).

A4.4 Multimodal interaction

While multimodal interaction can be very powerful in virtual reality interfaces, for instance using speech for descriptive gestures and gestures for commands (Cohen 1992), relatively few AR systems have taken ad-

vantage of multimodal input. Irawati et al. (2006) expands the VOMAR project to include simple speech commands supplementing the gesture-based inputs from the paddle. Speech commands such as "Copy this" or "Grab table" are fused with interpretations of the user's gesture input with the paddle into one single command, similar to the classic multimodal "put-that-there" interface (Bolt 1980), but in an AR setting.

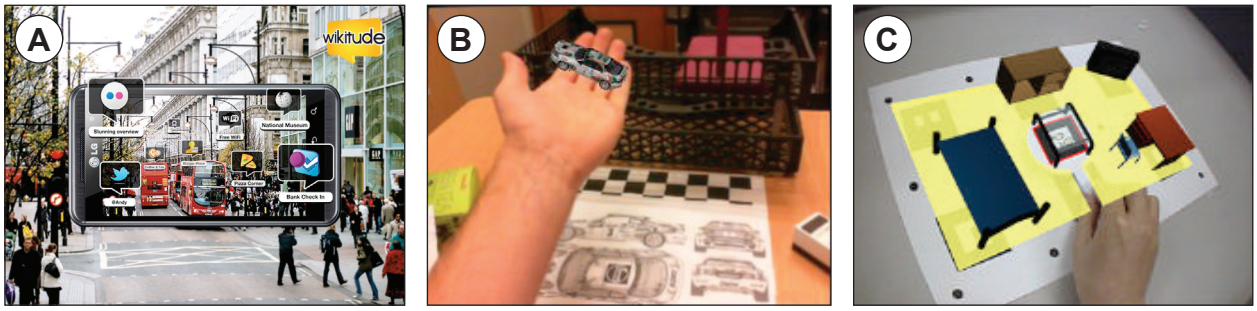


Figure 54: AR interaction methods. **A**: Mobile information browser in Wikitude (RVNews 2012). **B**: 3D interface using interpretation of a user's hand gestures (Piumsomboon et al. 2011). **C**: Tangible AR using a paddle-shaped object to manipulate digital content (Kato et al. 2000).

A5. Mobile/wearable augmented reality

Mobile and wearable AR is about providing services, which do not constrain the user's whereabouts to a specially prepared area, which is often the case for "standard" AR applications (Höllerer & Feiner 2004). They rely on combinations of a range of technologies, which until recently have not been mature enough to make them interesting from a commercial perspective. Below, a brief walktrough of the milestones in the history in mobile/wearable AR is provided.

A5.1 Wearable computing and mobile augmented reality systems

The first example of a mobile augmented reality system (MARS) was the Touring Machine (Feiner et al. 1997), which uses a see-through display with an integrated orientation tracker. The user wears a large backpack, which holds a computer and differential GPS, and a hand-held device with stylus and touchpad interface. Through the display, the user can see textual labels overlaid on campus buildings while walking at the Columbia University. The Touring Machine was used again in 1999 in an extended version, which presents "situated documentaries" - stories and multimedia presentations tagged to physical locations around the Columbia University campus.The interface guides users through a presentation containing audio, images videos and 3D graphics, and also provides users the freedom to follow their own trails through the material. (Höllerer et al. 1999). A wearable setup is also used by Thomas et al. (1998) in a system targeting terrestrial navigation guidance, and in ARQuake, an AR version of the classic Quake game (Piekarski & Thomas 2002).

A5.2 PDA's and camera-enabled phones

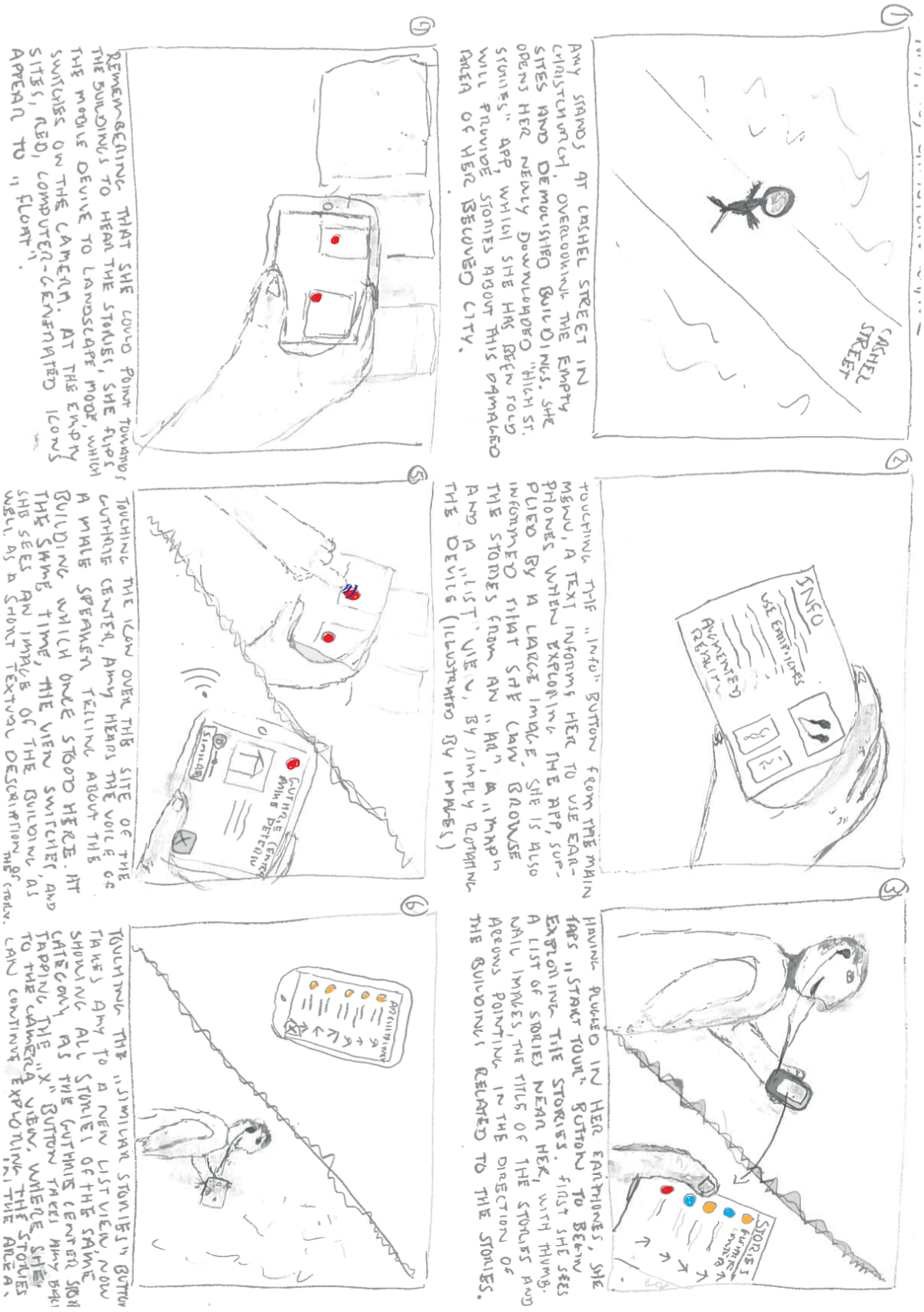
Because of the bulky nature of the early, wearable AR systems, a lot of research has been focusing on devices with smaller displays to make augmented reality more user-friendly. Early projects use PDA's to display location-based information. Fitzmaurice (1993) presents the Chameleon system, which uses a handheld

"palmtop" computer with a high-fidelity monitor and a 6 degrees-of-freedom tracking system to act as an information lens to nearby physical objects. Similar to the Chameleon project, Rekimoto & Nagao (1995) used a PDA as a video see-through to present annotations of physical objects in the real world, and relied on a seperate workstation for analysing the images and creating the augmented view. The first untethered mobile AR system was presented in 2003 by Wagner and Schmalstieg, whose AR system was self-contained in the PDA device itself in an indoor, marker-based navigation application (Wagner & Schmalstieg 2003). Also at this time, AR was seen in games on camera-enabled smartphones for the first time, for instance the Mozzies game, which superimposes virtual mosquitoes over the camera view of the Siemens SX1 phone. In 2005, the marker-based ARTennis game, which ported ARToolKit to the Symbian platform and supported collaboration between two mobile users, became highly popular (Henrysson et al. 2005).

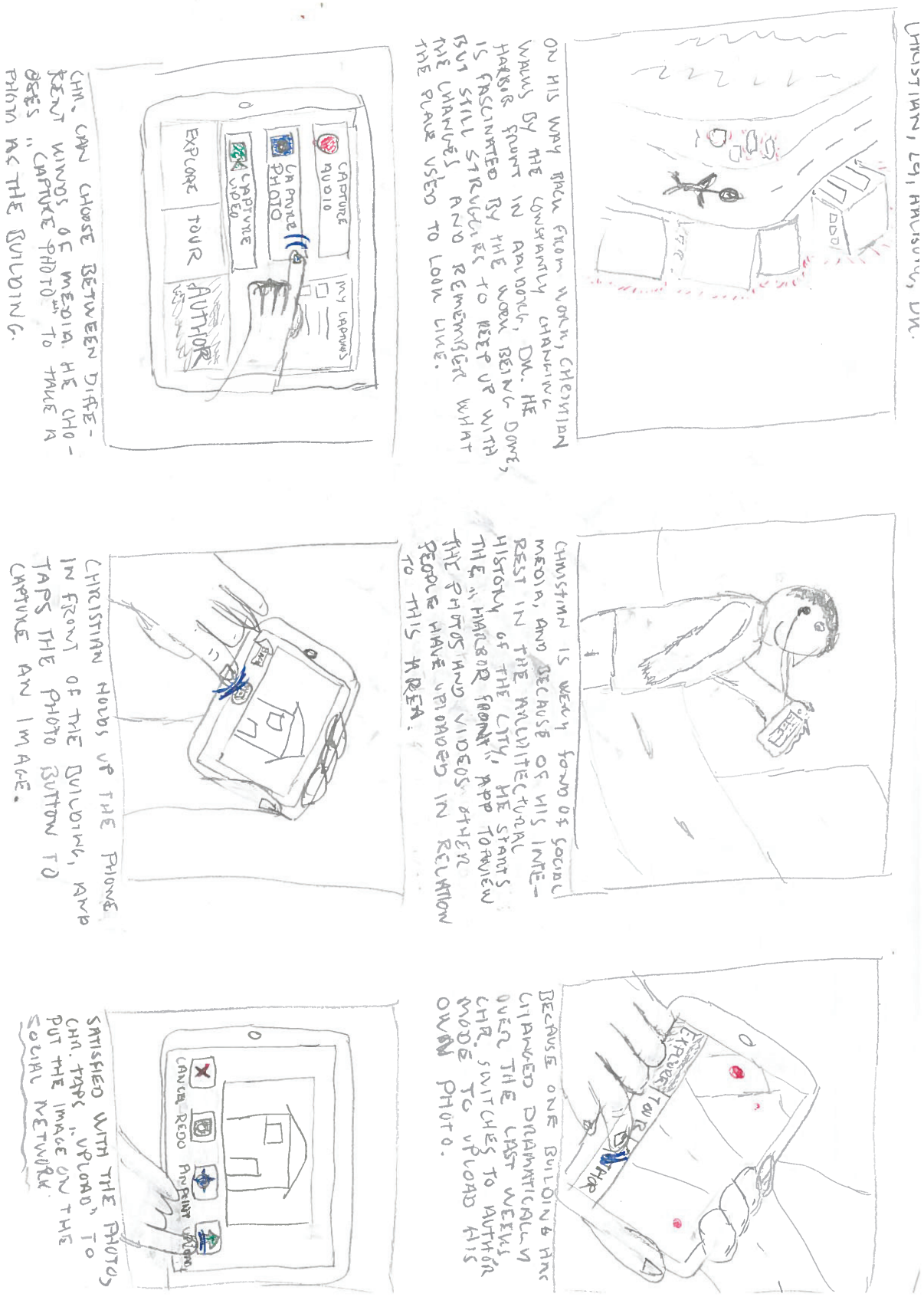
A5.3 Location-aware phones

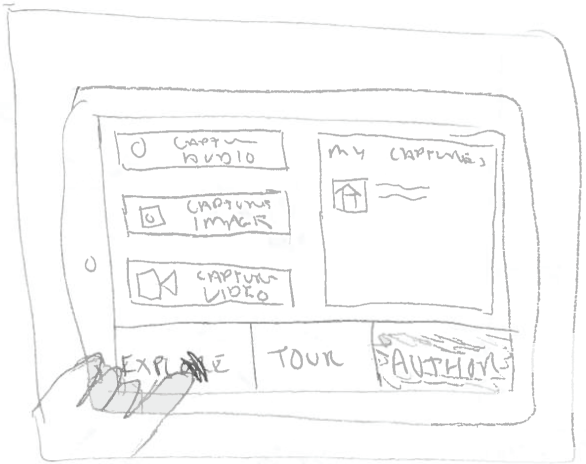
When location-aware phones turned mainstream around 2008, this started a wave of outdoor AR applications relying on GPS tracking and inertial sensors to superimpose information on top of real world locations. In particular, the so-called AR browsers have become popular, and junaio, Layar and Wikitude all provide good examples of the use of location-aware technology to tag real world locations with digital content. These systems let users author their own AR experiences and share these with large communities around the world. Nokia's City Lens has been mentioned several times throughout this report, because it emphasises an increasing awareness of the user's physical environment and engages interaction with physical locations in the cityscape.

B. Storyboard: Amy, Christchurch



C. Storyboard: Christian, Aalborg

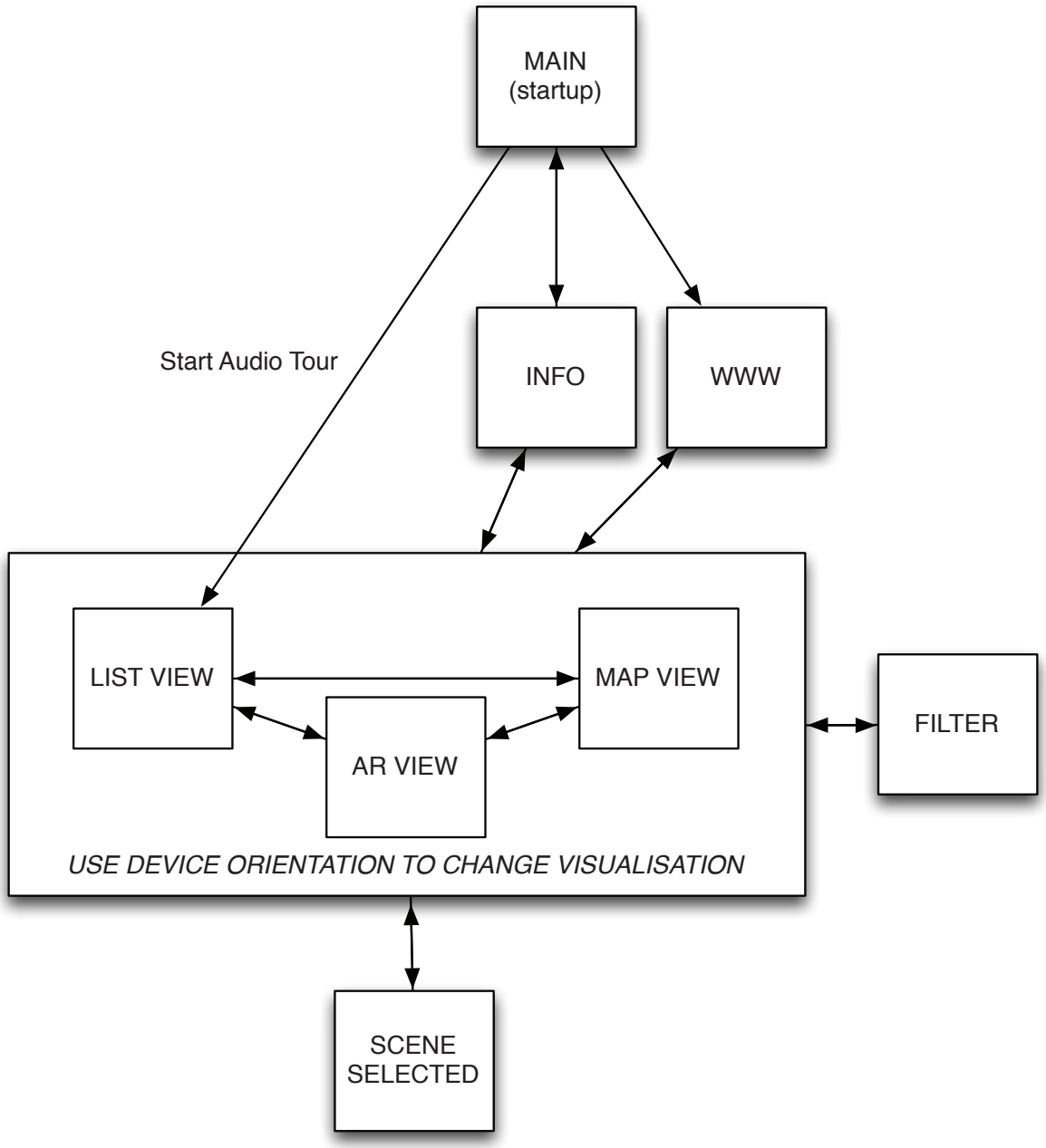




AFTER UPLOADING, CHRL IS TAKEN BACK TO THE „AUTHOR SCREEN. HIS PHOTO NOW APPEARS IN THE „MY CAPTURES" LIST. TAPPING „EXPLORE" TAKES HIM TO „EXPLORE" MODE.

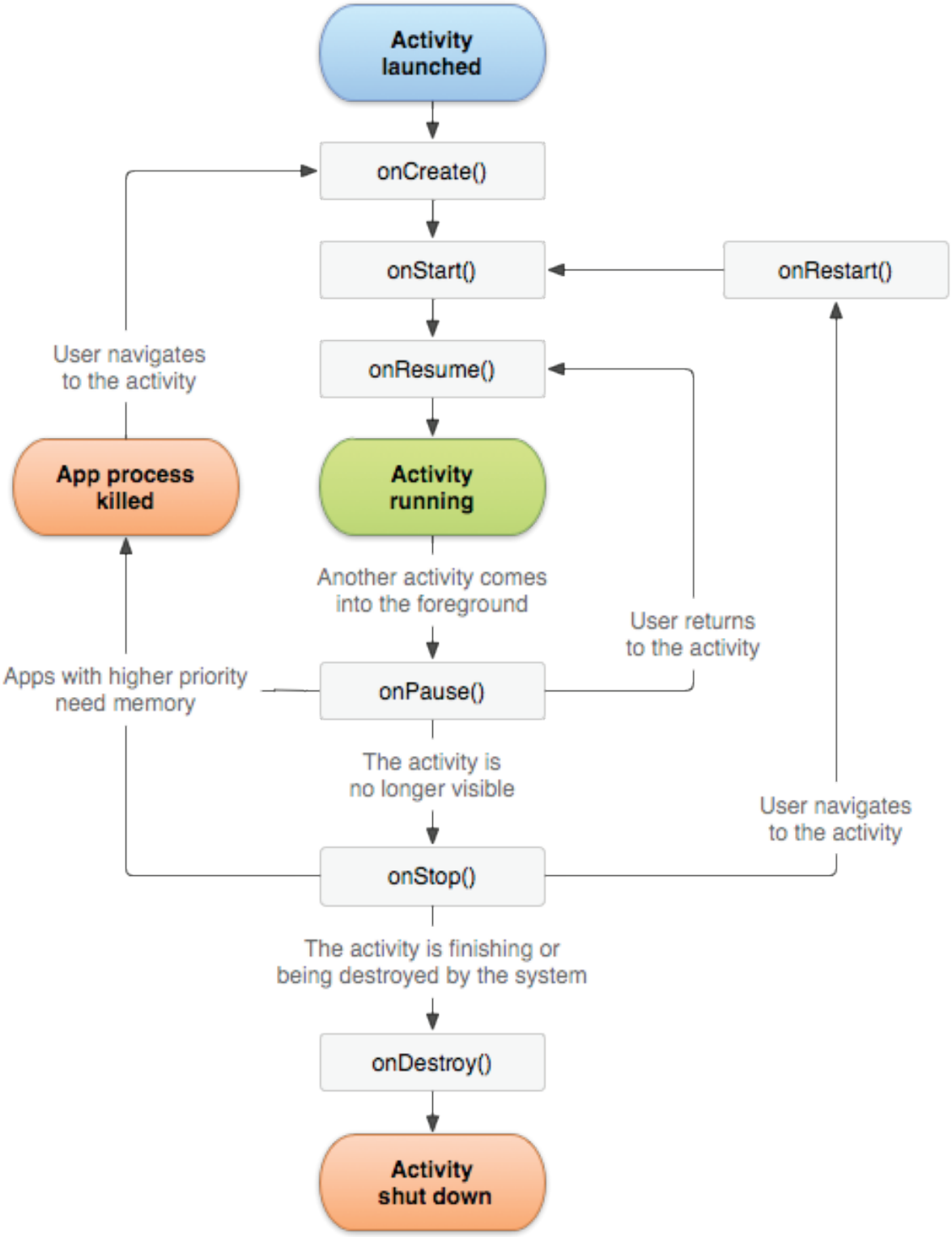
IN „EXPLORE" MODE, CHRL. CAN NOW SEE HIS IMAGE GEO-TAGGED, AND IN COEXISTENCE WITH OTHER ANNOTATIONS (INCLUDING PRE-DEFINED CONTENT AND FRIENDS' UPLOADS)

D. System navigation

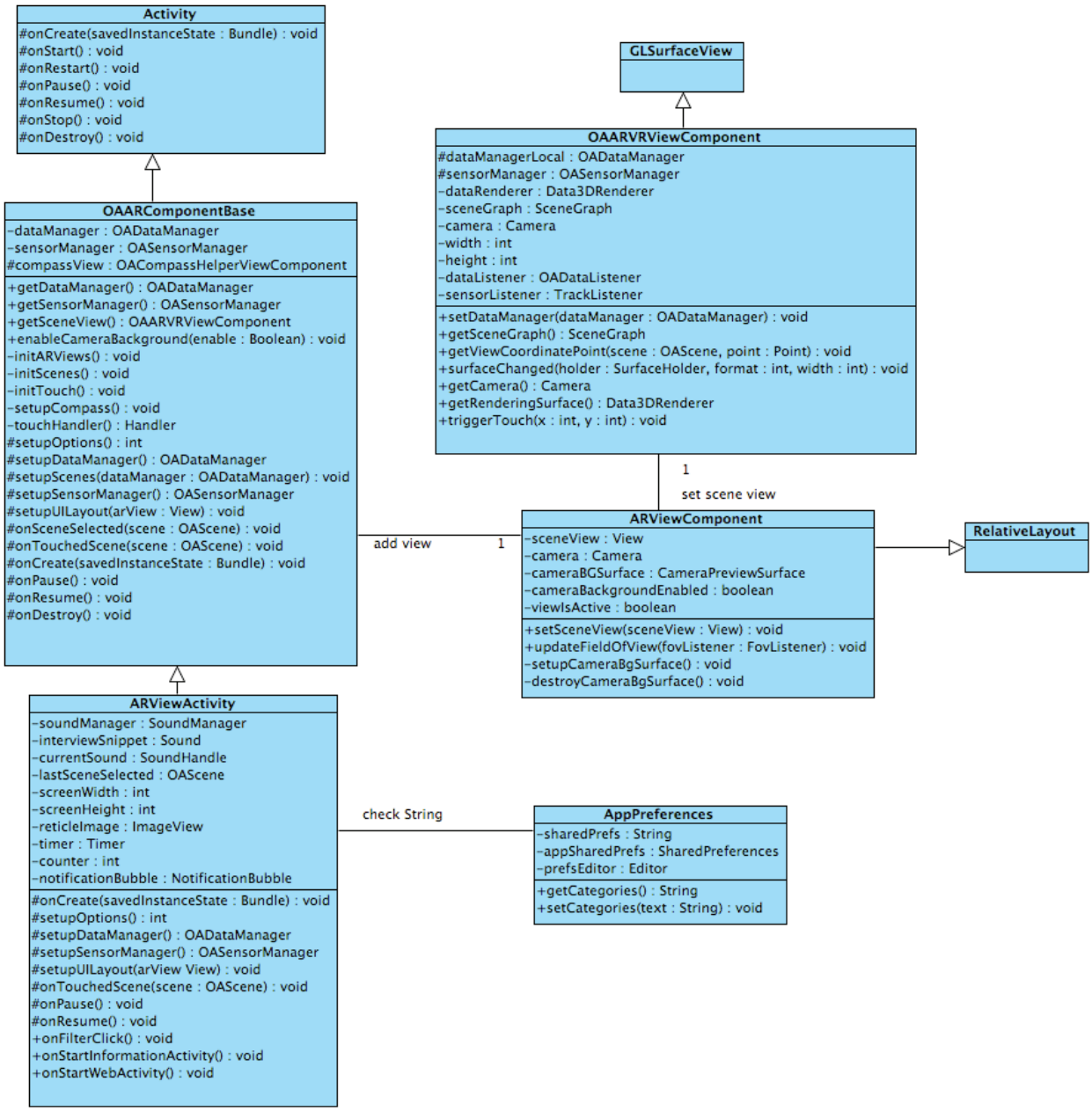


E. Android activity state paths

Source: Google (2013c)



F. UML class diagram of ARViewActivity



G. Stories for empirical field trial

This Appendix shows the written text of the speaks that was used for the fictive stories in the empirical field trial (Chapter 12). The speaks are here translated from Danish to English.

Aalborghus Castle (Subject: Kirsten Petersen)

"When I first heard of the history of Aalborghus Castle, I was surprised to hear that previously, an older castle was located here, a bit further to the south. So much happened in Aalborg in the years around the Count's Feud in the 16th century, and already around the year 1530, King Frederik I considered tearing down Aalborghus and decorate the city's Greyfriars monastery church as a castle. But then it was his son, Christian III, who came to realise the demolition, who instead raised a totally new, armed castle further to the North, close to the Limfjord. And that became the Aalborghus Castle as we know it today."

Utzon Center (Subject: Peter Carlsen)

"I knew Jørn Utzon as an amazing and passionate soul, who knew what he wanted and of course had a fantastic talent for architecture and design. Utzon designed for for humans, and his unique sense of detail shone through in all his buildings. Utzon himself was an Aalborg boy, and it is fantastic that this unique building would be his last masterpiece."

Nordkraft (Subject: Gitte Christensen)

"Even though Norkraft is first and foremost associated with cinema, concerts and venues, we also offer a wide range of sport activities. We house many sport clubs here at Nordkraft, and they have all become popular do to Nordkraft's aura and central location in the city. Amont others, Aalborg Kickboxing Club, Aalborg Sports Karate Club, the Handball Association of Jutland represented here, always looking for new members."

John Bull Pub (Subject: Christian Petersen)

"Here at John Bull, we are proud to be Aalborg's real English pub with charm, warmth and atmosphere. Here you will find the largest selection of Danish and international beer brands in North Jutland, as well as a number of famous and not so famous whisky brands. Further we have a large quality selection of rhum and cognac. Every saturday we have live music with Danish or international artists. Further, we have good music arrangements on thursdays and fridays."

Toldbod Square (Subject: Jakob Møller)

"Even though Toldbod Plads is primarily known for the water fountain, which goes hand in hand with Händel's "Watermusic", many people do not know that Aalborg businesses use the Square to restituation and relaxation. The connection between the Promenade and the newly renovated city center makes this neighborhood to a favorite place for lunch breaks. Among others, Jyske Bank, the law firm Hovmøller & Thorup as well as a wide range of restaurants and cafées are located in this area.

Damsø & Damsø Language School (Subject: Peter Damsø)

"Through decades, we have made a difference for our course participants, who cover people with linguistic troubles to people, who wish to learn new foreign languages. According to our participants, the lack of a linguistic barrier makes it possible to focus on the message and create good results. We are further proud to be located at Aalborg's lovely harbour fron with a view to the fjord and the great Utzon Center."

Sindal Church (Subject: Ole Dybro)

"In the city's church, we have served the inhabitants of Sindal for more than 100 years. We are responsible for confirmations, christenings, weddings and other ceremonial acts. The church was designed and projected by the architect Hother Paludan, and the fundament was laid in 1910. As is typical for Paludan's architectural style, the style of the church is calm and with focus on the robust implementation."

Sindal Stadium (Subject: Niels Christian Jensen)

"Sindal Sports Association was founded in 1936 when a selection of inhabitants wanted to start a local sports association. The club has always had a broad support from the city's inhabitants and the municipality. This support is expressed when big projects are to be carried out. When the club house was build in 1977 the city's inhabitants and craftsmen showed great will to help. The same goes for the municipality, whose support has made building and expansion finansially possible."

Sindal Station (Subject: Jens Petersen)

"For my whole life the station has been my home station. Actually Il live just on the other side of the rails. Often I stood here watching the old steam trains passing by in the good old days. The station building has changed a lot throughout the years. The old chimneys are gone, and the building has been painted. I still remember how they used to fire up the stoves in the old days. At that time, there was both a waiting lounge for smokers and non-smokers."

Mammas Pizza (Subject: Oktay Güven)

"Here at Mammas Pizza we always focus on delivering the best gastronomic experience. All our food is made in real, Italian stone ovens, and our special Italian pizzas are made using only the best ingredients. Please come inside and enjoy your meal." .

Jyske Bank (Subject: Lene Lauritsen)

"With us you get long-sighted advisory, whether you are a private or a business customer. In Jyske Bank we are interested in other things than your financial numbers, and you get counselling, which covers your whole economy - based on your wishes and needs."

Sindal School (Subject: Ole Sørensen)

"On Sindal School we have around 600 student from 0th to 10th grade. The school was build in 1953, and went through a large renovation between the years 1993 and 2000. Today we have a number of special offers for children with behavioural or linguistic issues, so it is fair to say that we are a school with enough space for differences."

H. Questionnaire : Post-task ratings

This questionnaire was handed to experiment participants after each of the seven tasks in the empirical evaluation of the application (Chapter 12).

A: I believe I made mistakes during the task.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

B: I believe I solved the task fast.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

C: I had trouble solving the task.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

D: I made no mistakes during the task completion.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

E: I feel that the task completion went slowly.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

F: The task was easy to solve.

Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree

I. System Usability Scale (SUS)

This questionnaire was handed to experiment participants after the empirical evaluation of the application. (Brooke 1996). In the experiment, the questionnaire was translated to Danish.

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5